AAEC 4984/5984 – Applied Economic Forecasting

Include your name here

Homework #2 – Spring 2020

So far, we have largely been focusing on the datasets included in the text. This assignment will offer you a real world example of how to utilize the tools we have learnt so far.

Instructions: In all cases, please ensure that your graphs and visuals have properly titles and axes labels, where necessary. For your convenience, I have posted my R markdown file on our course website so that you can open and alter as you see fit. Refer to the output, whenever appropriate, when discussing the results. Lastly, remember that creativity (coupled with relevance) will be rewarded.

US Natural Gas Consumption

Last week, I sat in on a presentation about the US Natural Gas industry. I was very intrigued but unfortunately, I do not have the time nor the tools to explore the dynamics in that market. Fortunately, I remembered that you have been taking the Applied Economic Forecasting class and can therefore help me with this. Can you use the tools you have been introduced to so far to make sensible future consumption forecasts? Of course you can, I will guide you through the process below.

1. Using the code below, import the NG consumption data from the U.S. Energy Information Administration (EIA) into `R`.

```
tmp <- tempfile(fileext = ".xls") # Storing the file to your computer's temporary memory
#Pull data from EIA
download.file(url = "https://www.eia.gov/dnav/ng/xls/NG_CONS_SUM_DCU_NUS_M.xls",destfile = tmp,
              mode="wb")
```

2. Using the readxl command, read the temporary file into R. Be sure to skip the first 2 rows.

```
ngdata <- readxl::read_excel(tmp, sheet = 2, skip = 2)
```

3. Drop the original date column and convert `ngdata` to a time series object starting at January 1973. Be sure to specify the proper frequency. Save this as `tsng`.
4. Now keep only the first column of `tsng` ["U.S. Natural Gas Total Consumption (MMcf)"] and use the window command to drop all observations before January 2001. Save this as `conng`
5. Convert the units of `conng` from MMcfs to Bcfs (it is ok to save this back into `conng`)
6. Present a time plot of `conng`.
7. As we usually do, use the tools you have learnt so far to **comment** on possible seasonality and trends in the consumption data. Would you say that the series appears to be white noise? *I anticipate seeing at least 3 graphs here.*
8. Split the data set into a training (ending December 2015) and testing set (starting January 2016). Name them `train.conng` and `test.conng`, respectively.
9. Show that the data is properly split by using the `autoplot` and `autolayer` function. Be sure to include `conng`, `train.conng` and `test.conng` in this plot. A title is not necessary.
10. Given your conclusion in Part 7, use the benchmark models to forecast the training set.
11. In a single graph, plot the training data and forecasted series.
12. In each case, compare the accuracy of the fitted values against the actual values in the test set.

- Extract the `RMSE`, `MAE`, `MAPE`, and `MASE` statistics.
- Keep only the `Test set` row

- Round your answers to 3 dps.
- Place into a table (see my code below)
- Which model is preferred under the each method?

```
#Remember to remove eval=FALSE in order for the results of your codes to show.
r1 <- round(........,3)
r2 <- round(........,3)
r3 <- round(........,3)
#Now row bind r1 - r3
accuracy.tab <- as.table(rbind("Mean" = r1, "Naïve"= r2, "Seasonal Naïve" = r3))
accuracy.tab
```

12. Comment on the residuals from each model. In particular,
    - Do they appear to be normally distributed?
    - Are the residuals uncorrelated? (Be sure to state the null and conclusion.)