HW2

DSA301

Due 2024 Week 9

Q1: ARIMA modelling

- Do question 11 from chapter 8.11 of the textbook:
  - https://otexts.com/fpp2/arima-exercises.html

Remarks

- Generally, code from in class case studies may be useful as starting point

- For part (f), treat the "latest figures from the EIA" as your "out of sample" period, and report accuracy statistics in the usual format for your forecast

- Similar as with HW1, repeat part (f) for around 3 "possible" models that you had shortlisted and estimated from parts (d) and (e).

- Hint: For part (g), consider the ratio of your numerical forecast to the width of the prediction interval

Q2: VAR estimation

1. Explore the "visnights" variable in fpp2. Look up online in R's documentation definition of variables. Put this description as the answer to this question

2. Plot the various columns. Does there appear to be:
   - A long term directional trend (upwards or downwards) in the data?
   - Seasonality?

3. Use ndiffs and nsdiffs to determine differences required to render 2 variables (from part 3) stationary. Convert variables to stationary form or explain why no transformation needed.

4. Build VAR model on the above 2 variables, and check residuals for time series patterns. Start by using VARselect, and then increase order of VAR model until residuals do not have significant information

Q3: ARIMA-X

1. Do Question 9.1.1 from the textbook (https://otexts.com/fpp2/dynamic-exercises.html Q1)

2. For part b where the question says to use tslm(), it is just a standard OLS regression.

   arima(0,0,0) only shifts the intersect as it only optimises for c

3. Question is essentially asking for difference between regression errors and innovation errors

Q4: ARIMA

1. Do Question 8.11 Question 6
    (https://otexts.com/fpp2/arima-exercises.html Q6)

2. For simulation of ARIMA time series, do **not** use the simulation code we used in class exercise.

3. However, the rnorm() function specified in the question should be sufficient.

Q5: Free form modelling (ARIMA / ARIMA-X/VAR/VECM)

1. Attached python file HW2_Q5.py will access Wikipedia API automatically and download # of times each Wikipedia page has been accessed per day. It will write data to a time series called 'output_timeseries.csv'

   - Although this step uses python, and we have not learnt python in class, purpose of this step is **only** to generate sample time series for modelling, and also to familiarize the class with Wikipedia API (for those who are doing this for term project). No further work in Python is required.

   - You can just copy and paste the attached python code in any python environment and press <enter> to run. That's it.

2. Wikipedia pages correspond to various popular contemporary musicians (I generated the list by searching google for "most popular singers" and taking the top few hits. Have no real idea who these people are)

3. Read in the csv file from the above preliminary step in R, convert it to a TS object, and build a forecasting model on each of the time series shown. As you are forecasting multiple time series, you can choose either single time series methods (e.g. ARIMA) or multiple time series methods such as VAR and VECM.

4. Remember to use out of sample performance as the gauge between different classes of models