# Khushboo Tekchandani

# CSE587 Homework 2

**Solution:**

This assignment requires us to forecast the stock prices of given set of stocks using various forecast models. There were three different models used to solve this problem, Arima, Linear Regression and Holt-Winters. The solution further calculates the Mean Absolute Error (MAE), which is popular method for statistical analysis. The MAE gives an idea of the difference between the forecasted data and the actual data.

$$MAE_i\ (each\ day) = |forecastData_i - testData_i|$$

$$sum\ of\ MAE = \sum_{i=1}^{10} MAE_i$$

To understand the variation of MAE, the solution calculates the sum of MAE for each stock. And for each model, it plots the top ten stocks with the minimum sum of MAE. These are the stocks which are closest to the forecast data.

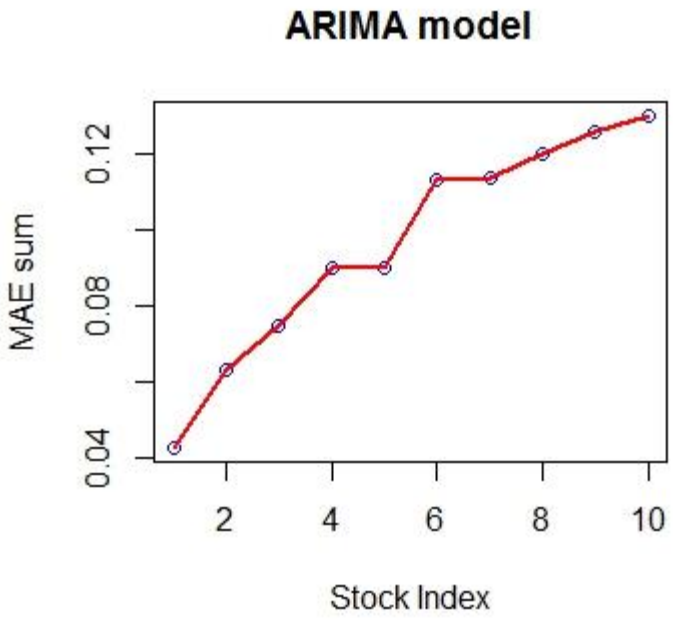
This code was built on the basis of the example provided by the TA.

**Results:**

The solution gave the following results:

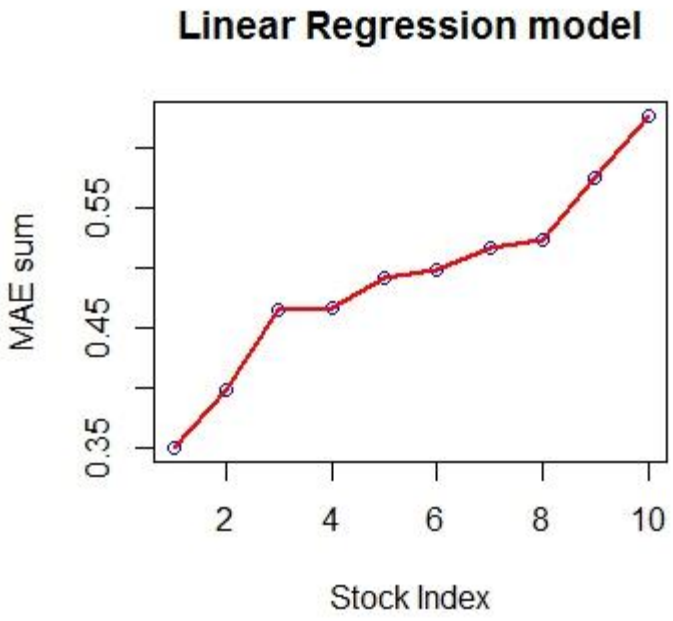
1. ARIMA model

| | sumMAE | stock |
|---|---|---|
| 469 | 0.04291029 | COCO.csv |
| 141 | 0.06308866 | APWC.csv |
| 822 | 0.07480337 | FREE.csv |
| 1043 | 0.09000000 | IKAN.csv |
| 1924 | 0.09000000 | SPU.csv |
| 675 | 0.11315610 | ELON.csv |
| 2161 | 0.11343679 | VLYWW.csv |
| 1283 | 0.12000000 | MFI.csv |
| 695 | 0.12096462 | ENZN.csv |
| 1356 | 0.12583623 | MTSL.csv |



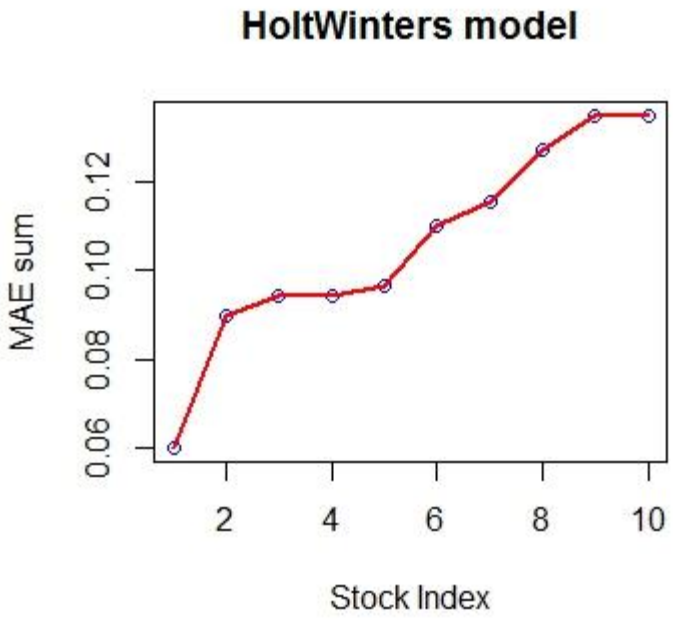

2. Linear Regression Model

| | sumMAE | stock |
|---|---|---|
| 1945 | 0.3500860 | STB.csv |
| 1770 | 0.3976781 | RVSB.csv |
| 871 | 0.4642998 | GIGA.csv |
| 220 | 0.4666093 | BAMM.csv |
| 491 | 0.4919287 | CPRX.csv |
| 2028 | 0.4976290 | TISA.csv |
| 1190 | 0.5170025 | LIOX.csv |
| 2026 | 0.5224693 | TINY.csv |
| 1626 | 0.5744103 | PRTS.csv |
| 328 | 0.6262162 | BYFC.csv |




3. Holt Winters Model

| | sumMAE | stock |
|---|---|---|
| 659 | 0.06022709 | EDS.csv |
| 2161 | 0.09000000 | VLYWW.csv |
| 1043 | 0.09451631 | IKAN.csv |
| 1128 | 0.09452480 | JOEZ.csv |
| 141 | 0.09639256 | APWC.csv |
| 1356 | 0.11008672 | MTSL.csv |
| 469 | 0.11565898 | COCO.csv |
| 974 | 0.12703413 | HNSN.csv |
| 2026 | 0.13458633 | TINY.csv |
| 1009 | 0.13481835 | IBCA.csv |




**Observations:**

1. The plots in increasing order of the best performing stocks in each model. These plots indicate the stocks, the forecast for which was the closest to the original data in an increasing fashion.
2. The ARIMA model is an extremely slow learning model as compared to the other two models. It takes an immensely long time to run. However, changing few parameters like fixing seasonal=FALSE, lambda=NULL, approximation=TRUE can hack it to work faster.
3. The values of sum of MAE for the ARIMA model ranges from 0.04 to 0.14, that for Holtwinters model ranges from 0.06 to 0.14 but for the Linear Regression model it ranges from 0.35 to 0.65. The Linear Regression Model has the highest range. So once can assume that MAE for each day would be larger for the stocks. Thus, the error for Linear Regression model would be the maximum of the three models.
4. ARIMA gives a tighter MAE to the forecasted data but the trade-off is that it has long execution hours. Depending on the requirements, if one can deal with a looser bound on the co-relation, Holt Winters is a good model to use and takes shorter time to execute.