Research Proposal: Understanding The Relationship Between Financial News and Stock Prices

Omair Shafi Ahmed {shafiahmed.o@northeastern.edu}

Summary

It is a well accepted fact that financial markets react to news stories and vice versa. However, the relationship between the two has been hard to quantify, empirically. This research proposes that we use financial news from Reuters and Bloomberg from October 2006 to November 2013, released by Ding et al. [2014][1], as well as market data from Kaggle[2] to unearth relationships between the two. The data released by Ding et al., contains 450,341 news from Bloomberg and 109,110 news from Reuters, enough a sample to create a general market sentiment. Specifically, this proposal intends to draw on previous methodology as described by 'News and narratives in Financial systems: exploiting big data for systemic risk assessment'[3]. The measures of sentiment have been derived from pre-defined common English language word lists representing two specific emotional groups capturing approach or avoidance. The words were selected through the lens of a social-psychological theory of "conviction narratives" (CNT) (Tuckett and Nikolic, 2017). The CNT theory describes a social context in which agents are influenced by other agents and the stock of narratives circulating around them. Approach and avoidance being the two basic groups of emotions interact with cognition to determine conviction. The key variables for us will be the anxiety for potential losses and excitement for potential gain. Assuming that the markets are not as efficient, and are likely to fall prey to irrational exuberance[4], collectively, this method hopes to provide us with a systematic way of modeling this 'irrationality', thereby providing us with another facet of financial market analysis.

Proposed Plan of Research

The research will constitute preprocessing the dataset obtained from Ding et al. [2014], to quantify the anxiety and for potential losses and excitement for potential gains. Having done that, we would use the new quantified dataset and to perform regression analysis on the financial market data and see if any of our models fit the data. As is standard, we would iteratively move to more complicated models and evaluate the performance of each model. We would start off with ETF's tracking indicators such as Dow Jones Industrial Average and NASDAQ. We will then, make our analysis more granular and perform regression analysis on each individual stock and build a performance table. This should allow us to understand the relationship between various stocks and the CNT.

References

1. [Ding et al., 2014] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Using structured events to predict stock price movement: An empirical investigation. In Proc. of EMNLP, pages 1415–1425, Doha, Qatar, October 2014. Association for Computational Linguistics.

2. Boris Marjanovic. https://www.kaggle.-com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs. Nov. 2017

3. Rickard Nyman, David Gregory, Sujit Kapadia, Paul Ormerod , David Tuckett &Robert Smith. News and narratives in financial systems: Exploiting big data for systemic risk assessment1. September 2016

4. Shiller, R (2000). Irrational Exuberance. Princeton University Press.