Topic: Impact of sentiment on LETF derivatives

Introduction & Background

Social medias are increasingly reflecting and influencing behavior of other complex system. The recent decade of technological revolution with widespread presence of computers and internet has created an unprecedented situation of data deluge, changing dramatically the way in which we look at social and economic science. Among the many fields of applications of data collection, analysis and modeling, we present here a research between social media and the financial system, such as the text on twitter or some other social media platform can cause effect on the financial market such as stock price. And for some popular social media such as Twitter or Facebook, and others like BBC news business sector or WSJ, they can also be used for the financial forecasting.

In our project, one line of research investigates the relation between the volume of tweets and financial markets, for example, studied whether the daily number of tweets predicts the S&P 500 stock indicators. And another line of research explores the contents of tweets or text data online. In a textual analysis approach to Twitter data, and other text data on other social medias, use sentiment analysis to figure out their correlation in financial market. For example, a recent study compares the information content of the Twitter sentiment and volume in terms of their influence on future stock prices. So we are going to study the relation between stock price returns and the sentiment expressed in financial tweets posted on Twitter. We will collect the 30 DJIA companies, and for each of them, we will build a time series of the sentiment expressed in the tweets, with daily resolution.

Data Source & Tools

For this time, we are going to collect our data mainly on Twitter, news websites like BBC, WSJ, and Bloomberg, WRDS. And we are going to use the Python packages such as “Requests” and “Beautiful Soup” to grab data on the websites, and also use the natural language process to deal with the positive or negative word. We will

also use the Twitter API to finish the text mining on Twitter. And finish the data cleaning and preprocessing, and use the time series to analyze the results. And use R to build the time series and fit the models.

Motivation & Insight

Despite the high quality of the datasets used, the level of empirical correlation between stock price derived financial time series and web derived time series remains limited, especially when a textual analysis of web messages is applied. So there supposed to be some complex factors to affect our results. Maybe our model is not fit enough to analysis a dynamic system. Therefore, it possible that the two systems dependent only at some moments of their evolution, and not over the entire time period.