The Role of Online Media in Shaping Investor Sentiment and Market Prices

Pengyu Shen

School of Mathematical Sciences, Anhui University, No. 111 Kowloon Road, Hefei, China
spy1222@ldy.edu.rs

Abstract. This paper summarizes the body of research on how social media affects investor mood and, in turn, stock market pricing, and the methods and models used to study and comparatively evaluate them. The methods and models can be applied to the analysis of other financial market price influences. This paper adopts horizontal and vertical comparison methods to analyze the better models and data processing methods in different scenarios, and at the same time, changes the parameters of the models and higher-order deformation of mathematical formulas in the methods, as well as arranging and deleting them in an organic way. The impact of Internet social media, traditional media on investor confidence and the correlation between market sentiment and share market returns, as well as the mutual exclusion of correlations under different stock market heats are found. The better raw data processing methods as well as regression and autoregression methods for specific scenarios are identified. It is of reference and practical value to generalize the related models and methods to other financial scenarios.

Keywords: Social media, Investor sentiment, Regression analysis, Stock prices.

1 Introduction

The financial market is no longer a simple peer-to-peer transaction, and elements related to financial transactions in various industries in society will also participate in and influence financial behavior. Traditional data such as stock prices, trading volumes, and financial statements are not comprehensive enough for the analysis of modern financial markets and economic trends, and alternative data such as social media, geolocation information, website search trends, and other new types of network footprints can be processed by artificial intelligence and other technologies to better help financial participants understand the market situation and make favorable decisions. At the same time, the concept of alternative data has also been widely studied by major credit reporting industries and scholars, among which the topics explored by researchers are also different for the different stages of impact communication. In traditional and emerging Internet social media, some scholars focus on the relevant behaviors of netizens in forums and official or unofficial articles to analyze the influencing factors of investor sentiment. Because sentiment data is directly related to financial market participants, in behavioral psychology, stock

J. Lu (ed.), *Proceedings of the International Workshop on Navigating the Digital Business Frontier for Sustainable Financial Innovation (ICDEBA 2024)*, Advances in Economics, Business and Management Research 315,
https://doi.org/10.2991/978-94-6463-652-9_80

prices will not only be affected by the stock value, but also will be interfered by their own psychological factors and behavioral patterns, so the textual analysis of the investor's their emotions will also play a certain role in guiding the allocation of information resources in the financial market [1]. In the investor and the financial market this emotional impact path, some scholars use a large range of large-scale data, through the indirect construction of sentiment indicators, to explore the correlation between sentiment and stock returns and to predict. At the same time, the research on investor sentiment analysis is also various, with different focuses on research objects, research platforms, research methods, and research purposes. For example, Li et al. detect investor sentiment based on massive data from Chinese local investor forums through customized financial vocabulary lists and supervised machine learning [2]. In terms of improvements from the widely used Harvard Dictionary, Loughran, T., and B. McDonald's analysis of negative words relies on textual contexts rather than a single vocabulary, avoiding negative misjudgments of positive emotional words [3]. Jiang, F.W. et al. summarized the manager's sentiment index for the company's annual financial statements, predicted the cross-sectional stock price and indirectly analyzed investor sentiment [4]. By exploring the similarities and differences between the research objects and research methods of related articles, this paper takes emerging and traditional social media as the starting point, synthesizes various literatures to obtain the optimal data processing methods and model types in specific situations, and gives a research direction on the impact of undiscovered media texts on stock prices and financial markets.

2 Factors Affecting Investor Sentiment

Alternative data has been widely used in the field of finance in recent years, in which the sentiment analysis of news text, emerging media carriers through machine learning, artificial intelligence, is also an important analysis method for predicting stock market prices. Various studies have shown that the investor sentiment and the text in the environment that the investor is exposed to, has different significant impacts in different situations. Merriman, L. examined how social media affects investors' copy trading information, arguing that the influence of visual information on attitude of investor as well as decision making is more significant [5]. Neal Michael, by comparing the efficiency of financial information dissemination between multiple “new media” channels and the news, based on companies on Twitter and the investors' personal financial disclosure, and found that information on new media is more likely to be trusted by investors [6]. Park, J. H. et al. studied the two major user social behaviors, information search and sharing, and found that in the weak link pattern, the two behaviors promote each other and their initial goals, while they repel each other [7]. At the same time, there are a large number of text sentiment analysis methods in the literature, and emotional text dictionaries are constructed under different text environments, such as forums and news reports. Loughran and McDonald created the Language Model Dictionary, or LM Dictionary, for market analysis, investment decision-making, and risk management, which includes words that are highly relevant to the fields of finance and economics to avoid the neglect or misclassification of some technical terms in sentiment analysis dictionaries [3].

In order to develop Chinese financial sentiment analysis, Jiang et al. constructed a Chinese financial sentiment lexicon based on financial news media reports and word2vec algorithm [8]. Generally, scholars will reduce the dimensionality of a large number of lexicon word matrices to form vectors that can compute cosine similarity based on the word2vec algorithm proposed by Mikolov et al. [9]. The word2vec algorithm includes Continuous Bag of Words (CBOW) model and Skip-gram model, the Skip-gram model functions as given a target word, the model tries to predict the words that appear around it, which is especially suitable for capturing word relationships in sparse data, and is therefore commonly used when dealing with large corpora. Jiang Fuwei set up sentiment unit which contains sentiment words and negative words to calculate text sentiment, based on the sentiment index of a single article in order to obtain daily and monthly text sentiment [8].

3 Factors that Affect Stock Prices

In the current literature on investor sentiment and stock returns, there are two main research directions: the effect of both established and new forms of influence of social networks on investor behavior, as well as the analysis of the relationship between stock prices and investor sentiment. This means that for the stock market, there is a direct factor i.e. investor sentiment, and indirect factors include traditional and emerging social media.

3.1 Construction of Investor Sentiment Variables

As far as investor sentiment is concerned, scholars generally use both direct and indirect measures. Direct measures, such as questionnaires, analyze the relevant standardized indicators of investor sentiment by counting the proportion of positive and negative investor sentiment towards the future market, such as the Cheung Kong Graduate School of Business's Cheung Kong Graduate School of Business Investor Sentiment Survey (CKISS). However, due to the general research using questionnaires with low data coverage, high cost, untimely feedback and the difficulty of processing the data obtained, it is not able to generally and objectively reflect the optimistic and pessimistic attitudes of investors towards the market expectations in that time period. Therefore, scholars generally collect and use the more popular alternative data, that is, some objective data reflecting different perspectives around investors and the market, with the aim of quantify the abstract concept of market outlook. For example, on the macroeconomic side, Haritha P. H. employed macroeconomic factors to gauge the extent of the impact of general investor sentiment, including the exchange rate, term spread (TS), net foreign inflow (FII), industrial production index (IIP), wholesale pricing index (WPI), distinguishing from the former indirect data variables, Hu et al. on the other hand, based on the data of the Chinese stock market to consider the Chinese stock market more microeconomic variables, five proxies for investor sentiment are selected, namely, monthly IPOs, market turnover, and the quantity of newly established share accounts, the first day return of IPOs, and closed-end fund discounts [10]. Using these sentiment proxies, a composite index of investor sentiment is constructed [11].

3.2 Pre-processing of Investor Sentiment Variables

Due to the existence of different and similar proxy variables, selecting a more significant impact, smaller number and wider coverage of the proxy variable set is a more concerned part of scholars. Meanwhile, since the stock market and investor sentiment are constantly changing processes, exploring the correlation between investor sentiment and stock returns at different particular points in time is more comprehensive for examining the cross-sectional performance of stock returns, e.g., Baker and Wurgler consider present as well as lagged cross-sectional proxies after screening the variables [12]. Due to the time as the standard re-division of variable types, the data dimension is more lengthy, scholars tend to use Principal Component Analysis (PCA) as a tool to reduce the sparse matrix, reduce the data dimension on the basis of retaining the original data, and construct market outlook indicators. Simultaneously, in order to ensure the significance of the influencing factors in the construction of investor sentiment variables, on the basis of standardization, the correlation analysis of the proxy variables as influencing factors and the investor sentiment indicators is carried out, and the emotion proxy variables with less correlation are deleted to complete the construction of the indicators [13]. In the noise trader model (DSSW), information traders and noise traders co-exist, and Bayes' law in traditional economics does not fit the real situation, investment traders are vulnerable to the influence of the environment and individuals, in order to eliminate the influence of the time series on the sentiment indicators, removing the autocorrelation is necessary, such as the use of the weighted least squares (GLS) method or conduct the Durbin- Watson test and adding lagged variables, i.e., AutoRegressive model.

3.3 Correlation Analysis Between Investment Sentiment and Security Prices

After completing the construction of investor sentiment indicators that are not affected by autocorrelation and heteroskedasticity using the selected influencing factors, the degree of Nexus between investor psychology and market prices is examined, while the prediction provides a way of predicting the stock market price under different parameters, and scholars usually introduce either a linear regression model or an autoregressive model. If the correlation analysis and the prediction of stock returns are conducted through time series based on the available time cross-section, autoregressive model is superior in terms of processing and use. For example, Wenzhao Wang et al. found that the first-order autocorrelations (The first-order autocorrelations) after using consumer confidence index (CCI) as a representative element of investor sentiment zero expectation and unit variance standardized CCI in the MSCI market classification framework The mean value is high, which causes the slope and standard error to be too far from the estimation, in order to eliminate this error, the regression analysis is performed using the moving-block bootstrap simulation procedure method [14]. In an early autoregressive model called Markov's autoregressive model containing groups of vectors, at this time the constant term as well as the autoregressive coefficients are susceptible to the influence of the environment and subsample dissimilarities, in order to eliminate this effect, Jiangshan Hu et al. consider p-order autoregressive model such that the state variables obey Markov-Switching process, and set m such variables, using MS (m)-VAR (p) model [11]. Also, to reduce the computational effort, Multiple Linear Regression (MLR) can be used to assess the correlation between the constructed investor sentiment indicators and stock returns.

4 Conclusion

This paper explores the impact of Internet social media, traditional media on investor sentiment and the relationship between mood and stock market performance with respect to the impact of traditional and emerging Internet influence of social media on share prices and investor sentiment. In exploring the literature on social media on market confidence, evidence suggests that changes in the sentiment level of noise traders cause fluctuations in the price of risky assets and that the sentiment indices generated by analyzing the texts are significantly correlated with stock returns. It is also found that when exploring sentiment texts, constructing a context-specific sentiment text dictionary is a more convenient and efficient method, based on the Word2vec algorithm as well as the continuous bag-of-words model and the Skip-gram model, which allows for the construction of a model dictionary that is more in line with the concept of a large model. Meanwhile, when exploring the literature on the correlation between market psychology and equity prices, principal component analysis was found to be the more used means of dealing with dimensionality and larger downscaling. According to the different dimensions and perspectives of the data contained in the data used, simple multidimensional regression models as well as first-order and higher-order autoregressive models are used, and each has its own unique advantages in different environments. Meanwhile, summarizing the conclusions of related literature, it is discovered that there is a negative correlation between investor confidence and share performance during a bad market and a positive correlation during a bull market. In both bull and bear markets, there is an inverse relationship between stock performance and market turbulence. Also, the model that separates periods of high and low sentiment has an improved adjusted R-squared than the model that does not separate them, and both positive levels of sentiment and upward changes in sentiment have a positive effect on stock returns. Negative sentiment has no significant effect on stock returns. In the future, the research ideas can be promoted which is related to the impact of traditional and emerging Internet social media's impact on equity prices and market psychology, and explore the analysis of online platform texts to user sentiment, and then to the predictive analysis of specific parameters in related fields, and market decision-making judgment. Other perspectives such as news text length, release time and frequency can be studied more deeply and comprehensively. At the same time for investor sentiment and stock return research methods and models lack of systematic organization, some methods in some scenarios to deal with certain data performance is more inefficient and incomplete, so summarize and organize the relevant methods and models to use, so that the subsequent research can be more efficient on similar factors affecting the price of the stock to analyze and calculate. At the same time, this paper is unable to traverse all known methodologies and models for processing data, so there are still shortcomings.

References

1. Wen, F., Xiao, J., Huang, C., et al.: Research on the impact of investors' emotional characteristics on stock price behavior. Journal of Management Science **17** (03), 60-69 (2014).
2. Li, J., Chen, Y., Shen, Y., Wang, J., Huang, Z.: Measuring China's stock market sentiment. Working Paper, (2019).
3. Loughran, T., McDonald, B.: When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. Journal of Finance **66** (1), 35-65 (2011).
4. Jiang, F. W., Lee, J., Martin, X. M., Zhou, G. F.: Manager sentiment and stock returns. Journal of Financial Economics **132** (1), 126-149 (2019).
5. Merriman, L. The impact of online peer influence and information modality on investor decision making. University College Cork, (2020).
6. Snow, N. M.: Retail investors' perceptions of financial disclosures on social media: An experimental investigation using Twitter. USF Tampa Graduate Theses and Dissertations, (2015).
7. Park, J.H., Gu, B., Leung, A.C.M., Konana, P.: An investigation of information sharing and seeking behaviors in online investment communities. Computers in Human Behavior **31**, 1-12 (2014).
8. Jiang, F.W., Meng, L.C., Tang, G.H. Media text sentiment and stock return prediction. Economics (Quarterly) **21**(4), 1519-1538 (2021).
9. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. Working Paper, (2013).
10. P. H., H., Rishad, A.: An empirical examination of investor sentiment and stock market volatility: evidence from India. Financial Innovation **6** (34), (2020).
11. Hu, J., Sui, Y., Ma, F.: The measurement method of investor sentiment and its relationship with stock market. Computational Intelligence and Neuroscience, 6672677 (2021).
12. Baker, M., Wurgler, J.: Investor sentiment and the cross-section of stock returns. Journal of Finance (4), 1645-1680 (2006).
13. Shi, Y., Tang, J., Guo, K.: The impact of social media investor attention, investor sentiment on Chinese stock market. Journal of Central University of Finance and Economics (07), 45-53 (2017).
14. Wang, W., Su, C., Duxbury, D.: Investor sentiment and stock returns: Global evidence. Journal of Empirical Finance **63**, 365-391 (2021).