7-8-2023

How does Digital Interactive Information Influence Stock Performances

Jinshui Huang
School of Management Science and Engineering, huangjinshui341000@163.com

How does Digital Interactive Information Influence Stock Performances

Short Paper

Jinshui Huang
Southwestern University of Finance and Economics
555 Liutai Avenue, Wenjiang, Chengdu, Sichuan, China
Huangjinshui341000@163.com

Abstract

Currently, a new form of media, called Digital Interactive Media, has been popular in Chinese stock market. Investors can interact with listed companies directly. Therefore, it might be necessary to deconstruct and extract features of the interactive platform's information to understand the influence mechanism of digital platform on stock price fluctuations. This paper uses web crawling technology to obtain massive interactive information from Panorama from 2015 to 2019, extracts some valuable influencing factors through emotional quantification technology, and creates a Multiple Regression Model to explore effect of different emotional factors on stock price volatility. The empirical results show that the sentiment in investors' questions will not significantly affect correspondent stock prices, but the tone and sentiment of the response of listed companies will significantly influence stock prices trends, especially the price of the current trading day or next trading day. The robustness test and endogeneity test supports the research conclusions.

Keywords: Stock market, digital interactive media, multiple regression

Introduction

Recent financial researches have verified that news information continuously affects the stock market(Wuthrich et al., 1998; Tetlock et al., 2008; Wang et al., 2020). External information especially investing media information, acts a vital role in the fluctuation of stock price (Hirshleifer et al., 2009). With the rapid update of IT applications, information media platforms in the stock market is constantly evolving. In the early stage of Internet technologies, authoritative news channels dominated by official sites have evolved to free grassroots information discussion regulated by investing social media, and then to the direct interaction between investors and listed companies. We called these platforms "digital interactive media". According to the Annual Report on Investor Relations of Listed Companies (2019) released by the Shanghai Stock Exchange, digital interactive media has become the main channel for information disclosure in Chinese stock market. The most representative are "E-Hudong", "Hudong Yi", "Asking Dongmi" and "Panoramic Network".

However, there exists huge differences in media characteristics between Digital Interactive Medias and news authoritative medias and investing social medias. First of all, the number of daily visits and information released by digital interactive media have increased explosively (Wang et al., 2022). According to the data of Hudong Yi Platform on June 30, 2020, daily visit volume of Hudong Yi exceeded 2.5 million. Therefore, it is crucial to deconstruct the massive interactive text based on the big data perspective and extract valuable feature factors. Secondly, the information released by the official news media is one-way transmission from the platform to investors(Tetlock et al., 2008) The discussion of investment social media is a single-subject exchange between investors and investors(Bollen 2011). However, digital interactive media is different from official news media and social media. It includes two subjects, namely listed companies and investors. Therefore, it is necessary to extract representative interactive features from massive texts to deeply analyze the characteristics of digital platform and its influence on stock market. In the study of traditional media, scholars usually build a multiple mapping model, taking stock returns, abnormal returns and other indicators as dependent variables, and taking the degree of attention and the number of news as independent variables, trying to study whether the release of news or the attention of investors will impact the stock price(Tetlock, 2007; Pinto & Asnani, 2011; Wang et al., 2012). However, digital interactive media includes two main bodies, and there are certain time intervals and content differences in the number of words(Huang et al., 2021). We not only need to consider whether the emotion of the listed companies or investors will affect the stock price, but also need to study whether the interactive characteristics between the listed companies and investors will also affect the trend of the stock price movement.

However, in the era of digital economy, digital interactive media has become the main information disclosure channel of listed companies in China. It is urgent to carry out research on how digital interactive media affect the price fluctuation and price trend of the stock market. This is essential to enhance listed companies' management level , increase the transparency of information disclosure, and maintain the stability of national financial market. Therefore, the research questions are:

RQ1: How does Digital Interactive Information Influence Stock Performances？

RQ2:What is the specific mechanism by which digital interactive media affects the stock market? What important factors play an important role?

Based on the emerging phenomenon, this paper uses web crawling technology to obtain massive data from open digital interactive platform, uses feature extraction technology to extract value factors and interactive features representing the emotions of listed companies and investors from the interactive data, and constructs a multiple regression model to look for the specific mechanism of new platform affecting stock performance. The empirical results show that the sentiment of investors asking questions to listed companies will not impact stock price, but the response sentiment and response attitude of listed companies would significantly affect the stock price. The subsequent endogeneity test, robustness test and heterogeneity test all support the research conclusion.

The innovations of this article are as follows:

- First of all, in terms of data expansion, this paper uses distributed multithreaded crawlers to crawl the massive interactive Q&A information text from the digital interactive platform to form the financial Q&A text database of the securities market, in order to reveal the influence of information interaction between listed companies and investors in the era of digital economy;
- Secondly, in the aspect of feature extraction, this paper introduces the interactive question and answer text into the study of stock price volatility by using the methods of emotional word matching, and quantifies the emotion of interactive information into more valuable and representative features, in order to understand the factors affecting stock price;
- Thirdly, in terms of the impact mechanism, this paper takes into account the dual-agent characteristics of digital interactive media, incorporates the interaction factors of investors and listed companies into Multiple Regression Model, and analyzes impact mechanism and operating mechanism of new investment media platform on the market share price from a new perspective.

Related Work

Text Information Feature Extraction

In the early stage of media development, the official news website controlled the authority of information release. During this period, scholars mainly studied influence of news information on stock market through proportion of emotional words contained in "authoritative" media information. The research results published by Tetlock (2007) show that there exists a strong correlation relationship between stock returns and proportion of emotional words. If the news media report on the listed company is highly pessimistic, it will indicate that the company's share price will fall. Li et al. (2014) used proportion of emotional words in the media news to calculate the public sentiment index, reflecting the investors' tendentious opinions, and fused with basic information to predict price trend of stock market.

The rapid development of information technology has led to a major revolution in the Internet. Investors can spread their thoughts through investing social media such as stock bars or discussing forums. With the increase in the complexity of social media texts and the development of natural language processing technology, scholars' analysis of emotional dimensions has also changed from simple positive and negative judgments to high-dimensional measurements. For example, Loughran & McDonald (2011) extracted eight dimensions of investor sentiment and discussed the impact of different dimensions of investor sentiment on the volatility of financial market. With the increase of investor sentiment dimension, the research and analysis methods began to expand from the initial econometric analysis model to the machine learning model. Scholars tried to study the advanced machine learning technology to deeply capture impact of news information on stock pricing (Nan, 2015). Dickinson & Hu(2015) used Deep Neural Networks (DNN) to find a strong correlation relationship between public sentiment in Twitter and stock prices. For consumer oriented companies such as Wal Mart and Microsoft, their stock prices are particularly affected by public sentiment. Huang et al. (2016) used the convolutional neural network (CNN) algorithm to predict stock price trend from the perspective of public sentiment on the stock price in Twitter, and found that CNN performed better in most cases.

In the digital economy era of information explosion, the research of digital interactive media cannot be limited to the previous research models. It is necessary to take into account the emotional characteristics contained in the massive text and the interactive characteristics based on the dual-agent characteristics, and build a diversified relationship mapping model, so as to truly reveal the impact mechanism.

Econometric Multiple Mapping Model

How to build a correlation mapping model is the core of studying the media effect of the securities market. In the current research, most scholars use basic data (such as stock price, trading volume, turnover rate, yield, etc.) and quantitative data of media information (such as quantity, keywords, emotion, etc.) to build a correlation mapping model to capture influence of market factors on the stock fluctuations.

To analyze causal relationship between influencing factors and stock fluctuations, econometric models are widely used to describe the development law of economic operation. Among them, the most classic model is Three-factor model (Fama & French, 1993) to study factors of return difference of various stocks. In the field of researching the influence of media text on stock market, many scholars have also used econometric models, considering text information as independent variables, and considering stock rate of return or price fluctuations as dependent variables. For example, Huang, Teoh and Zhang (2014) contrlled the fundamental characteristics of companies, investigated the emotional tone in listed company's press release and investors' response , used logistic regression model to estimate the abnormal positive tone (ABTONE), which is positively correlated with price immediate response, and negatively correlated with delayed response of 1st and 2nd quarter. Karabulut(2015) uses Vector Autoregressive Model (VAR) to verify that National Happiness Index (GNH) can predict rate of return and correspondent trading volume and other indicators. In the previous researches of digital platforms, scholars will focus on the number of asking questions (Cen, Li and Tong, 2014; Cen, Tong and He, 2016), the timeliness of replies (Zhang & Han, 2015), the clarity of replies (Zhang & Han, 2015), whether to open online interactive platforms (Tan, Kan and Cui, 2016), the number of interactive words (Ding, Lv and Chen, 2018; Ding, Lv and Huang, 2018) are used as explanatory variables to verify causal relationship between attention degree from investors and stock price responses. It can be seen that the research in the field of traditional finance mainly relies on econometric models to seek the correlation between media texts and financial market volatility. However, the econometric model can only input scalar data, and the process of converting high-dimensional data into dimensional data often induces the value loss of information, which affects the accuracy of the analysis of the price trend of the securities market.

Theoretical Analysis and Research Hypothesis

The "Panorama Network" platform is a direct interactive platform for the joint participation of listed companies and investors. Its interactivity is achieved by the listed companies answering questions for investors. In the "Panorama" platform, investors can log in to their accounts and directly ask some concerning questions to listed companies, and mentioned listed companies would answer investors' questions online. This interactive way of solving doubts of listed companies can not only optimize investors' information acquisition ability, but also have a positive impact on investors' information processing. The research of Piroli (2007) shows that effectively eliminating the information acquisition behavior that interferes with information can optimize the information acquisition results. On this platform, the communicative process between investors and listed companies is presented in the form of online text, which can be learned by investors. Panorama's Interactive Platform is currently the authoritative investor relations management platform in China, with more than 3000 listed companies joining, covering the Shanghai Stock Exchange, Shenzhen Stock Exchange, Hong Kong Stock Exchange. Based on its important position, this paper selects Panorama text data as research data.

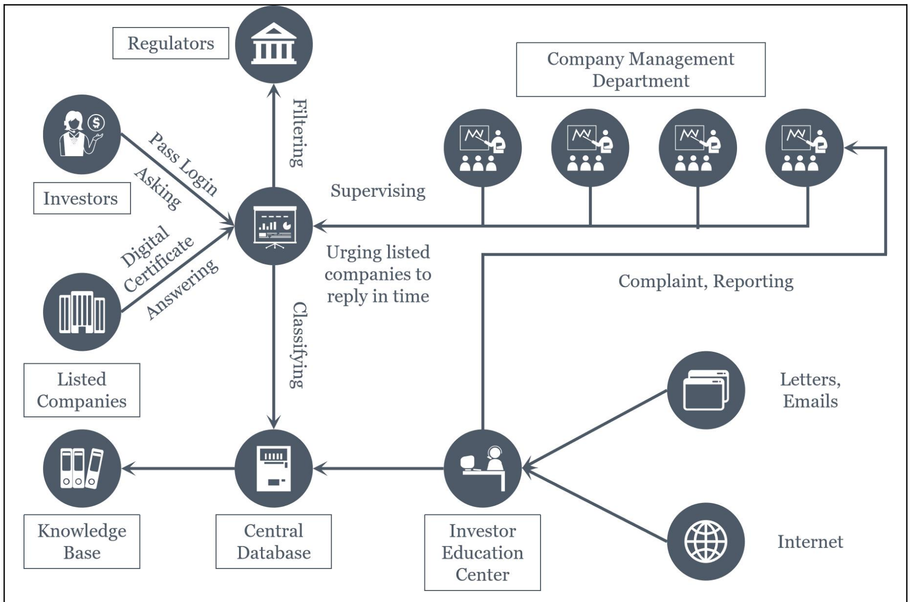

Figure 1. Panorama Platform Q&A Process

The emotion, attitude and tone of the response of listed companies can be obtained by investors, thus forming a price expectation close to the basic value. These micro processes can significantly affect the fluctuation of stock prices. The more investors' price expectations tend to the basic value, the smaller the risk of stock price fluctuations(Tan et al., 2016; Ding et al., 2018; Li & Lu, 2022). Relying on functional characteristics and institutional arrangements, the interaction of Panorama Network can enable investors to directly obtain replies from interested listed companies and make effective investment judgments. Therefore, the following assumptions are proposed in this paper.

H1: The more positive, timely and conscientious the listed companies respond to investors' questions, the more stable the stock price fluctuation will be and the better their performance will be.

In previous studies, most scholars mainly explored the correlation between the amount of external information and fluctuations of the stock market based on the information counting method (Jones et al., 1994; Ederington and Lee, 1993). However, the massive text content is very complex. Only measuring the degree of concern from the perspective of quantity can not fully reveal the impact mechanism, and some valuable information would be lost. At the same time, investors ask questions on the interactive platform. According to the herd effect of behavioral finance, other investors will also be affected, thus amplifying the impact caused by investors' questions (Morck, 2008). In this view, this paper constructs the corresponding emotional factors of investors' questions to reveal the impact of digital interactive media questions using emotional quantification. So this paper puts forward the following assumptions.

H2: The more positive the emotion investors ask is, the better the stock price will perform.

Research Method

Experimental Framework

Digital interactive media has become the main information disclosure channel for listed companies in China. This paper uses the web crawling technology to obtain Q&A text data from Panorama website from 2015 to 2019, and then uses the feature extraction technology to extract the factors that can represent the investors' asking emotion and the listed companies' answering emotion from the interactive data, as well as the factors that interact between them, and constructs a Multiple Regression Model with daily return rates as the dependent variable. Finally, endogeneity test, robustness test and heterogeneity test are used to further prove the reliability of conclusions. The research framework is shown in Figure 2.

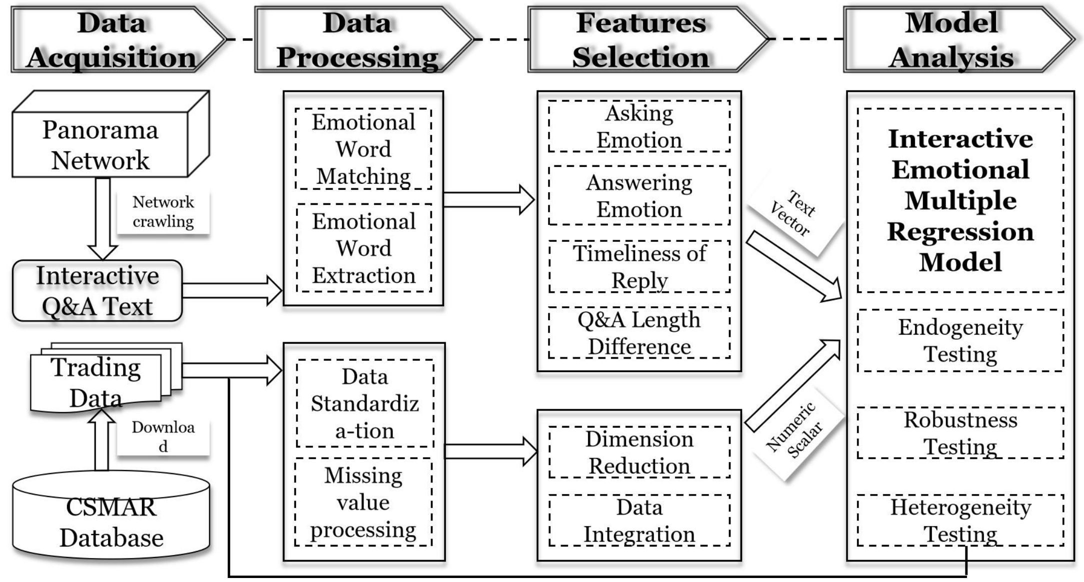

Figure 2. Main Work and Model Framework

Experimental Data

In terms of text information, we used automatic crawler programming to collect investors' questions and companies' replies to look for the specific mechanism of this platform on financial market. In the platform, we have collected 56847 Q&A pairs from 2015 to 2019. Before empirical analysis, the stock without trading records and some historical data are excluded. At last, we obtained 54328 Q&A communicative pairs, that is, 108656 textual information, which cover exactly 1970 companies.

In terms of transaction data, we collected some indicators such as daily return rate, turnover, current market value, debt-equity ratio from CSMAR Database. The time span of transaction data is the same as that of text information.

Text Analysis

We want to explore what features in the massive information would shock the stock price significantly. Therefore, it is vital to use Emotional Word Matching (EWM) Technology to extract the number and proportion of the emotional words. Before matching, we regularized interactive data. We uses Financial Emotional Dictionary of Sichuan Key Laboratory of Financial Intelligence and Financial Engineering (https://fife.swufe.edu.cn/sysgk/sysjj.htm). We wrote a program in C language. The logic is that if the words with positive and negative emotions contained in the emotion dictionary appear in each message, the algorithm will automatically count the number of negative and positive emotional words and cycle such operations. Due to space limitation, detailed procedures are not shown here. The specific matching process is shown in Figure 3. The green information bar represents the positive and negative emotion words set in the emotion dictionary, and the red and blue information bars represent the words of the interactive information in questions and replies. The positive and negative emotional words from Q&A text are respectively matched one by one. The calculation formula for quantifying emotional values is as follows:

$$S_{ij} = \left(P_{ij} - \mu_{ip}\right) \div \sigma_{ip} - \left(N_{ij} - \mu_{in}\right) \div \sigma_{in}$$

S represents the emotional value. The algorithm for calculating the emotional value of questions and replies is consistent. P represents the number of positive words in this text, and N represents the number of negative words in this text. I represents an individual, and here it refers to a listed company. Refer to operation of Wang et al. (2019), subtracting the number of standardized positive emotional words of question text from the number of standardized negative emotional words of question text, which is the Sentient Value of the question information. Subtracting number of standardized positive emotional words from number of standardized negative emotional words in each reply.

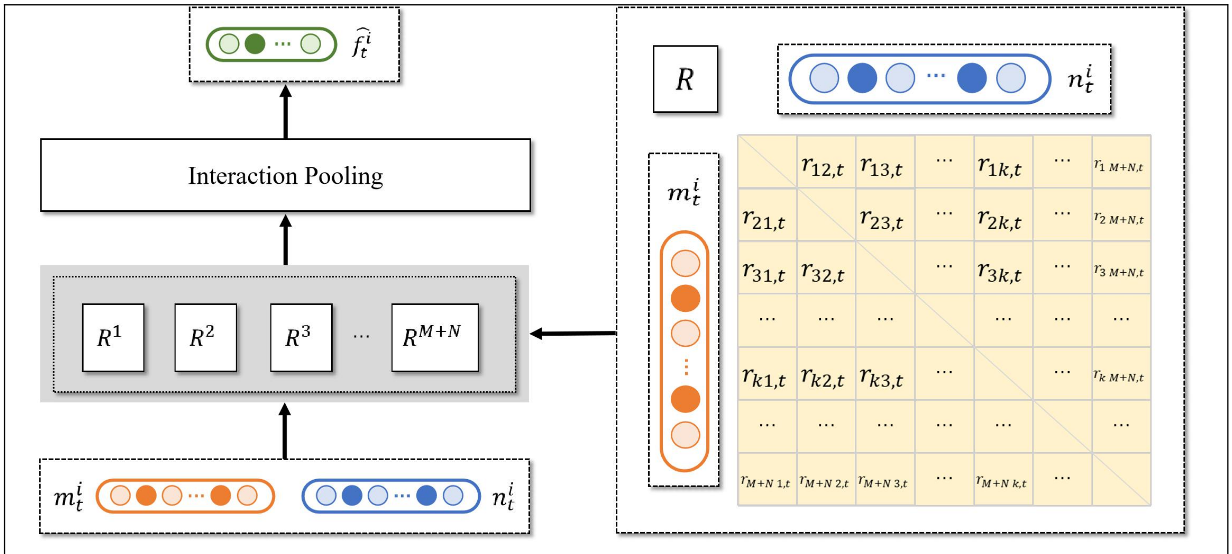

Figure 3. Q&A Information Interaction Factor Deconstruction Diagram

Empirical Analysis

After referring to a series of papers, we create a series of explanatory variables, such as emotional value of investors' questions, emotional value of listed companies' replies, the timeliness of listed companies' replies, and the difference in the number of words between questions and answers. We build a Multiple Regression Model to explore the influence mechanism of digital interactive platform.

$$R_{i,t} = \beta_0 + \beta_1 \times QUE_{i,t} + \beta_2 \times ANS_{i,t} + \beta_3 \times LEN_{i,t} + \beta_4 \times TIM_{i,t} + \beta_5 \times TR_{i,t} + \varepsilon_{i,t}$$

The explanations of each variable in the model are shown in the table below.

Variables Type	Abbreviation	Variables Meaning
Explained Variable	R	Daily correspondent stock return (Cen et al., 2016; Ding et al., 2018).
Explanatory Variable	QUE	The emotion value of each question (Wang et al., 2019).
	ANS	The emotion value of each reply (Wang et al., 2019).
	LEN	The difference between the words of the question and correspondent reply (Zhang & Han, 2015).
	TIM	Dummy variable. If the listed company replies to investors' questions within five days, then this variable is assigned a value of 1, otherwise it is 0 (Ding et al., 2018).
	TR	Daily turnover rate of stocks. According to traditional financial theory, trading volume can impact the stock price (Wang et al., 2019).
Control	ALR	The company's asset liability ratio for the current year. This indicator reflects

Variable (Ding et al., 2018; Li&Lu, 2022; Wang et al., 2019)		the company's capital structure.
	ROA	Return on assets. This indicator reflects profitability quality of the company.
	ROE	Return on equity. This indicator is a main component of DuPont analysis.
	EPS	Earnings per share. The ratio of post-tax profit to total equity.
	DMV	The market value of the stock on that day.
	YMV	Annual market value of the stock.
	IND	Industry of the company. Chinese latest practice is to divide the industry into 13 categories.
	AREA	Region of individual stock, divided into East, Central, Northeast, West.
	YEAR	Listing period of the stock

Table 1. Explanations of Variable in the Model

Table 2 shows the descriptive statistical analysis of the main variables. The average daily return of the stock is 0.00073, the minimum value is -0.1122, the maximum value is 0.2935. The large range indicates that the returns of each stock vary greatly. The mean of the emotion value of investors' questions is 0.00059, the maximum is 16.88, and the minimum is -12.485, which indicates that the number of questions with positive emotions is more than the number of questions with negative emotions. At the same time, we can see that there are large differences in questions, which is consistent with the investors' complex investment psychology. The mean emotional value of replies is 0.000064, maximum is 41.72, and minimum is -13.713. The absolute emotional difference of response is larger than that of questions. Mean value of the difference in the number of characters (LEN) is 45.77, maximum is 2077, and the minimum is -209. This reflects that the attitude of different listed companies to investors is quite different in stock market.

VARIABLES	N	mean	sd	min	max
R	54328	0.00073	0.0295	-0.1122	0.2935
TR	54328	0.00528	0.00205	0.00281	0.01521
QUE	54328	0.00059	1.767	-12.485	16.88
ANS	54328	0.000064	2.272	-13.713	41.72
LEN	54328	45.77	93.43	-209	2,077

Table 2. Descriptive Statistical Analysis

Table 3 shows uni-variate and multivariate regression models results. The variable, emotional value of response is significant and larger than 0, but the variable, emotional value of asking text is not significant. This implies that investors are very focused on the replies and take it into decisions. From the perspective of listed companies, positive emotional value of the replies of listed companies may reflect high level of corporate management.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Return	Return	Return	Return	Return	Return
Q_A	-0.0000243	— 0.0000274	—	—	—	—
	(-0.90)	(-1.02)	—	—	—	—
QUE	0.0000595	—	0.0001286*	—	—	—
	(0.80)	—	(1.80)	—	—	—
ANS	0.0001431**	—	—	0.0000816	—	—
	(2.28)	—	—	(1.47)	—	—

LEN	-0.00000427***	—	—	—	-0.000003**	—
	(-2.84)	—	—	—	(-2.23)	—
TIM	-0.0004046	—	—	—	—	-0.0003418
	(-1.51)	—	—	—	—	(-1.28)
TR	0.7215424***	—	—	—	—	—
	(11.63)	—	—	—	—	—
Constant	-0.003***	0.001***	0.001***	0.001***	0.001***	0.001***
	(-7.38)	(5.90)	(5.81)	(5.81)	(6.20)	(5.47)

Table 3. Uni-variate and Multivariate Regression Results

Robustness Testing

Deleting replies without substantive content. Some replies are platitudes with no substantive content. For example, "Thank you for your attention to the enterprise", "Thank you very much for your suggestions", "Thank you for your attention and support", "Hello, shareholders! Welcome to ask! Thank you!" and so on. On the whole, about 16000 responses were identified during the sample period. These responses with no substance each year were deleted from the cumulative response data and then re-inspected. The results are still stable. Due to space limitation, it will not be displayed .

Endogeneity Testing

Panorama began to use the online interactive communication platform in January 2015, which provides a relatively effective exogenous impact scenario for the adoption of the Difference-in-Difference Model. With reference to Tan et al.(2016), the opening of online interactive platform has significantly affected the fluctuations of stock prices, thus providing robust evidence for the interaction factors and emotional factors of the online platform to affect the stock prices.

Conclusions and Expected Contributions

This paper uses massive data from open digital interactive platform, uses feature extraction technology to extract value factors and interactive features representing the emotions of listed companies and investors from the interactive data, and constructs a multiple regression model to explore the specific mechanism of digital interactive media affecting stock performance. The empirical results show that the sentiment of investors asking questions will not shock stock price (H2), but the response sentiment and response attitude of listed companies would significantly affect the stock price (H1). The subsequent endogeneity test, robustness test and heterogeneity test all support the research conclusion.

Interactive information can be distinguished according to various categories. For example, based on the information attribute dimension, Q&A interactive information can be divided into repetitive information and non-repetitive information, true information and false information, as well as different content topics; based on attribute dimension of listed companies, interactiveinformation can also be divided according to the attributes of industry, region and sector. Q&answer information can also be divided according to the emotional polarity, detail, timeliness and other criteria of the text. Therefore, further works may be carried outto improve the heterogeneity test.

This paper is expected to have both theoretical and practical contributions. Firstly, this paper explores the specific mechanisms by which digital interactive media affects stock price performance, which will enrich the irrational theories related to behavioral finance. Secondly, the conclusions of this paper are helpful in providing insights to listed companies and investors in the financial market. It is conducive to helping listed companies improve their corporate governance level and information transparency, improving market information symmetry, and protecting the interests of small and medium-sized investors.

Acknowledgements

This work has been supported by Fundamental Research Funds for the Central Universities (JBK 2107015) and the Financial Innovation Center of the Southwestern University of Finance and Economics. I completed the initial draft in Nanchong City, Sichuan Province, and made revisions in Xi'an City, Shaanxi Province. This process is unforgettable to me. Thank you to my family and friends for their support.

References

Bollen, J., Mao, H. and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.

Cen Wei, Tong Naqiong, He Xiaoyue. (2016). Investor Attention and the Protection of the Interests of Small and Medium Shareholders: An Empirical Study Based on the Data of Shenzhen Stock Exchange's "Hudong Yi" Platform. *Securities Market Guide*, 21(2), 54-62.

Ding Hui, LV Changjiang, Chen Yunjia. (2018). Investor Information Capability: Differences of Opinion and Stock Price Crash Risk: Evidence from Social Media "Shanghai Stock Exchange E-Interaction". *Management World*, 34 (9), 161-171.

Hirshleifer, D., Lim, S.S. and Teoh, S.H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance*, 64(5), 2289-2325.

Huang, J., Qian, J., & Liu, Y. (2021). Research on the Impact of Digital Interactive Media on the Stock Market Based on Big Data Text. *Proceedings of 2021 4th International Conference on Artificial Intelligence and Big Data*, 17-21.

Huang X, Teoh SH, Zhang Y. (2013). Tone management. *The Accounting Review*, 89(3), 1083-1113.

Meng Qingbin, Huang Qinghua, Zhang Jinfan & Wang song. (2020). Does Internet communication between listed companies and investors have information content? -- Based on the research of "Hudong Yi" of Shenzhen Stock Exchange. *Economics (Quarterly),* 5(2), 637-662.

Morck, R. (2008). Behavioral finance in corporate governance: economics and ethics of the devil's advocate. *Journal of Management & Governance*, 12, 179-200.

Pinto, M. V., & Asnani, K. (2011). Stock price prediction using quotes and financial news. *International Journal of Soft Computing and Engineering*, 1(5), 266-269.

Tetlock P C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139-1168.

Tetlock, P.C., Saar‐Tsechansky, M. and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance,* 63(3), 1437-1467.

Wang, B., Huang, H., Wang, X. (2012). A novel text mining approach to financial time series forecasting. *Neurocomputing*, 83, 136-145.

Wang, J., Xie, Z., Li, Q., Tan, J., Xing, R., Chen, Y., & Wu, F. (2019). Effect of digitalized rumor clarification on stock markets. *Emerging Markets Finance and Trade*, 55(2), 450-474.

Wang, J., Zhang, X., Yu, G., Chen, Y., & Rao, S. (2022). DeeBERT-S3WD: Three-Way Multigranularity Decision for Interactive Information Sentiment Analysis Research. *Mathematical Problems in Engineering*, 64(5), 117-134.

Wuthrich, B., Cho, V., Leung, S., Permunetilleke, D., Sankaran, K., Zhang, J. (1998). Daily stock market forecast from textual web data. *Proceedings of 1998 IEEE International Conference on Systems, Man, and Cybernetics*, 2720-2725.

Yang Fan, Zhang Yuming. (2020). Can Internet communication reduce stock price synchronization? -- Evidence from "Shanghai Stock Exchange E interaction". *Journal of Central South University of economics and law*, 20 (6), 108-119.

Zhang Jixun, Han Dongmei. (2015). Timeliness and clarity of management's response and investors' investment decision-making in online interactive platform communication: an experimental evidence. *Management Review*, 27 (10), 70-83.