METU Dept. Of Business Administration
BA4318 Spring 2019

Homework Assignment #3 (Due 15.05.2019 23:50)

Please first take some time to study the accompanying data file HW03_USD_TRY_Trading.txt.

- The file is tab separated.
- This file gives USD to Turkish Lira exchange rates at an international trading site, **with 1 minute resolution**. For a whole day you get 24 hours x 60 minutes = 1.440 data entries.
- The file has opening, high, low and closing values as well as the trading volume for that particular minute.
- Because the data set has one minute intervals, these four values are in many cases exactly the same. However during times with high value fluctuations, they will differ.
- The trading volume also varies alot. Note that there are long durations such as weekends where there is no trading at all, therefore the value remains constant. The data points for those periods remain there, but do they provide any informational value for all purposes?
- The trading data ends at the 6th of May, 15:58 (Turkish time) during a day of relative increase in the value of USD.

You are basically expected to

1. Estimate the value of USD **for the next 2 minutes** (15:59 and 16:00) and for the **next day (May 7th) at 16:00.**
2. Check for the stationarity of your data set, prior to making any estimation.

In particular, do the following

1. Try to work with the raw 1-minute interval data set first. (30 points))
    a. Check the stationaritiy for the entire period, and then for the data subset of May 6th only. When you display the plot with the moving average of mean and standard deviation, **save the plot as a PNG file.**
    b. Try to estimate value for 15:57 and 15:58 (May 6th) with different methods, measure and display the RMSE. Then try to estimate value for 15:59 and 16:00 (May 6th) with the method with the smallest RMSE, and display the estimate. Do the same estimation (with the same technique) for 16:00 (May 7th).
    c. Using Python code, discard the rows with no trading data (i.e. volume as 0) and then repeat a and b. (also saving the plot)
2. Then try to work with end of hour data only (i.e. at :59 and :00 minutes, or for :57 and :58). (25 points)
    a. Check the stationarity for this data set as well. **Save the plot for this approach as an PNG file as well.**
    b. Try to estimate value for 15:57 and 15:58 (May 6th) this way and measure RMSE.
    c. Again discard rows with no trading data and then repeat a, and b. (also saving the plot)
3. Now try to smooth the data by first discarding data with zero trading, and then applying (15 points for each method)

a. A **simple (one-sided) moving average**, using the rolling() method of the DataFrame. Use the window size as 60 (minutes per hour).
    - Note that based on your windows size, there will be 59 NaN's for the first few data points. Since you are estimating the future, this should not pose much problem. You can simply discard those NaN values.
b. A **linearly weighted moving average**, weights based on the trading volumes. Again use the window size as 60 (minutes per hour).
    - Sum of weights for the window should be exactly 1. Therefore a good strategy would be to first calculate the total volume for that hour (i.e. sum up all volumes for the 60 minutes) and use that volume to calculate weight for each minute using that sum.
    - You might like to define an additional function to calculate the weights based on trading volumes.
c. Taking **one random (representative) pick from each hour** and replacing all values in that hour by that value.
    - So if the random pick is 15:45 then all values from 15:00 to 15:59 should be replaced by the value at 15:45.
d. For each smooting technique,
    - Again check stationarity (save the plots),
    - Measure RMSE by trying to estimate 15:57, and 15:58,
    - Then estimate 15:59 and 16:00 at May 6th, and 16:00 at May 7th.

Write a report showing the plots(you should have 7 of them), and the RMSE values(you should have 7 sets for estimating 15:57 and 15:58). Discuss in particular:

1. Which method seems to have the best results?
2. The effect of the smoothing approaches on stationarity and RMSE.
3. Choosing a window size of 60 (minutes) to smooth using moving averages, would impose an artificial lag of 60 minutes on observing changes in trend. Knowing this, would you change your window size? Would your decision be different for both estimations (next 2 minutes, next day)?

Upload your report next to your code on Github.

Notes:

1. If you feel frustrated by coding all this in a single file, you can have multiple Python programs such as HW03-Part1.py, HW03-Part2.py, etc.
2. Defining functions for most tasks and reusing them will be very important on managing your code in this homework assignment.
3. The following blog posts might help you to understand how to process data:
    - https://towardsdatascience.com/implementing-moving-averages-in-python-1ad28e636f9d
    - https://becominghuman.ai/introduction-to-timeseries-analysis-using-python-numpy-only-3a7c980231af