Scientific Research Publishing

Open Journal of Business and Management, 2019, 7, 799-816
http://www.scirp.org/journal/ojbm
ISSN Online: 2329-3292
ISSN Print: 2329-3284

R^2: Information or Noise? Textual Analysis Based on SSE E-Interaction

Ben Wang

Management School, Jinan University, Guangzhou, China
Email: BenAWang@126.com

How to cite this paper: Wang, B. (2019) R^2: Information or Noise? Textual Analysis Based on SSE E-Interaction. *Open Journal of Business and Management*, **7**, 799-816.
https://doi.org/10.4236/ojbm.2019.72054

Received: March 12, 2019
Accepted: April 12, 2019
Published: April 15, 2019

Abstract

The question whether R^2 represents information or noise is still a fundamental question in the study of stock price synchronicity. There are two main difficulties. Firstly, the trait information of a company is hard to measure; Secondly, the investors' sophistication is ignored when we discuss the effectiveness of market. Through the study of Chinese SSE E-interaction platform and the measurement by textual analysis method, this article argues that the improvement of investors' sophistication has negative association with stock price synchronicity. This result is more salient in companies with lower opacity or with higher corporate governance. This paper contributes to the deepen understanding of stock price synchronicity and new method of measuring trait information of companies.

Keywords

Investor Sophistication, Stock Price Synchronicity, E-Interaction, Textual Analysis, Chinese Capital Market

1. Introduction

Rising and Falling simultaneously is one of the most significant characters in China equity market for a long time. For example, in 2015, the Chinese stock market experienced the abnormally simultaneous fluctuations of "limit up" and "limit down" for nearly 2000 stocks, which seriously damaged the pricing efficiency of the Chinese stock market. Therefore, Chinese scholars pay much attention to the issue of stock price synchronism (*i.e.* R^2). However, when discussing the economic implications of R^2, scholars divide into two opinions: the "information efficiency view" and the "noise trading view". Scholars of the "information efficiency view" believe that the lower R^2 reflects the inclusion of more trait information in the stock price, which means that the market is more efficient [1]

[2] [3]. The "noise trading view" scholars argue that the lower R^2 reflects that the stock price contains more uncertainty and noise, which means that the market is less efficient [4]. The results of empirical research have not provided consistent evidence on earnings quality, corporate governance, information opacity, and analyst roles, making the basic issue of stock price synchronization research unsettled nowadays.

In order to solve of this problem, You Jiaxing (2017) believes that the discussion of R^2 needs to incorporate conditional considerations [5]. Under various non-ideal market conditions, the economic meaning of R^2 will be different. You argue that most of the existing literatures are from the perspective of information asymmetry, less consideration is given to investors' information acquisition and understanding capabilities. However, in Chinese stock market, small and medium-sized investors generally lack professional knowledges [6], energy and ability to understand and use information for investment decisions [7]. The purpose of this paper is to study the impact of investor ability of information acquisition and understanding on stock price synchronicity, and to answer whether R^2 is information or noise in a new site.

The "SSE e-Interaction" platform provides an ideal scenario for this paper to study this issue. On July 5, 2013, the Shanghai Stock Exchange "SSE e-Interaction" network platform was launched. The main function is to provide an Internet information platform, enabling listed companies to answer questions from investors online. The e-interaction platform is significantly different from other information disclosure channels. Firstly, compared with other information disclosure channels, the e-interaction platform is a network question and answer platform. Investors ask questions to listed companies, and listed companies have obligation to answer questions in time according to public information. The content of the information disclosure is initiated by the investors rather than the company, which makes it distinguished from the information disclosure channels such as announcements of listed companies and Weibo. Secondly, the SSE e-interaction platform forbid listed companies to disclose unpublished material information through the platform. They only explain and reply to the disclosed matters in this platform, which makes the e-interaction platform promotes market efficiency by improving investor information acquisition and understanding rather than by feeding new information. These two features help us examine the impact of investor information acquisition and understanding on stock price synchronicity.

This paper finds that the more trait information in the e-interactive platform, the lower the stock price synchronization is. Because the acquisition and understanding of trait information promotes the rational decision-making of investors, the negative correlation between trait information and stock price synchronicity supports the "information efficiency view". Further research has found that trait information in those companies with higher information transparency and better corporate governance have greater influence to the stock price syn-

chronization. These empirical results are consistent with the theoretical predictions of the "information efficiency view", that is, the view that stock price synchronicity is an information efficiency indicator.

The possible contributions of this paper are as follows: 1) In order to measure trait information, this paper uses the method of text analysis [8] [9] to propose a way of directly measuring heterogeneous information, which provide a new perspective in future studies in stock price synchronicity; 2) This paper proposes the theoretical reasoning and empirical results of investors' ability to acquire and understand information affecting stock price synchronicity, enriching the research on stock price synchronization "information efficiency view" [1] [2] [5].

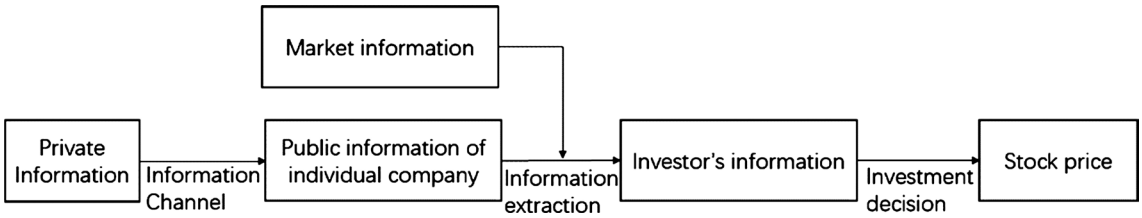
2. Theoretical Analysis and Hypothesis

2.1. "Three Jumps" in the Formation of Stock Price Synchronicity

This chapter summarizes the micro-path of "information-stock price" under the effective market hypothesis in **Figure 1**. Firstly, the company's private information is disclosed through various information channels in accordance with the requirements of securities laws and regulations. Then, the company's public information, mixing with other market information form the total set of public information in the market. Investors obtain this public information to form their own information sets, and then make investment decisions through their own risk preference models. The stock price information is the combination of all investors' investment decisions.

Roll (1988) believes that stock prices include information on the market, industry, and individual levels. The fitting coefficient (R^2) in the capital asset pricing model explains the market and industry information content reflected in individual stock returns [10]. The residual error explains the company's trait information content in individual stock returns, so the lower R^2 means higher market efficiency. This view, based on the classic theory of CAPM, gives R^2 the ability to interpret the real equity market practice and has received extensive attention from scholars.

What we should notice is that there are two basic assumptions in the CAPM classic theory, namely the completeness of information and the complete rationality of market participants [11]. However, in the real market, laws and regulations are not perfect. There are various forms of information asymmetry between insiders and outsiders. At the same time, investors can only achieve limited rationality,



Source: made by the writer.

Figure 1. "information-stock price" path in efficient market hypothesis.

but not complete rationality [12].

These real-world problems have caused three "thrilling jumps" in the process from information to stock price. The first jump is from the company's private information to the public information of their companies. this process will lead to a reduction in the company's public information due to differences in information transparency and information disclosure requirements [13]. The second jump is from public information to information extracted by investors. Because of the costs of extracting information, investors can only get part of public information and form their limited information sets to market investment decision [14]. The third jump is investors ability of acquisition and understanding information. When making decisions, the limited rationality of investors leads investment decisions not the optimal choice under their information set [7]. The result of three jumps, due to information asymmetry and bounded rationality, is that the stock price information cannot be as efficient as in the ideal situation, leading the residual of the CAPM regression model, include "noise and irrational factors", thus affecting the effectiveness of R^2 as a market efficiency indicator.

As an emerging capital market, the Chinese stock market still have problems in market efficiency and investors' rationale. These problems have brought an important empirical question: In the Chinese stock market, is the lower R^2 representing higher market efficiency or more noise trading?

In order to solve the above problems, foreign scholars mainly demonstrate the usefulness of R^2 as a market efficiency indicator from the perspectives of property rights protection and information transparency. Morck *et al.* (2000) support the "information efficiency view". After comparing R^2 in more than 40 economies around the world, they found that countries with poor property rights protection systems have higher stock price synchronies [1]. According to Morck *et al.*, the poor protection of private property rights is poor discourages arbitrageurs to arbitrage in the market because of the high cost of acquiring and using information. Therefore, stock participant use market information more often and individual information in the market is less included in the stock price. Based on the research, Jin and Myers (2006) show that when company information is opaque, company managers have the motivation and ability to gain private benefits by hiding individual information. In this case, individual information is less included in the stock price information. By following the empirical approach of Morck *et al.* (2000), Jin and Myers (2006) found that companies with lower information transparency have higher R^2, which also demonstrates the reliability of R^2 as an information efficiency indicator [2]. Since then, a large number of scholars at home and abroad have carried out a series of empirical tests in the Chinese market based on the theoretical explanations of these two articles, most of which support the interpretation of the "information efficiency view" [14] [15] [16], but there are still inconsistent results. For example, Zhou (2014) found that poor corporate governance brought about a decline in stock price synchronism [17]. Wang (2009) found that the higher the quality of ac-

counting information, the higher the stock price synchronicity [18].

2.2. "The Third Jump" and Stock Price Synchronicity

In these studies, scholars put most of their energy into the first hypothesis—the completeness of information, mainly discussing the first two jumps—the increase in information channels and the reduction in information acquisition costs. However, the current study is less involved in the second hypothesis: a fully rationale of investors. According to Pirolli's (2007) cognitive theory, investors make their investment decisions by repeating information acquisition and information understanding behaviors until their expectations are consistent with external evidence [12]. This means that not all market public information will influence investor decisions, and only information that investors can obtain, understand and verify is included in investment decisions. At the same time, empirical research has found that investors cannot fully and accurately acquire and understand all the information in the capital market. You Jiaxing and Wu (2012) used the "silent spiral" theory explain why media sentiment leads to mispricing of assets [19]. Xu and Chen (2016) analyzed the microblog of the listed company and found that Weibo disclosure is interfered by noise information and has the characteristics of hybridity [20]. Because news reports or social media disclosure of listed companies are voluntary disclosure issued from companies, it does not involve investors' information acquisition and understanding, and cannot directly influence investment decisions. In the "information-price" path described above, the formation of investment decisions through investor understanding is an essential part of stock price information.

Through the unique scenario of "SSE e-Interaction", this paper focuses on the reliability of R^2 as a market efficiency indicator from the perspective of the third jump—investor's acquisition and understanding of information. On July 5, 2013, the Shanghai Stock Exchange "SSE e-Interaction" network platform was launched. The main function is to provide an Internet information platform enabling listed companies to answer questions from investors online. The e-interaction platform is very different from other information disclosure channels of previous studies. Firstly, in prior literatures, company announcements, news media, and social media are voluntarily disclosed by listed companies or the media. Investors can only passively accept information. In e-interaction platform, investors actively obtain supplementary information where they feel doubtful in voluntary disclosure, thereby enhancing the understanding of the existing information. Secondly, e-interaction forbids listed companies to disclose unpublished material information through the platform, management only explains and responds to disclosed matters. This character makes the e-interaction platform promote the market efficiency only by explaining and emphasizing the existing information. This feature makes it possible to examine separately the impact of investor information acquisition and understanding on stock price synchronicity. Thirdly, the e-interaction platform is supervised by the Shanghai Stock Exchange. Investors

are not allowed to ask malicious questions. The listed company must ensure that their answers are valid and accurate. The official operation guarantees the credibility of the replies. These features help to enhance the support of e-interactive platform Q&A information for investment decisions. Therefore, this paper believes that the interactivity of e-interaction platform effectively increases the ability of investors to acquire and understand the public information of listed companies, without significantly affecting the access and acquisition costs of new information. With this scenario, we can discuss the impact of investor acquisition and understanding of information on stock price synchronicity when other information environments are consistent, thus providing a new perspective for answering whether R^2 is information or noise.

Because the investor's questions in the e-interaction platform must be targeted at a certain company, investors can ask questions about the company's specific governance issues, business strategy issues, and can also ask questions about the industry and market risks involved. Therefore, this paper implement the text analysis method of Hanley and Hoberg (2010) to divide the question and answer information in the e-interaction platform into general information and enterprise trait information that cannot be interpreted by the market and industry [8].

Compared with questions, the answer in the platform are more relevant with investors' decision-making. Investors ask questions in various language styles. Even though the questions will contain some factual judgments, they are often speculative and often have personal emotional color. In addition, some questions are left unanswered because of its ambiguity and offence. The listed companies' answers are of standard format, and the answer content is based on the disclosed company announcements and other information. Besides, the sentiment of answers is more restrained. Therefore, this paper believes that the trait information of the answer content can improve the information acquisition and understanding ability of investors compared with the trait information of the question content. According to the theory of "information efficiency", the lower R^2 can be used as an indicator of higher information efficiency, and the following hypothesis 1 can be obtained:

Hypothesis 1: Stock price synchronicity has negative association with the trait information of answers in e-interaction.

2.3. How "The Prior Two Jumps" Influence "The Third Jump"

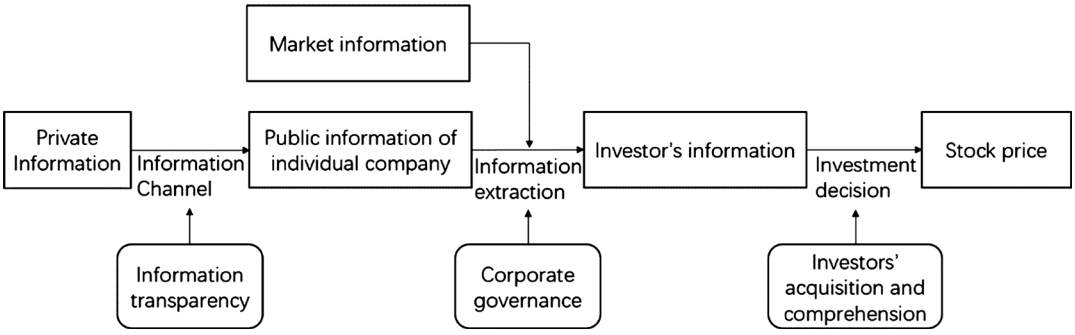
Because the path of "information-stock price" has experienced three consecutive jumps, if property rights protection and the corporate governance mechanism are absent, a large amount of company trait information will be lost in the "two previous jumps"; therefore without good property rights protection and information transparency in companies, even if the investor's information acquisition and understanding ability is improved, it is hard to reduce the stock price synchronicity. As shown in **Figure 2**, in order to comprehensively consider the

whole process of stock price synchronicity formation, we further discuss the impact of the "first two jumps" on the information efficiency of the "third jump". We refer to the predecessor literature to discuss the relationship between e-interaction traits and stock price synchronicity under different levels of information transparency and corporate governance.

The level of information transparency affects the information efficiency in the first jump. Jin and Myers (2006) argue that when companies have high transparency, the ability of insiders to use hidden information to obtain cash flow is weak [2]. In such a company, the value generated by the company's trait information is not used by insiders for personal gain, but is reflected in the stock price, resulting in lower stock price synchronicity. In a company with high transparency, the cost of obtaining reliable company information is low, and the trait information in the e-interaction question can be agreed with the information, thereby enhancing investor understanding and helping investors make investment decisions. On the contrary, in companies with low transparency, it is difficult for investors to obtain reliable company information, so instead of collecting the company's individual information at a cost, the company is priced through market and industry information. According to the existing literature research, this paper uses the accounting earnings quality and the number of announcements as substitute variables for corporate transparency (Hutton *et al.*, 2009). If the lower R^2 is indeed an indicator of higher information efficiency, then the hypothesis 2 can be obtained:

Hypothesis 2: The negative association is more prevalent in those companies which have high information transparency.

The level of corporate governance affects the information efficiency of the second jump. From the micro perspective of enterprises, the property rights protection of Morck *et al.* (2000) is the protection of small and medium shareholders by corporate governance mechanisms [1] [5]. When corporate governance is weak, the cost of acquiring individual information is higher. At this point, even if investors have decent ability in information acquisition and comprehension, it will not be able to successfully interpret the information and will not be able to form an optimal investment decision and obtain the expected return. As a result, investors have to turned to market information and industry



Source: made by the writer.

Figure 2. Major factors influencing the information efficiency in "information-stock price" path.

information, which has lower information costs for investment decisions. In this case, according to the "information efficiency view", the stock price synchronization is higher. Therefore, when companies have a better internal corporation governance, the trait information on the e-interaction platform enhances investors' ability to acquire and understand information and helps investors to rely on the understanding of trait information to form better investment decisions under existing information sets, thereby improving market efficiency. According to the existing literature research [21], this paper uses the largest shareholder shareholding and accounting firm as a substitute for corporate governance. If the lower R^2 is indeed an indicator of higher information efficiency, then the hypothesis 3 can be obtained:

Hypothesis 3: The negative association is more prevalent in those companies which have better corporate government.

3. Research Design

3.1. Sample Selection

This paper uses A-share companies listed on the Shanghai Stock Exchange from 2013 to 2017 as the research sample. According to the research design needs, this paper has made the following screenings for the sample: 1) Excluding the sample of financial service companies; 2) Excluding ST and ST* listed companies; 3) Excluding the first year of listing Observations; 4) Excluding observations with net assets below 0; 5) Excluding observations with incomplete financial data. Except for the manual calculated information content index, other data are downloaded from the CSMAR database or calculated through it. In addition, in order to prevent the influence of extreme values on the results, this paper performs a 1% winsorizing process on successive variables. Finally, a total of 3070 annual-company observations were obtained.

3.2. Information Content Index

The SSE e-interaction platform was officially opened on July 5, 2013. Since then, the number of questions and answers in the e-interaction platform has gradually increased. Until December 2017, there were 236,774 question records in the e-interaction platform. This paper uses python program to capture all pieces of Q & A records in "SSE e-interaction" website search module (http://sns.sseinfo.com/search.do) from July 5, 2013 to May 1, 2018. A total of 206,840 questions were recorded, involving 947 listed companies, of which 174,683 (84.45% of all questions) were answered. Unanswered questions are often problematic, such as wrong objects, repeated questions, malicious questions, or inconsistent questions. Therefore, this article only selects pieces that got answered by management in further analysis.

Based on the research ideas of Hanley and Hoberg (2010), this paper subtotal the question text and the answer text, by year and company and calculates the company-annual textual vector for questions $NormQ_{i,t}$ and company-annual

textual vector for answers $NormA_{i,t}$. Text vectorization is a commonly used technical means in computer textual analysis. Vectorization make it true for the quantitative expression and calculation of text features by expressing the frequency of each word in a piece of text in the form of a vector. For example, the text A is "*they sell potatoes and they sell the corn*", and the text B is "*they sold knives*". In the text vectorization, the ambiguous words "*they*", "*and*", "*the*" are removed, and then the vector is based on all the words ["*sell*", "*potato*", "*corn*", "*knife*"] of the two texts. Therefore, in this text there are wordA = {2, 1, 1, 0}, wordB = {1, 0, 0, 1}.

In English, readers judge words by spaces, while in Chinese, readers need to execute word segmentation with semantics. This paper uses the *ICTCLAS*2016 developed by the Institute of Computing Technology of the Chinese Academy of Sciences to conduct Chinese word segmentation and keyword judgment. Due to the social media nature of the e-interaction platform, the language style is rich and there are many self-made words, it is not suitable to use the ready-made dictionary as the complete information. The keyword extraction function provided by *ICTCLAS* can utilize the information entropy algorithm of context to delete ambiguous words and incorporate new words. In addition, because the length of the text is different, the result is incomparable merely by words segmentation. In this paper, the normalized vector is obtained by dividing the annual-company text vector by the total number of real words in the text. As in the above example, the normalized vectors are {0.50, 0.25, 0.25, 0} and {0.50, 0, 0, 0.50}, respectively.

According to the method of Hanley and Hoberg (2010) [8], Meng Qingbin *et al.* (2017) [9], this paper separates the general information such as market and industry information and the trait information related to individual company through linear regression.

In this paper, the market standardization information vector and the industry standardization information vector are calculated by the following methods: 1) According to the above method, the text normalization vector $Norm_{i,t}$ of the company t is obtained (the text of the question and answer is respectively processed $NormQ_{i,t}$ $NormA_{i,t}$); 2) Define the industry standardization vector $Norm_{I,t}$ as the arithmetic mean of the normalized vectors of all companies except the company in the industry where the company *i* is located; 3) Define the Market standardization vector $Norm_{M,t}$ as the arithmetic mean of the normalized vectors of all companies in the market except the companies in the industry of the company *i*. For example, in a certain year, there are *P* companies are in the same industry as company *i*, and there are *Q* companies in the whole market. The industry standardization vector and market standardization vector of the company are:

$$Norm_{M,t} = \frac{P}{P-1}\sum_{j=1 j\neq i}^{P} Norm_{j,t} \quad (1)$$

$$Norm_{I,t} = \frac{1}{Q-P}\sum_{j=1 j\neq i}^{Q-P} Norm_{j,t} \quad (2)$$

Then, the general information and trait information are separated by multiple linear regression, namely:

$$Norm_{i,t} = \alpha_0 + \alpha_1 \times Norm_{M,t} + \alpha_2 \times Norm_{I,t} + \mu_{i,t} \quad (3)$$

α_1 represents the part that can be explained by market information, the higher α_1 is, the more similar the company I is with the information of other companies in the market; α_2 represents the part that can be explained by other companies in the same industry, the higher α_2 is, the more similar the company *i* is with the information of other companies in the same industry; the residual $\mu_{i,t}$ is the information that cannot be explained by industry and market information. According to Hanley and Hoberg (2010) [8], this article will be defined $\alpha_1 + \alpha_2$ as the company's general information (questions and answers are recorded as *std_q* and *std_a* separately), and the sum of the absolute values of the residual is defined as trait information (question and answers are recorded as *inf_q* and *inf_a* separately).

3.3. Stock Price Synchronicity

With the reference of Morck *et al.* (2000) [1], Gul *et al.* (2010) [20], this paper calculate the fitted coefficient $R^2_{i,t}$ using the regression of Equation (4).

$$RET_{i,t} = \alpha_0 + \alpha_1 \times MRET_t + \alpha_2 \times IRET_t + \mu_{i,t} \quad (4)$$

$RET_{i,t}$ represents the return in week t of company *i*. $MRET_t$ and $IRET_t$ represents the average return in week t of the companies in the same industry or in the whole market separately. $R^2_{i,t}$ is the explainable part by market and industry volatility in stock price. $R^2_{i,t}$ has the range within (0, 1), we calculate the stock price synchronicity using the Equation (5).

$$SYN_{i,t} = \ln\left(\frac{R^2_{i,t}}{1 - R^2_{i,t}}\right) \quad (5)$$

3.4. Research Design

We use Equation (6) to study the association of trait information in e-interaction platform and stock price synchronicity.

$$SYN_{i,t} = \beta_0 + \beta_1 \times inf_{i,t} + \beta \times Control\ Variables_{i,t} + \sum Year + \sum Ind + \varepsilon_{i,t} \quad (6)$$

$SYN_{i,t}$ is stock price synchronicity, $inf_{i,t}$ is trait information of company i, including trait information of questions and answers. In addition, We control for company size (*size*), leverage (*lev*), profitability (*roa*), book-to-market ratio (*bm*), month turnover (*turnover*), listing history(*age*), and nature of property rights (*soe*). *Year* and *Ind* represent year and industry fixed effect. $\varepsilon_{i,t}$ represents the residual. The specific definition of variables is as followed (**Table 1**).

4. Empirical Result

4.1. Descriptive Statistics

Table 2 is descriptive statistics for the above main variables. In the following

Table 1. Definitions of variables.

Name	Definition
SYN	Stock price synchronicity, the co-movement of individual volatility and market volatility of stock price. Equation (5) demonstrates the calculation.
inf	Trait information, calculated from normalized text vector of questions and answers records in e-interaction platform. Equation (3) demonstrates the calculation.
std	General information, calculated from normalized text vector of questions and answers records in e-interaction platform. Equation (3) demonstrates the calculation.
size	The logarithm of the book value of total asset for the current year.
lev	Firm leverage, measured as total debt divided by total asset for the current year.
roa	Return of asset, measured as net profit divided by total asset for the current year.
bm	Book-to-market assets ratio, measured as book value of asset divided by market value of asset.
turnover	Monthly averaged turnover of company for the current year
age	Firm listing history, measured as the natural logarithm of the year number that firm is listed on A share market.
soe	Stated owned enterprise, measured as 1 if the controlling owner is State Asset Regulatory Commission, and 0 otherwise.
abs_da	Accounting information quality, measured through Modified Jones model
*top*1	Share proportion of the biggest shareholder
*big*10	Auditing quality, measured as 1 if annual report is audited by big10 auditor. The rank of auditors is based on the rating result of China Certified Public Accounting Association.
record	The amount of public disclosure issued by company in official publication channel.

Table 2. Descriptive statistics.

variables	size	mean	sd	min	median	max
r^2	3157	0.447	0.0389	0.00322	0.448	0.957
SYN	3157	−0.271	0.900	−3.002	−0.208	1.697
inf_q	3512	1.602	0.231	0.853	1.494	5.633
inf_a	3512	1.590	0.258	0.853	1.487	7.430
std_q	3512	0.999	0.147	0.00298	0.917	3.519
std_a	3512	0.996	0.170	−0.0222	0.922	4.443
size	3525	22.76	1.989	19.66	22.60	27.00
lev	3525	0.489	0.0412	0.0845	0.489	0.911
roa	3511	0.0351	0.00208	−0.129	0.0295	0.174
bm	3524	1.457	2.447	0.0927	0.929	9.419
turnover	3240	0.445	0.110	0.0383	0.354	1.744
age	3525	2.553	0.406	0.693	2.773	3.219
soe	3456	0.619	0.236	0	1	1
abs_da	3511	0.0646	0.0162	0	0.0384	3.015
*top*1	3525	0.372	0.0242	0.0803	0.354	0.771
*big*10	3525	0.631	0.233	0	1	1
record	3525	103.2	2379	22	93	490

table, I report the sample size (size), arithmetic mean (mean), standard deviation (sd), minimum number (min), median number (median),and maximum number (max) of main variables using in the following analysis. All continued variables are winsorized in 1% - 99% range. As can be seen from Table 2, the average of stock price synchronicity index (SYN) is −0.271, the median is 0.448, but the standard deviation is 0.9, indicating that there is a significant difference in stock synchronicity between companies. The average value of the trait information content of the question is 1.602, which is slightly higher than the average value of the trait information content of the answer of 1.590. This is because the language of the investors is more diverse, and the words chosen by different investors will be very different for the same question. The listed companies' answers are more official and standard, and the way of companies responding to the same type of question is similar. The mean value of state-owned enterprises is 0.619, indicating that 62% of the samples are state-owned enterprises. This empirical value is higher than that of the previous literature [6] [14]. This is because we only use the companies traded on the Shanghai Stock Exchange, excluding GEM companies with more private companies. The average shareholding ratio of the company's largest shareholder is 0.372, indicating that there is strong "one big shareholder" phenomenon in listed companies in China.

4.2. Empirical Result: Trait Information and Stock Price Synchronicity

Table 3 reports the impact of the trait information of the e-interaction platform questions and answers on stock price synchronicity. Column (1) reports the impact of trait information in question parts of SSE e-interaction platform. The ecoefficiency of our mainly concerned variable *inf_q* is -0.062 and t-statistics is −1.40, indicating that the trait information in questions have no significant association with stock price synchronicity. Column (2) reports the impact of trait information in answer parts. We can find that the ecoefficiency and t-statistics are −0.093 and −2.20 separately. It can be seen from the regression results that the trait information of answers reduces the stock price synchronism at the 5% significance level. This is consistent with Hypothesis 1, which illustrates the answer part of the e-interaction platform improves the ability of investors to acquire and understand trait information. The investment decision-making contains more company trait information and thus reduces the stock price synchronization. Because the answer is more informative than the question, in the subsequent analysis, this article focuses on the content of the trait information of answers.

4.3. Empirical Result: The Influence of Information Transparency

Table 4 reports the impact of information transparency on the information efficiency of the e-interactive platform. Columns (1) and (2) report the effects of trait information on stock price synchronicity in companies with abnormal accruals

Table 3. Main result.

variables	(1) Question	(2) Answer
inf_q	**−0.062**	
	(−1.40)	
std_q	−0.013	
	(−0.25)	
inf_a		**−0.093****
		(−2.20)
std_a		−0.048
		(−1.01)
size	0.136***	0.132***
	(7.83)	(7.52)
lev	−0.697***	−0.694***
	(−6.93)	(−6.93)
roa	−1.662***	−1.644***
	(−4.48)	(−4.49)
bm	0.013	0.014
	(0.91)	(0.94)
turnover	−0.408***	−0.407***
	(−6.86)	(−6.90)
age	−0.051**	−0.048*
	(−1.98)	(−1.85)
soe	0.212***	0.213***
	(6.26)	(6.31)
Industry	Yes	Yes
Year	Yes	Yes
N	3070	3070
adj. R^2	0.35	0.36

Note: Robust t-statistics adjusted for industry clustering effects are presented in parentheses below the coefficients. *, ** and *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

above the median and below the median. It can be seen from the regression results that in the group with lower abnormal accruals, the trait information reduces the stock price synchronism at a significant level of 1%, and the regression coefficient reaches −0.224, which is significantly higher than the coefficient of the main regression test. Columns (3) and (4) report the impact of trait information content on stock price synchronicity in companies with a higher number of company announcements than the median and below. It can be seen from the regression results that in the group with larger number of company announcements, the trait information content is negatively correlated with the stock price

Table 4. The influence of information transparency.

variables	(1)	(2)	(3)	(4)
	Abnormal accrual		announcement	
	High	Low	High	Low
inf_a	**0.017**	**−0.224*****	**−0.139****	**−0.098***
	(0.30)	**(−3.24)**	**(−2.09)**	**(−1.78)**
std_a	−0.164**	0.080	−0.056	−0.032
	(−2.39)	(1.16)	(−0.82)	(−0.48)
size	0.158***	0.098***	0.186***	0.115***
	(6.58)	(3.68)	(7.33)	(4.67)
lev	−0.734***	−0.658***	−0.815***	−0.427***
	(−5.44)	(−4.32)	(−5.27)	(−3.21)
roa	−1.633***	−1.678***	−2.116***	−1.137**
	(−3.57)	(−2.73)	(−3.64)	(−2.45)
bm	−0.015	0.041*	0.013	0.024
	(−0.72)	(1.83)	(0.70)	(0.96)
turnover	−0.369***	−0.459***	−0.363***	−0.377***
	(−4.63)	(−5.08)	(−3.96)	(−4.92)
age	−0.044	−0.058	−0.070*	−0.031
	(−1.19)	(−1.58)	(−1.84)	(−0.81)
soe	0.171***	0.264***	0.218***	0.064
	(3.57)	(5.39)	(4.31)	(1.36)
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
N	1546	1524	1478	1592
adj. R^2	0.37	0.34	0.36	0.36

Note: Robust t-statistics adjusted for industry clustering effects are presented in parentheses below the coefficients. *, ** and *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

synchronicity at a significant level of 5%. Although the coefficient is also negative in a group with a small number of company announcements, there is only a 10% significance level, and the coefficient absolute value is significantly smaller than the coefficient of a group with a higher number of company announcements. These regression results support Hypothesis 2, indicating that when information transparency is high, investors are more capable of using the e-interaction platform for information acquisition and understanding, and the stock price synchronization is more effective in indicating information efficiency.

4.4. Empirical Result: The Influence of Corporate Governance

Table 5 illustrates the impact of corporate governance on the information efficiency of the e-interactive platform. Column (1) and (2) report the impact of

Table 5. The influence of corporate governance.

variables	(1)	(2)	(3)	(4)
	Top1		Big10	
	High	Low	Yes	No
inf_a	**−0.055**	**−0.133****	**−0.169*****	**0.016**
	(−0.87)	**(−2.28)**	**(−3.22)**	**(0.22)**
std_a	−0.003	−0.095	−0.002	−0.104
	(−0.04)	(−1.36)	(−0.04)	(−1.21)
size	0.135***	0.100***	0.129***	0.116***
	(5.52)	(3.77)	(5.91)	(3.70)
lev	−0.805***	−0.571***	−0.544***	−0.774***
	(−5.42)	(−3.99)	(−4.22)	(−4.56)
roa	−2.161***	−0.880	−1.593***	−1.495**
	(−4.34)	(−1.57)	(−3.45)	(−2.40)
bm	0.013	0.028	−0.011	0.073***
	(0.71)	(1.03)	(−0.63)	(2.80)
turnover	−0.403***	−0.448***	−0.308***	−0.602***
	(−4.82)	(−5.06)	(−4.14)	(−6.05)
age	−0.075**	−0.010	−0.059*	−0.074
	(−2.06)	(−0.25)	(−1.79)	(−1.65)
soe	0.254***	0.199***	0.256***	0.178***
	(5.01)	(3.90)	(6.03)	(2.97)
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
N	1635	1435	1944	1126
adj. R^2	0.35	0.37	0.37	0.36

Note: Robust t-statistics adjusted for industry clustering effects are presented in parentheses below the coefficients. *, ** and *** indicate significance at the 0.10, 0.05 and 0.01 levels, respectively.

trait information content on stock price synchronicity among companies with a higher shareholding ratio than the median and below the median. In the group with higher share proportion of biggest shareholder, there is no significant association between *inf_q* and main dependent variable *SYN* (t-statistics is −0.87) while in the group with lower share proportion of biggest shareholder, *inf_q* has negative association with stock synchronicity, with t-statistics of −2.28 and ecoefficiency of −0.133, much lower than those in higher group. The regression results show that the trait information is significantly negatively correlated with stock price synchronism in the group with the lower shareholding ratio of the largest shareholder. Column (3) and (4) respectively report whether the content of trait information affects stock price synchronicity among companies audited by the top ten accounting firms. The results show that trait information has sig-

nificantly reduced stock price synchronicity in companies audited by the top ten accounting firms. This set of regression results supports Hypothesis 3, indicating that when corporate governance is high, investors are more capable of using the e-interaction platform for information acquisition and understanding, and the stock price synchronization is more effective.

5. Conclusion

Through the unique scene of the e-interaction platform, this paper constructs a direct measure of the ability of investors to acquire and understand the trait information in the e-interaction platform by using the text vectorization method. This paper finds that the higher proportion of the trait information in e-interactive platform, the lower the stock price synchronization. Further research finds that companies with high information transparency and high corporate governance have more individual information are more effective in reducing the stock price synchronicity. These results show that investor information acquisition and understanding ability can have a reducing impact on stock price synchronicity, supporting the theory of stock price synchronicity being an information efficiency indicator. This paper provides new empirical evidence for the "information efficiency view", deepening the research on stock price synchronization in China's stock market, and providing a useful reference for the future development of e-interaction platform.

6. Discussion

The possible contributions of this paper are as follows: 1) In order to measure trait information, this paper uses the method of text analysis [8] [9] to propose a way of directly measuring heterogeneous information, which provide a new perspective in future studies in stock price synchronicity; 2) This paper proposes the theoretical reasoning and empirical results of investors' ability to acquire and understand information affecting stock price synchronicity, enriching the research on stock price synchronization "information efficiency view" [1] [2] [5].

However, this paper is only a useful exploration of using textual data to study the problem of stock price synchronization. There are still inherent defects in vectorization technology [8] [9]. The current textual analysis stays in the word bag (bag of words) technology, cannot understanding the semantics of the text, but only capturing frequency and other statistical features to depict the content of the text. This leads to problems such as the inability to judge synonyms and the inability to judge the intrinsic nature of trait information based on the overall sentence. These problems have yet to be developed and perfected in the follow-up study.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Morck, R., Yeung B. and Yu, W. (2000) The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock Price Movements. *Journal of Financial Economics*, **58**, 215-260. https://doi.org/10.1016/S0304-405X(00)00071-4

[2] Jin, L and Myers, S.C. (2006) R^2 around the World: New Theory and New Tests. *Journal of Financial Economics*, **79**, 257-292. https://doi.org/10.1016/j.jfineco.2004.11.003

[3] Hutton, A.P., Marcus, A.J. and Tehranian, H. (2009) Opaque Financial Reports, R^2, and Crash Risk. *Journal of Financial Economics*, **94**, 67-86. https://doi.org/10.1016/j.jfineco.2008.10.003

[4] Hou, K.W., Wei, X. and Lin, P. (2006) R^2 and Price Inefficiency. Working Paper. http://ssrn.com/abstract=954559

[5] You, J.X. (2017) The Resurgence of R^2: Review and Prospect on the Studies on Stock Price Synchronicity. *Journal of Management Science and Engineering*, **20**, 63-79.

[6] Zhang, T.W., Wang, W. and Yu, C.T. (2016) Financial Knowledge, Risk Awareness and Investment Income: Based on the Small Investors' Rights and Interests Protection Questionnaire. *Accounting Research*, **7**, 66-73.

[7] Ding, H., Lyu, C.J. and Huang, H.J. (2018) Social Media, Investor Sophistication and Earning Expectation: Evidence from SSE EInteraction. *Economic Research*, **53**, 153-168.

[8] Hanley, K.W. and Hoberg, G. (2010) The Information Content of IPO Prospectuses. *Review of Financial Studies*, **23**, 2821-2864. https://doi.org/10.1093/rfs/hhq024

[9] Meng, Q.B., Yang, J.H. and Lu, B. (2017) The Informative Content of Management Discussion and Analysis and Stock Price Crash Risk: Based on Text Vectorization Method. *China Economist*, **12**, 132-150.

[10] Roll. R. (1988) R^2. *Journal of Finance*, **43**, 541-566. https://doi.org/10.1111/j.1540-6261.1988.tb04591.x

[11] Zhai, L.Y. (2004) Information, Investor Behavior and Capital Market's Efficiency. *Economic Research*, No. 3, 47-54.

[12] Pirolli, P. (2007) Information Foraging Theory: Adaptive Interaction with Information. Oxford University Press, Oxford, 153-156. https://doi.org/10.1093/acprof:oso/9780195173321.001.0001

[13] Li, Z.Q., Ye, Q. and He, H. (2011) Inter-Firm Relationship, Corporate Transparency and Stock Price Characteristics. *Accounting Research*, No. 1, 44-51.

[14] Wang, L.Z., Wang, Y.M. and Wang, Z.C. (2016) Stock Price Synchronicity and the Separation of Ownership between Control Rights. *Financial Research*, No. 5, 97-110.

[15] Zhang, B. and Wang, Y.T. (2014) Business Complexity, Industry Expertise of Independent Directors and Stock Price Synchronicity. *Accounting Research*, No. 7, 36-42.

[16] You, J.X. (2006) Institutional Construction, Trait Information and Stock Price Synchronicity. *China Economic Quarterly*, **23**, 18.

[17] Zhou, L.J. (2014) Corporate Governance, Institutional Ownership, and Stock Price Synchronicity. *Financial Research*, No. 8, 146-161.

[18] Wang, Y.P., Liu, H.L. and Wu, L.S. (2009) Information Opacity, Institutional Investor and Stock Price Synchronicity. *Financial Research*, **12**, 162-174.

[19] You, J.X. and Wu, J. (2012) Spiral of Silence: Media Sentiment and the Asset Mispricing. *Economic Research*, **47**, 141-152.

[20] Gul, F.A., Kim, J.B. and Qiu, A.A. (2010) Ownership Concentration, Foreign Share-

holding, Audit Quality, and Stock Price Synchronicity: Evidence from China. *Journal of Financial Economics*, **95**, 425-442.
https://doi.org/10.1016/j.jfineco.2009.11.005

[21] Xu, W. and Chen, D.H. (2016) Information Effect of "We Media" Disclosure: Evidence from Sina Weibo. *Financial Research*, No. 3, 157-173.