Assignment 3

Applied Forecasting in Complex Systems 2021

Week 4
November, 19, 2021

Background on how to work effectively on this assignment:

1. Follow the Guideline of Assignment 1 to 3.
2. Teamwork is highly recommended.
3. Each member of the team should be able to solve each problem.
4. Please explain all solutions, show the calculations and write down your foundations and reflections. When necessary, reduce your plot.
5. **MAXIMUM LENGTH OF THE REPORT is 15 PAGES**, excluding Appendix.

Use library(fpp3) to get the dataset unless specified as a file.

Exercise 1 (2 pts.)

Electricity consumption is often modelled as a function of temperature. Temperature is measured by daily heating degrees and cooling degrees. Heating degrees is 18°C minus the average daily temperature when the daily average is below 18°C; otherwise it is zero. This provides a measure of our need to heat ourselves as temperature falls. Cooling degrees measures our need to cool ourselves as the temperature rises. It is defined as the average daily temperature minus 18°C when the daily average is above 18°C; otherwise it is zero. Let y_t denote the monthly total of kilowatt-hours of electricity used, let $x_{1,t}$ denote the monthly total of heating degrees, and let $x_{2,t}$ denote the monthly total of cooling degrees.

An analyst fits the following model to a set of such data:

$$y_t^* = \beta_1 x_{1,t}^* + \beta_2 x_{2,t}^* + \eta_t$$

where

$$(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - B)(1 - B^{12})\eta_t = (1 + \theta_1 B)\varepsilon_t$$

and $y_t^* = \log(y_t)$, $x_{1,t}^* = \sqrt{x_{1,t}}$ and $x_{2,t}^* = \sqrt{x_{2,t}}$.

1.1) (0.5 pt.) What sort of ARIMA model is identified for η_t? Elaborate your comments.

1.2) (1 pt.) Write the equation in a form more suitable for forecasting.

1.3) (0.5 pt.) 4. Once you have a model with white noise residuals, produce forecasts for the next year.Describe how this model could be used to forecast electricity demand for the next 12 months.

Exercise 2 (2.5 pts.)

Monthly Australian retail data is provided in aus_retail. Select one of the time series as follows (but choose your own seed value):

```
set.seed(12345678)
myseries <- aus_retail %>%
  filter(`Series ID` == sample(aus_retail$`Series ID`,1))
```

2.1) (1 pt.) Develop an appropriate dynamic regression model with Fourier terms for the seasonality. Use the AICc to select the number of Fourier terms to include in the model. (You will probably need to use the same Box-Cox transformation you identified previously.)

2.2) (0.5 pt.) Check the residuals of the fitted model. Comment on how the residual series looks like.

2.3) (1 pt.) Compare the forecasts using ETS, ARIMA and dynamic regression.

Exercise 3 (2.5 pts.)

The `us_gasoline` series consists of weekly data for supplies of US finished motor gasoline product, from 2 February 1991 to 20 January 2017. The units are in "million barrels per day". Consider only the data to the end of 2004,to include more data and ARMA errors.

3.1) (0.5 pt.) . Fit a harmonic regression with trend to the data. Experiment with changing the number Fourier terms. Plot the observed gasoline and fitted values and comment on what you see. Select the appropriate number of Fourier terms to include by minimising the `AICc` or `CV` value. Check the residuals of the final model. Even though the residuals fail the correlation tests, the results are probably not severe enough to make much difference to the forecasts and prediction intervals. (Note that the correlations are relatively small, even though they are significant.)

3.2) (1 pt.) To forecast using harmonic regression, you will need to generate the future values of the Fourier terms by forecast newdata of your `(fourier(x,K,h)`. Forecast the next year of data. Plot the forecasts along with the actual data for 2005. What do you find? Using tslm(), fit a harmonic regression with a piecewise linear time trend to the full gasoline series. Select the position of the knots in the trend and the appropriate number of Fourier terms to include by minimising the AICc or CV value.

3.3) (1 pt.) Now refit the model using `auto.arima()` to allow for correlated errors, keeping the same predictor variables as you used with `tslm()`. Check the residuals of the final model. Do they look sufficiently like white noise to continue? If not, try modifying your model, or removing the first few years of data. Once you have a model with white noise residuals, produce forecasts for the next year.

Exercise 4 (3 pts.)

The NN3 (`NN3_REDUCED_DATASET_WITH_TEST_DATA.xls` attached as a file) reduced dataset consists of monthly time series drawn from homogeneous population of empirical business time series, and they are a subsample of 11 time series from the 111 time series (complete dataset). Source: NN3 Time Series Competition

4.1) (0.5 pt.) . Plot the data to look at some effect of the changing seasonality over time. What do you think is causing it to change so much? Do an STL decomposition of the data if necessary.

4.2) (1.5 pt.) . Forecast using simple methods (`Mean` and `Naïve`), (`Simple`, `Seasonal` & `Dampened Trend`) Exponential Smoothing, and `ARIMA`. Use `SMAPE`, `MSE`, `MAE`, `MAPE` and `RMSE` for your forecast accuracy.

4.3) (1 pt.) Compare the forecasts obtained. Elaborate and reflect based on the accuracy metrics.