UNSW ECON2209 Assessment
Project

2021

At the start of an R session for this course, remember to type `library(fpp3)` in the R Studio Console. This will then load (most of) the R packages you will need, including some data sets.

Details:

- Total value: 25 marks.
- Submission is due on **Friday of Week 9 (16 April), 5pm**.
- A submission link is on the Moodle site under Assessments.
- Submit your documents in PDF format.
- Your submitted answers should include the R code that you used.
- **Format:** No longer than 20 pages, including code, figures, tables and any appendices. Do not include a separate title page. At least 11 point font should be used, with adequate margins for comments. Any extra pages will not be marked.
- This project requires you to analyse time series data. The series will differ between students.
- The project is set out as containing four Parts, each with multiple subparts. This is mainly to guide you with your analysis of your data series. It is strongly recommended that you follow the given sequence in your analysis and in presenting your results.
- Unless approval for an extension is given on medical grounds (supported by a medical certificate submitted through the Special Consideration process) there will be an immediate late penalty of five marks from 5:01pm on 16 April, followed by additional penalities of five marks per calendar day.

Marking Criteria: Marks are not awarded by Part, but by overall achievement against the following criteria:

(a) Suitability of methods. (10)

(b) Interpretation of the results, arguments used and conclusions drawn. (10).

(c) Presentation: Appropriate style of graphs, tables, reporting and clarity of writing. (5)

Maximum marks: 25

Select the data series that you will analyse

Monthly Australian retail data is provided in aus_retail. This is a data set from the Australian Bureau of Statistics, catalogue number 8501.0, table 11. It contains data on retail turnover, in millions of dollars, for retail trade industries for each of the Australian states and territories.

You must use the following method for selecting your data series.

Use the seven digits of your UNSW student ID to get the data series that you will analyse in this project, as in the following example for the case when your student ID is Z1234567:

```r
set.seed(1234567)
myseries <- aus_retail %>%
  filter(`Series ID` == sample(aus_retail$`Series ID`,1))
```

The above code will select the retail industry "Cafes, restaurants and takeaway food services" for Victoria. It is possible that your student ID number will lead to the selection of the same series, but more likely that it will not. (Note while `sample()` takes a random sample, using the same "seed" through `set.seed()` will result in the same series being selected each time.)

Part 1: Data Exploration and Transformation

a. Plot your data using the following code:

```r
myseries %>%
  autoplot(Turnover) +
  labs(y = "Turnover (million $AUD)",
       title = myseries$Industry[1],
       subtitle = myseries$State[1])
```

b. Now explore your retail time series using the following functions, being sure to discuss what you find:

   ```r
   gg_season(), gg_subseries(), gg_lag(), ACF() %>% autoplot()
   ```

c. What Box-Cox transformation, if any, would you select for your data, and why?

Part 2: Forecasting

For your untransformed data series:

a. Create a training dataset (`myseries_train`) consisting of observations before 2011.
b. Check that your data have been split appropriately by producing a plot of `myseries_train`.
c. Calculate seasonal naïve forecasts using `SNAIVE()` applied to your training data.
d. Check the residuals. Do the residuals appear to be uncorrelated and normally distributed?
e. Produce forecasts for the test data.
f. Compare the accuracy of your forecasts against the actual values.

Part 3: Exponential Smoothing

For your full, untransformed data series:

a. Is a multiplicative seasonality method appropriate for this series?
b. Apply Holt-Winters' multiplicative method to the data. Then try the method making the trend damped. Plot and compare the 36-period ahead point forecasts and prediction intervals for both methods. Using `accuracy()`, compare the RMSE of the of the two methods; this comparison is based on one-step-ahead in-sample forecasts. Which method do you prefer and why?
c. Check that the residuals from the best method look like white noise.
d. Train the model to the end of 2010 and find the test set RMSE. Can you beat the seasonal naïve approach from Part 2?

e. Try an STL decomposition applied to the Box-Cox transformed series, followed by ETS on the seasonally adjusted data. How does that compare with your best previous forecasts on the test set?

Part 4: ARIMA Modelling

Finish b and c by 5:30am onwards

a. For your data series, find the appropriate order of differencing, after transformation if necessary, to obtain stationary data.

b. Select an appropriate seasonal ARIMA model. Explain your choice and report the results.

c. Using the test data set as before, compare the forecast performance with the models you obtained in Part 2 and Part 3. Try also an STL decomposition applied to the Box-Cox transformed series, followed by ARIMA on the seasonally adjusted data; that is, an STL-ARIMA model rather than the STL-ETS model used in Part 3(e).