G Model
INFMAN 2913 No. of Pages 9
ARTICLE IN PRESS

Information & Management xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Information & Management

journal homepage: www.elsevier.com/locate/im

The different impacts of news-driven and self-initiated search volume on stock prices

Xianwei Liu, Qiang Ye*

School of Management, Harbin Institute of Technology, Harbin, China

ARTICLE INFO

Article history:
Received 14 May 2015
Received in revised form 13 April 2016
Accepted 31 May 2016
Available online xxx

Keywords:
Internet search
Attention heterogeneity
Stock returns
Baidu Index

ABSTRACT

Big data benefits both Internet finance and behavioral finance research; Internet search frequency on stocks has been widely used to measure investor attention. In this study, we divide the search volume into news-driven and self-initiated by the online media coverage collected from Baidu Index. In a sample of CSI 300 stocks from 2009 to 2013, we find that self-initiated (news-driven) search volume is more likely to generate buy (sell) pressure, and media coverage can negatively moderate the impact of search volume on stock prices, suggesting that distinguishing search environment for investors can help improve the measure for investor attention.

1. Introduction

In this big data era, information plays the most important role. Data generated from social media, Internet search, and click stream grow exponentially. The rapid expansion of online-generated content creates many opportunities for both industry and academic research [29,18]. The most significant application of big data should be Internet finance, which has commercialized the online posts and searches to select stocks. Big data are closely related to investors, companies, and stock market.

The wisdom of crowds based on massive information has great power and possibility of influencing the financial market [7]. The usage and popularity of social media such as Twitter, Facebook, and Wikipedia has changed investors, companies, and stock market considerably in recent years. For instance, to companies, the usage of social media is associated with firm equity value; the transformative power of social media is crucial to company development [23]. To investors, the usage of Twitter can significantly reduce the information asymmetry and is associated with lower abnormal bid–ask spreads [6,9]. To the stock market, Wikipedia can improve the information environment in the financial market and moderate the timing of managers' voluntary disclosure of companies' bad news [34,33].

Furthermore, taking advantage of big data, Da et al. [8] collected the search volume data for Russell 3000 stocks from Google Trends and found "An increase in search volume index (SVI) predicts higher stock prices in the next 2 weeks." Subsequently, many studies further confirmed this conclusion [11,19,32]. Investors in the financial market have limited attention, and attention allocation has a profound impact on asset prices. An important step in empirically examining the impact of attention on prices is to measure investor attention in a direct and timely manner. Recent work has shown that Internet search frequency can achieve such an objective. The most commonly used Internet search frequency is the SVI from Google Trends, especially after Da et al. [8] showed that SVI can directly measure the attention of retail investors and predict short-term stock returns. However, Internet searches under different circumstances do not guarantee equal attention. This is especially true when there is an overabundance of information, which can lead to scarcity in attention.

Searches prompted by news headlines differ from searches motivated by research for trading ideas in the likelihood that attention will lead to action. News-driven search volume can be generated when there has been a news release, for example, earning announcements, mergers and acquisitions, and even rumors [2]. Many studies show that stocks with no media coverage earn higher returns than stocks with high media coverage [12,31,30,5]. Thus, news-driven search volume is more likely to induce lower returns. Self-initiated search volume is usually conducted by the investors who are searching for information to trade. As compared with news-driven search volume, which is passive, self-initiated search volume is more likely to generate buy pressure as initiative searching shows a demand for investment. Thus, different types of searches may have different impacts on stock prices. It is therefore an important exercise to explore the

* Corresponding author.
E-mail address: yeqiang@hit.edu.cn (Q. Ye).

http://dx.doi.org/10.1016/j.im.2016.05.009

heterogeneity of Internet search and investigate its varying impact on the asset prices.

Only when investors pay attention to the massive information can it influence the financial market. The current studies further confirmed the impact of investor attention on the stock market by virtue of big data; this is the first step as information is worthless without attention. However, an implicit assumption in existing studies is that investor attention (measured by proxies such as search volume) under different situations is supposed to be identical or equal. Our study extends the current studies to the second step, in search of attention heterogeneity, and contributes to the knowledge of how investors treat information differently under different situations: the same amount of search volume performed under different situations captures distinct attention, and results in different decision-making processes.

In this study, we set out to distinguish news-driven search volume from self-initiated search volume, and explored the moderating effect of media coverage. Here, we choose China's stock market and Baidu Index as our research sample for at least three advantages. First, the stock tickers in China are chosen to be unique (Chinese stock tickers are composed of six digits, defined as unique; searching for a six-digit stock code through Baidu is absolutely for the corresponding stock). Second, China's stock market has a higher proportion of retail investors than the US stock market; thus, it is better to explore the individual investor's behavior. Third, Baidu Index offers the online media coverage index (MCI) of each search term, which is more appropriate than the traditional newspaper news when studying the online search frequency, as online searching is more likely to be influenced by online news rather than the newspaper news. By contrast, Google Trends only offers the SVI of each search term and does not provide the online MCI (see Figs. 1 and 2).

2. Backgrounds and hypotheses

Big data are becoming an increasingly important asset for decision makers; four main features characterize big data: volume, variety, velocity, and veracity [24]. Large volumes of highly detailed data from various sources are rapidly generated, providing the opportunity to deliver significant benefits to both industry and academic research [14].

In practice, the significant and successful implication of big data helps many funds and Internet enterprises to stand out. Baidu is the largest Chinese search engine, which occupies >80% of the Chinese search market; the enormous search traces generated by innumerable users is a great wealth. Baidu combined with GF Fund established the first Internet big data fund (first released on 20 October 2014 and sold at 1.2 billion RMB on the first day) in China. This fund commercializes the search volume from Baidu Index, and considers it a key tool to select stocks (the stocks are updated monthly by the overall rating of fundamental financial factors and user-generated search volume).

In academic research, with big data, it is possible to conduct some studies that could not be performed before [3]; for instance, attention is difficult to be directly measured until innumerable search traces are recorded by search engine. However, information surplus is more serious along with the increase of observational data; thus, limited attention is exacerbated. The success of using search volume to measure investor attention is a big step in the research of behavioral finance.

2.1. Measures of investor attention

Empiricists face a substantial challenge in testing theories of attention as direct measures of investor attention are difficult to observe. Instead, researchers have resorted to indirect proxies for investor attention such as extreme returns [4], trading volume [4,15]), news and headlines [4,35], advertising expense [16,22], and price limits [27]. These proxies are indirect because of the reliance on the critical assumption that if a stock's return or turnover was extreme or its name was mentioned in the news media, then investors should have paid attention to it. This assumption can be especially problematic when investor attention is too scarce to cover all these events or when investor attention is not associated with these events.

The novel approach by Da et al. [8] introduces SVI from Google Trends as a timely and efficient proxy to measure investor attention. As compared with the traditional indirect measures, SVI at least has the following advantages. First, Internet users commonly use a search engine to collect information, and Google (Baidu) continues to be the favorite in the US (China). Indeed, up to now, Google (Baidu) has accounted >70% (80%) of all search queries performed in the US (China). The search volume reported by Google (Baidu) is thus likely to be representative of the Internet search behavior of the general population. Second, and more critically, search is a direct attention measure: if you search for a stock in Google (Baidu), you are undoubtedly paying attention to it. Third, Google Trends (Baidu Index) offers timely (weekly/daily) search data of any keywords, which can effectively measure the real-time attention from investors. Therefore, SVI from Google (Baidu) is a direct and unambiguous measure of attention. These advantages of using SVI to capture investor attention lead to widespread adoption in empirical studies.

2.2. Limited attention and asset prices

Traditional asset pricing models in finance and economics assume that information is instantaneously incorporated into prices when it arrives. This assumption requires that investors allocate sufficient attention to the asset. Many recent theoretical studies relax this assumption and provide a framework in which limited attention can affect asset prices (for instance Refs. [28,17,25]. The rationale behind this extension is that attention is a scarce cognitive resource in reality [20] and investors have a

Fig. 1. SVI from Google Trends.

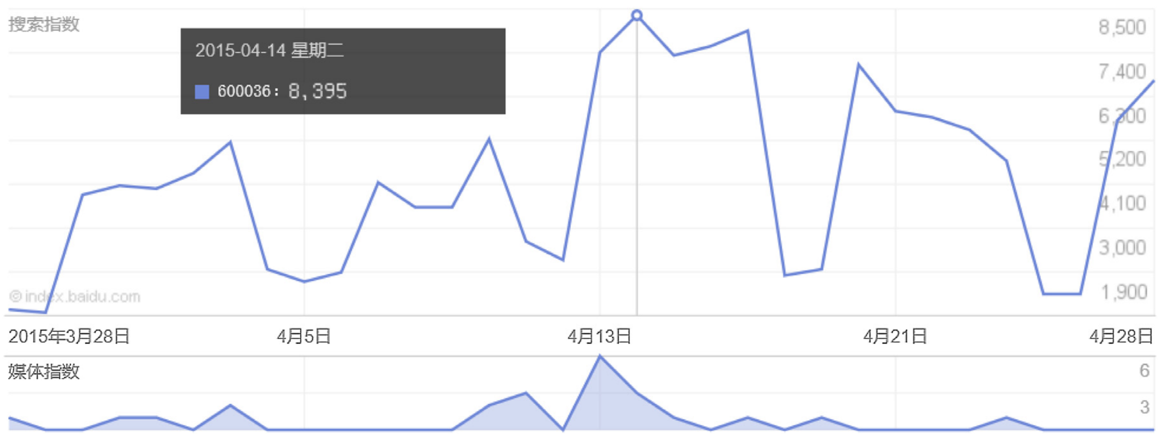

Fig. 2. SVI and MCI from Baidu Index.

limited amount of time and effort to process information. Intuitively, there is a large amount of information freely available relevant for decision making, but it takes time and mental attention for investors to incorporate these information into their decisions.

The constraint that limits the amount of information investors can process is especially tight for individual investors; limited attention of individual investors can have a significant impact on the stock returns [13]. According to the attention theory of Barber and Odean [4], individual investors are net buyers of attention-grabbing stocks and thus an increase in individual investor attention results in temporary-positive price pressure. The reasoning 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. This means that increased attention should lead, on average, to net buying from the individual investors.

The technological advance related to the Internet can make the effect of limited attention on asset prices more astute. This is especially true in the so-called big data age where “a wealth of information creates a poverty of attention.” Furthermore, the recent development in information technology has led to tremendous advances in Internet trading [36] especially among individual investors. According to Nielsen Online, the top 10 most popular online trading sites attracted about 20 million unique visitors per month from 2007 to 2009. The joint influx of information on financial securities and individual investors trading these securities can exacerbate the effect of limited attention.

2.3. Heterogeneity in search and decision-making

An implicit assumption in existing studies that used SVI is all searches are considered equal. However, Internet searches under different circumstances do not guarantee equal attention. For example, a quick glimpse of search results differs from a careful study in the amount of attention allocated by the investors conducting the search. Searches prompted by news headlines differ from searches motivated by research for trading ideas in the likelihood that attention will lead to investment decisions. Furthermore, it is known that limited attention affects the quality of decision making by investors [10].

2.3.1. Self-initiated versus news-driven search

What motivates an investor to conduct an Internet search on a particular stock? Generally, a search can be self-initiated or news-driven. Self-initiated search is generated when investors actively seek information for investment ideas. By contrast, news-driven search is reactive in nature, which may not directly relate to investment decisions. Although it is impossible to separate news-driven search volume from self-initiated search volume in practice, we can assume or justify that (1) the search volume is more likely news-driven rather than self-initiated when there has been news, and the search volume is more likely self-initiated rather than news-driven when there has been no news; (2) the proportion of news-driven search volume would be higher when there has been more news, and the proportion of self-initiated search volume would be higher when there has been less news.

The link between self-initiated search and investment is particularly apparent in China, where the high proportion of retail investors and information asymmetry existing for some companies can be severe. A typical individual investor would first seek useful information via search engine on a basket of stocks, and reach a decision on which stocks to buy. Search driven by news, from online news reports, social media, or word of mouth, can serve the purpose of confirming previous beliefs, acquiring in-depth analysis or making buy/sell decisions [26]. Therefore, news-driven search may not always lead to investment decision. Furthermore, previous findings on the impact of media coverage on stock returns suggest an opposite effect to that of limited attention, in that more news results in lower future stock returns. Therefore, news-driven search volume is more likely to induce lower future returns. We thus propose the following:

Hypothesis 1. *Self-initiated (news-driven) SVI is more likely to generate positive (negative) price pressure.*

2.3.2. Moderating effect of media coverage

Given the different impacts of self-initiated and news-driven search volume on the future stock returns, the moderating effect of media coverage is expected to be significant as the number of news in a week can influence the ratio between the two kinds of search volume. If there has been more news in a week, it would be more likely to generate ahigher (lower) proportion of news-driven (self-initiated) search volume than the weeks when there has been less news. Therefore, the number of news in a week can affect the ratio between self-initiated and news-driven search volume in a week. More news in a week indicates lower (higher) proportion of self-initiated (news-driven) search volume and results in lower future returns according to Hypothesis 1. Thus, media coverage can negatively moderate the impact of search volume on future stock prices.

Investors always face the constraint in the amount of time and cognitive resource when allocating attention across tasks. As

Please cite this article in press as: X. Liu, Q. Ye, The different impacts of news-driven and self-initiated search volume on stock prices, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.05.009

previous studies confirmed, investors have limited attention and they can be distracted by many factors [28,17,25]. For instance, announcements and news releases mostly occur during the weekdays. Furthermore, firms have an incentive to manage media coverage to influence their stock prices during important corporate events (for instance, merger negotiations and acquisitions); thus, the media coverage is not evenly distributed across time [1,21]. More news to be read leads to a tighter constraint for investors; therefore, this can result in large variation in attention level associated with each search, rendering search volume a noisy measure of attention. According to the theories of attention allocation, when there has been more news in a week, investors are more likely to be distracted and be influenced more seriously by the news, and hence, result in lower attention on their searching. Similarly, the quality and depth of each searching performed by investors is poorer when there has been more news release and result in lower-quality consideration for making investment decisions, hence leading to lower returns. We thus propose the following:

Hypothesis 2. *Media coverage can negatively moderate the impact of SVI on stock prices.*

3. Methodology

3.1. Data and variables

Baidu (NASDAQ: BIDU) is a dominant Internet search engine in China than Google (NASDAQ: GOOG) in the US. Baidu Index, like Google Trends (see Fig. 1, stock ticker: AAPL), provides the SVI. Fig. 2 is an example (stock ticker: 600036) of the SVI (the above curve) and MCI (the below curve). For our analysis, we download the daily SVI and MCI for individual stocks from January 2009 to January 2013. Following Da et al. [8], we focus on stocks in the CSI 300 Index containing the 300 largest listed companies of China, representing >70% of the total Chinese equity market capitalization. A crawler is developed to collect the daily data from Baidu Index, including SVI and MCI. The collected SVI from Baidu Index in this study is similar to the SVI from Google Trends, which is the aggregate search frequency from Baidu Index based on stock ticker, while the MCI is the aggregate media coverage from Baidu Index based on stock ticker. The data collection task was conducted in March 2013. Stock returns, turnover, market capitalization, number of analysts, institutional holdings, return on assets (ROAs), earnings per share (EPS), debt ratio (DR), and other related variables are obtained from RESSET (RESSET Financial Research Database: www.resset.cn).

To be consistent with Da et al. [8] and related studies when investigating investor attention, other variables related to investment attention/sentiment (see Table 1) are also used in this study. Variables reflecting firm performance (ROAs and EPS) and capital structure (DR) are also controlled. Table 1 defines all variables used in this study.

3.2. Empirical models

First, based on the advantages of using Chinese stock tickers discussed before, we expect to see stronger impact of search volume on stock prices. We thus replicate the results from Da et al. [8] based on model (1).

$$AR_{i(t+j)} = \alpha_i + \beta_1 ASVI_{it} + \beta_2 MCI_{it} + \beta_3 Cap_{it} + \beta_4 IH_{it} + \beta_5 Abs\mathrm{Ret}_{it} + \beta_6 Analyst_{it} + \beta_7 AbnTurnover_{it} + \beta_8 ROA_{it} + \beta_9 EPS_{it} + \beta_{10} DR_{it} + \mu_{it} \tag{1}$$

Second, in order to distinguish news-driven search volume from self-initiated search volume, we used a dummy variable that equals 0 if there has been no news in a week and 1 else. Subsequently, we ran model (2) two times, using abnormal search volume index (ASVI) when there has been no news and ASVI when there has been news, respectively:

$$AR_{i(t+j)} = \alpha_i + \beta_1 ASVI_{it} + \beta_2 MCID_{it} + \beta_3 Cap_{it} + \beta_4 IH_{it} + \beta_5 AbsRet_{it} + \beta_6 Analyst_{it} + \beta_7 AbnTurnover_{it} + \beta_8 ROA_{it} + \beta_9 EPS_{it} + \beta_{10} DR_{it} + \mu_{it} \tag{2}$$

where ASVI is defined as the log of SVI during this week minus the log of median SVI during the previous 8 weeks, which can capture the abnormal change of individual investors' attention; $\mathrm{AR}_{i(t+j)}$ is the abnormal return of stock$_i$ at week$_{(t+j)}$, calculated by market model; MCI is media coverage index collected from Baidu Index; Cap is market capitalization; IH is institutional holdings; Abs Ret is

Table 1
Variable definitions.

Variable	Definition
Variables from Baidu Index	
Search volume index (SVI)	Aggregate search frequency from Baidu Index based on stock ticker. SVI of a whole week from Monday to Sunday calculated by daily SVI
Abnormal search volume index (ASVI)	The log of SVI during the week minus the log of median SVI during the previous 8 weeks, following Da et al. [8]
Media coverage index (MCI)	The number of online news based on stock ticker calculated by Baidu Index (see Fig. 2), the log value is used in the regressions
Media coverage index dummy (MCID)	Dummy variable that takes the value of one if MCI is positive
Other variables related to investment attention/sentiment following Da et al. [8]	
Ret	Stock returns in RESSET
Abs Ret	Absolute value of stock returns
Abn Ret	Abnormal return, calculated by actual return minus expected return
Turnover	Turnover in RESSET
Abn Turnover	Abnormal turnover, calculated by log(turnover$_t$) minus log(turnover$_{t-1}$)
Cap	Market capitalization in RESSET
Analyst	Number of analysts in RESSET, the log value is used in the regressions
IH	Institutional holding in RESSET
Variables reflecting firm performance and capital structure	
ROAs	Return on assets in RESSET, evaluated by net income/total assets
EPS	Earnings per share in RESSET, evaluated by net income/outstanding shares
DR	Debt ratio in RESSET, evaluated by total debt/total assets

Please cite this article in press as: X. Liu, Q. Ye, The different impacts of news-driven and self-initiated search volume on stock prices, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.05.009

the absolute value of stock returns; Analyst is the number of analysts; and Abn Turnover is abnormal turnover.

Third, in order to explore the moderating effect of media coverage and search volume on future stock prices, the interaction term between ASVI and media coverage index dummy (MCID) is introduced in model (3). In addition, the number of news is used in model (4):

$$AR_{i(t+j)} = \alpha_i + \beta_1 ASVI_{it} + \beta_2 ASVI_{it} * MCI\boldsymbol{D}_{it} + \beta_3 Cap_{it} + \beta_4 IH_{it} + \beta_5 AbsRet_{it} + \beta_6 Analyst_{it} + \beta_7 AbnTurnover_{it} + \beta_8 ROA_{it} + \beta_9 EPS_{it} + \beta_{10} DR_{it} + \mu_{it} \quad (3)$$

$$AR_{i(t+j)} = \alpha_i + \beta_1 ASVI_{it} + \beta_2 MCI_{it} + \beta_3 ASVI_{it} * MCI_{it} + \beta_4 Cap_{it} + \beta_5 IH_{it} + \beta_6 AbsRet_{it} + \beta_7 Analyst_{it} + \beta_8 AbnTurnover_{it} + \beta_9 ROA_{it} + \beta_{10} EPS_{it} + \beta_{11} DR_{it} + \mu_{it} \quad (4)$$

4. Results

4.1. Online media coverage and SVI

We first present contemporaneous correlations among SVI and other common proxies for attention (see Table 1 for definitions), measurable at a weekly frequency in Table 2. Traditional measures such as news coverage, extreme returns, and trading volume are popular proxies that can be used for measuring investor attention; here, we mainly focus on the correlation between SVI and media coverage. According to the corresponding results from Da et al. [8], the correlation between SVI and media coverage is very low, ranging from 3.5% (Chunky News: number of news stories with multiple story codes in the Dow Jones news archive) to 5.0% (News: number of news stories in the Dow Jones news archive).

At least two reasons lead to the low correlation between SVI and newspaper news. First, SVI measures investor attention continuously over the year, while newspaper news of a typical firm is sporadic. [12] reported that >25% (50%) of NYSE (NASDAQ) stocks are not featured in the press in a typical year. Second, newspaper news does not guarantee attention unless investors actually read it, while search is a revealed attention measure: if you search for a stock in Google (Baidu), you are undoubtedly paying attention to it. Thus, newspaper news could not be used to classify the search volume.

Baidu calculates the MCI based on the Internet media reports and the keyword input. As stated by Baidu; the Internet media included are mainly official online publications of newspapers and magazines such as People's Daily; China Daily; and Securities Times; authoritative and popular financial websites such as Sina Finance; Tencent Finance; Netease Finance; Sohu Finance; and Phoenix Finance; and the official websites of the government and organizations. As compared with newspaper news; the online news collected from Baidu Index has a broader coverage and is more likely to be read by the online searchers; we thus expect to find a higher correlation between SVI and online media coverage. As reported in Table 2; the correlation coefficient between SVI and MCI is as high as 0.185; which is far more than that of using newspaper news; indicating that online media coverage can better measure investor attention and indeed leads to search volume. The positive relation between SVI and MCID suggests that more search volume is performed when there has been news and a certain proportion of SVI is induced by news; which is news-driven. Therefore; the online media coverage could be a valid proxy to differentiate search volume.

For instance, the news release of Steve Jobs' (in Chinese: 乔布斯) death (5 October 2011 US time) and iPhone6 can certainly lead to search volume from investors, consumers, and fans. As Fig. 3 shows, the search volume of "Steve Jobs" was very low before 5 October 2011, whereas it jumped rapidly thereafter and peaked on 6 October 2011. The search volume from Baidu Index presents similar trends: as Fig. 4 shows, the online media coverage and search volume of "乔布斯" peaked at the same time on 6 October 2011. Thus, it is clear that news-driven search volume can be generated along with online news release; in the online environment, they are almost synchronous.

Table 2
Correlations.

	SVI	Abs Ret	Abn Turnover	MCI	MCID
SVI	1				
Abs Ret	0.063***	1			
Abn Turnover	0.065***	0.262***	1		
MCI	0.185***	0.099***	0.098***	1	
MCID	0.150***	0.080***	0.073***	0.830***	1

Note: *** represents significance at 1% level.

4.2. The predictive power of SVI on China's stock market

Before further exploration of different types of search volume, we first test the validation of using SVI in China's stock market. As discussed before, we expect a stronger impact of SVI on China's stock market. We first investigate the empirical relations between ASVI and future stock returns (future 4 weeks) for CSI 300 stocks, firm-fixed panel regressions are used and standard errors are clustered by firms, and the results are reported in Table 3. We find strong evidence of positive price pressure in the next 2 weeks following an increase in individual attention as measured by ASVI (columns 1 and 2 in Table 3). Afterwards, the regression coefficients in weeks 3 and 4 become significantly negative, indicating a price reversal. The results are robust even if we control media coverage (columns 5–8), indicating the validation of using ASVI in China's stock market. As compared with the results of Da et al. [8] who only found significant impact of ASVI on Russell 3000 stocks in the first 2 weeks (weeks 3 and 4 are nonsignificant), we also find the significant negative impact in weeks 3 and 4, suggesting a more rapid and powerful impact of search volume on stock prices in China's stock market.

4.3. Self-initiated versus news-driven SVI

To explore whether different types of search volume have different predictive power on stock prices, we divide the search volume into two categories by the variable MCID that equals 1 when there has been news in a week, else 0. As discussed before, when there has been news, the search volume is more likely to be news-driven or may contain a greater proportion of news-driven search volume. Although there has been no news, the search volume is possibly spontaneously generated with the purpose of searching information for stock picking. According to the attention allocation theory, when there has been less news or no news, investors can have more time and effort on each searching, which can result in better consideration for making decisions. We thus expect the self-initiated rather than the news-driven search volume can generate positive price pressure. Based on the previous studies that confirmed the negative impact of media coverage on stock prices, we infer that the news-driven search volume is more likely to result in negative returns.

As is reported in Table 4 (columns 1–4), when there has been no news, the search volume is more likely self-initiated and is positively related to future stock returns in the first 2 weeks, and then reverses (nonsignificant at 5% level) in weeks 3 and 4. Although there has been news (the search volume is more likely

Please cite this article in press as: X. Liu, Q. Ye, The different impacts of news-driven and self-initiated search volume on stock prices, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.05.009

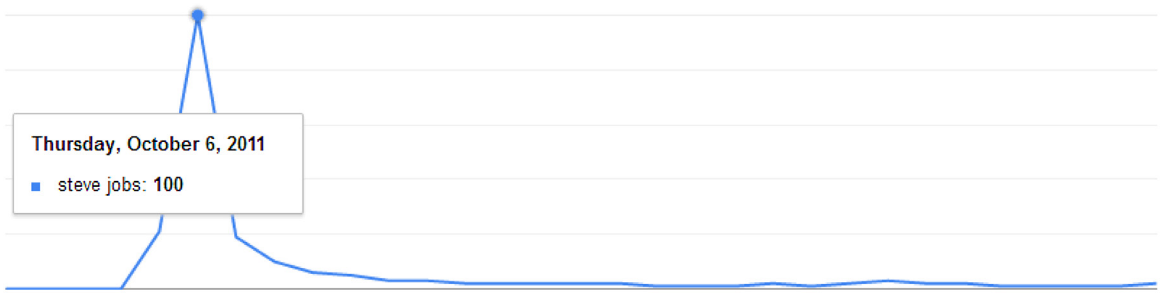

Fig. 3. The interest of "Steve Jobs" over time from Google Trends.

Fig. 4. The interest of "乔布斯" over time from Baidu Index.

news-driven), we can only find the negative relations between ASVI and future stock prices in weeks 3 and 4, and the positive impact of ASVI on stock prices in the first 2 weeks disappears (columns 5–8). This is to say, the self-initiated search volume is more likely to generate buy pressure and result in higher returns; however, the news-driven search volume is more likely to generate

Table 3
ASVI and CSI 300 stock returns.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Week 1	Week 2	Week 3	Week 4	Week 1	Week 2	Week 3	Week 4
ASVI	0.004***	0.010***	−0.003***	−0.004***	0.004***	0.010***	−0.003***	−0.003***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
MCI					−0.002***	−0.001	−0.001*	−0.001
					(0.001)	(0.001)	(0.001)	(0.001)
CAP	−0.015***	−0.014***	−0.017***	−0.020***	−0.015***	−0.014***	−0.017***	−0.020***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
IH	0.015***	0.014***	0.011***	0.010***	0.015***	0.014***	0.011***	0.010***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Abs Ret	−0.008	−0.023**	0.009	−0.028***	−0.006	−0.022**	0.010	−0.027***
	(0.010)	(0.009)	(0.008)	(0.009)	(0.010)	(0.009)	(0.008)	(0.009)
Analyst	0.000	0.000	−0.001**	−0.001**	0.000	0.000	−0.001**	−0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Abn Turnover	−0.002***	0.003***	0.001***	0.002***	−0.001***	0.003***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ROAs	0.009	0.013	0.023*	0.034***	0.009	0.013	0.023*	0.034***
	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)
EPS	0.001	−0.000	0.000	0.000	0.001	−0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
DR	−0.006	−0.012***	−0.013***	−0.004	−0.006*	−0.012***	−0.013***	−0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Constant	0.150***	0.144***	0.182***	0.202***	0.151***	0.144***	0.183***	0.203***
	(0.019)	(0.020)	(0.024)	(0.023)	(0.019)	(0.020)	(0.024)	(0.023)
Observations	46,864	46,207	46,381	46,054	46,864	46,207	46,381	46,054
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
cluster (firms)	299	299	299	299	299	299	299	299
R-squared	0.011	0.012	0.012	0.013	0.011	0.013	0.012	0.013

Note: The dependent variable is the abnormal return during the first 4 weeks, and independent variables are defined in Table 1. Robust standard errors clustered by firms are in parentheses. ***, **, and * represent significance at 1%, 5%, and 10% level, respectively.

Table 4
The impact of self-initiated versus news-driven search volume on stock prices.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Week 1	Week 2	Week 3	Week 4	Week 1	Week 2	Week 3	Week 4
ASVI	0.004***	0.011***	−0.002*	−0.002*	0.001	0.002	−0.012***	−0.015***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.004)	(0.003)	(0.004)	(0.004)
CAP	−0.016***	−0.016***	−0.021***	−0.022***	−0.014***	−0.012***	−0.014***	−0.018***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
IH	0.014***	0.012***	0.010***	0.008***	0.017***	0.015***	0.012***	0.012***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
Abs Ret	0.015	−0.017	0.014	−0.016	−0.026**	−0.022	0.012	−0.028**
	(0.013)	(0.012)	(0.011)	(0.012)	(0.013)	(0.014)	(0.013)	(0.012)
Analyst	0.001**	0.000	−0.000	−0.001*	−0.000	−0.000	−0.001*	−0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Abn Turnover	−0.002***	0.004***	0.002***	0.002***	−0.001	0.002**	0.001	0.002***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
ROAs	0.015	0.031*	0.025	0.011	−0.005	−0.013	0.018	0.058***
	(0.015)	(0.017)	(0.016)	(0.016)	(0.017)	(0.018)	(0.019)	(0.017)
EPS	0.002*	−0.001	0.001	0.002*	0.000	0.000	−0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
DR	−0.001	−0.016***	−0.014***	−0.001	−0.015**	−0.008	−0.012*	−0.012*
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Constant	0.154***	0.166***	0.218***	0.226***	0.152***	0.123***	0.153***	0.188***
	(0.021)	(0.020)	(0.023)	(0.024)	(0.027)	(0.028)	(0.033)	(0.032)
Observations	26,900	26,616	26,817	26,787	19,964	19,591	19,564	19,267
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
cluster (firms)	299	299	299	299	299	299	299	299
R-squared	0.016	0.019	0.019	0.018	0.020	0.019	0.018	0.025

Note: The dependent variable is the abnormal return during the first 4 weeks, and independent variables are defined in Table 1. Robust standard errors clustered by firms are in parentheses. ***, **, and * represent significance at 1%, 5%, and 10% level, respectively.

sell pressure and result in lower returns in the future. Thus, H1 is supported.

4.4. The moderating effect of media coverage

As the distinguishing impact of different types of search volume on stock prices is confirmed, media coverage may moderate the relations between search volume and stock prices as the number of news can influence the ratio between self-initiated and news-driven search volume in a week. For instance, if there has been more news in a week, the proportion of news-driven search volume would be higher than when there has been less news. In other words, more news in a week leads to a higher ratio between news-driven search volume and self-initiated search volume; thus, it is more likely to generate negative price pressure in the future. Two steps are taken to test the moderating effect of media coverage. First, we used the dummy variable of media coverage (MCID) to explore whether media coverage can moderate the impact of search volume on stock returns. Second, the number of media coverage (MCI) used to discover to what extent media coverage can moderate the predictive power of ASVI on stock returns.

The moderating effect of media coverage (dummy) on stock prices is reported in Table 5. ASVI is positively related to stock prices in the first 2 weeks, and then reverses. However, the moderating effect of media coverage lasts for 4 weeks and is consistently negative. Thus, media coverage can significantly moderate the impact of search volume on stock prices. When there has been no news, the search volume is more likely self-initiated, and investors are searching information for buying stocks. Although there has been news, the search volume is more likely driven by news headlines; thus, it is passively generated possibly for selling stocks. Table 6 reports the results of using the number of news instead of news dummy. In row 3 (Table 6), the interaction term (ASVI*MCI) is significant and negatively related to stock returns in 4 weeks, indicating that search volume is more influential on stock returns when there has been less news. Given the same amount of search volume in a week, less (more) news stands for a higher proportion of self-initiated (news-driven) search volume. When there has been more news, attention allocation is greater and results in low-quality consideration; conversely, when there has been less news, investors can concentrate more and think deeper of the information they are searching. Thus, H2 is supported.

Table 5
The moderating effect of media coverage (MCID).

	(1)	(2)	(3)	(4)
VARIABLES	Week 1	Week 2	Week 3	Week 4
ASVI	0.005***	0.012***	−0.002	−0.002*
	(0.001)	(0.002)	(0.001)	(0.001)
ASVI*MCID	−0.008**	−0.013***	−0.013***	−0.013***
	(0.004)	(0.004)	(0.003)	(0.004)
CAP	−0.015***	−0.014***	−0.017***	−0.020***
	(0.002)	(0.002)	(0.002)	(0.002)
IH	0.015***	0.014***	0.011***	0.010***
	(0.003)	(0.002)	(0.002)	(0.002)
Abs Ret	−0.005	−0.018*	0.014*	−0.022**
	(0.010)	(0.009)	(0.008)	(0.009)
Analyst	0.000	0.000	−0.001**	−0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Abn Turnover	−0.001***	0.003***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
ROAs	0.010	0.013	0.024**	0.035***
	(0.012)	(0.012)	(0.012)	(0.011)
EPS	0.001	−0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
DR	−0.006	−0.012***	−0.013***	−0.004
	(0.004)	(0.004)	(0.004)	(0.004)
Constant	0.151***	0.145***	0.183***	0.203***
	(0.019)	(0.020)	(0.024)	(0.023)
Observations	46,864	46,207	46,381	46,054
Fixed effects	Yes	Yes	Yes	Yes
luster (firms)	299	299	299	299
R-squared	0.011	0.013	0.012	0.014

Note: The dependent variable is the abnormal return during the first 4 weeks, and independent variables are defined in Table 1. Robust standard errors clustered by firms are in parentheses. ***, **, and * represent significance at 1%, 5%, and 10% level, respectively.

Table 6
The moderating effect of media coverage (MCI).

	(1)	(2)	(3)	(4)
VARIABLES	Week 1	Week 2	Week 3	Week 4
ASVI	0.005***	0.012***	−0.002*	−0.002**
	(0.001)	(0.002)	(0.001)	(0.001)
MCI	−0.002***	−0.000	−0.001	−0.000
	(0.001)	(0.001)	(0.001)	(0.001)
ASVI*MCI	−0.012**	−0.022***	−0.016***	−0.018***
	(0.005)	(0.005)	(0.005)	(0.005)
CAP	−0.015***	−0.014***	−0.017***	−0.020***
	(0.002)	(0.002)	(0.002)	(0.002)
IH	0.015***	0.014***	0.011***	0.010***
	(0.003)	(0.002)	(0.002)	(0.002)
Abs Ret	−0.003	−0.017*	0.014*	−0.022**
	(0.010)	(0.009)	(0.008)	(0.009)
Analyst	0.000	0.000	−0.001**	−0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Abn Turnover	−0.001***	0.003***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
ROAs	0.009	0.013	0.023*	0.035***
	(0.012)	(0.012)	(0.012)	(0.011)
EPS	0.001	−0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
DR	−0.006*	−0.012***	−0.013***	−0.005
	(0.004)	(0.004)	(0.004)	(0.004)
Constant	0.152***	0.146***	0.184***	0.204***
	(0.019)	(0.020)	(0.024)	(0.023)
Observations	46,864	46,207	46,381	46,054
Fixed effects	Yes	Yes	Yes	Yes
cluster (firms)	299	299	299	299
R-squared	0.012	0.013	0.012	0.014

Note: The dependent variable is the abnormal return during the first 4 weeks, and independent variables are defined in Table 1. Robust standard errors clustered by firms are in parentheses. ***, **, and * represent significance at 1%, 5%, and 10% level, respectively.

Table 7
Robustness test.

	(1)	(2)	(3)	(4)
VARIABLES	Week 1	Week 2	Week 3	Week 4
ASVI	0.008***	0.017***	0.001	−0.001
	(0.002)	(0.003)	(0.002)	(0.002)
MCI	−0.001***	−0.001**	−0.001***	−0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
ASVI*MCI	−0.004**	−0.008***	−0.004**	−0.003*
	(0.002)	(0.002)	(0.002)	(0.002)
CAP	−0.014***	−0.013***	−0.017***	−0.019***
	(0.002)	(0.002)	(0.002)	(0.002)
IH	0.015***	0.014***	0.011***	0.010***
	(0.003)	(0.002)	(0.002)	(0.002)
Abs Ret	−0.004	−0.019**	0.012	−0.025***
	(0.010)	(0.009)	(0.008)	(0.009)
Analyst	0.000	0.000	−0.001*	−0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Abn Turnover	−0.001***	0.003***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
ROAs	0.008	0.012	0.022*	0.034***
	(0.012)	(0.012)	(0.012)	(0.011)
EPS	0.001	−0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
DR	−0.006	−0.012***	−0.013***	−0.004
	(0.004)	(0.004)	(0.004)	(0.004)
Constant	0.148***	0.143***	0.181***	0.201***
	(0.019)	(0.020)	(0.023)	(0.023)
Observations	46,864	46,207	46,381	46,054
Fixed effects	YES	YES	YES	YES
cluster (firms)	299	299	299	299
R-squared	0.012	0.013	0.012	0.013

Note: The dependent variable is the abnormal return during the first 4 weeks, and independent variables are defined in Table 1. Robust standard errors clustered by firms are in parentheses. ***, **, and * represent significance at 1%, 5%, and 10% level, respectively.

4.5. Robustness

The media coverage of a stock can be identified by either stock ticker or stock name, but the stock ticker is a more strict and accurate criteria; however, the stock name is more extensively used in financial news reports. Thus, we also collected the MCI from Baidu Index based on stock name, and we replicate the results from Table 6 to check the robustness. The regression results are reported in Table 7, which are qualitatively similar to that of Table 6; the interaction term remains negative and significant at the 5% level in 3 weeks. Therefore, the empirical results are robust whether the media coverage is defined by stock ticker or stock name.

5. Conclusion

Both enterprises and scholars have noticed the great value of big data. Baidu has already closed its service of offering SVI and MCI for most stocks (besides financial-related search terms and other search terms such as movies are still available), and regards it as confidential data. With the progress of Internet technology and the arrival of big data, Da et al. [8] successfully found a direct proxy (weekly SVI from Google Trends) to measure retail investors' attention in a timely manner. The current related studies confirmed the impact of investor attention on the stock market, and this is the first step as information is useless without attention. The implicit assumption in existing studies is that attention measured under different situations is considered as equal. Our study extends the current studies to the second step: attention heterogeneity.

In search of attention heterogeneity and its impact on stock prices, we use a dataset from China's stock market and Baidu Index. We find a certain proportion of search volume that is news-driven and divide the search volume into news-driven and self-initiated based on the presence of news. The empirical results suggest that an increase in self-initiated search volume predicts higher stock prices in the next 2 weeks, while an increase in news-driven search volume predicts lower stock prices in the future. In other words, self-initiated (news-driven) search volume is more likely to generate buy (sell) pressure. The negative moderating effect of online media coverage indicates that the same amount of search volume is more powerful when under lower media coverage.

This study contributes to the knowledge of attention heterogeneity and attention allocation in the context of SVI, and optimizes the measures of investor attention. The new findings of this study enrich the attention theories by showing the existence of attention heterogeneity in a direct manner, as search is a revealed attention measure. The findings of this study also revealed the limitation of using Internet search to measure investor attention: search volume under different situations captures distinct attention level and leads to different decision-making processes; thus, researchers should be cautious when using search volume to measure investor attention. Furthermore, online search and online media coverage is closely connected in the age of big data; online news report can be immediately transmitted into searching behavior and affect the stock market. This study also reveals how retail investors process information differently under different situations.

Beyond testing theories of attention, the findings of this study have potential implications in practice and can significantly benefit the funds' rating systems for selecting stocks. Search volume is an objective and timely means of revealing and quantifying the interests of investors; therefore, it has many potential applications in Internet finance in this big data era. Given the wide and rapid application of search volume in Internet finance such as big data funds, which depend heavily on the search volume when selecting stocks, we suggest the fund companies to differently treat the search volume performed under different situations as distinguishing search environment for investors can help improve the measure for investor attention.

This study also has some limitations. First, the immaturity of China's stock market indeed creates uncertainties for conducting this study as some other forces may influence the stocks invisibly and largely, this systematic risk is inevitable when using datasets from emerging stock markets. Second, though Baidu Index states that it has worked considerably toward anticheating and established a set of corresponding anticheating technology systems to reduce it to a minimum, cheating may still occur. This is also a considerable risk when datasets collected from websites such as online reviews/ratings, online searching, and other user-generated content are used, as these valuable online data are always concerned with profits.

Acknowledgments

This research is partially supported by NKBRPC (2014CB340506) and NSFC (71225003, 71532004, 71490720).

References

[1] K.R. Ahern, D. Sosyura, Who writes the news? Corporate press releases during merger negotiations, J. Finance 69 (1) (2014) 241–291.
[2] K.R. Ahern, D. Sosyura, Rumor has it: sensationalism in financial media, Rev. Financ. Stud. 28 (7) (2015) 2050–2093.
[3] R. Agarwal, V. Dhar, Editorial—big data, data science, and analytics: the opportunity and challenge for IS research, Inf. Syst. Res. 25 (3) (2014) 443–448.
[4] B.M. Barber, T. Odean, All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors, Rev. Financ. Stud. 21 (2) (2008) 785–818.
[5] G. Birz, J.R. Lott, The effect of macroeconomic news on stock returns: new evidence from newspaper coverage, J. Bank. Finance 35 (11) (2011) 2791–2800.
[6] E. Blankespoor, G.S. Miller, H.D. White, The role of dissemination in market liquidity: evidence from firms' use of Twitter, Account. Rev. 89 (1) (2013) 79–112.
[7] H. Chen, P. De, Y.J. Hu, B.H. Hwang, Wisdom of crowds: the value of stock opinions transmitted through social media, Rev. Financ. Stud. 27 (5) (2014) 1367–1403.
[8] Z. Da, J. Engelberg, P. Gao, In search of attention, J. Finance 66 (5) (2011) 1461–1499.
[9] S.R. Das, M.Y. Chen, Yahoo! for Amazon: sentiment extraction from small talk on the web, Manage. Sci. 53 (9) (2007) 1375–1388.
[10] S. Dellavigna, J.M. Pollet, Investor inattention and Friday earnings announcements, J. Finance 64 (2) (2009) 709–749.
[11] M.S. Drake, D.T. Roulstone, J.R. Thornock, Investor information demand: evidence from Google searches around earnings announcements, J. Account. Res. 50 (4) (2012) 1001–1040.
[12] L. Fang, J. Peress, Media coverage and the cross-section of stock returns, J. Finance 64 (5) (2009) 2023–2052.
[13] L.H. Fang, J. Peress, L. Zheng, Does media coverage of stocks affect mutual funds' trading and performance? Rev. Financ. Stud. 27 (12) (2014) 3441–3466.
[14] G. George, M.R. Haas, A. Pentland, Big data and management, Acad. Manage. J. 57 (2) (2014) 321–326.
[15] S. Gervais, R. Kaniel, D.H. Mingelgrin, The high-volume return premium, J. Finance 56 (3) (2001) 877–919.
[16] G. Grullon, G. Kanatas, J.P. Weston, Advertising, breath of ownership, and liquidity, Rev. Financ. Stud. 17 (2) (2004) 439–461.
[17] D. Hirshleifer, S.H. Teoh, Limited attention, information disclosure, and financial reporting, J. Account. Econ. 36 (1) (2003) 337–386.
[18] J. Jin, Y. Li, X. Zhong, L. Zhai, Why users contribute knowledge to online communities: an empirical study of an online social Q&A community, Inf. Manage. 52 (7) (2015) 840–849.
[19] K. Joseph, M. Babajide Wintoki, Z. Zhang, Forecasting abnormal stock returns and trading volume using investor sentiment: evidence from online search, Int. J. Forecasting 27 (4) (2011) 1116–1127.
[20] D. Kahneman, Attention and Effort, Englewood Cliffs, NJ, Prentice-Hall, 1973 (p. 246).
[21] B. Liu, J.J. Mcconnell, The role of the media in corporate governance: do the media influence managers' capital allocation decisions? J. Financ. Econ. 110 (1) (2013) 1–17.
[22] D. Lou, Attracting investor attention through advertising, Rev. Financ. Stud. 27 (6) (2014) 1797–1829.
[23] X. Luo, J. Zhang, W. Duan, Social media and firm equity value, Inf. Syst. Res. 24 (1) (2013) 146–163.
[24] M.L. Markus, New games, new rules, new scoreboards: the potential consequences of big data, J. Inf. Technol. 30 (1) (2015) 58–59.
[25] L. Peng, W. Xiong, Investor attention, overconfidence and category learning, J. Financ. Econ. 80 (3) (2006) 563–602.
[26] J. Peress, The media and the diffusion of information in financial markets: evidence from newspaper strikes, J. Finance 69 (5) (2014) 2007–2043.
[27] M.S. Seasholes, G. Wu, Predictable behavior, profits, and attention, J. Empir. Finance 14 (5) (2007) 590–610.
[28] C.A. Sims, Implications of rational inattention, J. Monetary Econ. 50 (3) (2003) 665–690.
[29] K. Singh, S.C. Guntuku, A. Thakur, C. Hota, Big data analytics framework for peer-to-peer botnet detection using random forests, Inf. Sci. 278 (2014) 488–497.
[30] R.F. Stambaugh, J. Yu, Y. Yuan, The short of it: investor sentiment and anomalies, J. Financ. Econ. 104 (2) (2012) 288–302.
[31] P.C. Tetlock, Giving content to investor sentiment: the role of media in the stock market, J. Finance 62 (3) (2007) 1139–1168.
[32] N. Vlastakis, R.N. Markellos, Information demand and stock market volatility, J. Bank. Finance 36 (6) (2012) 1808–1821.
[33] B. Xu, D. Li, An empirical study of the motivations for content contribution and community participation in Wikipedia, Inf. Manage. 52 (3) (2015) 275–286.
[34] S.X. Xu, X.M. Zhang, Impact of Wikipedia on market information environment: evidence on management disclosure and investor reaction, MIS Q. 37 (4) (2013) 1043–1068.
[35] Y. Yuan, Market-wide attention, trading, and stock returns, J. Financ. Econ. 116 (3) (2015) 548–564.
[36] X.M. Zhang, L. Zhang, How does the internet affect the financial market? An equilibrium model of internet-facilitated feedback trading, MIS Q. 39 (1) (2015) 17–38.

Xianwei Liu is a PhD candidate at the School of Management, Harbin Institute of Technology, China. His research interests include E-commerce, Internet finance, and behavioral finance. His papers have appeared in *International Journal of Hospitality Management*, *Asia Pacific Journal of Tourism Research*, and others.

Qiang Ye is a professor of MIS and the dean of School of Management, Harbin Institute of Technology, China. He received his PhD from the School of Management, Harbin Institute of Technology. He is currently serving as a senior editor for *Journal of Electronic Commerce Research* and an associate editor for *Electronic Commerce Research and Applications*. His research interests include E-commerce, Internet finance, and data mining. His papers have appeared in *Production and Operations Management*, *International Journal of Hospitality Management*, *Decision Support Systems*, and others.

Please cite this article in press as: X. Liu, Q. Ye, The different impacts of news-driven and self-initiated search volume on stock prices, Inf. Manage. (2016), http://dx.doi.org/10.1016/j.im.2016.05.009