ETW3420

Principles of Forecasting and Applications

Topic 5 Exercises - Part 1

Question 1

Data set `books` contains the daily sales of paperback and hardcover books at the same store. The task is to forecast the next four days' sales for paperback and hardcover books.

(a) Plot the series and discuss the main features of the data.

(b) Use the `ses` function to forecast each series, and plot the forecasts.


```
#Forecast hardcover books
fcast1 <- ses(books[,"Hardcover"], h=4)



#Forecast paperback books
fcast2 <- ses(books[,"Paperback"], h=4)


#Plot forecasts



#Print outputs
```

(c) Compute the RMSE values for the training data in each case.

(d) Compute a 95% prediction interval for the first forecast of hardcover books using $\hat{y} \pm 1.96s$ where s is the standard deviation of the residuals. Compare your interval with the interval produced by R.

```r
#Obtain standard deviation of residuals
sd <- fcast1 %>% residuals() %>% var() %>% sqrt()

#Obtain first point forecast
f1 <- fcast1$mean[1]

#Calculate prediction intervals; print as a vector
```

Will notice that the intervals are close but not identical. This is because R estimates the variance of the residuals differently, taking account of the degrees of freedom properly (i.e. $T -$ no. of parameters).


```r
#Obtain standard deviation of residuals
sd <- sqrt(sum(residuals(fcast1)^2)/(nrow(books)-2))
```



```r
#Calculate prediction intervals; print as a vector
```


(e) Print and plot the values of ℓ_t of the `hardcover` series estimated by SES.

Question 2

(a) Now apply the Holt's linear method to the `paperback` and `hardcover` series and compute 4-day forecasts in each case.

```
#Produce forecasts using Holt's linear method
fcast1 <- holt(books[,"Hardcover"], h=4)
fcast2 <- holt(books[,"Paperback"], h=4)

#Print output



#Plot forecasts
```


(b) Compare the RMSE measures of Holt's method for the two series to those of simple exponential smoothing in Question 1. (Remember that the Holt's method is using one more parameter than SES). Discuss the merits of the 2 forecasting methods for these data sets.

(c) Compare the forecasts for the two series using both methods. What do you observe, and which do you think is better?

(d) Calculate a 95% prediction interval for the first forecast for each series, using the RMSE values and assuming normal errors. Compare your intervals with those produced using `ses` and `holt`.

```
#Prediction interval for hardcover
s <- sqrt(fcast1$model$mse)

High <- fcast1$mean[1] + 1.96*s
Low <- fcast1$mean[1] - 1.96*s
```

```
c(Low = Low, High = High)
fcast1


#Prediction interval for paperback (do it yourself)
```

(e) Print and plot the values of ℓ_t and b_t of the `hardcover` series estimated by the Holt's linear method.

```
#Extract l_t and b_t
holt.l <- fcast1$model$states[, "l"]
holt.b <- fcast1$model$states[, "b"]

#Plot l_t and b_t
```

Question 3

For this exercise, use the data set `eggs`, the price of a dozen eggs in the United States from 1900 - 1993. Experiment with the various options in the `holt()` function to compare forecasts across the different permutations. Use `h=100` when calling `holt()` so that you can clearly see the differences between the various options when plotting the forecasts.

Which model gives the best RMSE?

(a) Estimate the smoothing parameters of various exponential smoothing methods and produce a 100 period forecast by considering various options in `holt()`.

(b) Produce plots of the forecasts. Any problems that you can identify?

(c) Produce the plot of b_t under the damped trend method and observe its values.

(d) Which model gives the best RMSE? Plot the model's fitted values and forecasts.

Question 4

The "retail.xls" file contains data on the retail sales in various categories for different Australian states. For this question, we will consider the retail turnover for New South Wales (A3349873A) and apply the Holt-Winters Seasonal Method.

(a) Why is multiplicative seasonality necessary for this series?

```
#Import data and convert into ts
myts <- readxl::read_excel("retail.xlsx", skip=1)[,"A3349873A"] %>%
  ts(frequency=12, start=c(1982,4))

#Plot data
autoplot(myts)
```


(b) Apply Holt-Winters' multiplicative method to the data. Experiment with making the trend damped.

(c) Compare the RMSE of the one-step forecasts from the two methods. Which do you prefer?

(d) Check that the residuals from the best method look like white noise.

(e) Now find the test set RMSE, while training the model to the end of 2010. Can you beat the seasonal naive approach?