ABS State space workshop

Lab session 1 30 May 2014

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

1. Consider the data set `books` — the daily sales of paperback and hardcover books in the same store.

   (a) For the paperback series, use the `ses` function in R to find the optimal values of α and ℓ_0, and generate forecasts for the next four days of sales.

   (b) For the paperback series, use the `holt` function in R to find the optimal values of α, β, ℓ_0 and b_0, and generate forecasts for the next four days. Compare the results with those obtained from `ses`.

   (c) Try other non-seasonal exponential smoothing methods for the paperback series. Which method do you think is best?

   (d) Try various non-seasonal exponential smoothing methods to forecast the next four days of sales for the hardcover series. Select the one you think is best.

   (e) Compare your models with those obtained automatically using `ets`.

2. For this exercise, use the monthly Australian short-term overseas visitors data, May 1985–April 2005. (Data set: `visitors`.)

   (a) Forecast the next two years using Holt-Winters' multiplicative method.

   (b) Why is multiplicative seasonality necessary here?

   (c) Experiment with making the trend exponential and/or damped.

   (d) Compare the RMSE of the one-step forecasts from the various methods. Which do you prefer? (The `accuracy` function will be useful here.)

   (e) Now use the `ets()` function to select a model automatically. Does it choose the same model you did?

3. Use `ets` to model and forecast time series selected from the `fma`, `expsmooth` or `fpp` packages.

   (a) Experiment with different options in the `ets` function and see what effect they have.

   (b) Check the residuals of the fitted model to ensure they look like white noise using `Acf(residuals(fit))`

   (c) Can you find an example where the forecasts are obviously poor?

4. For this exercise, use the monthly Australian short-term overseas visitors data, May 1985–April 2005. (Data set: `visitors` in `expsmooth` package.)

   (a) Use `ets` to find the best model for these data and record the training set RMSE. You should find that the best model is ETS(M,A,M).

   (b) We will now check how much larger the one-step RMSE is on out-of-sample data using time series cross-validation. The following code will compute the result, beginning with four years of data in the training set.

```
k <- 48 # minimum size for training set
n <- length(visitors) # Total number of observations
e <- visitors*NA # Vector to record one-step forecast errors
for(i in 48:(n-1))
{
  train <- ts(visitors[1:i],freq=12)
  fit <- ets(train, "MAM", damped=FALSE)
  fc <- forecast(fit,h=1)$mean
  e[i] <- visitors[i+1]-fc
}
sqrt(mean(e^2,na.rm=TRUE))
```

   Check that you understand what the code is doing. Ask if you don't.

   (c) What would happen in the above loop if I had set `train <- visitors[1:i]`?

   (d) Plot `e`. What do you notice about the error variances? Why does this occur?

   (e) How does this problem bias the comparison of the RMSE values from (4a) and (4b)? (Hint: think about the effect of the missing values in `e`.)

   (f) In practice, we will not know that the best model on the whole data set is ETS(M,A,M) until we observe all the data. So a more realistic analysis would be to allow `ets` to select a different model each time through the loop. Calculate the RMSE using this approach. (Warning: it will take a while as there are a lot of models to fit.)

   (g) How does the RMSE computed in (4f) compare to that computed in (4b)? Does the re-selection of a model at each step make much difference?