INVESTOR ATTENTION, MARKET LIQUIDITY AND STOCK RETURN: A NEW PERSPECTIVE

Check for updates

(iD) **BinWang¹**
(iD) **Wen Long²⁺**
(iD) **Xianhua Wei³**

¹,²School of Economics & Management, University of Chinese Academy of Sciences, Beijing, P.R.China; Research Center on Fictitious Economy & Data Science, Chinese Academy of Sciences, Beijing, P.R.China; Key Laboratory of Big Data Mining & Knowledge Management, Chinese Academy of Sciences, Beijing, P.R.China
¹E-mail: wangbin750@126.com   Tel: +86-18611430500
²E-mail: longwen@ucas.ac.cn Tel: +86-13811734056
³School of Economics & Management, University of Chinese Academy of Sciences, Beijing, P.R.China; Research Center on Fictitious Economy & Data Science, Chinese Academy of Sciences, Beijing, P.R.China
³E-mail: weixh@ucas.ac.cn   Tel: +86-18600107077

(+ Corresponding author)

ABSTRACT

We propose a new method to measure the investor attention paid to a specific industry using search data from search engine. Instead of taking company names or stock codes as keywords, we select keywords from a corpus of texts concerning a given industry by text-analysis technique such as TextRank algorithm. Two indices were constructed by principal component analysis method, including a positive index and a negative index. The empirical analysis demonstrates that the influence of investor attention on market liquidity is coincident and significant, and the effect on industry stock index return is less significant.

Contribution/ Originality: The paper's primary contribution is finding that investor attention have both positive and negative effect on stock market, and the influence of investor attention on market liquidity is coincident and significant, but that on industry stock index return is less significant.

1. INTRODUCTION

This paper focus on the topic of investor attention. Scholars have used various kinds of data as a proxy of investor attention because of immeasurability of attention. Grullon *et al.* (2004); Chemmanur (2009); Vorkink *et al.* (2010) think that advertising expenditure is a proxy of investor attention, the more advertising expenditure is, the more attention is paid by investors. Peress (2008) indicates that more attention brings higher return and larger volume to a corporate stock by employing the frequency of a company's reports by The Wall Street Journal as a proxy of investor attention. Similar to Peress (2008); Rao *et al.* (2010) use the amount of report in *Sina News* as a proxy, and their result shows that a stock grabbing more attention tends to undertake a relative lower return, which supports the "over attention underperformance" hypothesis. Gervais *et al.* (2001); Berkman *et al.* (2012)

341

choose volume as a proxy of investor attention. The former article demonstrates that abnormal volume enables a stock draw more attention from investors, which may cause price rising in a month. The latter article constructs a profitable trading strategy based on their findings that stocks with more attention are inclined to undertake excessive open price. Seasholes and Wu (2007) consider that when a stock reaches its limit up, investors would pay more attention to it, hence limit up can be a good proxy of investor attention. Similarly, Yu and Hsieh (2010) think extreme return to be a good proxy of investor attention. Loh (2010) takes turnover rate as an investor attention proxy and concludes that high attention induces high information efficiency.

All the indicators mentioned above are indirect measurement for investor attention. During the last three decades, internet has become an important component of the society. Internet is affecting almost every aspect of our life. Now we can get information on almost everything through internet, and meanwhile, what we do on internet is all recorded down by web logs. These online behavior records are amazing data sources for research. In financial field, many researchers employ these data to analyze investor attention. On internet, investors may express their interest to a certain stock by searching related keywords with a search engine or by joining in online discussion forums. Since the online behavior of these investors can be recorded, these records, if properly obtained, can be used as research data for investor attention. Such data from search engines and internet forums are widely used nowadays as a good proxy of investor attention. Employing Google Trends as data source and using stock symbol as keyword, Da *et al.* (2011) find that, regarding retail investors, investor attention can be represented by search data and high investor attention will generate temporal return abnormal and long-run return reversal. Dimpfl and Jank (2016) select Google search data of the keyword "Dow" as a proxy of investor attention to stock market, indicate that investor attention is the Granger cause of realized volatility of the Dow Jones. Aouadi *et al.* (2013) find significant correlation between investor attention and stock liquidity as well as stock volatility by using Google search data with company names as keywords.

Previous studies mainly focus on two kind of investor attention: market-wide attention and specific attention. For market-wide attention, researchers aim to find out the relationship between stock market index and investor attention paid to stock market. In these studies, researchers employed some market-wide measurements as attention proxy, such as search volume of keyword "stock market", "S&P500", "DJI". These will imply how many persons are interested in stock market and how is the passion of investors. For specific attention, researchers are interested in a cross-sectional problem, whether change of attention to a certain stock affects its return, volatility and other market indicators. Proxy variables used in these researches are related to some specific stocks, such as search volume of company names or stock codes. As indicated by Barber and Odean (2008) when investors pay attention to a certain stock, they tend to purchase it. Because of these investors' behavior, the stock price would experience temporary rising and then mean reversion. In previous studies, the method to measure investor attention selects stock index codes, company names and stock codes as searching keywords, and then takes search volume of these keywords as proxy variable. This method is so direct, therefore so narrow. In our view, search volume of other keywords is also informative to measure investor attention. As an example, an investor can also pay his attention to stock market or stock of a specific company by searching for name of a policy related to stock market or main product of this company. Based on this thought, we propose a new method to measure investor attention using search data of alternative keywords except traditional keywords like company names and stock codes.

There are so many factors related to the whole stock market, including economic growth, inflation, technical innovation, monetary policy and so on. Because of difficulty to find out most of keywords related to the whole stock market, we limit the scope of our research to industry level. On the other hand, industry is also an important aspect of stock market research. Industry factor can explain a significant portion of stock returns, especially in Chinese stock market. King (1996) found significant evidence supporting existence of industry effect. In empirical study of Livingston (1977) it is indicated that industry factor accounts for 26% of stock return change. By utilizing restricted regression model, Xiong and Yang (2006) make an empirical research on industry and style effects, and find that

Chinese stock market has obvious industry and style effects, and industry effects are more significant. Therefore, an industry-oriented study is also meaningful for investment practice.

Benefitting from development of technology in NLP (natural language processing), we are able to choose words selected by a text-analysis technique as search keywords, such as industry policy, main product or relevant concept, to acquire their search volume as a more comprehensive proxy for investor attention. A possible concern about this approach is that search volume of keywords is not only generated by investors, which may result in that search volume of those keywords is not an appropriate proxy. However, investor attention is a complex concept, which can be generated from the interaction between different investors, or from the interaction between investors and the public. The attention paid by some investors to a certain stock or a certain industry will itself draw the attention of other investors, and the attention paid by the general public, which we define as social attention here, will also have a similar effect. Many investors will consider social attention distribution as a reference item when deciding their portfolio. They tend to select stocks that get more social attention. In this sense, the indicator generated by our approach is still a good proxy for investor attention with acceptable noise.

The difference between our more comprehensive measure for investor attention and previous direct and simple measure using company names or stock codes leads us a new finding about investor attention. Theoretically, investor attention is a question about the allocation of finite cognitive resource by investors. When the allocation is unbalanced, investor attention would affect market. In previous studies, researchers focus on two kinds of unbalance, one on horizon, and the other on cross-section. The former induces that the market portfolio return goes up or down along with attention paid by investors to the whole stock market. The latter induces that stocks with more attention will achieve higher return than others. These two kinds of unbalance are among different entities. However, attention can not only be paid to stock market or certain stock, and can be paid to information like new law, government policy, product and many relevant concepts. This allocation of finite cognitive resource would decide the speed of stock price adjusting as a result of some new information. For example, the issue of a more rigorous national energy-conserving criterion will decrease income of energy enterprise and potential value of this enterprise. In extreme situation, no investor knows this new criterion, and the stock price will not change immediately. The extent that investors pay attention to this new criterion decides speed and extent of stock price adjusting, resulting in under-reacting or over-reacting. On the other hand, keywords will have different effect on stock market performance, just thinking the difference effects on energy industry between an energy-saving policy and a more efficient refining technology emerging. We not only focus on investor attention on information around stocks, but also distinguish their different effect on stock market performance. Therefore, we separate investor attention into two different sides, a positive one and a negative one, which is the most important innovation of this paper. We investigate how positive and negative investor attention affect market through empirical studies focus on market liquidity and stock return.

In this paper, we take four industries as examples, including Real Estate, Energy, Health-Care and Automobile & Components. The paper is organized as follows. Section 2 describes the method to construct the industry investor attention index. Section 3 presents the main results on how these indices affect stock market performance. Section 4 concludes the paper.

2. INDEX CONSTRUCTION

In order to construct the index of investor attention, firstly, we select alternative keywords. The simplest way is to employ frequency criterion, however, this method usually leads to a poor result. Then many other methods are explored, such as GenEx developed by Turney (2002) combining an automatic system for keywords extraction with a genetic algorithm, or a supervised learning system suggested by Hulth (2003). Mihalcea and Tarau (2004) indicated that TextRank algorithm is a successful method for keywords extraction. TextRank algorithm is a graph-based algorithm derived from PageRank algorithm, which can be used to extract keywords from an article. Its basic

idea is that to convert text to a weighted graph firstly, and then find central vertices in the graph by computing a defined score associated with every vertices. Words represented by these central vertices are important in text, which are keywords of the article.

The main steps of TextRank algorithm are:

(1) Identify a text unit best aiming at analysis target, and then make each word in the unit as a vertex;

(2) Designate any two words as a pair of vertices, if there is any relationship between the pair, define the relationship as an edge, then map every edge out to hold up a whole graph;

(3) Assign arbitrary initial score to every vertex, then iterate until convergence;

(4) Use the final score of every vertex to select keywords.

In step 1, we remove stop words from our target articles firstly, and then we employ NLPIR, a Chinese Word Segmentation System, for elementary text analysis to acquire meaningful words and phrases as vertices in graph from Chinese articles. Chinese is different from English, because it has no space as natural separator, and the meaningful lexical units in Chinese are always words or phrases consisting of multiple characters rather than one single character, which usually make the borderlines between characters, words or phrases not so clear. Characters occurring in a particular order may have different meanings only by different word segmentation. This makes Chinese Word Segmentation a challenge. NLPIR is the leading Chinese Word Segmentation System with high accuracy and usability. In step 2, any relationship between two words/phrases is a potential connection that can be used as an edge in graph. In this paper, in order to define the relationship, we use co-occurrence criterion, which is, two vertices are considered linked with each other when the words/phrases they represented co-occur within a window of maximum N words/phrases, and the co-occur times is the weight assigned to the edge between two vertices. N varies from 2 to 10, and we set it to 10. After step 1 and step 2, we will get a graph like Fig. 1.
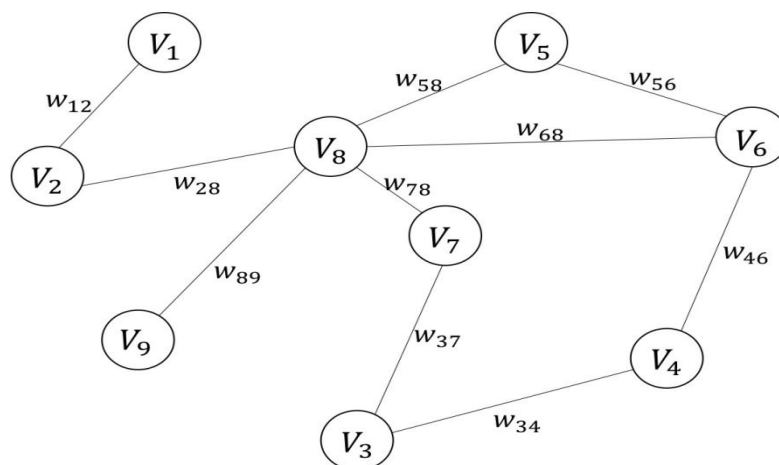

Fig-1. An example of graph of TextRank

In Fig. 1, $V_i, i = 1, 2, \ldots, 9$, is the vertex representing a word or phrase in articles, w_{ij} is the weight assigned to the corresponding edge of the graph. To find out central vertices in the graph, we adopt a basic idea of "voting" or "recommendation". When a vertex is linked to another vertex, it votes for the other one, and the weight assigned to the edge connecting them is the number of votes. The more votes casting to a vertex, the more importance of this vertex. Moreover, the importance of one vertex will determine its voting weight to other vertices, which is also taken into account by employing TextRank algorithm. Therefore, the score of one vertex depends on the number of votes from other vertices, as well as the score assigned to them.

So we let $WS(V_i)$ be the score of V_i in the weighted graph, $LINK(V_i)$ be the set of vertices that link to V_i, then we define $WS(V_i)$ as:

$$WS(V_i) = (1 - d) + d \times \sum_{V_j \in LINK(V_i)} \frac{w_{ij}}{\sum_{V_k \in LINK(V_j)} w_{jk}} WS(V_j) \tag{1}$$

Where d is a damping parameter that can be set between 0 to 1.

In step 3, we assign arbitrary initial score, usually 1, to every vertex, and decide a threshold. Then the algorithm will be iterated until convergence satisfying (1). Finally, we sort words or phrases by scores assigned to them and choose key words from the article.

For selected industries, we totally collect 158 articles from the news database Wind Financial by searching keyword "Real Estate", "Energy", "Health-Care" and "Automobile & Components" in Chinese. These articles constitute our corpus. Then we apply TextRank algorithm to the corpus to extract keywords for every industry, sorted by their importance in text. By this algorithm, we obtain an aggregate of alternative keywords. After this step, we get selected keywords by means of calculating correlation coefficient between weekly return of industry stock index and weekly search volume of these alternative keywords. When calculating correlation coefficient, we use four CSI 300 industry segmentation indices, which are, index of Real Estate industry, as well as index of Energy industry, of Health-Care industry and of Automobile & Components industry. In china, Baidu is the search engine with the highest market share, so we download daily search data of keywords on Baidu search engine by web-crawler. The sample period is from Jan.6, 2013 to Apr.5, 2015.

In data process and data analysis, we find that daily search data of keywords is full of noise and this data has a behavior pattern in a period of a week. For eliminating noise and periodicity from daily data, we convert it to weekly data by summation. Then, we filter out weak correlated keywords by the criterion that absolute value of correlation coefficient mentioned above is smaller than a threshold. We set the threshold 0.2, because larger or smaller are not appropriate. If we set the threshold 0.1, there are too many keywords left, in addition, full of words of nonsense. On the contrary, few keywords' correlation coefficient can exceed 0.25 or furthermore 0.3. After filtering out by correlation coefficient, there are still some keywords left without any economic meaning to Energy industry, like stop words and general words. Therefore, we select practical keywords manually in order to ensure that these keywords are truly related to these four industries, rather than only correlated to them on data.

According to the signs of correlation coefficients, these keywords can be divided into two opposite categories: positive keywords for positive investor attention, and negative keywords for negative investor attention. Table 1 shows these two types of selected keywords.

Table-1. Keywords of four industries

Industry	Positive keywords	Negative keywords
Real Estate	Asset, residence, finance, property developers, inventory, house tax, urbanization	Public finance, investment, trust company, broker
Energy	Nuclear power, shale oil, crude oil, crude oil price, shale	Motor, environmental problem, car market, solar energy, energy conservation and emission reduction, coal mine, energy conservation, car, motor vehicle, chemurgy
Health-Care	Public hospital reform, internet, regroup	Registration, drug, traditional Chinese medicinal materials, pharmaceuticals industry, medicinal materials, medical treatment
Automobile & Components	Electromobile, imported car, 4s store	Domestic car, motor, taxi, power, city, internet

After obtaining weekly search data of those keywords from Jan.6, 2013 to Apr.5, 2015, we construct positive indices using principal component analysis (PCA) to search volumes of positive keywords. We take the first principal component as the positive attention index, denoted by PI. We can also get negative attention index NI using the same method. According to Da Z. et al. (2011), our key variables in the paper, API, ANI, are defined as

$$API_t = PI_t - Med(PI_{t-1}, ..., PI_{t-8}) \qquad (2)$$

$$ANI_t = NI_t - Med(NI_{t-1}, ..., NI_{t-8}) \qquad (3)$$

Where $Med(...)$ denotes the median value. The median over a longer time window (approximately 8 weeks) captures the "normal" level of attention that is robust to recent jumps. These indices will also remove time trends.**3.**

3. EMPIRICAL ANALYSIS

3.1. Data

We choose four industries from CSI300 industry system, including Real Estate industry, Energy industry, Automobile & Components industry and Health-Care industry. CSI300 is a stock index compiled with share prices of the top 300 stocks with the largest market capitalization and best stock liquidity in Chinese share market. CSI300 aims to measure the performance of all stocks in both Shanghai stock exchange and Shenzhen stock exchange, so its components stocks should have no evidence of serious financial or legal problems and show no trace of stock manipulation or insider trading. These four industries, Real Estate industry, Energy industry, Automobile & Components industry and Health-Care industry, the first three of which are cyclical industries covering finance, energy and consumption areas, and the last of which is non-cyclical industry.

From our data resource Wind database, we download daily return data of corresponding indices of these four industries, and daily return, close price, trading volume and market value of component stocks in these indices. The sample period is from 2013.1.6 to 2015.4.5. Fig. 2 shows the trend of the four industry indices.
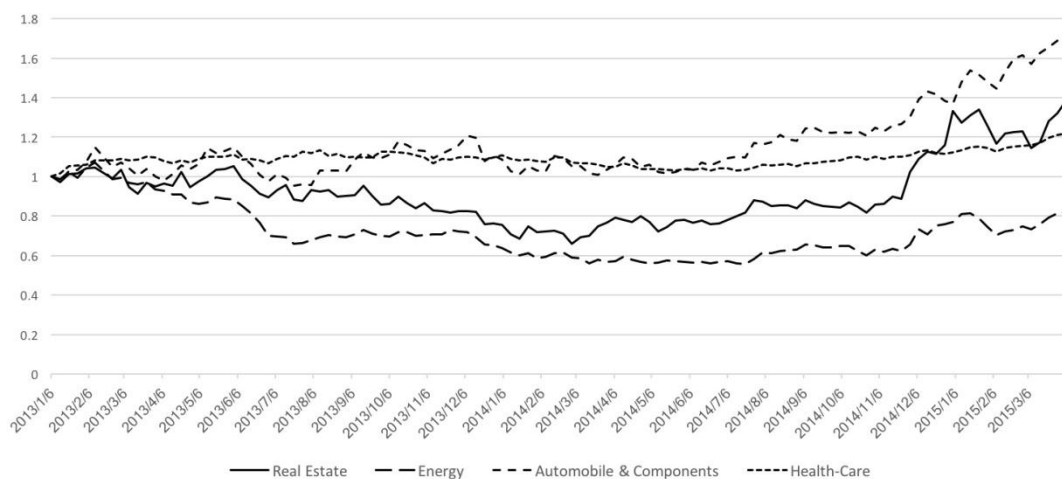

Fig-2. Four industry indices of CSI300

Source: Wind Database.

3.2. Investor Attention and Market Liquidity

Previous studies have demonstrated that investors tend to trade stocks receiving more attention. Since we separate investor attention into two different sides, a positive one and a negative one, do these two opposite

investor attention have different effect on stock market liquidity? In our analysis frame, the answer is yes. As a result of difficulty in short-selling, investors especially individual investors incline to trade stocks that are likely to rise up in their imagination. The positive investor attention index rising means that investor pay more attention to positive information, so they think corresponding stock's price will go up and participate in trading of this stock actively. In addition, this kind of investor activity makes liquidity of corresponding stock better. On the contrary, when investors pay more attention to negative information and the negative investor attention index rises, investors tend to avoid trading corresponding stocks, which makes liquidity of these stocks exhausted gradually. In order to test the hypothesis above, we apply Amihud (2002) measure, which is calculated for each stock i for each week j as follows:

$$Amihud_{ij} = \frac{1}{T}\sum_t^T \frac{1000000 \times |return_{it}|}{price_{it} \times volume_{it}} \quad (4)$$

Where T is the number of trading days in week j. The Amihud measure is a price-impact measure of illiquidity and reflects the degree to which price move in response to trading volume changes. It is important to note that a higher Amihud measure indicates lower liquidity. We then aggregate the Amihud illiquidity measure by taking the market value weighted average of the individual stock illiquidity measure for those stocks contained in a given set. At last, we take the log of average illiquidity measure as stock market illiquidity indicator as follows:

$$Amihud_j = \ln\left(\frac{\sum_{i \in I} Amihud_{ij} \times market\_value_{ij}}{\sum_{i \in I} market\_value_{ij}}\right) \quad (5)$$

Where I is a set of stocks. If I is the set of constituent stocks of a certain industry stock index, we can obtain the illiquidity indicator of this industry. Similarly, when I is the set of all stocks listed in the market, we obtain the illiquidity indicator of the whole stock market.

The model used to investigate how investor attention affect industry liquidity is:

$$Amihud_t = c + \vec{a} \cdot \overrightarrow{IV} + \beta_1 \cdot Amihud_{mt} + \beta_2 \cdot res_{t-1} + \varepsilon_t \quad (6)$$

where $Amihud_t$ is the illiquidity indicator of a certain industry at time t, $\overrightarrow{IV} = (API_t, API_{t-1}, API_{t-2}, API_{t-3}, API_{t-4}, ANI_t, ANI_{t-1}, ANI_{t-2}, ANI_{t-3}, ANI_{t-4})^T$ which have been defined above, $Amihud_{mt}$ is the simultaneous illiquidity indicator of the whole stock market, res_{t-1} is one-period-lag residual obtained from a model regressing $Amihud_t$ on $Amihud_{mt}$. We think that res_{t-1} can represent other effective factors except the whole market illiquidity $Amihud_{mt}$. We estimate model (6) by step-wise method under AIC criterion. Table 2 represents estimation results of these industries according to model (6).

Table-2. Relationship between the illiquidity of industry and investor attention

	Real Estate	Energy	Automobile & Components	Health-Care
API_t				
API_{t-1}		–0.21410*** (0.03080)	–0.07916*** (0.02361)	–0.04156 (0.02851)
API_{t-2}	–0.14812 (0.09119)			0.10223*** (0.03296)
API_{t-3}				
API_{t-4}				
ANI_t	0.04923 (0.03373)		0.04809** (0.02309)	0.07682*** (0.02522)
ANI_{t-1}	–0.07088* (0.03609)			
ANI_{t-2}	0.15637* (0.09346)		0.04364* (0.02410)	
ANI_{t-3}				–0.07391** (0.03366)
ANI_{t-4}			0.04823** (0.02203)	0.04417 (0.02528)*
$Amihud_m$	1.11401*** (0.09405)	1.24394*** (0.08331)	1.04007*** (0.06363)	0.85997*** (0.07224)
res_{t-1}	0.65900*** (0.08653)	0.57677*** (0.06885)	0.69908*** (0.07766)	0.28780*** (0.09328)
Observations	111	111	111	111

*indicates significance at the 10% level
**indicates significance at the 5% level
***indicates significance at the 1% level

In all of these four models，the parameters of $Amihud_m$ and res_{t-1} are significant, which demonstrates that market-wide and industry-specific factors have significant influence on industry liquidity.

For Energy, Automobile & Components and Health-Care industry, API has significant effect on their illiquidity. For all of these four industries, the parameters of API are negative in a short time, which means that positive investor attention rising makes industry liquidity better in short run. ANI has significant influence on Real Estate, Automobile & Components and Health-Care industry illiquidity. The short-term positive parameters of ANI show that negative investor attention rising induces lower market liquidity. These empirical results support our analysis above, and they confirm that positive investor attention is positively associated with market liquidity positively in short run. In addition, the correlation is negative for negative investor attention. We can find another phenomenon that market revise the effect caused by investor attention in subsequent time horizon. For example, in Health-Care industry, ANI_t and ANI_{t-3} have opposite effect on market liquidity. This phenomenon also occurs in Real Estate and Automobile & Components industry models.

3.3. Investor Attention and Stock Return

According to previous studies, stock market index goes up with investor attention rising, and stocks that receive more investor attention is able to achieve higher return in a short time. In our analysis frame, investor attention has influence on the speed that new information integrates into stock price. Therefore, the positive

investor attention index rising means that investor pay more attention to positive information about a certain industry, and this should push up the industry stock index. Similarly, rising of the negative investor attention index should bring down the industry stock index.

To test how our positive attention index and negative attention index affect the return of industry segmentation index, we construct a model based on Fama-French three-factor model:

$$IIR_t = c + \vec{\alpha} \cdot \overrightarrow{IV} + \beta \cdot RMRF_t + s \cdot SMB_t + h \cdot HML_t + \varepsilon_t \qquad (7)$$

Where IIR_t is the excess return of industry segmentation index at time t, $\overrightarrow{IV} = (API_t, API_{t-1}, API_{t-2}, API_{t-3}, API_{t-4}, ANI_t, ANI_{t-1}, ANI_{t-2}, ANI_{t-3}, ANI_{t-4})^T$, $RMRF_t$, SMB_t and HML_t are the three factors in Fama-French three-factor model. $RMRF$ is the difference between return of the market portfolio and the risk-free return rate, SMB is the difference between return of small market capitalization portfolio and that of big market capitalization portfolio, and HML is the difference between return of high book-to-market ratio portfolio and that of low book-to-market ratio portfolio. Using this model, we analyze the influence of positive or negative investor attention on the return of corresponding industry segmentation index. We estimate model (7) also by step-wise method under AIC criterion. Table 3 represents estimation results of these industries according to model (7).

Table-3. Relationship between the return of industry index and investor attention

	Real Estate	Energy	Automobile & Components	Health-Care
API_t		-0.0056** (0.0026)		
API_{t-1}	0.0063** (0.0027)			
API_{t-2}			0.0054** (0.0021)	
API_{t-3}			-0.0038 (0.0022)	
API_{t-4}				
ANI_t				
ANI_{t-1}			-0.0079** (0.0024)	
ANI_{t-2}				-0.0023** (0.0010)
ANI_{t-3}				0.0034*** (0.0010)
ANI_{t-4}				
$RMRF$	1.1222*** (0.0029)	1.0388*** (0.0531)	1.1502*** (0.0879)	0.3309*** (0.0315)
SMB		-0.4755*** (0.0801)	-0.5768*** (0.1548)	
HML	0.4562 (0.2476)		-0.7239*** (0.2276)	-0.3209*** (0.0675)
Observations	111	111	111	111

*indicates significance at the 10% level
**indicates significance at the 5% level
*** indicates significance at the 1% level

In all of these four models，the parameters of RMRF are significant, which demonstrates that RMRF is the most powerful factor among Fama-French three factors. The parameter of RMRF is greater than 1 in model of Real Estate industry, as well as in model of Energy industry and Automobile & Components industry. On the contrary, it is less than 1 in model of Health-Care industry. These results agree well with the common perception that the first three industries are cyclical, and the last one is non-cyclical. For Real Estate industry index and Energy industry index, only positive attention index has significant influence on their returns. However, the parameter of API_{t-2} is positive significantly in the model of Real Estate industry, while API_t is negative significantly in the model of Energy industry. For Health-Care industry index，ANI_{t-2} and ANI_{t-3} both have significant effect on its return, yet it should be noted that the former is negative and the latter is positive. The return of Automobile & Components industry index is affected significantly by both positive attention index and negative attention index simultaneously. ANI_{t-1} has negative effect on the return of Automobile & Components industry index.

We can summarize as follows. Firstly, among these four industries, the impact of investor attention on index returns is less significant than on market liquidity. Secondly, in Health-Care industry and Automobile & Components industry, we can also find out the evidence that market revises price fluctuation caused by investor attention in subsequent time horizon. For Health-Care industry, the second order lag negative index will depress the index return, and the third order lag negative index has a positive influence on the index return conversely. These two effects are both significant. Similar phenomenon occurs in Automobile & Components industry with a little difference, that is, the second order lag positive index will have a positive influence on the index return, but the third order lag positive index will depress the index return.

4. CONCLUSION

In this paper, we focus on how to measure the investor attention with a more comprehensive approach. Just like previous studies, we employ search data from search engine as proxy of investor attention. Our innovation is that, instead of company names or stock codes, we choose representative keywords from a corpus of texts concerning a given industry by text-analysis technique, and then construct two indices with PCA method, a positive one and a negative one, according to different relations between keywords and stock market performance.

In empirical analysis, we choose four representative industries, including Real Estate industry, Energy industry, Automobile & Components industry and Health-Care industry, the first three of which are cyclical industries covering finance, energy and consumption areas, and the last of which is non-cyclical industry. We construct industry-oriented investor attention indices to describe the meso-level type of investor attention by our new approach to analyze the influence of investor attention on stock market. Our empirical analysis demonstrates that positive investor attention has positive influence on stock market liquidity and return, and negative investor attention have the opposite influence. However, the influence is temporary, market would revise it later. In addition, the influence on market liquidity is more coincident and significant. These results indicate that individual investors' trading activities are more likely affected by attention factor in Chinese stock market, as their relatively large contribution to trading volume but weak power on market pricing.

In our research, we find that investor attention is still a complex concept. In our analysis frame, investor attention is a question about how investors allocate their finite cognitive resource; allocation of finite cognitive resource would decide the speed of stock price adjusting as a result of some new information. Therefore, investor attention is not only associated with stocks, but also information around stocks. With this new thinking, we extend measure method for investor attention to a more comprehensive and complex level. Then we are planning to apply

our method to more industries and optimize the algorithm for keywords selection. Then we will use our indices as a weighting factor in research of investment decision-making and financial risk management in the near future.

REFERENCES

Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets, 5(1): 31-56. *View at Google Scholar | View at Publisher*

Aouadi, A., M. Arouri and F. Teulon, 2013. Investor attention and stock market activity: Evidence from france. Economic Modelling, 35: 674-681. *View at Google Scholar | View at Publisher*

Barber, B.M. and T. Odean, 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies, 21(2): 785-818. *View at Publisher*

Berkman, H., P.D. Koch, L. Tuttle and Y.J. Zhang, 2012. Paying attention: Overnight returns and the hidden cost of buying at the open. Journal of Financial and Quantitative Analysis, 47(04): 715-741. *View at Google Scholar | View at Publisher*

Chemmanur, T., 2009. Advertising, attention, and stock returns. (Doctoral Dissertation, School of Business Administration, Fordham University).

Da, Z., J. Engelberg and P. Gao, 2011. In search of attention. Journal of Finance, 66(5): 1461-1499. *View at Google Scholar*

Dimpfl, T. and S. Jank, 2016. Can internet search queries help to predict stock market volatility? European Financial Management, 22(2): 171-192. *View at Google Scholar | View at Publisher*

Gervais, S., R. Kaniel and D.H. Mingelgrin, 2001. The high-volume return premium. Journal of Finance, 56(3): 877-919. *View at Google Scholar | View at Publisher*

Grullon, G., G. Kanatas and J.P. Weston, 2004. Advertising, breadth of ownership, and liquidity. Review of Financial Studies, 17(2): 439-461. *View at Google Scholar | View at Publisher*

Hulth, A., 2003. Improved automatic keyword extraction given more linguistic knowledge. In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics. pp: 216-223.

King, B.F., 1996. Market and industry factors in stock price behavior. Journal of Business, 39(S1): 139-190. *View at Google Scholar | View at Publisher*

Livingston, M., 1977. Industry movements of common stocks. Journal of Finance, 32(3): 861-874. *View at Google Scholar | View at Publisher*

Loh, R.K., 2010. Investor inattention and the underreaction to stock recommendations. Financial Management, 39(3): 1223-1252. *View at Google Scholar | View at Publisher*

Mihalcea, R. and P. Tarau, 2004. Textrank: Bringing order into text. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing.

Peress, J., 2008. Media coverage and investors' attention to earnings announcements. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2723916.

Rao, Y.L., D.F. Peng and D.C. Cheng, 2010. Does media attention cause abnormal return?-Evidence from China's stock market. Systems Engineering-Theory & Practice, 30(2): 287-297.

Seasholes, M.S. and G. Wu, 2007. Predictable behavior, profits, and attention. Journal of Empirical Finance, 14(5): 590-610. *View at Google Scholar | View at Publisher*

Turney, P.D., 2002. Learning to extract keyphrases from text. arXiv preprint cs/0212013.

Vorkink, K., E. DeRosia, G. Christensen and G.R. McQueen, 2010. Advertising, visibility, and stock turnover. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1572097.

Xiong, S.J. and C.J. Yang, 2006. An empirical research on industry effects and style effects in Chinese stock markets. Systems Engineering-Theory & Practice, 26(4): 44-49.

Yu, H.Y. and S.F. Hsieh, 2010. The effect of attention on buying behavior during a financial crisis: Evidence from the Taiwan stock exchange. International Review of Financial Analysis, 19(4): 270-280. *View at Google Scholar | View at Publisher*