Clustering and Statistical Modeling of Multivariate Financial Time Series

Data Science for Energy Transition

Hawa Desai, Navya Bagireddy, Nagendra Palagulla, Kartheek Reddy

University of Houston - Victoria, University of Houston

Dr. Kim

Professor of Statistics,

Sam Houston State University

U.S. National Science Foundation

Introduction

S&P 500 - what is it?

- S&P 500, or Standard and Poor's 500, is "market-capitalization-weighted index of 500 leading publicly traded companies in the U.S." (INVESTOPEDIA)
- Essentially it follows 500 of the performance of the largest publicly traded companies, all from a variety of sectors
- The more market capitalizations a company has the more impact it has on the index

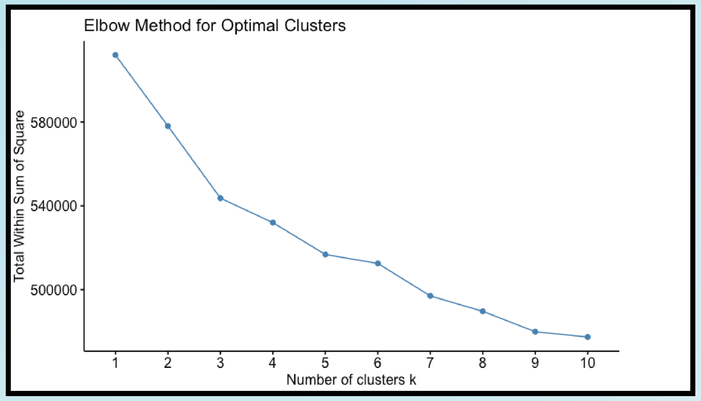
Objectives

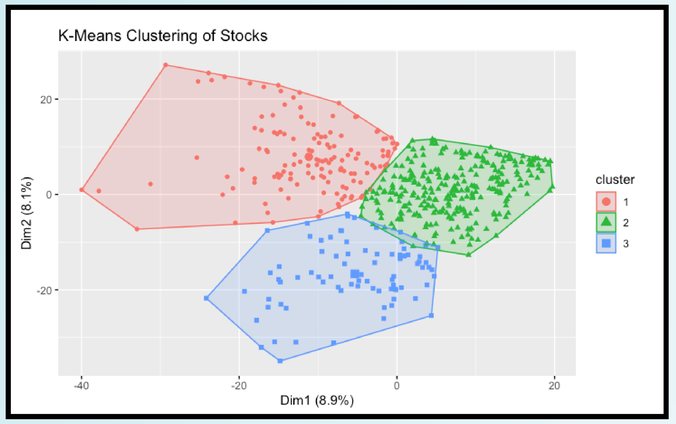
Our objective was to implement time series analysis, by using the closing prices of each company ticker in the S&P 500 Index over the course of 5 years, cluster the companies using K-means clustering techique, and develop a VAR model that would forecast future cloing prices for companies within the S&P 500

Methodology

Time Series Analysis Using K-means Algorithm

The optimal number of clusters is 3, based on within cluster SS





The elbow method for K-means clustering involved taking the rate of return and the log return of each ticker's closing price

- **Input the log return and data associated with it into a matrix**
- **Scaled the data accordingly**
- **Used R clustering function to build an elbow graph to determine the optimal clusters**

VAR(1) Model

$$
\begin{aligned}
\mathrm{AAPL}_t &= -0.0936\,\mathrm{AAPL}_{t-1} - 0.0464\,\mathrm{DELL}_{t-1} + 0.0070\,\mathrm{TSLA}_{t-1} \\
&\quad +0.0015 - 8.384 \times 10^{-7} \cdot t + \varepsilon_{1t} \\
\mathrm{DELL}_t &= -0.1185\,\mathrm{AAPL}_{t-1} + 0.0453\,\mathrm{DELL}_{t-1} - 0.0063\,\mathrm{TSLA}_{t-1} \\
&\quad +0.0013 - 1.529 \times 10^{-7} \cdot t + \varepsilon_{2t} \\
\mathrm{TSLA}_t &= -0.0918\,\mathrm{AAPL}_{t-1} + 0.0519\,\mathrm{DELL}_{t-1} - 0.0039\,\mathrm{TSLA}_{t-1} \\
&\quad +0.0034 - 2.838 \times 10^{-6} \cdot t + \varepsilon_{3t}
\end{aligned}
$$

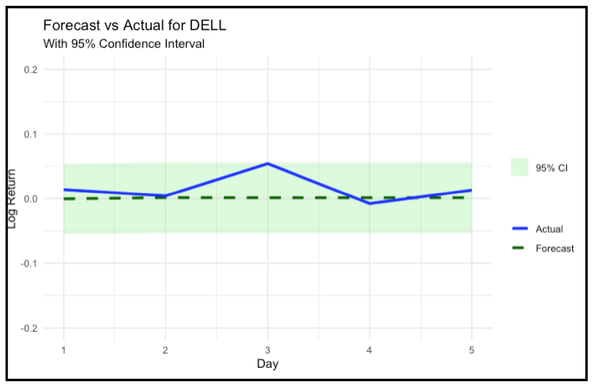
Built a VAR model for three companies within the third cluster

We chose three large companies - Dell, Tesla and Apple - to build a VAR model on and to forecast future closing price data

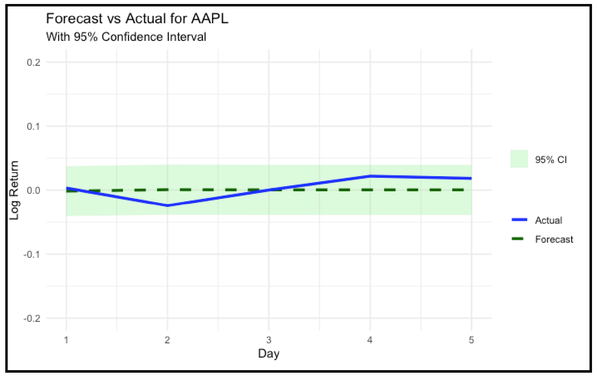
Forecasting

***expected line is straight due to forecast values being very close to 0**

DELL



APPLE



TESLA