AAEC 4984/5984 – Applied Economic Forecasting

Your name Here

Homework #3 – Spring 2020

Instructions: Where necessary, please ensure that your graphs and visuals have properly titles and axes labels. **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.

Explaining US Residential Natural Gas Consumption

In the last homework, you were introduced to the USNG data from the U.S. Energy Information Administration (EIA). This week, your job is to explore the relationship between Residential NG consumption and a few predictors.

1. Using the codes from your last homework, import the NG consumption data from the EIA into `R`. Be sure to:

i. skip the first 2 rows
ii. drop the original date column and declare the variable as a ts object at a monthly frequency.
iii. declare this variable as `tsng`.
iv. keep only the column called `U.S. Natural Gas Residential Consumption (MMcf)` and subset the data, keeping only those **after** December 2000. Store this as `residng`.

2. Present a time plot of `residng`. Use the `geom_point` command to add the data points to your graphs

3. Using **two** appropriate graphs, **comment** on possible seasonality and trends in the consumption data.

A possible predictor for the dynamics noted in the NG consumption data is the weather. To this effect, the code below pulls the data for average temperature for the Contiguous United States from the National Oceanic and Atmospheric Administration (NOAA). I will assign this to a variable called `temp`. *Temp is collected in* $^\circ F$.

```
# Remove the eval argument and add echo=FALSE to mute the code in your
# output!
tmp2 <- tempfile(fileext = ".csv")
download.file(url = "https://www.ncdc.noaa.gov/cag/national/time-series/110-tavg-all-1-2001-2020.csv?base_pr
    destfile = tmp2)

# Assign the temperature variable to a variable called temp.
temp <- read.csv(tmp2, skip = 4)
```

4. Convert `temp` data to ts. *It is OK to save this back into* `temp`.

i. Keep only the `Value` column.
ii. Subset the data to match the time period of `residng`
iii. Use the `head` (`tail`) commands to view the first (last) 5 observations to ensure that the data has been properly imported.

5. Produce a scatter plot of `residng` (y-axis) against `temp` (x-axis) and **comment** on the strength of the association. Also, use the `geom_smooth` argument to fit the regression line to the plot.

6. Using the `ggpairs` function formally confirm your comments in Part 5. *Recall that the* `ggpairs` *command require that you supply the variables* `as.data.frame`. *Also, please ensure that your columns are appropriately titled. You can suppress the message about the library by using* `message=FALSE` *in your r-chunk preamble*

7. **Model 1:** Estimate the time series regression of `residng` on `temp`.

i. Store the model results in `fit.model1`.
ii. Store the model coefficients in `coef.model1` and round to 3 dps.

8. Produce a plot that shows the original data and the regression line superimposed.

i. Be sure to properly name each series

ii. As the graph title, add the formula for the estimated regression.

9. Does this simple linear model seem appropriate for the data? It would also prove helpful to extract and interpret the R^2 of **Model 1**.

10. Conduct a check of the residuals and comment on any discernible pattern in the residuals of **Model 1**. Specifically, are any of the assumptions possibly violated?

A number of economic studies have explored the relationship between NG consumption and other economic variables. Some possible predictors are (i) the price of NG, (ii) the price of alternative energy sources, (iii) GDP growth and (iv) inflation.

11. Use the `quantmod` package to grab the following data series from `FRED`. The symbol codes are presented in parenthesis below.

i. Henry Hub Spot Prices (MHHNGSP)
ii. WTI oil prices (MCOILWTICO)
iii. U.S Industrial Production Index (INDPRO): This will be used as a proxy for US GDP since GDP is not available monthly.
iv. U.S Consumer Price Index (CPIAUCSL): This will help us to calculate inflation.

12. Convert each series in Part 11 to ts and subset the data to match the time frame of `residng`.

- In the order listed in Part 11 call these: `ngprice`, `oilprice`, `ip`, `cpi`. A shorter way of subsetting and converting the `xts` objects is suggested in the R chunks below.

```
#Remove the eval argument to run
ngprice <- ts(coredata(MHHNGSP["2001::2019"]), start = c(2001,1), frequency=12)
... Complete as you see fit!
```

13. Compute the logged first difference of `ip` and `cpi` and multiply by 100. This will give us the growth rate of production and inflation, respectively. Store them as `out.growth` and `infl`, correspondingly.

The code chunk below combines all the variables into a matrix ts object called `ng.full`. I will drop the first row since we lost the first observation to differencing earlier.

```
#Remove the eval argument to run
ts.union(residng,temp,ngprice,oilprice,out.growth,infl) %>% window(start= c(2001,2)) -> ng.full
```

14. Produce a correlation plot of all variables in `ng.full` using `ggpairs`.

15. Estimate the regression of `residng` against all the predictors included in `ng.full`.

i. Store the model results in `fit.model2`.
ii. Store the model coefficients in `coef.model2` and round to 3 dps.

16. Produce a plot that shows the original data and the regression line superimposed.

i. Be sure to properly name each series
ii. As the graph title, add the formula for the estimated regression. *You can include the `atop` command in the `bquote` function to break the equation if required.*

17. For the two models estimated above, compare the accuracy of the fitted values against the actual values.

- Conduct the comparisons based on the `RMSE`, `MAE`, `MAPE`, and `Adjusted` R^2 statistics.
- Round your answers to 3 dps.
- Represent these results into a table.
- Explain which model is preferred under each method?