12-7-2022

Social Media Sentiment and Stock Return: A Signalling Theory Explanation for Application of the Natural Langrage Processing Approaches

Chuan Qin
University of Newcastle, chuanqinau@gmail.com

Shah J Miah
University of Newcastle, shah.miah@newcastle.edu.au

David Shao
University of Newcastle, david.shao@newcastle.edu.au

Social Media Sentiment and Stock Return: A Signaling Theory Explanation for Application of the Natural Langrage Processing Approaches

Research-in-progress

Chuan Qin
Newcastle Business School
The University of Newcastle
Callaghan, NSW
Email: chuanqinau@gmail.com

Dr. Shah J Miah
Newcastle Business School
The University of Newcastle
Callaghan, NSW
Email: Shah.Miah@newcastle.edu.au

Dr. David Xuefeng Shao
Newcastle Business School
The University of Newcastle
Callaghan, NSW
Email: david.shao@newcastle.edu.au

Abstract

Social media, especially microblogs, have potentials to develop unavoidable factors in investment decision-making, because of its use for capturing human sentiment. In this paper, by applying signaling theory and Natural Language Processing (NLP) technique, we concern social media sentiment as a signal to stock return which is based on human the sentiment, which may lead to price fluctuation in the market. We take the strength of signal into consideration, introducing the sentiment of traditional media to compare with social media sentiment in different industry. The empirical result of this paper will prove the relationship between social media sentiment and stock return. It will also reflect on analyzing the changes of stock price given different strength of signals in both positive and negative way. The entire study will be viewed as a guideline for investors to filter and smartly use the huge numbers of information when making investment decision.

Keywords: Social media sentiment, Natural Language Processing, Signaling theory, Stock return analysis

Social Media Sentiment and Stock Return: A Signaling Theory Explanation for Application of the Natural Langrage Processing Approaches

1 Introduction

Investors make judgment based on not only company's historical performance but also their knowledge, intuition, and different signals of varying strength that indicate the company's economic prospects (Spence 1973; Stuart et al. 1999). Many of the core concepts and constructs of signaling theory grew out of the finance and economics literatures (Riley 2001). A considerable body of research has investigated how single and strong signals influence investor decisions. Those signals include team characteristics, intellectual property, and alliance partnerships (Connelly et al. 2011), et cetera. For instance, firm debt (Ross 1973) and dividends (Bhattacharya 1979) represent signals of firm quality. Only high-quality firms can make interest and dividend payments over the long term. In contrast, low-quality firms will not be able to sustain such payments. Consequently, such signals influence outside observers' (e.g., lenders, investors) perceptions of firm quality (Brian et al. 2011). Whereas for many years finance and economics research in signaling theory has mainly focused on signals that are argued to be strong. Few research has argued that signals are not uniformly strong but rather exist on a continuum of varying strengths (Vanacker et al. 2019). In recent years, signals from social media sentiment plays an unignorable role in decision making of financial market. For example, a novel social networks sentiment analysis model, which has the accuracy of 97.87%, is proposed based on Twitter sentiment score (TSS) for real-time prediction of the future stock market price without the knowledge of historical data (Guo et al. 2019). Social media sentiment also benefits to measure the financial risks. When social media sentiment indicates low liquidity, investors try to reduce the negotiation volume, which has a positive impact on risk (Paraboni et al. 2018). It is also important for firms to differentiate and leverage the unique impact of various sources of media outlets in implementing their social media marketing strategies (Yu et al. 2013). Opinion mining of social media especially microblog messages has become a popular application of business analytics. Despite of casual evidence of investors' use of social media sentiment, few research demonstrate whether and to what extent investors incorporate such signals in their decision-making.

In this article, we will extend the socio-cognitive perspective of signaling theory which the perspective was published by Rindova (2012). Social media sentiment, therefore, represents a weak and indirect signal on stock return in financial secondary market. Signals from social media sentiment about company's operation or economic prospects could be relatively weak signals for investment decisions. Regarding the effect of microblog sentiment on return, most previous studies did not address the difference between positive sentiment and negative sentiment at the same time in one model. The different roles of positive emotions and negative emotions in influencing human behaviour have been comprehensively discussed in the IS literature (Zhang 2013) and supported by empirical findings (Luo et al. 2013). Extending this perspective, we will analyse the effects of social media sentiment in both direction-positive effects and negative effects. We will also address the void in the literature and contributes of signaling theory to investigate multiple strong and weak signals in one model and their consequence on stock market valuation, introducing social media sentiment as a weak signal as well as traditional news as a strong signal when analysing stock return of listed companies.

Stock valuation by NLP technology is the pioneer method in the research field of finance compared to the traditional methods such as Discounted Cash Flow Model (DCF), Dividend Discount Model (DDM), mathematical modelling et cetera. Discounted cash flow (DCF) is a valuation method that estimates the value of an investment using its expected future cash flows. It attempts to determine the value of an investment today, based on projections of how much money that investment will generate in the future. Dividend Discount Model (DDM) is a quantitative method used for predicting the price of a company's stock based on the theory that its present-day price is worth the sum of all of its future dividend payments when discounted back to their present value. (Gordon 1959) Our article combines NLP technology with the signaling theory, empirically proves the existence of relationship of social media sentiment and stock price.

Unlike existing studies mostly focus on either optimistic or pessimistic sentiment, in this article, social media sentiment is quantified to a range of numbers from negative to positive, which means the sentiment gives different direction of signals that impact on the market. It extends the edge of the application of NLP in finance as well as bring NLP technology into management research.

2 Theoretical Background and Hypothesises

The theory we propose to use in this article is signaling theory, which exams the communication between different factors given certain situation. We set sentiment from social media and traditional media as the signal to predict the change of stock price. We also assume that different level of signal strength from media sentiment will lead to different result of the change of stock price.

Prior research of signaling theory on market valuation investigates that the positive effect of microblog sentiment on listed company is positive correlation to its stock return. Meanwhile, the negative effect of microblog sentiment on listed company is negative correlation to its stock return. In this article, what we want to address is to compare the reaction of market given different direction of two signals, positive sentiment and negative sentiment in one model. If stock return fractures bigger under negative sentiment, which means stock return is more sensitive to negative sentiment, the negative sentiment will be a stronger signal for investors compare to positive sentiment on microblog. Negative sentiment on microblogs may have a larger impact on stock return than positive sentiment.

Hypothesis 1. *The negative effect of microblog sentiment on stock return is stronger than the positive effect of stock return. Stock return is more sensitive to negative sentiment.*

Traditional news used to be the only channel for investors to extract useful investment hints to reduce information asymmetry in past times. Sentiment reflected in news can be considered as a proxy for the fundamental values of a company's prospect that has not been incorporated into prices (Tetlock 2007). It is a strong signal on stock return in investment decision-making process because investors trust the authenticity of official news agency. In recent years, social media is growing fast in financial market. The most used social media are microblog messages (e.g., Twitter, Facebook, Weibo, StockTwits) because microblog is high volume, high velocity, and quick reaction in real-time. Compared to the restriction of republication in news agency, microblog is more like a communication tool that everyone could post their opinion to the public. As a result, investors have to manually filter the information from microblog before investment analysis. It makes the effect of microblog sentiment weaker than news, concerning the effect is positive effect.

Hypothesis 2. *The positive effect of microblog sentiment on stock return is weaker than the positive effect of news on stock return.*

3 Research Methodologies

The data pool in this article is from Stocktwits.com, which is the largest social network for investors and traders focusing on financial market discussion. It shares over 200,000 messages per day. We will use NLP technology to filter and refine those high volume of messages in the time frame of 10 years from selected companies. The application of NLP technology has developed over 50 years, which is widely used in the research areas of economics, finance, business administration, international business and sustainability. Using NLP technology, the communication data can be categorized by venue, daily activities (Miah et al. 2022), company name (Zwilling 2022), company logo (Srivastava et al. 2022), industry (Andrawos et at 2022).

The methodology we propose to use is Ordinary Least Squares (OLS) regression as well as multivariate statistical analysis by applying either SAS or Stata. One technique can be used as first-time calculation and the other technique can be used as second-time test. The regression results will be discussed in detail about the relationship of all variables.

4 Measurements and Data

4.1 Independent Variables

4.1.1 StockTwits sentiment

Sentiment analysis (or opinion mining) is a natural langrage processing (NLP) technique used to determine whether data is positive, negative, or neutral. There are three domain methods for sentiment analysis which are SentiStrength, VADER (Valence Aware Dictionary and sEntiment Reasoner) and Harvard General Inquirer word list (GI). While some previous studies manually classified sentiment (Aggarwal et al. 2012b), it is more practical to use a computational method for this task given the volume of the data (Luo et al. 2013). The method we propose to use is SentiStrength. It is a sentiment analysis (opinion mining) program, which can automatically analyze sentiment up to 16,000 social web texts per second with up to human level accuracy for English. SentiStrength estimates the strength of positive and negative sentiment in short texts, even for informal language. It reports two sentiment strengths:

-1(not negative) to -5(extremely negative)

1(not positive) to 5(extremely positive)

SentiStrength can also report binary (positive/negative), trinary (positive/negative/neutral) and single scale (-4 to +4) results.

4.1.2 Traditional news

For the news data, we propose to download all AP news articles from the news archives of apnews.com from June 2012 to June 2022, which discussed about the listed company mentioned on StockTwits. News sentiment will be extracted by SentiStrength or Harvard General Inquirer word list (GI). GI is primarily preferred because this method provides high accuracy in formal documents especially in classify the tone in financial reports and news (Garcia 2013; Tetlock 2007). A recent study benchmarks 20 popular Twitter sentiment analysis tools and shows that SentiStrength is among the best in terms of classification accuracy of the language styles used in microblogs (Abbasi et al. 2014). Given the similarity between StockTwits and Twitter, SentiStrength is well-suited for the sentiment classification task in terms of microblog sentiment.

4.2 Dependent Variables - Stock Return

One of the first researchers to predict the effect of microblog sentiment on stock return is Bollen et al. (2011). They found that Twitter sentiment predicts Dow Jones Industrial Average (DJIA) return at a directional accuracy of up to 86.7% which has inspired the institutional traders and analytics professionals (Bollen et al. 2011). S&P 500 will be the primary option for this study taking the sample size of stock return into consideration because it will be a larger and more comprehensive data set spanning multiple years compared to DJIA which has only 30 listed companies. Stock return is the daily closing prices on DJIA or S&P500.

4.3 Moderate Variables - Industry

We propose to use industry category as one of the moderate variables in this study. The industry in DJIA or S&P 500 includes high-tech, traditional energy, biomedicine, et cetera. High-tech companies are more likely to share information on public because they have high motivation to explore their new products and draw public attention especially potential investors. Marketing strategy on social media is an efficient way to reach as many as people in a short period. Our assumption is:

The positive effect of microblog on stock return of high-tech company is stronger than that of traditional energy company.

4.4 Data

Microblog dataset consists of all messages on StockTwits.com between June 2012 and June 2022. StockTwits is a microblog website for investors and investment professionals to share information and ideas about financial markets (StockTwits 2014). Similar to Twitter messages, each posting on StockTwits is limited to 140 characters. Compared to Twitter messages, StockTwits are closely related to the stock market and contain much less noise. In StockTwits, listed company is symbolized by Cashtag which can accurately allocate a message to its stock.

Traditional news dataset consists of all news article according to the listed company mentioned on StockTwits from the same period, June 2012 to June 2022, on apnews.com. The Associated Press (AP) is an American non-profit news agency headquartered in New York City, founded in 1846, which operates 248 news bureaus in 99 countries. Its authority allows to source adequate news articles related to listed company mentioned on StockTwits backtracking decades ago.

Dow Jones Industrial Average (DJIA) is commonly used in prior research as stock returns standard. It is a price-weighted measurement stock market index of 30 prominent companies listed on stock exchanges in the United States. The dataset could be all daily closing prices on DJIA of Cashtaged companies on StockTwits from June2012 to June 2022. On the other hand, since the empirical study has not been finished yet, if taking the sample size of stock return dataset into consideration, Standard and Poor's 500 (S&P 500) is also an option when choosing stock return dataset because it is a stock market index tracking the performance of 500 large companies listed on stock exchanges in the United States.

The process of data collection is still at primary state. While collecting data from StockTwits, It is also necessary in next empirical chapter to balance the sample sizes between time frame and stock return.

5 Discussion and Next Step

The first contribution of this article is to empirically prove the existence of the relationship between microblog and stock return. Inconsistent findings have been reported in precious studies of signaling theory in finance and economics. For instance, using econometric models on a panel data set over a 3-month period, neither positive nor negative sentiment on Twitter significantly predicts the return of individual stocks (Yu et al. 2013). Yu also investigated that the effect of forum sentiment is significant. This finding is contradictory to Das and Chen (2007), who did not find the relationship between forum sentiment and stock return to be significant. Oliveira et al. (2013) developed a predictive model using StockTwits sentiment and compared it with traditional equity models that use fundamental and technical indicators in forecasting individual stock returns. They found no evidence to support the predictive power of StockTwits sentiment. Sprenger et al. (2014) found that sentiment in Twitter messages has a contemporaneous association, but not a predictive relationship, with individual stock return.

It is also to analyse the reaction of stock return given different strength of signal, which combines stock valuation in finance and signaling theory in management. By verifying the above hypothesises with followed regression results, this will be one of the first researches to empirically investigate that microblog sentiment could affect stock return in both ways, positive and negative in one model. This article also extends signaling theory by investigating multiple signals of different strength. Signal from microblog sentiment is generally weaker than the traditional strong signal such as news. This article will also contribute to finance literature by introducing a new factor, industry category as a moderate factor. The strength of signal of microblog sentiment changes in different industry. Advanced technology listed company is more sensitive to microblog sentiment compared to traditional energy listed company (to be proved). In terms of practical implications, this article will be a guide for investors to balance and filter information in different situation in investment decision-making process.

We analyze the reaction of stock return given both positive and negative effect of social media especially microblog sentiment in certain time frame. Microblog sentiment is treated as exogenous when using it to predict stock return. To extend this study, in future research, the endogeneity of microblog sentiment needs to be investigated because microblog users post not only their opinions about the future, but also their response to past social and economic activities, such as historical stock returns. To extend our study, those social media data that talked about historical information of stock return will be avoided in the model to reduce the bias from past to future stock price. We need to carefully pay attention to the relationship between microblog sentiment and stock return when collecting data because those two could be double directed.

References

Abbasi, A., Hassan, A., and Dhar, M. 2014. "Benchmarking Twitter Sentiment Analysis Tools," Proceedings of the Ninth International Conference on Language Resources and Evaluation, Reykjavik, Iceland.

Aggarwal, R., Gopal, R., Gupta, A., and Singh, H. 2012a. "Putting Money Where the Mouths Are: The Relation Between Venture Financing and Electronic Word-of-Mouth," Information Systems Research (23:3-Part-2), pp. 976-992.

Aggarwal, R., Gopal, R., Sankaranarayanan, R., and Singh, P. V. 2012b. "Blog, Blogger, and the Firm: Can Negative Employee Posts Lead to Positive Outcomes?" Information Systems Research (23:2), pp. 306-322.

Bhattacharya, S. 1979. An exploration of nondissipative dividend-signaling structures. Journal of Financial and Quantitative Analysis, 14: 667-668.

Bollen, J., Mao, H., and Zeng, X. 2011. "Twitter Mood Predicts the Stock Market," Journal of Computational Science (2:1), pp. 1-8.

Das, S. R., and Chen, M. Y. 2007. "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," Management Science (53:9), pp. 1375-1388.

Garcia, D. 2013. "Sentiment during Recessions," The Journal of Finance (68:3), pp. 1267-1300.

Gordon, Myron J. (1959). "Dividends, Earnings and Stock Prices". Review of Economics and Statistics. The MIT Press. 41 (2): 99–105.

Guo, X., Li J. 2019. "A Novel Twetter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency," Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS).

Luo, X., and Zhang, J. 2013. "How Do Consumer Buzz and Traffic in Social Media Marketing Predict the Value of the Firm?," Journal of Management Information Systems (30:2), pp. 213-238.

Luo, X., Zhang, J., and Duan, W. 2013. "Social Media and Firm Equity Value," Information Systems Research (24:1), pp. 146-163.

Oliveira, N., Cortez, P., and Areal, N. 2013. "On the Predictability of Stock Market Behavior Using StockTwits Sentiment and Posting Volume," in Progress in Artificial Intelligence. Springer, pp. 355-365.

Paraboni, A.L., Righi, M.B., Vieira, K. M., Silveira, V. G. 2018. "The Relationship between Sentiment and Risk in Financial Markets," in Brazilian Administration Review (15:1)).

Riley, J. C. 2001. Silver signals: Twenty-five years of screening and signaling. Journal of Economic Literature, 39: 432-478.

Rindova, V.P., Reger, R.K., Dalpiaz, E., 2012. The mind of the strategist and the eye of the beholder: the socio- cognitive perspective in strategy research. In: Dagnino, B.G. (Ed.), Handbook of research on competitive strategy, pp. 147–164.

Ross, S. 1973. The economic theory of agency: The principal's problem. American Economic Review, 63: 134-139.

Spence, M. 1973. "Job market signaling." Quarterly Journal of Economics, 87: pp, 355-374.

Sprenger, T. O., Tumasjan, A., Sandner, P. G., and Welpe, I. M. 2014. "Tweets and Trades: the Information Content of Stock Microblogs," European Financial Management (20:5), pp. 926-957.

Srivastava, S., Varma, O., Gayathri. M. 2022. “Electronic Invoicing Using Image Processing and NLP,” in Innovative Data Communication Technologies and Application (96), pp.843-856

Stuart, T.E., Hoang, H., Hybels, R.C., 1999. Interorganizational endorsements and the performance of entrepreneurial ventures. Adm. Sci. Q. 44 (2), 315–349.

Tetlock, P. C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," The Journal of Finance (62:3), pp. 1139-1168.

Vanacker, T.R., Forbes, D.P., Knockaert, M., Manigart, S., 2019. Signal strength, media attention, and resource mobilization: evidence from new private equity firms. Academy of Management Journal, Advanced online publication. https://doi.org/10.5465/amj.2018.0356.

Yu, Y., Duan, W., and Cao, Q. 2013. "The Impact of Social and Conventional Media on Firm Equity Value: A Sentiment Analysis Approach," Decision Support Systems (55:4), pp. 919- 926.

Yu, Y., Duan, W., Cao Q. 2013. “The Impact of Social and Conventional Media on Firm Equity Value: A Sentiment Analysis Approach,” in Decision Support Systems (55:4), pp.919-926.

Zhang, P. 2013. "The Affective Response Model: a Theoretical Framework of Affective Concepts and their Relationships in the ICT Context," MIS Quarterly (37:1), pp. 247-274.

Zwilling. M. 2022. “Big Data Challenges in Social Sciences: An NLP Analysis,” in Journal of Computer Information Systems.