Quantifying the effect of investors' attention on stock market

Zhen-Hua Yang[1,3]**, Jian-Guo Liu**[2,4]*****, Chang-Rui Yu**[1]*****, Jing-Ti Han**[2]

1 School of Information Management Engineering, Shanghai University of Finance and Economics, Shanghai 200433, PR China, **2** Data Science and Cloud Service Research Centre, Shanghai University of Finance and Economics, Shanghai 200433, PR China, **3** Business School, Huzhou University, Huzhou 313000, PR China, **4** Department of Physics, Fribourg University, CH-1700 Fribourg, Switzerland

* liujg004@ustc.edu.cn (JGL); yu.changrui@mail.shufe.edu.cn (CRY)

Abstract

The investors' attention has been extensively used to predict the stock market. Different from existing proxies of the investors' attention, such as the *Google trends*, *Baidu index* (*BI*), we argue the collective attention from the stock trading platforms could reflect the investors' attention more closely. By calculated *the increments of the attention volume for each stock* (*IAVS*) from the stock trading platforms, we investigate the effect of investors' attention measured by the *IAVS* on the movement of the stock market. The experimental results for *Chinese Securities Index 100* (CSI100) show that the *BI* is significantly correlated with the returns of CSI100 at 1% significance level only in 2014. However, it should be emphasized that the correlation of the new proposed measure, namely *IAVS*, is significantly at 1% significance level in 2014 and 2015. It shows that the effect of the measure *IAVS* on the movement of the stock market is more stable and significant than *BI*. This study yields important invest implications and better understanding of collective investors' attention.

Introduction

The investors' attention plays an important role in predicting the movement of the stock market, which has attracted much attention recently [1–4]. The attention is not only a scarce cognitive resource [5] but also the hard currency of cyberspace [6]. In fact, it is hard for most investors, especially retail investors, to access the market information timely and accurately. Therefore, most investors would like to pay more attention to the attracted information to adjust their investment behavior, leading to the movement of the stock market [7, 8].

The proxies of investors' attention for predicting the movement of the stock market could be roughly classified into two categories. The indirect proxies of investors' attention mainly include extreme return [9], trading volume [10–14], turnover [15, 16], etc. [17–20], which have been extensively analyzed by professional investors for many years. On the other hand, direct proxies of investors' attention, such as *search volume index* (*SVI*) [1, 21–34], social network (Twitter feeds, blogs, forum, Wikipedia etc.) [35–42], news [43–51], etc. have been introduced to predict the movement of the stock market.

In particular, the massive data sources resulting from human interaction with the Internet have offered a new perspective on the behavior of market participants besides investors in the stock market. For example, the *SVI* has been used to predict the movement of the stock market [1, 21–26,30–34]. By introducing the search volume in Google of a sample of Russell 3000 stocks, Da *et al.* [1] found that the increase in the *SVI* could successfully predict higher stock prices in a short term and eventual price reversal. Based on the *Google trends* or *Baidu index*, the similar results of French, Japanese and Chinese stock markets have been found [32–34]. Preis *et al.* [21] found that the search volume of the *Google trends* for financial related words could be used to predict the stock market volatility. Besides the *SVI*, the sentiment detected from the social network users also could affect the movement of the stock market [35–40]. By the inclusion of specific public mood dimensions, Bollena *et al.* [35] found that the prediction accuracy of the Dow Jones Industrial Average can be significantly improved. These works help us to predict the movement of the stock market in the new perspectives. However, in the data generated from search engine, different users always input the same keywords with different motivations. Even for the exact same motivation, two users always input different keywords. For the sentiment of the social networks, the sentiment need to be calculated based on the ontology or domain dictionaries, which could not updated dynamically in terms of the stock market, leading to biased prediction. In this paper, we argue that the data generated from the stock trading platforms can reflect the investors' attention more accurately and effectively.
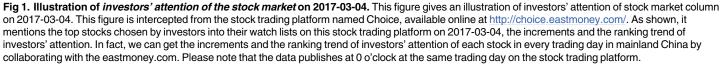
Inspired by the idea, by taking into account the increments of the attention volume for each stock on the stock trading platforms, we present a new measure *IAVS* to quantify the investors' attention for predicting the movement of the stock market. The measure *IAVS* introduced in this paper refers to the daily increments of the aggregate number of each stock which is chosen by investors into their watch lists in a stock trading platform. Fig 1 presents an illustration of *investors' attention of stock market* on 2017-03-04. Firstly, we quantify the effect of the investors' attention on the returns of CSI100 using the *BI* and the *IAVS* to measure investors' attention respectively. The results show that the *BI* is significantly correlated with the returns of CSI100 at 1% significance level only in 2014. However, the correlation of the new proposed measure, namely *IAVS*, is significantly at 1% significance level in 2014 and 2015. Secondly, we investigate the relationship between the investors' attention and the trading activities in the stock market. The results indicate that the measure *IAVS* is correlated with the turnover of CSI100 at 5% significance level in 2014. In addition, we test the robustness with other indices in the stock market, including CSI500 and CSI-ALL, the results show that the *IAVS* is correlated with the returns of CSI500 and CSI-ALL at 1% significance level in 2014 and the *IAVS* is correlated with the returns of CSI-ALL at 5% significance level in 2015. Different from existing proxy *BI* of the investors' attention, the effect of the measure *IAVS* on the stock market is more stable and significant.

Theoretical background and hypotheses development

We study the effect of investors' attention on the stock market with different measures at three levels in this work. Firstly, we investigate the effect of investors' attention on the returns of CSI100, CSI500, CSI-ALL at the same trading day. Secondly, we uncover the impact of investors' attention on the returns of CSI100, CSI500, CSI-ALL in the following two trading days. Finally, we testify the relationship between the investors' attention and the trading activities in the stock market at the same trading day. Fig 2 gives the illustration of the hypotheses structures presented in this paper.

Fig 1. Illustration of *investors' attention of the stock market* on 2017-03-04. This figure gives an illustration of investors' attention of stock market column on 2017-03-04. This figure is intercepted from the stock trading platform named Choice, available online at http://choice.eastmoney.com/. As shown, it mentions the top stocks chosen by investors into their watch lists on this stock trading platform on 2017-03-04, the increments and the ranking trend of investors' attention. In fact, we can get the increments and the ranking trend of investors' attention of each stock in every trading day in mainland China by collaborating with the eastmoney.com. Please note that the data publishes at 0 o'clock at the same trading day on the stock trading platform.

https://doi.org/10.1371/journal.pone.0176836.g001

2.1 The effect of investors' attention on stock index returns at the same trading day

How does an increase in retail investors' attention affect the stock index returns? Barber and Odean [9] argued that individual investors are net buyers of attention-grabbing stocks and thus an increase in individual investors' attention results in temporary positive price pressure. The reason behind their argument goes as follows. When individual investors are buying, they have to choose from a large set of available alternatives. However, when they are selling, they can only sell what they own. Within the framework of Barber and Odean [9], a positive investors' attention should predict higher stock prices in the short term. This assumption has been verified in the study of Da *et al.* [1] with the American stock market. Thus, we propose the following hypotheses:

H_{1a}: Increased investors' attention measured by the *IAVS* will lead to higher returns of corresponding stock index at the same trading day.

H_{1b}: Increased investors' attention measured by the *BI* will lead to higher returns of corresponding stock index at the same trading day.
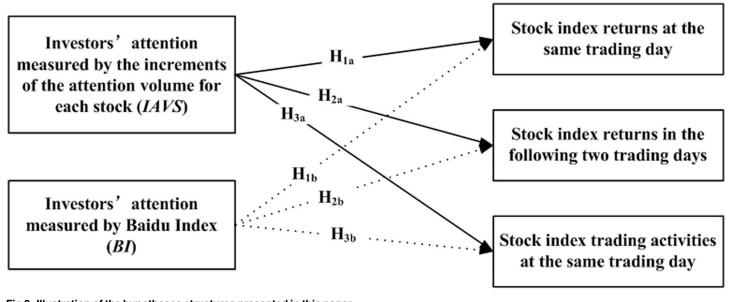
Fig 2. Illustration of the hypotheses structures presented in this paper.

2.2 The impact of investors' attention on stock index returns in the following trading days

After the retail investors have made a purchase, professional investors and institutional investors holding high-attention stocks may be aware of the price premium that is included in their price, and they may sell the stock for a premium. When the retail investors' attention is distracted by other stocks, the price of high-attention stocks may reverse. The results of relative studies [32–34] confirmed that while an increase in investors' attention brings higher stock prices, the price may reverse in a short term with the samples of French, Japanese and Chinese stock markets. Thus, we propose the following hypotheses:

H_{2a}: Increased investors' attention measured by the *IAVS* will impact on the returns of corresponding stock index significantly in the following trading days.

H_{2b}: Increased investors' attention measured by the *BI* will impact on the returns of corresponding stock index significantly in the following trading days.

2.3 The relationship between the investors' attention and trading activities in stock market

Over recent years, researchers have found that psychological factors including attention appear to have significant influences on the trading decisions of investors within the financial markets, and thus, on both returns and trading activities [52–54]. Trading volume and turnover are two important indicators of the trading activities to describe the liquidity of the stock market. Once an investor pay attention to some stocks, he or she would probably put these stocks into their watch list termed, which will inevitably affect their expectations and behavior of investment. And the final performance is there will be correlation between the investors' attention and trading activities in the stock market. Yu and Zhang [49] argued that the stocks with active performance of trading activities will cause more investors' attention, thereby

strengthen the correlation between the investors' attention and trading activities [33]. Li argued that the stocks with high trading volume bring more attention in news reports. Thus, we propose the following hypotheses:

H$_{3a}$: The investors' attention measured by the *IAVS* has associated significantly with the trading activities of corresponding stock index.

H$_{3b}$: The investors' attention measured by the *BI* has associated significantly with the trading activities of corresponding stock index.

Data collection and measures for variables
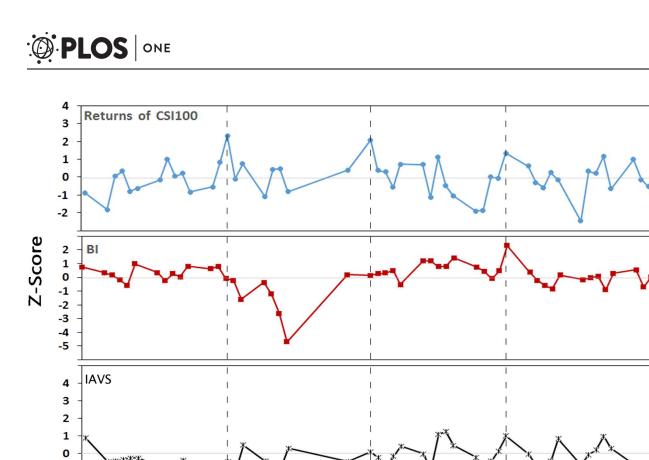
3.1 Data collection and analysis

In this paper, we introduce three databases, including search volume index of *Baidu Index* (http://index.baidu.com/) termed *BI* as benchmark, investors' attention on stock trading platform (http://choice.eastmoney.com/) termed the *IAVS*, and stock transaction data from the stock trading platform (http://choice.eastmoney.com/). The Baidu search engine is the largest Chinese search engine in the world at present, and eastmoney.com is one of the largest and most influential financial trading platform in mainland China with approximately 1.6 million active users daily and more than one hundred million users so far. The *BI* is generated on *Baidu Index* for the keywords of the names and codes of each stock; the measure *IAVS* is generated from eastmoney.com for the increments of attention volume of each stock. To match the *BI* and the measure *IAVS* to the respective time series of the stock market we only consider trading days of the stock markets in this paper.

This work focuses on the three major Chinese stock indices, including CSI100, CSI500, and CSI-ALL, in the Shanghai Stock Exchange(SSE) and Shenzhen Stock Exchange(SZSE) in mainland China from January 3, 2014 to December 31, 2015, when is the only period that we can get the data for CSI100 from Baidu Index due to the service limitations of Baidu Index. CSI100 consists of the largest 100 stocks in mainland China; CSI500 reflects the overall situation of small cap companies in Chinese stock markets; CSI-ALL consists of all securities in mainland China. We use CSI100 to quantify the effect of the investors' attention on the movement of the stock market and compare the results between the *IAVS* and *BI*. Then we use CSI500 and CSI-ALL to test the robustness of the results based on the *IAVS* as a measure of the investors' attention. Fig 3 gives the illustration of the correlation between the CSI100 returns and the investors' attention respectively measured by the *BI* and the *IAVS*. We also provide explanation about the measures of this work in more detail in the rest of this section (The minimal data set is included in S1 and S2 Data).

Stationarity tests. To cover various combinations of relationships, we initially study standard transformations of the original series, i.e. the first differences. For each of the series, we test their stationarity using the Augmented Dickey-Fuller test (ADF) [55] and the KPSS [56] test. As both tests have opposite null and alternative hypotheses, they form an ideal pair for the stationarity vs. unit-root testing [42]. For the $BI_{i,t}$, we find the original series to be non-stationary and to contain the unit-root. Correspondingly, its first differences are stationary. The same results are found for the $Volume_{i,t}$, $Turnover_{i,t}$, we find the unit-root only for the first differences transformation of the series mentioned above. For this reason and also for more convenient interpretation, we opt for the first differences series. And the original series $IAVS_{i,t}$ is not always stationary in different indices, so we opt for the first differences of the series $IAVS_{i,t}$ (see detailed discussion in S1 Text and S1 Table).
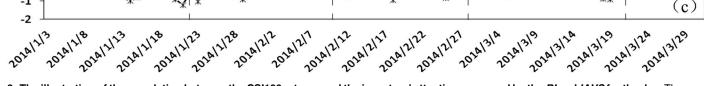
Fig 3. The illustration of the correlation between the CSI100 returns and the investors' attention measured by the *BI* and *IAVS* for the day. The subplots (a)-(c) show the time series of Z-score of the returns of CSI100, the *BI* and the measure *IAVS* respectively, from which one can find that the measure *IAVS* correlates with the returns of CSI100 more closely. The Z-score is calculated in the way $z = (x - \mu)/\sigma$, where x, μ and σ denote the original, mean value and standard deviation. We only consider trading days of the stock market, so there is no data in weekends and holidays.

Serial correlation tests. Turning now to the analysis of the serial correlation. At the beginning, the autocorrelation coefficient, partial autocorrelation coefficient and Q statistic of the residual series are calculated. If the corresponding p-value of the test is less than 0.05, the null of no serial correlation is rejected and, therefore, it can be concluded that there might be serial correlation in the returns [57, 58]. If the residual series of the regression equation has serial correlation, it is necessary to modify the autocorrelation of the residuals in a correct way. There is no serial correlation among all the modified regression equations, the estimated results of which are valid (see detailed discussion in S2 Text, S1, S2, S3 and S4 Figs)

3.2 Independent variables

$BI_{i,t}$. We construct a time series of daily aggregate search volume of each stock listed in the index i on *Baidu Index* as the following way:

$$BI_{i,t} = \sum_{j=1}^{n}(BIname_{j,i,t} + BIcode_{j,i,t}), \tag{1}$$

where $BIname_{j,i,t}$ and $BIcode_{j,i,t}$ are the aggregate search volume through its name or code of

stock j in the list of stock index i on day t via *Baidu Index*, n is the total number of the stocks listed in stock index i. We use $D\_BI_{i,t}$ indicate the first difference of the series $BI_{i,t}$.

IAVS$_{i,t}$. We construct a time series of daily attention increments of stocks in the index i on the stock trading platform as the following way:

$$IAVS_{i,t} = \sum_{j=1}^{n} IAVS_{j,i,t}, \tag{2}$$

where $IAVS_{j,i,t}$ is the daily attention increments of stock j which is in the list of stock index i on day t via the stock trading platform. That is to say if one investor add stock j into his watch list on day t, the attention increments of stock j will be added by 1 on day t on the stock trading platform. And n is the total number of stocks listed in index i. Also, we use $D\_IAVS_{i,t}$ indicate the first difference of $IAVS_{i,t}$.

3.3 Dependent variables

Return$_{i,t}$. We construct a time series for daily returns of the stock indices without deduction of charges as the following way:

$$Return_{i,t} = \frac{(Closing\ price_{i,t} - Closing\ price_{i,t-1})}{Closing\ price_{i,t-1}} \times 100\%, \tag{3}$$

where *Closing price*$_{i,t}$ is the closing price of index i on day t, and day t-1 is the trading day before day t. Also, we use $D\_Return_{i,t}$ indicate the first difference of the series $Return_{i,t}$.

Volume$_{i,t}$. We use this variable to describe the trading volume of index i on day t. Trading volume measures the total number of shares or indices transacted for a specified security during a specified time period. It includes the total number of indices transacted between a buyer and seller during a transaction. When indices are more actively traded, their trade volume may high, and when indices are less actively traded, their trade volume may low. Similarly, we use $D\_Volume_{i,t}$ indicate the first difference of the series $Volume_{i,t}$.

Turnover$_{i,t}$. We use this variable to describe the turnover rate of index i on day t. The turnover rate is the percentage of a mutual fund or other investment's holdings that have been replaced in a given period. Generally speaking, high turnover rate of index means that the index has strong liquidity. Similarly, we use $D\_Turnover_{i,t}$ indicate the first difference of the series $Turnover_{i,t}$.

Table 1 gives the variable description used in this paper.

Table 1. Description of variables.

Variable	Description
BI$_{i,t}$	The aggregate search volume number of all stocks listed in stock index i on day t released at 24 o'clock at the same trading day on Baidu Index
IAVS$_{i,t}$	The Increments of the Attention Volume of all Stocks in stock index i in all investors' watchlist termed on day t released at 0 o'clock at the same trading day on stock trading platforms
Return$_{i,t}$	The returns of stock index i on day t without deduction of charges
Volume$_{i,t}$	The trading volume of stock index i on day t
Turnover$_{i,t}$	The turnover rate of stock index i on day t

Results

By using Single-Variable Regression Analysis in time series, we test the relationship between the investors' attention and the movement of the stock market. The investors' attention was measured by two different proxies, the *BI* and the *IAVS,* to compare with each other. First, we estimate the Model I to testify the effect of the *BI* on the returns of stock index.

$$Return_{i,s} = \beta_1 + \beta_2 D\_BI_{i,t} + \varepsilon_{i,s}. \tag{I}$$

Also, we estimate the Model II to testify the effect of the *IAVS* on the returns of stock index.

$$Return_{i,s} = \beta_3 + \beta_4 D\_IAVS_{i,t} + \varepsilon_{i,s}, \tag{II}$$

where *s* vary from *t* to *t+1, t+2*, that is, we use the variables of investors' attention on day *t* to predict the returns of the very trading day *t* and the following two trading days day *t+1*, day *t+2*.

Then, we estimate the Model III to testify the relationship between the investors' attention measured by the *BI* and the *Trading_Activities.* The *Trading_Activities*$_{i,t}$ includes the *D_Turnover*$_{i,t}$ and *D_Volume*$_{i,t}$.

$$Trading\_Activities_{i,t} = \beta_5 + \beta_6 D\_BI_{i,t} + \varepsilon_{i,t}. \tag{III}$$

Also, we estimate the Model IV to testify the relationship between the investors' attention measured by the *IAVS* and the *Trading_Activities.* Similarly, the *Trading_Activities*$_{i,t}$ includes the *D_Turnover*$_{i,t}$ and *D_Volume*$_{i,t}$.

$$Trading\_Activities_{i,t} = \beta_7 + \beta_8 D\_IAVS_{i,t} + \varepsilon_{i,t}. \tag{IV}$$

4.1 Results for CSI100

The sample in this section includes 73 stocks with no missing data in CSI100 with 485 daily observations from January 3, 2014 to December 31, 2015. Particularly, we test the impact of the investors' attention for day *t* on the the stock index returns for *day t, day t+1, day t+2* in order to find whether the investors' attention has prediction effect on the movement of the stock market in the short term.

Table 2 shows the results of the regression modeling with CSI100. Firstly, the results show that the coefficients of *Return*$_{i,t}$ are positive and significant at 1% significance level for

Table 2. Results of the regression modeling with CSI100.

Variables	$D\_BI_{i,t}$ (2014)	$D\_IAVS_{i,t}$ (2014)	$D\_BI_{i,t}$ (2015)	$D\_IAVS_{i,t}$ (2015)
$Return_{i,t}$	0.201** (0.009)	0.419** (0.000)	-0.147 (0.138)	0.261** (0.000)
$Return_{i,t+1}$	0.533** (0.000)	-0.360** (0.000)	-0.058 (0.597)	-0.141* (0.011)
$Return_{i,t+2}$	0.181 (0.087)	0.193* (0.045)	-0.006 (0.783)	0.031 (0.639)
$D\_Turnover_{i,t}$	-0.000 (0.983)	0.234* (0.026)	-0.032 (0.289)	0.181 (0.087)

Note:

**$p < 0.01$,

*$p < 0.05$

$D\_IAVS_{i,t}$ and $D\_BI_{i,t}$ in 2014, supporting the hypotheses H_{1a} and H_{1b}, which indicate that the more investors' attention of the stocks, the higher returns of the corresponding stock index at the same trading day. Specifically, the $D\_IAVS_{i,t}$ is released at 0 o'clock on day t which is before the opening time of the stock market on day t. However, the $D\_BI_{i,t}$ is generated at 24 o'clock on day t which is after the closing time of the stock market on day t. So the $IAVS$ may has prediction effect on the movement of stock market at the same trading day.

Secondly, the coefficients of $Return_{i,t+1}$ are significant at 1% significance level for $D\_IAVS_{i,t}$ and $D\_BI_{i,t}$ in 2014, supporting the hypotheses H_{2a} and H_{2b}. Specifically, only the coefficients of $Return_{i,t+2}$ are significant at 5% significance level for $D\_IAVS_{i,t}$ in 2014, supporting the hypotheses H_{2a}, which suggest that increased investors' attention measured by the $IAVS$ will impact on the returns of corresponding stock index significantly in the following two trading days.

Thirdly, the coefficients of $D\_Turnover_{i,t}$ are significant at 5% significance level for $D\_IAVS_{i,t}$ in 2014, supporting the hypotheses H_{3a}. Different from the BI, the measure $IAVS$ has associated significantly with the $Turnover$ of corresponding stock index at the same trading day. According to the results, we find the effects of investors' attention with different proxies on the movement of the stock market in the following trading days are different. A reason may be the different proxies are generated by different users, the $IAVS$ is generated on stock trading platform where the users are more professional in investment than the users on search engine web.

It should be noticed the coefficients of $Return_{i,t}$ are significant at 1% significance level for $D\_IAVS_{i,t}$ in 2015, providing support for H_{1a}. And the coefficients of $Return_{i,t+1}$ are significant at 5% significance level for $D\_IAVS_{i,t}$ in 2015, providing support for H_{2a}. However, the results from the BI are not significant in 2015. The possible reason is that the stock market is changed a lot between 2014 and 2015, and it is in the bullish market in 2014 and volatile market in 2015 in mainland China. From the empirical results, we find the measure $IAVS$ contributes to the forecasting for the movement of CSI100 persistently no matter how the market changes. Generally speaking, with the sample in this work, the results from the $IAVS$ outperforms the results of the BI.

4.2 Results for CSI500 and CSI-ALL

We explore the effect of investors' attention on the movement of the stock market by using different stock indices, including CSI500 and CSI-ALL, to provide a cross-category robustness check. The data sets contain 485 daily observations from January 3, 2014 to December 31, 2015.

Empirical analysis based on CSI500

Table 3 shows the results of the Regression Modeling with CSI500. The coefficients between the independent variables $D\_IAVS_{i,t}$ and the variables of the stock index, including $Return_{i,t}$, $Return_{i,t+1}$, $Return_{i,t+2}$, $D\_Volume_{i,t}$ and $D\_Turnover_{i,t}$ for CSI500 are all significant at 5% significance level for $D\_IAVS_{i,t}$ in 2014, supporting the hypotheses H_{1a}, H_{2a}, H_{3a}. The results indicate that the more investors' attention on the stocks of CSI500, the higher returns in the first three trading days and active trading activities at the same trading day of CSI500 in 2014. Although the coefficients between the $D\_IAVS_{i,t}$ and the movement of the stock index in 2015 are not significant, the direction is the same as its in 2014.

Comparing the results of CSI500 with CSI100, we find that the measure $IAVS$ is more efficient to the stock index with large firms than small firms in volatile market. This results are different from the argument posited by GM Mian and Sankaraguruswamy(2008) that the

Table 3. Results of the regression modeling with CSI500.

Variables	$D\_IAVS_{i,t}$ (2014)	$D\_IAVS_{i,t}$ (2015)
$Return_{i,t}$	0.118**	0.164
	(0.009)	(0.155)
$Return_{i,t+1}$	0.284*	0.104
	(0.022)	(0.370)
$Return_{i,t+2}$	0.214*	0.102
	(0.013)	(0.482)
$D\_Volume_{i,t}$	0.878**	0.076
	(0.000)	(0.332)
$D\_Turnover_{i,t}$	0.146**	0.175
	(0.000)	(0.346)

Note:

**$p < 0.01$,

*$p < 0.05$

influence of sentiment on the stock price response is especially pronounced for small stocks [51]. Both results tell us that the investors' attention is not the same as the sentiment of investors, and the detail research need to be explored in future studies.

Empirical analysis based on CSI-ALL

Table 4 presents the results of the Regression Modeling with CSI-ALL. The coefficients between the independent variables $D\_IAVS_{i,t}$ and the variables of the stock index, including $Return_{i,t}$, $Return_{i,t+1}$, $D\_Volume_{i,t}$ and $D\_Turnover_{i,t}$ for CSI-ALL are all significant at 1% significance level for $D\_IAVS_{i,t}$ in 2014, providing support for H_{1a}, H_{2a}, H_{3a}. The results indicate that the more investors' attention on the stocks of CSI-ALL, the higher returns in the first three trading days and active trading activities at the same trading day of CSI-ALL in 2014.

In addition, the coefficients between the independent variables $D\_IAVS_{i,t}$ and the variables of the stock index, including $Return_{i,t}$, $D\_Volume_{i,t}$ and $D\_Turnover_{i,t}$ for CSI-ALL are significant at 5% significance level for $D\_IAVS_{i,t}$ in 2015, supporting the hypotheses H_{1a}, H_{2a}, H_{3a}. The results indicate that the more investors' attention on the stocks of CSI-ALL, the higher returns and active trading activities at the same trading day of CSI-ALL in 2015. Comparing

Table 4. Results of the regression modeling with CSI-ALL.

Variables	$D\_IAVS_{i,t}$ (2014)	$D\_IAVS_{i,t}$ (2015)
$Return_{i,t}$	0.109**	0.090*
	(0.000)	(0.019)
$Return_{i,t+1}$	0.115**	0.006
	(0.000)	(0.783)
$Return_{i,t+2}$	0.107**	-0.015
	(0.000)	(0.742)
$D\_Volume_{i,t}$	0.641**	0.256*
	(0.000)	(0.010)
$D\_Turnover_{i,t}$	0.175**	0.089*
	(0.000)	(0.020)

Note:

**$p < 0.01$,

*$p < 0.05$

the results between 2014 and 2015, we find the *IAVS* is efficient to the movement of CSI-ALL, although it has experienced a volatile period in Chinese stock market in 2015.

Among the results of CSI-ALL, CSI100 and CSI500, we could find the consistent results in most cases as follows: (1) the effect of investors' attention measured by the *IAVS* on the returns of the stock index at the same trading day are positive and significant; (2) the relationship between the investors' attention measured by the *IAVS* and the trading activities of the corresponding stock index at the same trading day are positive and significant; (3) the impact of investors' attention measured by the *IAVS* on the movement of the stock index with small firms are the same as the stock index with all firms.

Conclusion and discussions

To accurately predict the movement of the stock market, different with the traditional measures of investors' attention, including the *Google trend* and *Baidu Index*, we empirically investigate the effect of the investors' attention, namely *IAVS* generated by investors on the stock trading platforms on the movement of the stock market. The measure *IAVS* is updated rapidly, providing first-hand information to reflect the investors' attention ahead of other sources. Thus, the attention volume of each stock provides us with a good context to study how to quantify the investors' attention effectively and what are the influences of investors' attention on the movement of the stock market.

In this paper, we propose the hypotheses that the investors' attention will impact on the returns and the trading activities of corresponding stock index in a short time. By using Single-Variable Regression Analysis in time series, we test the relationship between the investors' attention, measured by two different proxies the *BI* and *IAVS*, and the movement of the stock market. The samples includes CSI100, CSI500 and CSI-ALL with 485 daily observations from 2014 to 2015. The findings in this work show that the results of the measure *IAVS* generated from the stock trading platform is more significant and stable than the *BI* generated from the search engine. The experimental results mainly show that compared with the *BI*, the measure *IAVS* is significant related to the returns of the corresponding stock indices CSI100 and CSI-ALL at the same trading days in 2015; the measure *IAVS* is significant related to the returns of the corresponding stock index in the first three trading days in the bullish market in 2014; the measure *IAVS* has associated significantly with the trading activities of corresponding stock index at the same trading day. From the results, we also get the findings that the effect of the measure *IAVS* on the movement of the stock market may differ depending on the firm's size, which show that the measure *IAVS* is more significant to the movement of stock index containing large firms; and the effect of the measure *IAVS* on the movement of the stock market is also influenced by the market environment, which show that the results of the bullish market in 2014 are more stable than those of the volatile market in 2015 in mainland China.

From a theoretical perspective, our study enriches extant research by focusing on the relationship between the investors' attention and the movement of the stock market. Specifically, we developed the measure *IAVS* to quantify the investors' attention, which also can be readily applied to predict the financial risks or understand the collective investors' behavior. From a practical perspective, our findings may help retail investors better predict the stock index trends, based upon which they can make better investment decisions. Stock price prediction is an important and challenging problem for studying financial markets, which has been investigated in different ways. Recently, Lin et al. [59] found that the friendship networks and information asymmetry play an important role for the P2P lending system. Therefore, how to construct the trader social network [60] in terms of their trading behavior, and how to construct the stock network in term of the movement of stock pricing [61–63], as well as how to

measure the effect of the network structure on the prediction performance are open questions, which would be investigated in our future research direction. Besides the social network structure, for the online traders, the collective holding behaviors [64, 65] and social influence [66] would affect their trading decision. More importantly, to identify the mechanism of the online trading system, predict the movement of stock market, and build the recommendation system [67], we need to investigate the above questions from the viewpoint of complex systems, which could bring new insights and knowledge for deeply understanding the evolution mechanism of the stock market.

Supporting information

S1 Text. The stationarity tests.
(PDF)

S2 Text. The serial correlation tests.
(PDF)

S1 Table. Stationarity and unit-root tests.
(PDF)

S1 Fig. Serial correlation tests between the $D\_BI_{i,t}$ and $Return_{i,t}$.
(JPG)

S2 Fig. Serial correlation tests between the $D\_IVAS_{i,t}$ and $Return_{i,t}$.
(JPG)

S3 Fig. Serial correlation tests between the $D\_BI_{i,t}$ and $Turnover_{i,t}$.
(JPG)

S4 Fig. Serial correlation tests between the $D\_IVAS_{i,t}$ and $Turnover_{i,t}$.
(JPG)

S1 Data. The minimal data set for CSI100 in 2014.
(XLSX)

S2 Data. The minimal data set for CSI100 in 2015.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: ZHY JGL CRY JTH.

Data curation: ZHY.

Formal analysis: ZHY JGL CRY JTH.

Funding acquisition: JGL.

Investigation: ZHY.

Methodology: ZHY JGL CRY JTH.

Project administration: JGL.

Resources: ZHY JGL.

Software: ZHY.

Supervision: ZHY JGL CRY JTH.

Validation: ZHY JGL CRY JTH.

Visualization: ZHY JGL CRY JTH.

Writing – original draft: ZHY JGL.

Writing – review & editing: ZHY JGL CRY JTH.

References

1. Da Z, Engelberg J, Gao P. In search of attention. The Journal of Finance. 2011; 66: 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

2. Lou D. Attracting investor attention through advertising. Review of Financial Studies. 2014; 644: 1797–1829. https://doi.org/10.1093/rfs/hhu019

3. Wei Q, Zhao X, Zhang Z Q, et al. Stock price prediction and trading strategy design based on investors' attention: Cross patterns perspective between stocks. Systems Engineering—Theory & Practice. 2016; 36: 1361–1371.

4. Shen D, Zhang W, Xiong X, et al. Trading and non-trading period Internet information flow and intraday return volatility. Physica A: Statistical Mechanics and its Applications. 2016; 451: 519–524. https://doi.org/10.1016/j.physa.2016.01.086

5. Kahneman D. Attention and effort. Englewood Cliffs. NJ: Prentice Hall. 1973.

6. Mandel T, Van der Leun G. Rules of the net: On-Line operating instructions for human beings. The US: Hyperion Books. 1996.

7. Daniel K, Hirshleifer D, Teoh S H. Investor psychology in capital markets: Evidence and policy implications. Journal of Monetary Economics. 2002; 49: 139–209. https://doi.org/10.1016/S0304-3932(01)00091-5

8. Merton R C. A simple model of capital market equilibrium with incomplete information. The Journal of Finance. 1987; 42: 483–510. https://doi.org/10.1111/j.1540-6261.1987.tb04565.x

9. Barber B M, Odean T. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies. 2008; 21: 785–818. https://doi.org/10.1093/rfs/hhm079

10. Kenett D Y, Ben-Jacob E, Stanley H, E. How high frequency trading affects a market index. Scientific Reports. 2013; 3: 2110. https://doi.org/10.1038/srep02110 PMID: 23817553

11. Gervais S, Kaniel R, Mingelgrin D H. The high-volume return premium. The Journal of Finance. 2001; 56: 877–919. https://doi.org/10.1111/0022-1082.00349

12. Huang Z, Heian J B, Zhang T. Differences of opinion, overconfidence, and the high-volume premium. Journal of Financial Research. 2011; 34: 1–25. https://doi.org/10.1111/j.1475-6803.2010.01283.x

13. Statman M, Thorley S, Vorkink K. Investor overconfidence and trading volume. Review of Financial Studies. 2006; 19: 1531–1565. https://doi.org/10.1093/rfs/hhj032

14. Kaniel R, Ozoguz A, Starks L. The high volume return premium: Cross-country evidence. Journal of Financial Economics. 2012; 103: 255–279. https://doi.org/10.1016/j.jfineco.2011.08.012

15. Avramov D, Chordia T, Goyal A. Liquidity and autocorrelations in individual stock returns. The Journal of Finance. 2006; 61: 2365–2394. https://doi.org/10.1111/j.1540-6261.2006.01060.x

16. Loh R K. Investor inattention and the underreaction to stock recommendations. Financial Management. 2010; 39: 1223–1252. https://doi.org/10.1111/j.1755-053X.2010.01110.x

17. Corwin S A, Coughenour J F. Limited attention and the allocation of effort in securities trading. The Journal of Finance. 2008; 63: 3031–3067. https://doi.org/10.1111/j.1540-6261.2008.01420.x

18. Jacobs H. The role of attention constraints for investor behavior and economic aggregates: what have we learnt so far? Management Review Quarterly. 2015; 65: 217–237. https://doi.org/10.1007/s11301-015-0112-5

19. Huberman G. Familiarity breeds investment. Review of Financial Studies. 2001; 14: 659–680. https://doi.org/10.1093/rfs/14.3.659

20. Lim K P, Hooy C W, Chang K B, et al. Foreign investors and stock price efficiency: Thresholds, underlying channels and investor heterogeneity. The North American Journal of Economics and Finance. 2016; 36: 1–28. https://doi.org/10.1016/j.najef.2015.11.003

21. Preis T, Moat H S, Stanley H E. Quantifying trading behavior in financial markets using Google Trends. Scientific Reports. 2013; 3: 1684. https://doi.org/10.1038/srep01684

22. Bordino I, Battiston S, Caldarelli G, et al. Web search queries can predict stock market volumes. PLoS One. 2012; 7: e40014. https://doi.org/10.1371/journal.pone.0040014 PMID: 22829871

23. Ap Gwilym O, Kita A, Wang Q. Speculate against speculative demand. International Review of Financial Analysis. 2014; 34: 212–221. https://doi.org/10.1016/j.irfa.2014.03.001

24. Dzielinski M. Measuring economic uncertainty and its impact on the stock market. Finance Research Letters. 2012; 9: 167–175. https://doi.org/10.1016/j.frl.2011.10.003

25. Dimpfl T, Jank S. Can internet search queries help to predict stock market volatility? European Financial Management. 2016; 22: 171–192.

26. Bank M, Larch M, Peter G. Google search volume and its influence on liquidity and returns of German stocks. Financial Markets and Portfolio Management. 2011; 25: 239–264. https://doi.org/10.1007/s11408-011-0165-y

27. Zhang W, Shen D, Zhang Y, et al. Open source information, investor attention, and asset pricing. Economic Modelling. 2013; 33: 613–619. https://doi.org/10.1016/j.econmod.2013.03.018

28. Curme C, Preis T, Stanley H E, et al. Quantifying the semantics of search behavior before stock market moves. Proceedings of the National Academy of Sciences. 2014; 111(32): 11600–11605. https://doi.org/10.1073/pnas.1324054111

29. Kristoufek L. Can Google Trends search queries contribute to risk diversification? Scientific Reports. 2013; 3: 2713. https://doi.org/10.1038/srep02713

30. Joseph K, Wintoki M B, Zhang Z. Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. International Journal of Forecasting. 2011; 27: 1116–1127. https://doi.org/10.1016/j.ijforecast.2010.11.001

31. Preis T, Moat H S. Early signs of financial market moves reflected by google searches. The Germany: Springer International Publishing. 2015: 85–97.

32. Takeda F, Wakao T. Google search intensity and its relationship with returns and trading volume of Japanese stocks. Pacific-Basin Finance Journal. 2014; 27: 1–18. https://doi.org/10.1016/j.pacfin.2014.01.003

33. Yu Q J, Zhang B. Limited attention and stock performance: An empirical study using baidu index as the proxy for investor attention. Journal of Financial Research. 2012; 386: 152–165.

34. Aouadi A, Arouri M, Teulon F. Investor attention and stock market activity: Evidence from France. Economic Modelling. 2013; 35: 674–681. https://doi.org/10.1016/j.econmod.2013.08.034

35. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. Journal of Computational Science. 2011; 2: 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

36. Nofer M, Hinz O. Using twitter to predict the stock market. Business & Information Systems Engineering. 2015; 57: 229–242. https://doi.org/10.1007/s12599-015-0390-4

37. Ackert L F, Jiang L, Lee H S, et al. Influential investors in online stock forums. International Review of Financial Analysis. 2016; 45: 39–46. https://doi.org/10.1016/j.irfa.2016.02.001

38. Zheludev I N. When can social media lead financial markets? University College London. 2015. http://discovery.ucl.ac.uk/id/eprint/1467246

39. Wu D D, Zheng L, Olson D L. A decision support approach for online stock forum sentiment analysis. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2014; 44: 1077–1087. https://doi.org/10.1109/TSMC.2013.2295353

40. Moat H S, Curme C, Avakian A, et al. Quantifying Wikipedia usage patterns before stock market moves. Scientific Reports. 2013; 3: 1801. https://doi.org/10.1038/srep01801

41. Zhang W, Li X, Shen D, et al. Daily happiness and stock returns: Some international evidence. Physica A: Statistical Mechanics and its Applications. 2016; 460: 201–209. https://doi.org/10.1016/j.physa.2016.05.026

42. Kristoufek L. BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. Scientific Reports. 2013; 3: 3415. https://doi.org/10.1038/srep03415

43. Ranco G, Bordino I, Bormetti G, et al. Coupling news sentiment with web browsing data improves prediction of intra-day price dynamics. PLoS one. 2016; 11: e0146576. https://doi.org/10.1371/journal.pone.0146576 PMID: 26808833

44. Piškorec M, Antulov-Fantulin N, Novak P K, et al. Cohesiveness in financial news and its relation to market volatility. Scientific reports. 2014; 4:5038. https://doi.org/10.1038/srep05038 PMID: 24849598

45. Zhang J L, Härdle W K, Chen C Y, et al. Distillation of news flow into analysis of stock reactions. Journal of Business & Economic Statistics. 2016; 34: 547–563. https://doi.org/10.1080/07350015.2015.1110525

46. Zhang Y, Feng L, Jin X, et al. Internet information arrival and volatility of SME price index. Physica A: Statistical Mechanics and its Applications. 2014; 399: 70–74. https://doi.org/10.1016/j.physa.2013.12.034

47. Zhang Y, Song W, Shen D, et al. Market reaction to internet news: Information diffusion and price pressure. Economic Modelling. 2016; 56: 43–49. https://doi.org/10.1016/j.econmod.2016.03.020

48. Alanyali M, Moat H S, Preis T. Quantifying the relationship between financial news and the stock market. Scientific Reports. 2013; 3:3578. https://doi.org/10.1038/srep03578 PMID: 24356666

49. Li Q, Wang T J, Li P, et al. The effect of news and public mood on stock movements. Information Sciences, 2014; 278: 826–840. https://doi.org/10.1016/j.ins.2014.03.096

50. Azuma T, Okada K, Hamuro Y. Is no news good news? The streaming news effect on investor behavior surrounding analyst stock revision announcement. International Review of Finance. 2014; 14: 29–51. https://doi.org/10.1111/irfi.12027

51. Sankaraguruswamy S, Mian G M. Investor sentiment and stock market response to corporate news. Ssrn Electronic Journal. 2008; 87: 1357–1384.

52. Daniel K, Hirshleifer D, Subrahmanyam A. Investor psychology and security market under and overreactions. The Journal of Finance. 1998; 53: 1839–1885. https://doi.org/10.1111/0022-1082.00077

53. Hirshleifer D. Investor psychology and asset pricing. The Journal of Finance, 2001; 56: 1533–1597. https://doi.org/10.1111/0022-1082.00379

54. Seasholes M S, Wu G. Predictable behavior, profits, and attention. Journal of Empirical Finance. 2007; 14: 590–610. https://doi.org/10.1016/j.jempfin.2007.03.002

55. Dickey D A, Fuller W A. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association. 1979; 74(366a): 427–431. https://doi.org/10.1080/01621459.1979.10482531

56. Kwiatkowski D, Phillips P C B, Schmidt P, et al. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics. 1992; 54(1-3): 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

57. Humala A, Rodriguez G. Some stylized facts of return in the foreign exchange and stock markets in Peru. Studies in Economics and Finance. 2013; 30(2): 139–158. https://doi.org/10.1108/10867371311325444

58. Epaphra M. Modeling exchange rate volatility: application of the GARCH and EGARCH models. Journal of Mathematical Finance. 2017; 7(01): 121. https://doi.org/10.4236/jmf.2017.71007

59. Lin M, Prabhala N R, Viswanathan S. Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. Management Science. 2013; 59: 17–35. https://doi.org/10.1287/mnsc.1120.1560

60. Guo Q, Shao F, Hu Z L, et al. Statistical properties of the personal social network in the Facebook. Europhysics Letters. 2013; 104: 28004. https://doi.org/10.1209/0295-5075/104/28004

61. Sun X Q, Shen H W, Cheng X Q. Trading network predicts stock price. Scientific Reports. 2014; 4: 3711.

62. Casnici N, Dondio P, Casarin R, et al. Decrypting financial markets through e-joint attention efforts: online adaptive networks of investors in periods of market uncertainty. PLoS one. 2015; 10: e0133712. https://doi.org/10.1371/journal.pone.0133712 PMID: 26244550

63. Liu J G, Hou L, Pan X, et al. Stability of similarity measurements for bipartite networks. Scientific Reports. 2016; 6: 18653. https://doi.org/10.1038/srep18653 PMID: 26725688

64. Hou L, Pan X, Guo Q, et al. Memory effect of the online user preference. Scientific Reports. 2014; 4: 6560.

65. Guo Q, Han J T, Liu J G. Collective behaviors of book holding durations. Physics Letters A. 2016; 380: 3460–3464. https://doi.org/10.1016/j.physleta.2016.08.043

66. Liu J G, Hu Z, Guo Q. Effect of the social influence on topological properties of user-object bipartite networks. The European Physical Journal B. 2013; 86: 478. https://doi.org/10.1140/epjb/e2013-40328-4

67. Zhou T, Kuscsik Z, Liu J G, et al. Solving the apparent diversity-accuracy dilemma of recommender systems. Proceedings of the National Academy of Sciences. 2010; 107: 4511–4515. https://doi.org/10.1073/pnas.1000488107