Information Bias and Stock Market*

Hang Dong†
Jie Ren‡

Abstract

This paper hypothesizes and tests information bias in the stock market. We propose that information in the stock market may be systematically biased as irrational investors, or noise traders, affect information distribution in the stock market and rational investors, or arbitrageurs, choose to conveniently use the available but biased information. Our empirical evidence shows (i) investors tend to exhibit confirmation bias, i.e., read information that is consistent with their prior beliefs, at the aggregate level; (ii) information channels provide information that is slanted towards investors' prior beliefs, and therefore, information for noise traders are biased; (iii) arbitrageurs are also subject to biased information, and as such, they are not able to correct asset pricing errors caused by biased information.

Keywords: information bias; stock market; noise trader; confirmation bias; stock return; financial decisions; limits to arbitrage.

---

*The authors are indebted to Yuanping Wang, Xikun Liu, and Lu Liu for providing the Sina data. The authors also thank Xuezhi Cai and Shoufu Luo for their kind help on programing.

†Corresponding author. IE University, Calle Cardenal Zuñiga 12, 40003 Segovia, Spain, hang.dong@ie.edu

‡Gabelli School of Business, Fordham University, 140 West 62nd Street, New York, NY 10023, jren11@fordham.edu

The role of the news media in the stock market is not, as commonly believed, simply as a convenient tool for investors who are reacting directly to the economically significant news itself. The media actively shape public attention and categories of thought, and they create the environment in which the speculative market events we see are played out. (Shiller 2015, p.121)

1 Introduction

Information is critical to stock investors. To make optimal investment decisions, investors spend tremendous amounts of money and effort to acquire stock-related information that is timely, complete and accurate. Bloomberg Terminal, a commonly used financial information channel, charges US$24,000 per terminal per year (Seward, 2013). The price of Eikon Premium (Thomson Reuters) is around US$19,000 per year.[1] Marketwatch.com, a free stock-information platform, had 16 million unique visitors and 84.1 million views in the 30-day period from October 7 to November 5, 2016 (Quantcast, obtained in November 2016).[2] Seeking Alpha, a user-generated opinion platform, had between 500,000 and 1 million unique visitors per day in August 2013 (Chen et al., 2014). In this paper, we hypothesize and examine that information in the stock market is systematically biased, which leads to asset pricing error.

Previous literature in behavioral finance recognized the importance of information transfer in finance (Fang and Peress, 2009; Engelberg and Parsons, 2011; Tetlock, 2014). They argued that investors' irrational decision-making process (e.g., investors fail to search or process information) is responsible for asset pricing error. Following De Long et al. (1990); Baker and Wurgler (2006), one strand of literature argued that irrational investors, or noise traders, incorrectly react to contents in the information channels (e.g., news reports). For example, Barber and Odean (2008) documented that individual investors are subject to limited attention: they are more likely to buy stocks in the news. Tetlock (2007); Dougal et al. (2012); García (2013) argued that media coverage on stock is associated with investor sentiment, i.e., beliefs that cannot be justified by facts, and consequently, it affects aggregate stock market outcome (e.g., stock return and trading volume). Antweiler and Frank (2004); Bollen et al. (2011); Siganos et al. (2014); Chen et al.

[1] The price is obtained by the authors as of February, 2016.

[2] As a comparison, the website of New York Times, www.nytimes.com, reaches over 23 million people monthly in the United States (Quantcast, obtained in November 2016).

(2014) showed similar evidence that information in the social media, an information channel, can affect stock return. Another strand of literature theoretically and empirically documented evidence that some information can predict future stock return. They interpret this anomaly as a result of investors systematically ignoring some information in the stock market: both individual and professional investors are subject to limited cognitive abilities (Hirshleifer and Teoh, 2003; DellaVigna and Pollet, 2007; Dellavigna and Pollet, 2009; Fang et al., 2014).

A right financial decision requires both unbiased information input and rational decision making process. Previous studies focused on decision-making process (e.g., investors fail to react to information correctly or they have limited cognitive ablities to search for information), leaving an implicit assumption that information input is always available and free-of-bias, and therefore, is uncorrelated to asset pricing error. To the best of our knowledge, few previous study has discussed the role of information input to explain asset pricing error. This paper aims to fill in this research gap. We argue that asset pricing error, as described in section 3, results from systematic information transfer defect, i.e., information bias, in the stock market.

There are at least two related mechanisms through which systematic information bias appears. First, irrational investors, or noise traders, preferentially value information that is consistent with their previous beliefs, i.e., confirmation bias (Nickerson, 1998; Park et al., 2013). Consequently, information channels may overvalue some information and undervalue other information to maximize their viewership (Mullainathan and Shleifer, 2005). As such, information in the information channels is biased. Second, it is both difficult and economically inefficient for rational investors, or arbitrageurs, to find unbiased information (French, 2008). Investors, including both noise traders and arbitrageurs, may conveniently use available but biased information instead. If investors make stock trading decisions based on biased information, one expects that their decisions are not optimal and stock price can be wrong.

It is worth noting that the term "information bias" here does not necessarily imply that the information is fake. In this paper, we define "information bias" as information channels create and/or distribute information NOT based on its relevance to investor optimal decisions. Information bias may include but not limit to two meanings: first, information channels may selectively report (or ignore) the information that is favorable (or unfavorable) to their interests. Second, information channels may explain the same information in a tone favorable to their interests. For example, Mullainathan and Shleifer (2005) illustrated that the same piece of information may be described using different tones, suggesting either positivity or negativity.

We test our hypotheses using the data from Sina Finance and Sina Weibo for 2013 and 2014. Sina Finance is an Internet news platform in China. It includes almost all public financial information available on the Chinese stock market. Sina Weibo is a microblogging service (similar to Twitter) in China. Our original dataset includes 43.17 million stock-related weibos and 4.27 million pieces of stock news. Our dataset has several appealing features: (1) it includes all financial news and weibos associated with stocks for the years of 2013 and 2014. In other words, we did not sample Sina users; (2) the Chinese stock market is a suitable market to study because most of the information recipients in this market are individuals. As of the end of 2014, more than 99.6% of investors, or 72.71 million, in the Chinese stock market are individuals. (3) The Chinese stock market is one of the most important and understudied markets. The Chinese stock market accounted for 37.7% and 42.6% of the trading volumes and the transactions of trading in the world in 2014, respectively (World Federation of Exchanges, 2015). Given the importance of the Chinese stock market, it is surprising to find very few studies on Chinese stock market.

In the empirical tests, we first identified the cause of information bias. Following Mullainathan and Shleifer (2005) and Park et al. (2013), we conjectured that information bias may result from confirmation bias, i.e., investors over-value the information that confirms their prior beliefs while undervaluing or ignoring the information that is inconsistent with their beliefs. We examined whether confirmatory information, *ceteris paribus*, receives more viewership at the aggregate level. We then scrutinized the presence of information bias in the stock market. Specifically, we used investor sentiment to proxy for investor belief.[3] We provided evidence that investor sentiment can predict future information sentiment. Finally, we indirectly tested whether arbitrageurs are subject to biased information. We argued that if their financial decisions are based on biased information, they cannot correct the asset pricing error caused by noise traders. The results supported our hypotheses.

Our paper contributes to literature in three ways. First, we provide an information-based explanation for asset pricing error. Previous behavioral finance literature argued that noise traders make irrational investment decisions that deviate stock price from its fundamental values, yet, arbitrageurs are unable to correct these stock errors due to limits to arbitrage (De Long

[3]In this paper, investor belief and investor sentiment are interchangeable.

et al., 1990; Baker and Wurgler, 2006, 2007). Empirical studies further supported this argument by showing that investor sentiment can predict stock return (Baker and Wurgler, 2006; Tetlock, 2007; García, 2013; Da et al., 2015; Huang et al., 2015). These studies neglected to examine the role of information in the stock market and implicitly assumed information is unbiased and free to be obtained. Instead, we argue that if information is systematically biased, decisions by both noise traders and arbitrageurs will be tainted, leaving noise traders with more chances to deviate stock prices and arbitrageurs fewer chances to correct asset pricing error.

Second, our paper also contributes to the information system (IS) literature on the role of information channels in the stock market. Specifically, is stock information transfered through information channels noisy or "value-relevant"? Many studies documented the association between information channel (e.g., mass media and social media) and stock return (Antweiler and Frank, 2004; Luo et al., 2013; Sprenger et al., 2014). However, studies generate inclusive results whether such association was based on noisy information or value-relevant information. For example, Chen et al. (2014) explicitly interpreted the association between opinions of social media and future stock return as that social media contains "value-relevant information, which, as of the article publication date, are not fully factored into the price" (p. 1369). Fang and Peress (2009); Tetlock (2010) argued that information channels help to alleviate informational friction in the stock market, implicitly implying information transferred in information channels is value-relevent. Contrarily, Tetlock (2007) showed evidence that information in mass media is noisy. Fang et al. (2014) documented that media coverage predicts a negative future return, consistent with the theoretical model of noise traders. Our paper provides further evidence on this topic.

Third, our findings enrich the literature on investors' information-search behaviors. Park et al. (2013) documented that investors exhibit confirmation bias in the stock market at the individual level. As we discussed in Section 3, their conclusions did not necessarily imply (1) that confirmation bias exists at the aggregate level, and (2) that confirmation bias is able to affect the stock market with the presence of arbitrageurs. Our empirical evidence confirms their findings at the aggregate level, and more importantly, shows the collective information-search behaviors play a role in affecting the stock market. As such, our paper highlights the importance of information system in the finance studies and calls for future research.

The remainder of this paper is structured as follows. Section 2 reviews the previous explanation of asset pricing error in behavioral finance literature. In Section 3, we provide an information-based view and develop our hypotheses. Sections 4 - 7 describe the methods used in the study, offer analysis, present results, and, finally, discuss the implications for future research, limitations of the study, and the outlook.

2 Traditional Explanation of Asset Pricing Error

From the Great Crash of 1929 to the subprime mortgage crisis of 2007 to 2009, finance history is full of asset pricing errors that the Efficient Market Hypothesis (EMH) cannot explain (Baker and Wurgler, 2007). One of the explanations lies in investors' irrational decision-making process. Keynes (1936) argued that the market is not only driven by rational factors but also by "animal spirits", and Black (1986) argued that some investors act on noise and cited them as "noise traders". Further, De Long et al. (1990) showed theoretically that noise traders may be a cause of asset pricing error. These authors argued that noise traders are subject to sentiment, and, as such, they aggregately generate sentimental demand shocks for stocks in the market.

However, the presence of noise traders does not necessarily imply that asset prices are erroneous in the market. The EMH argued that arbitrageurs can reasonably respond to information and correct these price deviations so that price can reflect the true value of corporate fundamentals (Barberis and Thaler, 2003). Yet, behavioral finance research documented that in reality arbitrageurs may not always be able to correct prices to match fundamentals due to limits to arbitrage (e.g., due to risks that cannot be fully eliminated, and transaction costs) (Shleifer and Vishny, 1997; Barberis and Thaler, 2003; Baker and Wurgler, 2006, 2007). To sum up, previous behavioral studies explicitly made two assumptions of (1) sentimental demand shocks and (2) limits to arbitrage, and consequently concluded that asset pricing error can result from irrationality of noise traders (De Long et al., 1990; Brown and Cliff, 2005; Baker and Wurgler, 2006).

The follow-up empirical studies indirectly tested whether noise traders can affect stock market following two approaches. First, Brown and Cliff (2005); García (2013); Da et al. (2015) examined a reversal effect in stock return as predicted by the assumptions: investor sentiment is positively associated with contemptuous stock return as noise traders overreact or underreact to information; yet the association between investor sentiment and future stock return reverses to be negative as market prices return to fundamental values. Second, the literature also predicted that some stocks (e.g., small, young, unprofitable, extreme-growth, and high-beta) are relatively easy for noise traders to speculate on. Since arbitrage forces are weaker for these stocks, they are more vulnerable to investor sentiment. As such, a few studies tested the heterogeneity of the impact of investor sentiment on stock returns across stocks (Baker and Wurgler, 2006; Stambaugh et al., 2012, 2014; Huang et al., 2015).

In summary, these studies explained asset pricing error as a result of (1) irrational decision-making processes of noise trader and (2) limits to arbitrage. However, the correctness of a financial decision depends not only on decision making processes as argued in previous literature, but also on the information input, which is less studied by previous literature. In the next section, we focus on this understudied factor, information, and argue that information in the stock market is systematically biased. As the input of decision-making process, information can affect both noise traders and arbitrageurs' financial decisions.

3 New Explanation: An Information-Based View

In a simple decision making model, the final decisions, or "action", requires both information input and "decision" process (Forrester, 1992). A correct financial decision is based on both the fact that information obtained is unbiased and the fact that "decision" process is rational. Any of prerequisites may affect the final decisions. Unlike the previous explanation focusing on the "decision" process, we provide an information-based view to explain asset pricing error in this section.

Investor information-search behavior is not always rational. In this paper, following Nickerson (1998); Mullainathan and Shleifer (2005); Park et al. (2013), we assume that some investors (e.g., noise traders) are subject to confirmation bias. They overvalue the confirmatory information while devaluing other information. One prediction is that noise traders desire to read information that is consistent with their prior beliefs more than other information. Relating to the stock market, Park et al. (2013) measured confirmation bias as the number of times an investor clicks on a message that supports their sentiment. They found that investors with strong beliefs about a stock's return performance are more likely to read the messages that confirm their sentiment. More concisely, they showed confirmation bias exists at the individual level in the context of stock information seeking.

However, the presence of confirmation bias at the individual level does not imply noise traders, as a group, exhibit confirmation bias. For example, if noise traders hold prior beliefs randomly, their desires to read confirmatory information may be cancelled out and there is no aggregated viewership shift.

We argue that at the aggregate level noise traders may hold convergent prior beliefs and seek confirmatory information. First, individual noise traders' prior beliefs are in fact established and developed based on the same events and/or company announcements. These events and/or company announcements may lead to similar investor sentiment. As Shleifer and Summers (1990) puts, individual investors "tend to make the same mistakes; they do not make random mistakes." (p23). Second, investors exchange their private information and opinions through various channels, such as social media (Antweiler and Frank, 2004; Chen et al., 2014). With exposure to others and others' opinions, people tend to converge their own opinions to follow the majority of opinions - people are subject to social influence (Asch, 1955; Muchnik et al., 2013). Similarly, noise traders' beliefs may be influenced by each other and later converge to be similar or the same as a group. As such, we expect that a confirmation bias would exhibit itself also at the aggregate level. Accordingly, we propose our first hypothesis:

Hypothesis 1 (H1) *Stock information attracts more viewership when its sentiment is consistent with investors' prior beliefs.*

Information channels are compensated by its viewership, either in the form of subscription revenue or in the form of advertising revenue. In order to maximize their viewership and profit, information channels may want to provide information in a way that caters to readers' belief. As a result, the provided information may be biased. Mullainathan and Shleifer (2005) provided evidence that information channels' profit-maximizing choice of catering to the preferences of readers leads to information bias. Moreover, these authors showed that if readers share common beliefs, competition among information channels does not help to eliminate the information bias but rather to generate it. Particularly, as we argued in Hypothesis 1, noise traders share similar beliefs in the stock market. So, one may expect that information channels "slant" stories towards noise traders' prior beliefs. Since the information in the information channels is biased, all noise traders can only get access to the biased information. We propose the following hypothesis, which suggests information bias:

Hypothesis 2 (H2) *Investors' prior beliefs about stocks affect stock information sentiment.*

In the following section, we hypothesize that arbitrageurs are subject to biased information that stems from information channels catering to noise traders' collective information searching behaviors. Ideally, we could test this hypothesis by observing whether information received by arbitrageurs is biased or not or whether final financial decisions made by arbitrageurs are optimal or not. However, we can only observe the aggregated outcome of both noise traders' decisions and arbitrageurs' decisions, i.e., stock return. As such, we follow two steps to examine this hypothesis. First we examine whether information sentiment is associated with asset pricing error, which suggests that arbitrageurs fail to correct the asset pricing error (H3a). However, both traditional behavioral finance explanation and our information-based explanation predict this result. So, in the second step, we argue that asset pricing error exists with or without limits to arbitrage, which is consistent with our information-based explanation but is inconsistent with traditional explanation (H3b).

In H2, we showed that information provided by information channels is systematically biased. Consequently, noise traders' decisions are biased as information input is biased and decision process is irrational. However, H2 is not a sufficient condition for that arbitrageurs use biased information and make suboptimal decisions. It is possible that information channels are divided into two segments: public and professional, which may target noise traders and arbitrageurs respectively. Information channels that target noise traders may have information bias to cater to readers' confirmation bias. Other channels targeting arbitrageurs may provide bias-free information that is based on its relevance to stock fundamental value. In this case, arbitrageurs are NOT subject to information bias, and consequently, information bias does not affect stock market if there are no limits to arbitrage.

However, this concern may not be correct for two reasons. First, noise traders outnumber arbitrageurs. Public information channels generate most of information. In practice, almost all major public information channels (e.g., *CNN Money, New York Times, Baron*, and *Yahoo Finance*) target all investors, a large fraction of which are noise traders. Moreover, information, biased or unbiased, is not free to be obtained. It is NOT economically efficient for information channels to generate, verify or diffuse information if they only target arbitrageurs. For example, Bloomberg, a common professional information channel, provides data services to professionals. Yet, other information provided by Bloomberg, including its news services and TV, are also open to public. Second, information flows from one channel to another. No information channels, either public or professional, contains all information related to the stock market. To obtain complete sets of information, arbitrageurs need to seek information from public channels, which is full of biased information. Also, information channels distribute information from other channels. In particular, public information channels report information (e.g., cite or reproduce information) from professional channels and professional channels include information from public channels. Therefore, it is evitable that information bias is contagious and can transfer from one information channel to another. Fang et al. (2014) provided empirical evidence that professional investors, who are more likely to be arbitrageurs, are influenced by the reports form public channels (e.g., mass media coverage). These authors documented some professional investors persistently buy the stocks that are in the public media coverage.

Arbitrageurs, by definition, make decisions rationally. More straightforward, unbiased information can be attractive to them if the benefit of finding unbiased information exceeds the cost. It may be economically inefficient for arbitrageurs to obtain or search for, at least some, unbiased information (French, 2008). Arbitrageurs may need to collect all information and then evaluate the accuracy of the information. They also need to judge whether the information is relevant to stock fundamental value. As such, arbitrageurs may conveniently use available but biased information instead. Empirical evidence also supported this argument. For example, arbitrageurs are subject to limited attention and use information from mass media, which leads to a lower performance (Fang et al., 2014). This choice is probably rational and optimal as the cost of finding unbiased information exceeds the benefit. Thus, we conjecture that arbitrageurs also suffer from information bias. Based on biased information input, arbitrageurs, even if their decision-making process is rational, are not able to correct price error caused by noise traders. Specifically, we can observe that information sentiment is associated with stock return.

Hypothesis 3a (H3a) *Stock information sentiment affects stock return.*

Previous behavioral finance literature also predicts the same results in H3a (García, 2013). In other words, asset pricing error can be either a result of limits to arbitrage, as in previous explanation, or a result of biased information input. To further examine the conjecture that

H3a is driven by our information-based explanation rather than by traditional explanation, we scrutinize the assumptions in two explanations. Studies of traditional behavioral finance argued that arbitrageurs may not be able to correct asset pricing errors because of limits to arbitrage while our information-based view argues that stock pricing error results from biased information input for investors' decisions. Given unbiased information, previous explanation predicts arbitrageurs cannot (can) correct asset pricing errors in the presence (absence) of limits to arbitrage. Consequently, the effect of noise investor varies across stocks with different difficulties of arbitrage (Baker and Wurgler, 2006; Da et al., 2015). In contrast, our information-based explanation argues that given biased information, arbitrageurs are not able to correct pricing error, with or without the presence of limits to arbitrage. In other words, the impact of information bias on stock returns does not vary across stocks with different levels of arbitrage difficulties. Table 1 summarizes and compares the predictions of two explanations. Accordingly, we propose our last hypothesis:

Hypothesis 3b (H3b) *The effect of information sentiment on stock returns does NOT vary across stocks with different difficulty levels of arbitrage.*

Table 1: Effect of limits to arbitrage on stock price based on two explanations

	Information Input	
	Unbiased	Biased
No limits to arbitrage	Without errors	Errors
Limits to arbitrage	Errors	Errors

4 Data

To test our hypotheses and theoretical model, we analyzed data from *Sina Finance* - a finance news platform that creates and distributes its own stock news and mainly distributes stock news from other sources - and *Sina Weibo*, a microblogging service in China. In the following sections, we detail the data-collection process and our measurement.

4.1 Investor Sentiment

We use investor sentiment obtained by Sina Weibo to proxy for previous belief of noise traders. Da et al. (2015) argued that "investor sentiment can be directly measured through the Internet search behavior of households" (p. 2). Chen et al. (2014) also showed that social media play an important role in transmitting investment advice.

We collected microblogging data from Sina Weibo from 2013 and 2014. As reported by 2015 annual report of Sina, Sina Weibo's operating company, as of December 2015, Sina Weibo had 235.7 million monthly active users and an average of 106.3 million daily active users. Weibo is considered as the most influential microblogging platform in China (Harwit, 2014).

Sina provided us with the original sentiment data. The company first extracted all weibos that mention Chinese stocks by using Ticker and *Jiancheng*, i.e., short name of stock in Chinese, during 2013 and 2014. Then the company used its proprietary sentimental dictionary to measure the sentiment of weibos. The company used 1, 0 and -1 to indicate positive, negative, and neutral sentiment, respectively.

As stated earlier, our original dataset includes more than 43.17 million weibos from January 2013 to December 2014, each of which refers to at least one stock. We calculated investor sentiment following Antweiler and Frank (2004):

$$B_{i,t} = ln(\frac{1 + n_{i,t}^P}{1 + m_{i,t}^N}) \tag{1}$$

where $B_{i,t}$ is the proxy for investor sentiment on the stock i at day t, and $n_{i,t}^P$ ($m_{i,t}^N$) is the number of weibos with positive (negative) sentiment for stock i at day t. Finally, our database contains 1,436 million stock-day observations.

4.2 Information Sentiment

Sina provided us with more than 6.2 million information items for 2013 and 2014, 4.27 million of which mention at least one stock. Some news provided by Sina may mention several stocks. For example, every morning, Sina Finance creates and displays a piece of news that summarizes all stock-related announcements from companies. In these summaries, a different sentiment may be conveyed about each stock: some are positive and some are negative. We suspect that the sentiment of one stock may affect the sentiment of another stock in the same news. As such, we retained only information that mentions one stock, narrowing down our database to 3.46 million information items[4]. It is also important to note that "information" in our dataset refers not only to news reports from published mass media but also to company announcements and financial analyst reports. We define "information" in our dataset as all published information regarding each stock.

The company of Sina provided information sentiment in a way similar to investor sentiment of weibo. We calculated the "information sentiment" ("InfoSent") for each stock mentioned, following Equation 1.

4.3 Verification of Investor Sentiment and Information Sentiment

We verified Sina's sentiment measure to make sure it was robust. Like other commercial sentiment databases (e.g., Thomson Reuters MarketPsych indices and Facebook's Gross National Happiness index), Sina does not release its sentiment word dictionary to the public, so we examined the company's sentiment measurement indirectly. Specifically, we associated information sentiment with information content. We randomly selected 40,000 articles published on Sina in December 2014. Then we used third-party commercial software, *BosonNLP*, to measure the sentiments of these news articles and compared them with the sentiment measure provided by Sina. As BosonNLP does not generate a neutral value, we only examined the news marked as positive and negative. We found that 82.67% of the sample provided by Sina has the same sentiment as that determined by the third-party software.

Many practitioners have scrutinized Sina's sentiment measure. Specifically, we verified, through public information (e.g., practitioners' webpages and news reports), that the Shenzhen Stock Exchange, Nanfang Mutual Fund Management Company, and Dongxin Mutual Fund Management Company purchased a similar sentiment database from Sina to edit indices or manage assets.

[4]This choice does not affect our conclusions.

4.4 Information Viewership

Following Park et al. (2013), we used the natural logarithm of 1 plus the number of clicks on the information to measure information viewership, that is,

$$IV_i = ln(1 + PV_i) \tag{2}$$

where PV_i is the number of clicks on information i.

4.5 Abnormal Return

We followed Carhart (1997) measure to estimate abnormal stock return. First, we regressed the daily stock return on Fama and French (1993) three factors (MKT, SMB, and HML for details see below) and the momentum factor for the previous quarter as follows:

$$r_{i,t} = \alpha_i + \beta_{1,i}MKT_t + \beta_{2,i}SMB_t + \beta_{3,i}HML_t + \beta_{4,i}MOM_t + \epsilon_{i,t} \tag{3}$$

where $r_{i,t}$ is net return in month, t, in excess of the daily interest rate derived from three months of the Shanghai Interbank Offered Rate; MKT_t is the market portfolio return in excess of the risk-free rate; SMB_t is the return on a portfolio of small-size stocks minus large-size stocks; HML_t is the return on a portfolio long high book-to-market stocks and short low book-to-market stocks; MOM_t is the return difference between stocks with high and low past returns in the past three months.

Then, we calculated abnormal return as the difference between the return and the realized risk premium, that is:

$$\alpha_{i,t} = r_{i,t} - (\hat{\beta}_{1,i}MKT_t + \hat{\beta}_{2,i}SMB_t + \hat{\beta}_{3,i}HML_t + \hat{\beta}_{4,i}MOM_t) \tag{4}$$

where $\alpha_{i,t}$ is the abnormal return. $\hat{\beta}_{1,i}$, $\hat{\beta}_{2,i}$, $\hat{\beta}_{3,i}$ and $\hat{\beta}_{4,i}$ are the loadings estimated by Equation 3, over the past quarter. The abnormal return, $\alpha_{i,t}$, captures the returns that cannot explained by these four factors.

4.6 Control Variables

Our control variables are size, PB, ROE, past stock return, and past information attitude. “Size” is defined as the natural logarithm of 1 plus the number of shares times the closing price; “PB ratio” is the price-to-book ratio, estimated as market capitalization over company net assets; “ROE” is return on equity, calculated as net profit over the shareholders’ equity; past stock return, “CumRet,” is cumulative return over the past five days.[5] Past information sentiment, “InfoSent”, is the n day lagged information sentiment.

5 Empirical Strategies and Results

5.1 Confirmation Bias and Viewership (H1)

According to confirmation bias documented in the previous literature (Nickerson, 1998; Mullainathan and Shleifer, 2005; Park et al., 2013), we argued that noise traders at the aggregate level exhibit confirmation bias. As such, information attracts more clicks if its sentiment is consistent with that of investors’ prior beliefs. To test this hypothesis, we regressed information viewership (“InfoView”) on investor sentiment (“InvSent”) at t-1, information sentiment (“InfoSent”), and their interaction term - that is:

$$\text{InfoView}_i = \alpha + \beta_1 \text{InvSent}_i + \beta_2 \text{InfoSent}_i + \beta_3 \text{InterTerm}_i + \beta_4 \text{Control Variables}_i \qquad (5)$$

where “InfoView_i” is the natural logarithm of 1 plus the number of clicks on the information i; “InvSent_i” is the investor sentiment for the stock mentioned in information i at day *t-1*; “InfoSent_i” is the information sentiment for information i; “InterTerm_i” is the product of InvSent_i and InfoSent_i; “Control Variable” includes size, PB ratio, ROE, past stock return, and lagged “InfoSent”. Column (1) in Table 2 shows the results.

Our main variable of interest is the interaction term. The interaction term measures consistency of lagged investor sentiment and information sentiment. When both variables have the same direction, i.e., both are positive or negative, the value of interaction term is positive, and vice versa. The coefficient on the interaction term is significantly positive. The result shows that

[5] We have used cumulative return over past one day, three days, five days and fifteen days. The choice does not affect our conclusions.

a piece of information receives more clicks when its attitude is consistent with one-day-lagged investor sentiment. This result supported our Hypothesis 1.

In addition, Column (1) in Table 2 also presents an extra interesting result: "InfoSent" is negatively and significantly correlated with the number of clicks on the information, suggesting a negativity bias effect: on average, investors value negative information more than positive information.

Table 2: Confirmation Bias and Information Viewership

	(1)		(2)	
	Coef.	t-stat	Coef.	t-stat
Belief	-0.0003	-0.05	0.0013	0.16
InterTerm	0.0777***	15.72	0.0759***	13.67
InfoSent	-0.1356***	-10.32	-0.0424***	-3.53
CumRet	1.5587***	14.87	1.5789***	14.64
Size	0.1529***	5.76	0.1507***	5.61
PB	-0.0004**	-2.35	-0.0004**	-2.52
ROE	-0.0001	-1.70	-0.0001*	-1.77
Cons	-1.2321**	-2.07	-1.1752*	-1.94
Num of Obs	350,277		329,731	
Adjusted R2	0.0312		0.0292	

Notes: The table reports the results of OLS regression of equation 5. "Belief" is investor sentiment on this stock at t-1, i.e., "InvSent", at column (1) and the residual from a regression of "InvSent" on lagged "InfoSent" at column (2) ; "InfoSent" is the information sentiment; "InterTerm" is the interaction term of Belief and InfoSent; "Size" is defined as the natural logarithm of 1 plus the number of shares times the closing price; "PB" is the price-to-book ratio, estimated as market capitalization over company net assets; "ROE" is return on equity, calculated as net profit over the shareholders' equity; past stock return, "CumRet", is cumulative return over the past five days. The regressions incorporate daily fixed effect. *, **, and *** denote significance at the 0.10, 0.05, and 0.01 levels, respectively.

There was a concern that both information sentiment and investor sentiment are driven by the lagged information sentiment in previous periods. To address this concern, we first regressed investor sentiment at t on information sentiment from the past five periods. Then we obtained the residual from the regression as the proxy for our new measure of investor sentiment. This new measure excludes the impact from previous news sentiment. We re-ran the regression in equation (5). Column (2) in Table 2 shows the results. The results suggest the same as in Column (1).

5.2 Information Bias in Information Channels (H2)

Hypothesis 2 argues that information in the stock market is biased; in other words, that in order to attract more clicks information channels purposely create, display or explain information in a way consistent with investors' previous sentiment. As such, the hypothesis predicts that investor sentiment for stock i at day *t-1* forecasts information sentiment for stock i at day t. We regressed information sentiment ("InfoSent") for stock i on day t on the one-day-lagged investor sentiment ("InvSent") and other control variables:

$$\text{InfoSent}_{i,t} = \alpha + \beta_1 \text{InvSent}_{i,t-1} + \beta_2 \text{Control Variables}_{i,t-1} \tag{6}$$

where "$\text{InfoSent}_{i,t}$" is the information sentiment for stock i at day t; "$\text{InvSent}_{i,t-1}$" is the investor sentiment for stock i at day *t-1*. We controlled for the information sentiment during the past seven days. The other control variables include size, PB ratio, ROE, cumulative stock return over the past five days and lagged information sentiment.

Table 3 shows the results. In Column (1), the coefficient on investor sentiment at *t-1*, $\text{InvSent}_{i,t-1}$, is significantly positive, suggesting a positive association between lagged investor belief and information sentiment. The information sentiment tends to be consistent with lagged investor sentiment. The results support H2.

5.3 Sentiment: Biased or "Value-relevant"?

An issue faced by empirical studies is whether proxies for investor sentiment reflect the information that is (1) on fundamental values of stocks and (2) not currently incorporated into prices (Chen et al., 2014). If this is true, the association between investor sentiment and information sentiment is driven by the "value-relevant" information. Specifically, investor sentiment is correlated with information sentiment because both sentiments reflect "value-relevant" information, not because investor sentiment affects information sentiment as we suggested. As such, we conducted three robustness tests to verify that (some of) investor sentiment is noisy (i.e., not related to fundamental values of stocks) in our dataset. In the first two robustness tests, we follow previous literature, as described below, and test whether the association between investor sentiment and information sentiment has reversal effect in a longer horizon and whether this association is weaker using Monday samples. In the last robustness test, we test a joint

Table 3: Information Sentiment and Investor Sentiment

	(1)		(2)		(3)	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
InvSent (t-1)	0.0102***	13.9	0.0110***	17.68	0.0191***	12.1
InvSent (t-2)			-0.0006	-1.06		
InvSent (t-3)			0.0000	-0.07		
InvSent (t-4)			0.0006	0.95		
InvSent (t-5)			-0.0005	-0.84		
InvSent (t-6)			-0.0019***	- 3.39		
InvSent (t-7)			-0.0014***	-2.67		
CumRet	0.0609***	5.5	0.0675***	6.06	0.0720***	3.56
Size	0.0544***	19.46	0.0546***	19.36	0.0540***	17.75
PB	0.0000	1.28	0.0000	1.22	0.0000	1.44
ROE	0.0000	0.81	0.0000	0.77	0.0000	0.85
InfoSent (t-1)	0.2431***	82.06	0.243***	82.38	0.3445***	28.02
InfoSent (t-2)	0.0679***	19.05	0.0679***	19.01	0.1218***	19.49
InfoSent (t-3)	0.0800***	25.28	0.0801***	25.11	0.0928***	21.76
InfoSent (t-4)	0.0647***	20.97	0.0646***	20.85	0.0620***	15.24
InfoSent (t-5)	0.0641***	21.88	0.0642***	21.76	0.0541***	13.60
InfoSent (t-6)	0.0539***	20.53	0.0544***	20.52	0.0462***	12.47
InfoSent (t-7)	0.0665***	22.23	0.0672***	22.20	0.0738***	14.83
Cons	-1.2248***	-19.63	-1.2337***	-19.43	-1.2338***	-17.82
Num of Obs	1,074,393		1,074,393		205,896	
Adjusted R2	0.1728		0.1729		0.1796	

Notes: The table reports the results of OLS regressions of information sentiment on investor sentiment. Column (3) only includes the samples on Mondays. The dependent is information sentiment for stock i at day t. "InvSent" is investor sentiment; "InfoSent" is the information sentiment; t-1 to t-7 show the lagged periods; "Size" is defined as the natural logarithm of 1 plus the number of shares times the closing price; "PB" is the price-to-book ratio, estimated as market capitalization over company net assets; "ROE" is return on equity, calculated as net profit over the shareholders' equity; past stock return, "CumRet", is cumulative return over the past five days. The regressions incorporate daily fixed effect. The standard errors are clustered at stock level. *, **, and *** denote significance at the 0.10, 0.05, and 0.01 levels, respectively.

hypothesis of that information in the information channels is biased and that this bias results from that information channels trying to maximize their viewership.

Previous studies rejected that information only contains "value-relevant" information by showing reversal effect: Tetlock (2007) and García (2013) argued that if investor sentiment were noisy, one would predict a positive association between investor sentiment and stock return in a short horizon, and predict a reversal effect - that is, a negative association - in the long run. Following this logic, we regressed information sentiment on investor sentiment for a longer horizon. We hypothesized that our investor sentiment may positively predict news sentiment in the short term and negatively predict it in the long term. A simple example is that a company may later deny some rumors that were reported several days before. Column (2) in Table 3 shows a clear reversal effect from *t-2* to *t-7*, even though some of the coefficients are not significant.

Some other studies examined the association between investor sentiment on Sundays and stock returns on Mondays, to examine whether no or less value-relevant information is published during weekends (García, 2013; Siganos et al., 2014). If investor sentiment captures value-relevant information, which is less likely to arrive in weekends, no association or a weaker association would be observed between investor sentiment on Sundays and information sentiment on Mondays. Yet, our hypothesis predicts an unchanged or stronger association. Following this logic, we tested whether investor sentiment on Sunday affects information sentiment on Monday. Specifically, we repeated the regression in equation 6 with a sample from Mondays. Column (3) in Table 3 shows the results. The coefficient of investor sentiment at *t-1* is 0.0191, which is larger than 0.0102 in Column (1). The results are consistent to our prediction.

To further examine that information is biased and this bias is driven by the pursuit of viewership, we examine whether information bias varies across the information channels with different sources of revenue. If some information channels do not depend on the viewership to obtain revenue, we expect those channels have less severe information bias. Some information creators in our database-*China Securities Journal, Shanghai Securities News, Securities Times, Cninfo*, and their affiliated entities-are Designated Information Revealing Media (DIRM), as determined by the regulator, the China Securities Regulatory Commission. All listed companies and financial products traded in the stock market are obligated to publish announcements through these media. A large portion of those information creators' revenue comes from publishing these announcements. Considering that other information channels' revenue mostly comes from viewership, we expect that those DIRM have less motivation to cater to readers than other information creators.

We divided our sample into two categories according to whether the information creators are DIRM and then calculated information sentiment for both categories. Our hypothesis predicts that aggregated investor sentiment has a better predictive ability for information sentiment estimated by the non-DIRM information creators than that by the DIRM information creators. Table 4 presents the results.

The dependent variables in Columns (1) and (2) describe information sentiment estimated by the DIRM; and those in Columns (3) and (4) are estimated by the non-DIRM information creators. Following Huang et al. (2015), we compared the R-squares in those regressions to examine the predictive ability of investor sentiment. The R-squares in Columns (3) and (4) are much larger than those in Columns (1) and (2), suggesting that investor sentiment has a stronger ability to predict the information sentiment of DIRM information than that of non-DIRM information. The results are consistent with our predictions.

5.4 Information Bias and Stock Return

In this section, we examine the effect of information bias on stock return. Information in the stock market can be unbiased or biased. Both types of information would affect the stock market. Empirically, it is difficult to directly distinguish the impact of biased information from that of unbiased information: we can only observe the aggregated effect of news sentiment on stock return, which combines both the effect of unbiased information and that of biased information. However, the effects of unbiased and biased information are different. Following the logic of Tetlock (2007) and García (2013) as described in the previous section, we examined whether there is a reversal effect: a negative association between stock return and information attitude in a longer horizon following a positive association in a short horizon. If our information sentiment measure only incorporates unbiased information, we are not able to observe the reversal return. The regression is as follows:

$$\alpha_{i,t} = \alpha + \beta_1 L7(\text{InfoSent}_{i,t}) + \beta_2 \text{Control variables}_{i,t-1} + \epsilon_{i,t} \tag{7}$$

where α is the abnormal return for stock i at day t; $L7(\text{InfoSent}_{i,t})$ is the news sentiment

Table 4: Information Sentiment and Investor Sentiment

	(1)		(2)		(3)		(4)	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
InvSent (t-1)	0.00319***	9.11	0.0034***	11.40	0.0041***	6.61	0.0046***	9.21
InvSent (t-2)			-0.0001	-0.41			-0.0006	-1.26
InvSent (t-3)			-0.0006**	-2.02			0.0002	0.34
InvSent (t-4)			0.0004	1.26			0.0005	1.01
InvSent (t-5)			-0.0002	-0.85			-0.0019***	-4.48
InvSent (t-6)			-0.0008***	-2.69			-0.0002	-0.38
InvSent (t-7)			-0.0007***	-3.00			-0.0014***	-3.47
CumRet	0.0454***	10.17	0.0489***	11.37	0.0357***	3.97	0.0415***	4.49
Size	0.0119***	11.47	0.0120***	19.36	0.0315***	12.28	0.0316***	12.23
PB	-0.0000	-1.18	-0.0000	-1.21	-0.0000	- 0.48	-0.0000	- 0.52
ROE	-0.0000	-1.55	-0.0000	-1.52	-0.0000	-0.71	-0.0000	-0.68
InfoSent (t-1)	0.0653***	43.45	0.0652***	43.55	0.1192***	41.08	0.1190***	41.25
InfoSent (t-2)	0.0055***	3.89	0.0055***	3.9	0.0422***	12.96	0.0423***	13.00
InfoSent (t-3)	0.0128***	8.57	0.0129***	8.52	0.0464***	16.07	0.0464***	15.91
InfoSent (t-4)	0.0079***	6.38	0.0078***	6.31	0.0394***	14.18	0.0393***	14.07
InfoSent (t-5)	0.0090***	6.89	0.0090***	6.86	0.0374***	13.12	0.0375***	13.14
InfoSent (t-6)	0.0076***	7.25	0.0079***	7.34	0.0314***	14.45	0.0320***	14.54
InfoSent (t-7)	0.0072***	6.90	0.0076***	7.02	0.0357***	15.10	0.0365***	15.05
Cons	-0.3000***	-13.03	-0.3039***	-12.68	-0.7141***	-12.65	-0.7206***	-12.48
Num of Obs	1,050,959		1,050,959		1,050,959		1,050,959	
Adjusted R2	0.0572		0.0573		0.1179		0.1180	

Notes: The table report the results of OLS regressions of information sentiment at t on investor sentiment using different samples. The dependent variable of Columns (1) and (2) is InfoSent at t, estimated from information generators of Designated Information Revealing Media (DIRM). The dependent variable of Columns (3) and (4) is InfoSent at t, estimated by information from non-DIRM generator. t-1 to t-7 show the lagged periods. "InvSent" is investor sentiment; "InfoSent" is information sentiment; "Size" is defined as the natural logarithm of 1 plus the number of shares times the closing price; PB is the price-to-book ratio, estimated as market capitalization over company net assets; "ROE" is return on equity, calculated as net profit over the shareholders' equity; past stock return, "CumRet", is cumulative return over the past five days. The regressions incorporate daily fixed effect. The standard errors are clustered at stock level. *, **, and *** denote significance at the 0.10, 0.05, and 0.01 levels, respectively.

for the past seven days from *t-1* to *t-7*.

Column (1) in Table 5 shows our results. The results are consistent with our prediction. The coefficient of news sentiment, "InfoSent", at t-1 is positive and significant ($p < 0.001$). This coefficient captures effect from both unbiased and biased information. The coefficients for other lagged information attitudes from *t-2* to *t-7* are mostly negative (except for t-3). Those coefficients show the reversal effect. The sum of these coefficients is -0.0012. This number measures the effect resulting from biased information. We can conclude that a one-standard-deviation increase of information attitude, 0.3843, at *t-1* would lead to a 4.6-basis-point (0.0012×0.3849) increase in abnormal return at t, while the effect of unbiased information would be only 2.6 basis points. The results support H3a. Our results are consistent with Dougal et al. (2012), who showed that media coverage has a causal effect on aggregate market outcomes.

Essentially, our regression of column (1) is similar to the regressions reported in Tetlock (2007) and García (2013). Traditional behavioral finance literature, assuming information is unbiased as described in Section 2, also predicts our results in Column (1). To distinguish our results from the previous literature, we started with the difference in assumptions between the previous literature and our information-based explanation in this paper. The former assumes that information in the market is unbiased and arbitrageurs can estimate fundamental value. However, because of limits to arbitrage, arbitrageurs are not able to correct asset-pricing error. Based on our biased-information view, asset pricing error results from biased information. Both noise traders and arbitrageurs would make suboptimal decisions under the erroneous information set. Table 1 summarizes the extent of asset-pricing-error correction given limits versus no limits to arbitrage and given biased versus unbiased information. Previous studies argued that asset pricing error emerges when information is unbiased and there are limits to arbitrage. Yet, our information-based view predicts that stock price is always biased given biased information with or without the presence of limits to arbitrage. Empirically, we examine whether the effect of news sentiment on stock return varies with difficulty levels of arbitrage, as we argued in H3b.

Our database provides a valuable opportunity to test H3b. Not all stocks in the Chinese stock market can be arbitraged. The regulator decides whether one stock is permitted to be arbitraged depending on stock transaction and market conditions. The previous literature predicts that the effect of information sentiment on stock return will be different between the stocks that are in the arbitrage list and those that are not in the list. Our hypothesis predicts this difference

Table 5: Information Sentiment and Stock Return

	(1)		(2)	
	Coef.	t-stat	Coef.	t-stat
InfoSent (t-1)	0.0019***	17.11	0.0025***	15.12
InfoSent (t-2)	-0.0003***	-3.69	-0.0003**	-2.61
InfoSent (t-3)	0.0001	0.74	0.0001	0.65
InfoSent (t-4)	-0.0003***	-3.45	-0.0004***	-2.83
InfoSent (t-5)	-0.0001	-1.47	-0.0001	-1.15
InfoSent (t-6)	-0.0002*	-1.82	-0.0001	-0.59
ArbInfoSent (t-1)			-0.0016***	-7.30
ArbInfoSent (t-2)			0.0001	0.55
ArbInfoSent (t-3)			-0.0000	-0.04
ArbInfoSent (t-4)			0.0001	0.78
ArbInfoSent (t-5)			0.0000	0.29
ArbInfoSent (t-6)			-0.0002	-1.19
ArbInfoSent (t-7)			0.0003	1.11
Size	-0.0001	-0.98	-0.0001	0.35
PB	0.0000	0.02	0.0000	0.01
ROE	0.0000	0.68	0.0000	0.67
Cons	0.0034	1.19	0.0021	0.48
Num of Obs	1,087,180		1,087,180	
Adjusted R2	0.0123		0.0124	

Notes: The table reports the results of regressions of stock i abnormal return at t on information sentiment. The dependent variables of Columns (1) and (2) are stock abnormal return, α, at t. "DumArb" is the dummy variable that equals to 1 if the stock, i, is permitted to be short sold at day t, and equals 0 otherwise; "ArbInfoSent" is the interaction term between "DumArb" and "InfoSent"; "InfoSent" is the proxy for news sentiment; "Size" is defined as the natural logarithm of 1 plus the number of shares times the closing price; "PB" is the price-to-book ratio, estimated as market capitalization over company net assets; "ROE" is return on equity, calculated as net profit over the shareholders' equity. t-1 to t-7 show the lagged periods. The regressions incorporate daily fixed effect. The standard errors are clustered at stock level. *, **, and *** denote significance at the 0.10, 0.05, and 0.01 levels, respectively.

will be marginal. We regressed stock abnormal return on lagged information attitude, dummy variable for being in arbitrage list and their interaction terms; that is:

$$\alpha_{i,t} = \alpha + \beta_1 L7(\text{InfoSent}_{i,t}) + \beta_2 L7(\text{InfoSent}_{i,t} \times \text{DumArb}_{i,t}) \\ + \beta_3 \text{DumArb}_{i,t} + \beta_4 \text{Control variables}_{i,t-1} + \epsilon_{i,t} \quad (8)$$

where "DumArb$_{i,t}$" is the dummy variable that equals 1 if the stock i, is permitted to be short sold at day t, and equals 0 otherwise.

Column (2) in the Table 5 shows the results. All coefficients of the interaction terms are insignificant except for that for "ArbInfoSent" (the interaction term between "DumArb" and "InfoSent") at *t-1*. The result is consistent with the prediction from our information-based view but inconsistent with that from traditional financial literature. Therefore, we find evidence supporting H3b.

6 Discussion and Implications

To our knowledge, our paper is the first attempt to explain asset pricing errors (i.e., market inefficiency) from the perspective of information bias. However, we do not argue that our explanation can replace the classic explanation of market inefficiency; rather, we suggest that ours can complement the classic explanation. Based on prior studies, we detected information bias at the aggregate level in the stock market: due to the pursuit of viewership, information channels provide biased information systematically to cater to investors' confirmation bias. At the same time, it is not economically efficient for arbitrageurs to find the unbiased information. They may conveniently use the biased information to make financial decisions. Thus, investors' decisions (both noise traders and arbitrageurs) are inevitably tainted due to the biased information input, and consequently, asset pricing errors emerge.

This paper highlights the importance of information transfer in the stock market, an understudied market in the domain of information systems. The prior information system literature focused on markets like the consumer market (Duan et al., 2008). The information transfer in the stock market is different than that in other markets. For example, the stock market is more time sensitive to information than other markets (e.g., the consumer market). In the stock market, information that is valuable at one second can be worthless at the next. Moreover, none of the participants (either sellers or buyers) in the stock market can have perfect information. It is difficult for them to discover the fundamental price of an asset (Shiller, 1984). Therefore, the efficiency of collecting information is crucial for investors to ensure a better investment decision.

Our results have practical implications for investor education. Current investor education emphasizes an understanding of finance market products and risk management. Our research highlights that it is also important for investors to understand how to apply their critical thinking to differentiate stock-market-related information based on usefulness.

7 Limitations and Outlook

As a pioneer in explaining market inefficiency from the information-based view, our paper has several limitations. First, the paper does not empirically differentiate individual investors from institutional investors. Typically, individual investors have less capacity to obtain and analyze unbiased information than institutional investors. As such, in the future, we aim to examine whether these two categories of investors respond to biased information differently.

Second, following Mullainathan and Shleifer (2005) and Gentzkow and Shapiro (2010), we assume that the information bias in the stock market comes from the demand side, i.e., information channels pursuing their readers. However, information bias may also come from the supply side (e.g., information channels may reflect the preferences of owners). Specifically, even though biased information in information channels results from viewership as in H2, it is possible that this identified biased information is not the only type of information bias to affect stock returns, as suggested by H3a and H3b. However, as far as we are concerned, the prior literature did not identify any empirical or theoretical evidence supporting other types of information bias regarding stock. For example, Gentzkow and Shapiro (2010) showed no or very little evidence supporting that information bias comes from protecting the interests of owners of information channels. This discussion calls for studies to identify other types of information bias and to explore how these different sources of information bias interact with each other to influence the stock market.

Third, we tested our hypotheses using data from the Chinese stock market. One reason for this choice of market is that individual investors are dominant in the Chinese stock market. In contrast, most investors in the United States (US) market are institutional investors, who are more likely to be sophisticated. As such, investors in the Chinese market are, on average, less sophisticated than those in the US market. Our research calls for studies that aim to replicate our results in different stock markets and, more importantly, examine the extent to which information bias exists in, and how different the impact of such bias is on, these stock markets.

References

Antweiler, W. and Z. M. Frank (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *Journal of Finance 59*(3), 1259–1294.

Asch, S. E. (1955). Opinions and Social Pressure. *Readings about the social animal 193*(5), 31–35.

Baker, M. and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance 61*(4), 1645–1680.

Baker, M. and J. Wurgler (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives, American Economic Association 21*(2), 129–152.

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

Barberis, N. and R. Thaler (2003). A Survey of Behavioral Finance. In *Handbook of the Economics of Finance*, pp. 1052—-1128. Elsevier.

Black, F. (1986). Noise. *Journal of Finance XLI*(3), 529–543.

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

Brown, G. W. and M. T. Cliff (2005). Investor Sentiment and Asset Valuation. *The Journal of Business 78*(2), 405–440.

Carhart, M. M. (1997, mar). On Persistence in Mutual Fund Performance. *The Journal of Finance 52*(1), 57–82.

Chen, H., P. De, Y. Hu, and B. H. Hwang (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies 27*(5), 1367–1403.

Da, Z., J. Engelberg, and P. Gao (2015). The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies 28*(1), 1–32.

De Long, B., A. Shleifer, L. Summers, and R. Waldmann (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy 98*(4), 703–738.

DellaVigna, S. and J. M. Pollet (2007). Demographics and industry returns. *American Economic Review 97*(5), 1667–1702.

Dellavigna, S. and J. M. Pollet (2009). Investor inattention and friday earnings announcements. *Journal of Finance 64*(2), 709–749.

Dougal, C., J. Engelberg, D. García, and C. A. Parsons (2012). Journalists and the stock market. *Review of Financial Studies 25*(3), 640–679.

Duan, W., B. Gu, and A. B. Whinston (2008). Do online reviews matter? An empirical investigation of panel data. *Decision Support Systems 45*(4), 1007–1016.

Engelberg, J. E. and C. a. Parsons (2011). The causal impact of media in financial markets. *The Journal of Finance 66*(1), 67–97.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fang, L. and J. Peress (2009). Media Coverage and the Cross-section of Stock Returns. *The Journal of Finance LXIV*(5), 2023–2052.

Fang, L. H., J. Peress, and L. Zheng (2014, dec). Does media coverage of stocks affect mutual funds' trading and performance?

Forrester, J. W. (1992). Policies, decisions and information sources for modeling. *European Journal of Operational Research 59*(1), 42–63.

French, K. R. (2008). The Cost of Active Investing. *Journal of Finance 63*(4), 1537–1573.

García, D. (2013). Sentiment during Recessions. *Journal of Finance 68*(3), 1267–1300.

Gentzkow, M. and J. M. Shapiro (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica 78*(1), 35–71.

Harwit, E. (2014, dec). The Rise and Influence of Weibo (Microblogs) in China. *Asian Survey 54*(6), 1059–1087.

Hirshleifer, D. and S. H. Teoh (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics 36*(1), 337–386.

Huang, D., F. Jiang, J. Tu, and G. Zhou (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies 28*(3), 791–837.

Keynes, J. M. (1936, feb). *The General Theory of Employment, Interest, and Money.* London: Macmillan.

Luo, X., J. Zhang, and W. Duan (2013). Social media and firm equity value. *Information Systems Research 24*(1), 146–163.

Muchnik, L., S. Aral, and S. J. Taylor (2013). Social Influence Bias: A Randomized Experiment. *Science 341*(6146).

Mullainathan, S. and A. Shleifer (2005, aug). The Market for News. *American Economic Review 95*(4), 1031–1053.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology 2*(2), 175–220.

Park, J., P. Konana, B. Gu, A. Kumar, and R. Raghunathan (2013). Information valuation and confirmation bias in virtual communities: Evidence from stock message boards. *Information Systems Research 24*(4), 1050–1067.

Seward, Z. M. (2013). This is how much a Bloomberg terminal costs (May 15), Accessed November 21, 2016.

Shiller, R. J. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity 2*(1), 457–498.

Shleifer, A. and L. Summers (1990). The Noise Trader Approach to Finance. *Journal of Economic Perspectives 4*(2), 19–33.

Shleifer, A. and R. W. Vishny (1997, mar). The Limits of Arbitrage. *The Journal of Finance 52*(1), 35–55.

Siganos, A., E. Vagenas-Nanos, and P. Verwijmeren (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior and Organization 107*, 730–743.

Sprenger, T. O., A. Tumasjan, P. G. Sandner, and I. M. Welpe (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management 20*(5), 926–957.

Stambaugh, R. F., J. Yu, and Y. Yuan (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics 104*(2), 288–302.

Stambaugh, R. F., J. Yu, and Y. Yuan (2014). The long of it: Odds that investor sentiment spuriously predicts anomaly returns. *Journal of Financial Economics 114*(3), 613–619.

Tetlock, P. C. (2007). Giving Content to Investor Sentiment : The Role of Media in the Stock Market. *Journal of Finance 62*(3), 1139–1168.

Tetlock, P. C. (2010). Does public financial news resolve asymmetric information? *Review of Financial Studies 23*(9), 3520–3557.

Tetlock, P. C. (2014). Information Transmission in Finance. *Annual Review of Financial Economics 6*(1), 365–384.

World Federation of Exchanges (2015). 2015 Market highlights. Technical report, World Federation of Exchanges.