Predicting Volatility in Equity Markets Using Macroeconomic News

Technocolabs project by; Anefu P. Oche

INTRODUCTION

In this project, we investigate how macroeconomic sentiment immediately impacts the volatility in liquid markets, by measuring market volatility through the VIX (volatility index of the S&P 500). Using data pulled from Twitter, we are researching how breaking economic news affects the markets, and subsequently how to predict which news stories can increase volatility. We will consider a tweet 'significant', in that the news presented in the tweet contributes to volatility in the market, if within 30 minutes of the tweet being tweeted, the volatility of the asset increases by one-fifth of a standard deviation.

BUSINESS PROBLEM

On June 26, 2015, months of debt negotiation between the Greek government, headed by Prime Minister Alexis Tsipras, and its creditors, including the IMF and fellow Eurozone countries, broke off abruptly. Tsipras announced a snap referendum regarding the terms of the pending bailout. By the following morning, S&P500 was down significantly. As political and economic uncertainty grew in Europe, investors across the world were moving their funds away from risky assets that could be negatively affected by a Greek default, and towards safer assets. Days later, when the Greek situation had 'resolved', the process reversed as equities rallied.

These market movements are not uncommon; just weeks later, crisis erupted in the Chinese equity markets, and asset values were responding to the uncertainty this provoked.

DATASET AND FEATURES

1. Twitter Data; features include;
   - Date: date of the tweet
   - Number of positive tweets
   - Number of negative tweets
   - Number of neutral tweets
   - Total number of tweets.

2. Financial Data; features include;
   - Date: The period of the stock
   - High: maximum price in a given time
   - Close: price at which stock end in a given period
   - Open: price at which stock starts in a given period
   - Low: minimum price in a given time

METHODOLOGY

Supervised learning approaches will be employed;

- Naive bayes
- Support vector machine (SVM)
- Logistic regression and principal component analysis (PCA)

PERFORMANCE METRICS AND RESULTS

- Compare the three Regression models
- Use hold-out cross validation (70% for training, 30% for testing)
- Compare the accuracy, precision and recall of each model

CONCLUSION

- Accuracy of the three supervised learning techniques falls between 57% - 67%.
- Logistic regression is the best fit model.
- Accuracy and precision can be increased by eliminating non macroeconomic headlines, modifying and shortening the dictionary etc.

REFERENCE

- https://github.com/AnefuIII/Predicting-Volatility-in-Equity-Markets-Using-Macroeconomic-News/blob/main/PREDICTING%20MARKET%20VOLATILITY%20USING%20MACRO%20HEADLINES-Paper.pdf