Applied Economic Forecasting

Include your name here

Homework #2 - Spring 2021

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

In this exercise, we are using data on the consumption of Natural Gas (NG) in the U.S. The data is collected directly from the U.S. Energy Information Administration (EIA) in `R`.

Our aim is to estimate and evaluate the performance of our basic forecasting model in and out of sample. Throughout, we will be building on the topics addressed thus far in the course.

At the end, you will select an optimal model for modeling the residential segment of U.S. consumption.

1. Using the code below, import the NG consumption data from the EIA into `R`.

```
tmp <- tempfile(fileext = ".xls") # Storing 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 `readxl::read_excel()`, read the temporary file, `tmp` into R. Be sure to skip the first 2 rows since they are not our column headings. Store this as `ngdata`.

```
# Remember to set eval to TRUE in order to compile
...... <- readxl::read_excel(tmp, sheet = 2, skip = 2)
```

3. Drop the original date column (column 1) 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 fifth column ["U.S. Natural Gas Residential Consumption (MMcf)"] of `tsng` and use the `subset` command to drop all observations before January 2001. Save this as `ng2`.

Hint: To find the appropriate starting value, you can use `which(time(tsng)==2001)`. This will tell us the position of 2001 in the time element of `tsng`.

5. Convert the units of `ng2` from MMcfs to Bcfs (save this back into `ng2`).

6. Present a time plot of `ng2`.

7. As we usually do, use the tools you have learned so far to **comment** on possible seasonality and trends in the consumption data. Do you observe any strange year? Would you say that the series appears to be white noise? *I anticipate seeing at least 3 graphs here.*

8. Let us proceed with our forecasting of the `Residential` series. Using the `subset` command again, split the data set into a training (`train.res`) and testing set period (`test.res`). Assign the last **4 years** of data to the testing period.

9. Confirm that the data is properly split by using the `autoplot` and `autolayer` function. Be sure to include `ng2`, `train.res` and `test.res` in this plot. *A title is not necessary.* **If your data is properly subsetted then the training and test series should span the entire sample period. You can ignore a small line segment that connects the training and testing datasets.**

10. Give your conclusion regarding seasonality and trends, use the appropriate benchmark models from class to forecast the training set. **Use a horizon of 4 years.**

11. In a single graph, plot the training data and the point predictions of your forecasted series.

12. Comment on your observations from your graph in 11.

13. Compare the accuracy of the fitted values of the training period against the hold out values (those in the test set).

- Extract the `RMSE`, `MAE`, `MAPE`, and `MASE` statistics.
- Keep only the `Test set` row
- Use `digits = 3` in the `kable` command.
  - Be sure to have appropriate column and row labels.
- Which model is preferred under the each method?

14. Using the `checkresiduals` command, comment on the residuals from each model. In particular,
    - Do they appear to be normally distributed?
    - Are the residuals uncorrelated? In other words, is there evidence of serial correlation? (Be sure to state the null of the test and conclusion.)