Assignment

Data Science in Financial Markets

Total = 20 Marks: 10 Marks for Submission + 10 Marks for Viva

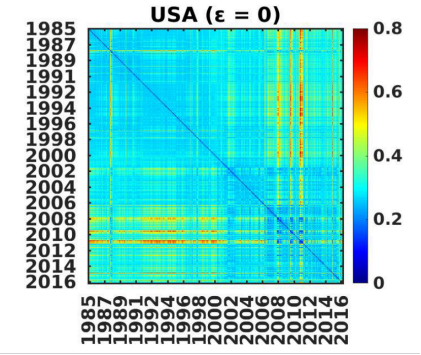
Question 1: Use the following stocks from the given list (Link: FTSE 250 market three stocks under the columns *Eth, Fth, and Gth*) and perform a detailed time series analysis. The objective is to demonstrate how prediction accuracy improves as we move from simple AR/MA models to more advanced ARCH and GARCH models.

Key points to cover:

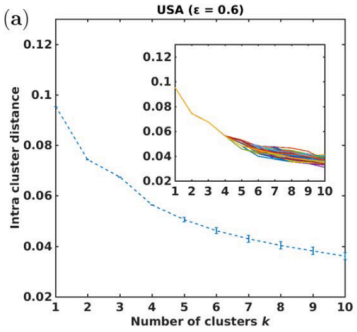
1. Show that raw stock prices are not suitable for prediction and explain why log returns are preferred.
2. Use ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots to identify appropriate values of p and q for ARMA models.
3. Compare the performance of AR, MA, ARMA with volatility models like ARCH and GARCH.
4. Conclude by highlighting how volatility clustering in financial data makes ARCH/GARCH more effective for modeling log returns.

Question 2:

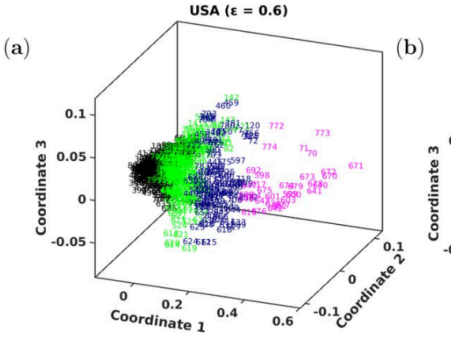
1. Download stock data from the **FTSE 250 markets for 10 years** based on the starting date mentioned (Dth column) in the following DSFM groups link FTSE 250 market.
2. Construct one CSV file of all the stocks for close prices and arrange them in sectors.
3. Plot all the closing time series in 3X3 subplots, i.e., 9 time series each, as discussed in class.
4. Remove all the stocks with more than **two** consecutive days' NAN values.
5. Plot all the log return-time series [$r(t)=lnP(t)-lnP(t-1)$] in 3X3 subplots as discussed in class.
6. Remove all the stocks with $|r(t)| > 0.8$.
7. Construct a correlation matrix for the full-time horizon of 10 years using log returns.
8. Construct correlation matrices using $r(t)$ of epoch size 20 working days and with a shift of 10 days.
9. Construct a similarity matrix $S(t1,t2)= <| C(t1) - C(t2) |>$.
10. Construct a 3D Multidimensional scaling plot using the Similarity matrix “S”.



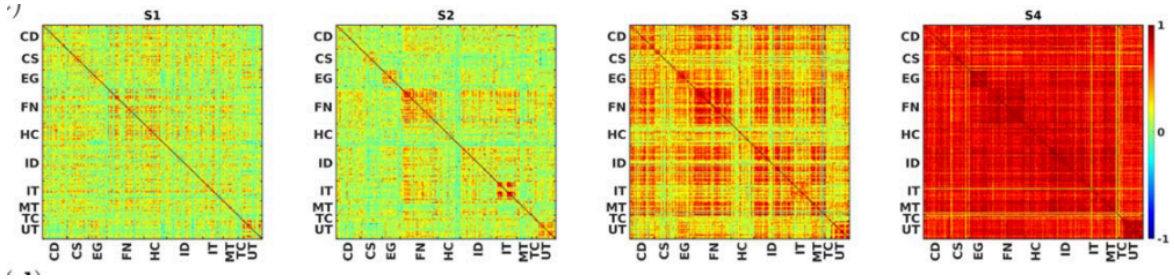
11. Find an optimum number of clusters using 1000 different conditions for the elbow method of *k*-means clustering. eg.



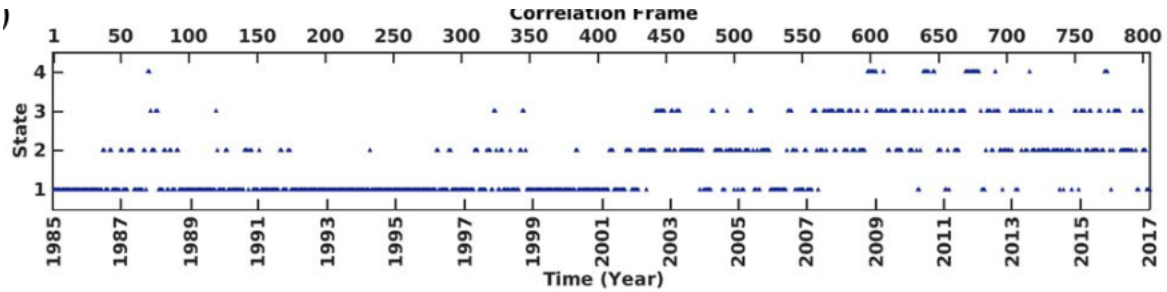
12. Plot *k*-means clustering results with optimum value of clustering found in No. 11 (above) eg.



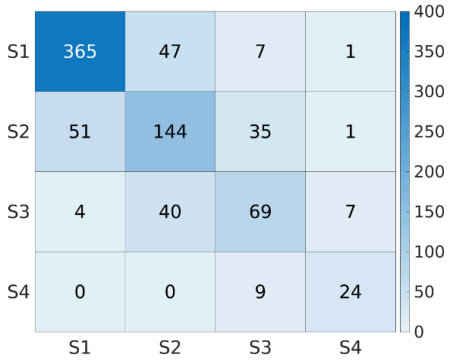
13. Plot typical correlation structures of different states, eg.



14. Arrange all the correlation matrices for the market states (y-axis) and time (x-axis), eg.



15. Show the transition of consecutive market states, eg.



16. Upload the Python code with the assignment and mention the total number of lines of the code.

Note: Ensure that all x-labels and y-labels must be the same as those given in the figures.