Online Sentiment Contagion in China

Xu FENG[a], Xue-zhong HE[b], Shen LIN[a,1], Jianxin WANG[b]

[a] China Center for Social Computing & Analytics
College of Management & Economics
Tianjin University
Tianjin 300072, China

[b] Finance Discipline Group
Business School
University of Technology, Sydney
NSW 2007, Australia

First Draft: August 11, 2015

Abstract

This study provides evidence that the internet is an important channel through which sentiment contagion in the Chinese stock markets takes place. We use the number of clicks of online messages as a proxy of sentiment contagion, and use two filters to remove messages containing hard news or information. We find that the number of clicks has predictability on stock returns, order types, order imbalance of individual investors, and the total trading volume. In addition it is found that sentiment-based portfolios in China can generate excess returns.

Keywords: Sentiment Contagion; Online message; Natural language processing; Predictability; Chinese stock market

[1] Corresponding Email: linshen@tju.edu.cn

1. Introduction

Whether sentiment of one investor affects others is debated for many years. In rational models, agents are fully rational and information is conveyed through price or price function which are observed by all agents (Grossman and Stiglitz [1980], Kyle [1985]). Sentiment cannot affect others and has impact on stock price because of rational and arbitraged investors. However, behavioral finance studies suggested that optimistic or pessimistic mood brought by investors can spread and infect others in stock market, so called "sentiment contagion" (Shiller [1984]; Musumeci and Sinkey [1990]; Hirshleifer and Teoh[2003]; Collins and Gavron [2004]). Under the assumption of irrational agents, investors imitate others' sentiments and make decisions, and drive stock price deviating from its fundamental value. Therefore, bubbles and crashes emerged in stock market (Lux[1995]; Shiller[2000]).

There is indeed evidence that sentiments are contagious and affect individuals' behavior in social and psychological studies (Neumann and Strack [2000]; Barsade and Sigal [2002]; Francesco and Semin [2009]). In financial market, two ways are considered as how sentiments are contagious in stock markets. The first one is through capital flows. For example, investors in one country might be optimistic or pessimistic, which leads the capital flows shifting into risky assets more broadly, including international equities. Sentiment of one country will affect prices in another country (Baker, Wurgler and Yuan [2012]). The second one is through social influence. Investors communicate with others through different social interaction channels. If some of the investors is optimistic or pessimistic, others may converge to a similar sentiment because of herding behavior or peer group effects (Shiller [1984]; Glaeser et al.[1996]).

Although social influence was seen as an important channel of sentiment contagion, it is hard to show explicit evidence that sentiment is contagious through social interaction in stock market because of the absence of the measurement of investors' sentiment affected by others. In this paper, we measure sentiment contagion by using data from online stock message board. Investors express their optimistic or pessimistic opinions by posting online messages. If another investor read it, there is an interaction between two investors and the reader might be affected by the message. Actually, it is impossible to know how many investors be influenced by a message exactly. But we use the click number of a message as a proxy of one message deliver sentiment to other investors. The idea is intuitively: if an investor clicks a message and read it, he is probably interested in the content and influenced by the message. If the sentiment is contagious through interaction, the more the message is clicked, the more the investors might be affected.

There has rapidly growing literatures (Tumarkin and Whitelaw [2001]; Antweiler and Frank [2004]; Das and Chen[2007]; Kim and Kim [2014]; Chen et al. [2014]) used online messages data and natural language process method to measure investor sentiment. Most of these studies, e.g. Tumarkin and Whitelaw (2001), Antweiler and Frank (2004) and Das and Chen (2007), find that social media outlets cannot predict stock presumably. Tumarkin and Whitelaw (2001) detect no association; Antweiler and Frank (2004) find a statistically significant, yet economically meaningless, association; Das and Chen (2007) find "no strong relationship from sentiment to stock prices on average across the individual stocks"; and Kim and Kim (2014) find online message have no predictability both on return, volatility and volume. On contract, Chen et al. (2014) observed investors' messages transmitted through social media predict future stock returns and earnings surprises. Chen et al. (2014) get online message from the website "Seeking Alpha", which has an incentive mechanism for investors releasing useful information and accumulating reputation.

Our study draw a conclusion consist with Chen et al. (2014) that we find online massages has predictability on stock returns. Three main differences distinguish our study from those above:

First of all, our study make a difference for "sentiment" and "information". If a message is about fundamental of the company, the spread of the message is more likely an information diffusion process rather than sentiment contagion. Although lots of studies above also find that online messages can predict stock returns or trading volume, it is hard to say the predictability of online messages is because of information or sentiment. In this paper, we employ two filters to remove the message might contain macro / industry news and firm specific news. Besides, we give more robustness test to see if the data sample we used is a proxy of information or sentiment.

Secondly, previous studies use number of posts as a measurement of online sentiment. Our study is more focus on the question whether sentiment contagion take place. We use number of clicks as a proxy of sentiment contagion because it captures how many investors might be affected by others, which cannot be captured by the number of posts. We built sentiment contagion index as the number of clicks divided number of posts, which can better describe the contagion process and avoid the situation that a large click number caused by lots of investors post messages but only a few of them be viewed. If one post has a large click number, which indicated the content of the post is more attractive for investors. Therefore, the sentiment contagion is likely to happen. Additionally, our results also suggested that the number of clicks has more predictability than number of posts on stock returns.

Last but not least, our sample comes from Chinese stock market, where individual traders contribute 80% of daily trading volume. If sentiment contagion takes place, it will have more impact on individual investors rather than on institutions. Chinese internet message board is more active than other countries. Our message sample contains more than 10 million messages during Jan 2011 to July 2014 for 290 stocks. Sentiment contagion might have more predictability in Chinese stock market than other developed markets.

Our paper answered several questions related with sentiment contagion. First of all, this paper tests whether sentiment contagion through internet happened. We find number of clicks of online messages can predicts stock return, which support the existence of the online sentiment contagion. Moreover, the impacts of positive and negative sentiment contagion on stock return are different. Positive sentiment contagion caused a higher return in short term, and a lower return in the future, while the negative sentiment contagion caused an opposite impact. These results are consistent with the noise trader model developed by De Long et al. (1990a) and herding behavior model developed by Lux (1995).

The second question answered in this paper is: How does sentiment contagion affect stock trading? Previous studies related investors sentiment with stock price (Lee, Jiang, and Indro [2002]; Brown and Cliff [2004]; Baker and Wurgler [2006]), while few of them linked sentiment with investors' trading behavior. We tested whether sentiment contagion can predict directions of trades. We find positive (negative) sentiment contagion generates buying (selling) initiated trades. The order imbalance can also be explained by the aggregate sentiment contagion. These results provide more evidence that investors affect by others' sentiment.

This study also finds that different sentiment contagion is associated with higher trading volume. We suppose different sentiment contagion generate different opinions. Antweiler and Frank (2004) use online posting data to find different sentiment positively predict trading volume, while Tetlock (2007) find a negative relationship between sentiment and trading volume. We test the impact of sentiment contagion on the volume. The result shows that sentiment contagion causes the disagreement opinion among the investors and increases trading volume. This result supports both the theory of different opinion hypothesis and noise trader hypothesis, and inconsistent with the results in Tetlock (2007).

It is important to understand what drives sentiment contagion process. Feedback trading theory suggests irrational traders tend to buy securities when prices rise and sell when prices fall. Therefore, market status might accelerate sentiment contagion process. We further use stock return and trading volume to predict sentiment contagion and find that these variables can significantly predict sentiment contagion, which consistent with Feedback trading theory that historical trading and market status play a role in the process of sentiment contagion.

Finally, this study tests whether we can get profits from sentiment contagion. A portfolio is built based on abnormal sentiment contagion. Without consideration of transaction costs, the portfolio could gain excess return by holding the portfolio with maximum abnormal relative sentiment contagion and selling out portfolio with minimum abnormal relative sentiment contagion.

This study contributes to the investor sentiment literatures in two ways. It provides evidence to demonstrate that investors' sentiment can affect others through social interaction. The number of clicks as the proxy of contagion and the filters of sentiment posts help us to better estimate the number of investors that affected by the sentiment of posters. Moreover, this study provides new insights on the relationship between sentiment contagion and market order, trading volume, and historical price and volume. Through these work, it helps to describe the mechanism how does sentiment contagion impact the stock price and trading volume and how does sentiment contagion occurred.

The rest of the paper is organized as follows. Section 2 discusses the relative theories and our basic hypothesis. Section 3 describes the internet message board and stock market data. The details of sentiment extraction and sentiment contagion index building are also given in section 3. Section 4 shows the examination of online sentiment contagion. Several robustness tests are given in section 5. We conclude in Section 6.

2. Theories and Hypotheses

2.1 Online sentiment contagion and stock return

Noise trader theory (De Long et al.[1990a]) assumes that two kinds of investors in stock market: noise traders and arbitrageurs. When noise investors is optimistic or pessimistic, they will buy or sell risky assets to arbitrageurs and caused a temporary pressure on the price of risky assets. Therefore, sentiment contagion should predict the stock return in short time if it really happened. Herding model (Lux [1995]) also suggested that investor follows their neighbors or the sentiments of the market. A buy or sell pressure will appears on stock price and followed a reversal because of the non-sustainable of bubbles.

It is important to distinguish “sentiment” with “information”. In this study, we define online sentiment as the optimistic or pessimistic mood delivered by the online message which does not contain fundamental information of the listed companies. If online messages contain fundamental information, information diffusion process lead a pressure on stock return, and also follow a reversal because of overreaction trading by investors (Hong and Stein [1999]).

The information diffusion process and sentiment contagion process appears similar results on stock return. One way to identify the sentiment contagion process is whether the fully reversal appears. For information diffusion process, because of the fundamental of the companies changed, the impact of information on stock return will persistent indefinitely. Reversal caused by overreaction trading will not lead the price back to the status before information impact. Therefore, we can make hypothesis as follows:

Hypothesis 1: The online sentiment cognation predicts stock return in short term and following a fully reversal in the long-term.

If we find the online sentiment cognation predict stock return followed by a part reversal or no reversal, which means the message data we used contain information. Another impossible is that that the online contagion did not happen and the click number is just noise. In this case, our proxy should have no predictability on stock return. Our hypothesis 1 should be rejected.

2.2 Online Sentiment Contagion and trading behavior

If online sentiment contagion reflected the irrational investors’ demand and cause prices to deviate from underlying fundamentals, it should also has a link with trading behavior. Studies on individual investors’ behavior (Kumar and Lee [2006]; Barber, Odean and Zhu [2009]) tried to use individual trading to infer sentiment. Kumar and Lee (2006) find that the individual trades are systematically correlated and explains return co-movements for stocks which is costly to arbitrage. The results of Kumar and Lee (2006) are consistent with noise trader models and support a role for investor sentiment in the formation of returns. Barber, Odean and Zhu (2009) find individual trades appears herding behavior and the order imbalance of small trade can forecasts future returns, which is also consistent with noise trader theory. We give a directly test on the relationship of investor sentiment contagion and investors' trading behavior. Noise trader model suggests sentiment should induce investors' trading activity. We make hypothesis as follows:

Hypothesis 2: Positive sentiment cognation predicts buy initiate trading, and vice versa. Moreover, aggregate sentiment contagion predicts order imbalance.

2.3 Online Sentiment Contagion and trading volume

Online Sentiment Contagion might also be related with trading volume. According to noise trader theory (De Long et al.[1990a]), unusually high or low values of sentiment will generate high trading volume. Another possible relationship between sentiment contagion and trading volume is through investors' disagreement. Disagreement hypothesis (Hirshleifer [1977], Diamond and Verrecchia [1981], Hong and Stein [2007]) suggested that trading volume is induced by the opposing opinions on the future price held by investors in the market. Antweiler and Frank (2004) test disagreement hypothesis and make conclusion consistent with it. We test both noise trader theory and disagreement hypothesis to find in which way online sentiment contagion affect trading volume. Our hypothesis is:

Hypothesis 3a: The positive / negative sentiment contagion predicts the trading volume.

Hypothesis 3b: The disagreement caused by different sentiment contagion predicts the trading volume.

2.4 What drives online sentiment contagion?

Some studies on social influence established on the reputation mechanism (mutual trust and familiarity of individuals), like neighborhood (Hong, Kubik and Stein [2004]; Brown et al. [2008]), working in the same industry (Hong, Kubik and Stein [2005]), or sources have incentive to release high quality information (Chen et al. [2014]). Our data is very different with these studies because the posters in our sample are anonymous. Feedback trading model (De Long [1990b]) might be an alternative explanation if online sentiment contagion exists in our sample. Feedback trading studies (Kurov [2008]; Chau, Deesomsak and Lau [2011]) find that Positive feedback trading appears to be more active in periods of high investor sentiment. This evidence implies that market status might have impact on sentiment contagion process, which is verified by Tetlock (2007) with the data from Wall Street Journal. With feedback trading theory, we make hypothesis as follows:

Hypothesis 4: Pervious stock price and trading behavior predicts investors' sentiment contagion process.

2.5 Portfolios based on online sentiment contagion

We are also trying to build portfolios to check whether we can get profits using online sentiment contagion. We build the Abnormal Sentiment Index (ASI) based on the abnormal online sentiment contagion value of each stock compare with past time, and then divide stocks into 5 portfolios according to their ASI. If the sentiment contagion exits and can predict the future stock price, then stocks with abnormal positive sentiment contagion should face a higher up-ward pressure in the short term. Therefore we can profit from buying portfolio with high ASI and selling out the portfolio with low ASI. We make hypothesis as follows:

Hypothesis 5: Buying portfolio with higher ASI and selling portfolio with lower ASI can bring a significant excess profit.

3. Data and Method

3.1 Data description

Our data come from Chinese stock market where individual trading account for about 80% of the total trading volume (Ng and Wu [2007]). The inexperienced individual investor is more likely than the professional to be subject to sentiment (Baker and Wurgler [2007]). We choose 300 stocks of the constituent stocks of China Securities Index 300 (CSI 300) in January 1st 2011 as our sample stocks. The CSI 300 is selected as our sample due to its representativeness of Chinese stock market. The constituent stocks of CSI 300 make up 60% market capitalization of all the listed stock in two Chinese stock exchanges, Shanghai Stock Exchange and Shenzhen Stock Exchange. Because of the selection criteria of CSI 300, our sample is unlikely affected by small size effect and liquidity effect. The sample period is from 1 January 2011 to 1 July 2014. 10 stocks are excluded because of delisting or a longtime trading halt. Stock data including daily price, daily return, trading volume and book-to-market is downloaded from Wind Database. Our order data come from historical tick trading in Sina website, which record trading volume and value of each trade, also including a tag to specify the trading is buy initiated or sell initiated.

For online data, message data of 290 stocks come from online message board: Eastmoney (guba.eastmoney.com). Eastmoney is the largest and most active stock message board in China. For better estimate the daily sentiment contagion, we delete the messages which have posting day and last reply day are not the same day. Total of 10,137,691 messages of the 290 stocks are posted in the Eastmoney. We download these posts by using self-writing java program. The data contain the message title, content, the date of posting (accurate to days), the account name and the number of clicks on each posts. Table 1 shows a sample for the stock Vanke (000001) on 2014/07/23. In this sample, we can find the title, which means ‘VK appears peak signal again!!’, date of posting (July 23st 2014), the account name of poster (an anonymous investor from Zhejiang Province) and the click number (207) is given in the first line.

[INSERT TABLE 1 HERE]

The summary statistics are given in table 2. Colum 1 to column 6 show statistics of the individual stock return, market return, capitalization of each stock, book to market of each stock, daily message posts and daily clicks of each stock. As shown in Table 2, the market from January 2011 to July 2014 is in a range-bound stage and the mean of individual stock return and the market return is closely to 0. All the selected stocks are sizeable and the biggest one is Industrial and Commercial Bank of China (ICBC) which capitalization is 1.48 trillion. The Eastmoney stock message board is also active. Average daily posts are 12.33 thousands and Average daily clicks are 12.65 million for 290 stocks, which means one post is viewed by 1026 other investors on average. During the most active period, the posts is 2.09 times than normal times and the clicks also reached 2.51 times.

[INSERT TABLE 2 HERE]

3.2 Sentiment Classification and filter

Because of the large number of online messages in our sample, we cannot classify messages into different kinds of sentiment manually. We employ the Natural Language Process method to classify the messages into three type: positive, neutral, negative.

Chinese sentiment classification is consists of two steps: word segmentation and sentiment classification. Word segmentation is usually unnecessary as English sentiment classification because the words in English sentence are separated naturally.

However, Chinese sentences are composed by Chinese characters. Characters usually have meaning itself such as '好' means good. But generally, one word composed by 2 to 4 characters, and the meaning of characters is not always same as the word. For example, the word '多头' means long side of the market, neither the character '多' nor '头' has the same meaning. Therefore, we must divide the sentences into the words appropriately and the meaning must remain the same. We employ the 'FudanNLP-1.6.1' software as our word segmentation instrument, which is also widely used in other studies of natural language processing for Chinese text (Li, Wang and Yan [2015]).

A key factor for word segmentation and sentiment classification is the sentiment dictionary. The dictionary included in 'FudanNLP-1.6.1' contains a large number of simple and basic words. However, for our study, the dictionary we need must contain more professional word in finance and stock market. Therefore, we build a dictionary special for Chinese online stock message board and classify the messages by following steps:

I. Choosing the training sample. We randomly choose 5000 messages which contain more than 5 Chinese characters from our online messages sample. The 5 characters limitation is added because the short messages lack of sentiment content and have no value for training.

II. Dictionary building and word segmentation of training sample. We segment our 5000 training sentences with the dictionary included the default dictionary of FudanNLP, HowNet Chinese sentiment dictionary, 219 terminology of stock market in MBAlib, and all stock names of Chinese stock market. HowNet is a frequently-used dictionary in the study of Chinese sentiment classification (Dong and Dong 2003) and MBAlib which is the biggest encyclopedia website on economics and management in China.

III. Manual classification. The Training sample is classified manually by 10 master students who have majored in finance and have experience in stock trading.

We ask the students to classify the messages into 3 different types: positive, neutral and negative and choose the key words which support their judgments. For example, a sentence “中行的跳水原因 (the reason of the BOC price collapses)” is divide into words as “中行 (BOC), 的 (of), 跳水 (price collapses) and 原因(reason)”. The word “跳水” (price collapses) is chosen by the student as the key word which make the sentence sentiment negative.

For excluding the manmade errors, each message is classified by 3 different students. When a message is classified into positive and negative at the same time, we remove it from our sample. 10 messages are removed and most of our sample can be classified into the same type. Especially, when a message is classified in to positive/neutral or negative/neutral, we choose the majority one. If the result of all three students is identical, the message will be regarded as the type directly.

IV. Naive Bayesian Classification. We collect the sentiment key words identified by the students and remove the meaningless words. Finally, 1043 words are considered as the Key Sentiment Words of Chinese stock message board. We employ Naive Bayesian Classification (NBC) for sentiment classification.

The NBC has an assumption that the occurrences of words are independent of each other. The conditional probability of one message contains the keyword W_I which is included in Key Sentiment Words belongs to the sentiment group T_c, $C \in \{Positive, Neutral, Negative\}$, is:

$$P(T_c|W_I) = \frac{P(W_I|T_c)P(T_c)}{P(W_I)} = \frac{P(T_c)\prod_{k=1}^{I} P(w_k|T_c)}{\prod_{k=1}^{I} P(w_k)} \quad (3.1)$$

Where w_k is a word from the sequence W_I. I is the total number of W_I. Based on equation 3.1, we can calculate the sentiment probability of each message and we choose the one with the maximum probability as its sentiment type.

$$Type(W_I) = Max\{P(T_C|W_I)\}, C \in \{Positive, Neutral, Negative\} \quad (3.2)$$

[INSERT TABLE 3-A HERE]

V. In-sample and Out-sample results. We train 5000 training samples with NBC and get the in-sample accuracy. In Antweiler and Frank (2004), the same NBC has been used to classify the English messages and their in-sample accuracy is 88.1% with 1000 manual samples. Our result is given in Table 3-A. The in-sample accuracy is 85.4%, which is a little lower than the accuracy in Antweiler and Frank (2004). The reason probably is that the tone and sarcasm contained in Chinese is difficult to be classified. Moreover, we randomly choose and manually classify another 1000 messages for out-sample test. Table 3-B shows the accuracy of out-sample classification. The accuracy is declining from 85.4% to 77.9%, which is also keep in a high level. We cannot compare our out-sample accuracy with Antweiler and Frank (2004) because they do not report it. However, our accuracy is higher than other studies with English classification (e.g. Das and Chen [2007]; Kim and Kim [2014]). Most importantly, the situation of messages has been classified to opposite sentiment (the positive messages is classified as negative, and vice versa) keep a low percentage (0.4% and 0.2%). This result ensures our classification does not have systematic error.

[INSERT TABLE 3-B HERE]

Moreover, we employ two filters to remove the information related with fundamental of the companies. We assume the fundamental information of one company consistent of three parts: macro / market information, industry information and firm specific information. Public news and informed trader might be two ways of these information transmitted. The first filter we used is for removing public news. We got public news and the dates came from the Securities Times (www.stcn.com), which is a financial newspaper sponsored by People's Daily (the largest newspaper in China) and appointed by the Chinese Securities Regulatory Commission as one of the official media for listed companies' information disclosing. We get macro / market news from the column of "Important News" and "Oversea News", get industry the column "Industry and Economy" and get firm specific news and announcement from the column "Company News" of Securities Times.

We compare the tittle of the public news and tittle of online message 1 day before and after the date that the news published. For example, a piece of news is published on 15 August, We selected the online messages in the period of 10 August to 20 August. Then we segment the tittles of public news and online message, and filter online message if the tittle of the message has at least two words (we restrict words only in noun, verb and numeral and composed by at least two characters) overlapped with the tittle of public news.

The second filter we used is the account name of posters. Eastmoney allow investors to post messages using account name or anonymous. If the poster is anonymous, the account name only show where the poster come from according to her IP address. The posters with an account name are more easy to be recognized. More importantly, they can accumulate reputation through historical messages with account name. Therefore, account name users have incentive to release private information to get social norms and influences. We tried our best to keep our sample consistent of sentiment rather than fundamental information, and removed all the message posted by users have account name 1 day before and after the date that the firm specific news published, and only kept messages posted by anonymous during this period.

[INSERT TABLE 3-C HERE]

Table 3-C provided statistics of the messages deleted by two filters. The deleting messages happened in 107 of 845 trading days. Del is the number of posts/clicks which had been deleted of all stocks. SumofNewsday is the total number of posts/clicks during the 107 trading days of all stocks. Mean, Max and Min are reported as the average statistics of each stock. Name is the number of posts/clicks posted by users who had account name of all stocks. In Table 3-C, We can find that the deleted posts take a small proportion of total posts (7.74%), The number of clicks take more proportion than posts means that the message contain fundamental information attract more readers. It is also noticed that the deleted posts take a small proportion of total posts during the days of news happened. The proportion is 40.52%

for deleted posts and 62.15% for deleted clicks, which implied that online message board also play an important role on information diffusion. For account name, we find that about 1/3 users post messages with account name and attract more readers than anonymous users (51.45%). The deleted posts with account name take a large part of the total deleted messages.

3.3 Variables and Controls

After the sentiment classifying and filtering, we assumed that the contagion of sentiment c on a certain stock k in day t is defined as the sentiment contagion index $\mathrm{SCI}^{c}_{k,t}$ as follows:

$$SCI^{c}_{k,t} = log\,(\frac{1+\sum_{i=1}^{N^{C}_{k,t}} click^{c}_{i,k,t}}{1+N^{C}_{k,t}})\ ,\ C \in \{Positive, Neutral, Negative\} \tag{3.3}$$

Where $\mathrm{click}^{c}_{i,k,t}$ is the number of clicks of message i. and $\mathrm{N}^{C}_{k,t}$ is the total number of messages of sentiment C. In 3.3, we measure the contagion of sentiment C in stock k at day t as the total clicks of all messages which is classified as group of sentiment C. It might also get a large SCI in the case that lots of investors post messages but only a few of them be clicked. This case might happen because of investors have certain sentiment affect by other factors, such as newspapers or television, rather than online social interaction. We use the total clicks divide total number of posts in sentiment group C to ensure SCI only measure the sentiment contagion through online social interaction channel.

We also define the relative sentiment contagion index (RSCI) and opposite sentiment contagion index (OSCI) of stock k at day t. RSCI describe the relative difference between the positive sentiment contagion and negative sentiment contagion. The definition of RSCI is:

$$RSCI_{k,t} = \frac{SCI^{positive}_{k,t} - SCI^{negative}_{k,t}}{1+SCI^{positive}_{k,t} + SCI^{negative}_{k,t}} \tag{3.4}$$

As in 3.4, RSCI can be positive or negative. Especially, when the positive SCI is equal to the negative SCI, the RSCI will be zero.

The opposite sentiment contagion index (OSCI) describes the degree of disagreement sentiment contagion among the investors. The OSCI will decrease if either the positive or negative sentiment dominates the most sentiment contagion. The definition of OSCI is given by:

$$OSCI_{k,t} = \sqrt{1 - \left(\frac{SCI_{k,t}^{positive} - SCI_{k,t}^{negative}}{1 + SCI_{k,t}^{positive} + SCI_{k,t}^{negative}}\right)^2} \quad (3.5)$$

OSCI is between 0 and 1. It is impacted by the sum and difference of the positive SCI and negative SCI. If the contagion of positive sentiment is equal to negative sentiment, the disagreement caused by sentiment contagion is strong and the OSCI will be 1. On the contrary, if only one type sentiment is diffused, the investors will be impacted by same sentiment contagion and OSCI will be 0.

We consider control variables as the Daily Market Return Rm_t, Daily Volatility of stock k $V_{k,t}$, stock return anomalies such as: Monday Effect $M_{k,t}$, Weekend Effect $W_{k,t}$, and Investors' Attention Effect $Att_{k,t}$. The control variables calculated as follows:

Daily Market Return: We define market return Rm_t as the daily return of CSI 300.

Daily Volatility: We use the method suggested by Garman and Klass (1980) to estimate the Daily Volatility $V_{k,t}$ as:

$$V_{k,t} = 0.5 * \left(p_{k,t}^{h} - p_{k,t}^{l}\right)^2 - (2 * ln2 - 1)\left(p_{k,t}^{cl} - p_{k,t}^{op}\right)^2 \quad (3.6)$$

Where $p_{k,t}^{h}$, $p_{k,t}^{l}$, $p_{k,t}^{cl}$, and $p_{k,t}^{op}$ are the high, low, close and opening price of stock k at day t respectively.

Monday Effect: We define a dummy variable equal to 1 if the trading day is Monday.

Weekend Effect: We define a dummy variable equal to 1 if the trading day is Friday.

Investors' Attention: We estimate the investors' attention on a certain stock k in day t as the total number of clicks on three sentiments:

$$Att_{k,t} = \sum_C log(1 + \sum_{i=1}^{N_{k,t}^C} click_{i,k,t}^c), \ \mathrm{C} \in \{\mathrm{Positive, Neutral, Negative}\} \qquad (3.7)$$

We give a preliminary statistics on SCI, RSCI, OSCI and other variables. The results given by Table 4-A. The mean of $\mathrm{SCI}_{k,t}^{positive}$ and $\mathrm{SCI}_{k,t}^{negative}$ are very close (6.559 and 6.329) , which means neither of the sentiments can dominate market during sample period. The value of $\mathrm{SCI}_{k,t}^{positive}$ and $\mathrm{SCI}_{k,t}^{negative}$ also keep stable (with Std. Dev = 0.604 and 0.514). The range $\mathrm{RSCI}_{k,t}$ of is from -0.677 to 0.576, which also draw the same conclusion that the no extreme sentiment contagion happened in sample period.

[INSERT Table 4-A HERE]

Table 4-B statistics the correlation coefficients of the variables and control variables cross stocks in sample period. All of the coefficients are significant at 1% level. The correlation between $\mathrm{SCI}_{k,t}^{positive}$ and $\mathrm{SCI}_{k,t}^{negative}$ is 0.391, which indicated that the contagion process of positive sentiment and negative sentiment might be simultaneously. For stock return, we can observed $\mathrm{SCI}_{k,t}^{positive}$ is positively correlated with stock return and market return, while $\mathrm{SCI}_{k,t}^{negative}$ is negatively correlated with stock return and market return. $\mathrm{RSCI}_{k,t}$ also has a positive relationship with stock return and market return. For trading volume $\mathrm{Volm}_{k,t}$, $\mathrm{SCI}_{k,t}^{positive}$, $\mathrm{SCI}_{k,t}^{negative}$ and $\mathrm{RSCI}_{k,t}$ have a positive relationship with volume. However, $\mathrm{OSCI}_{k,t}$ has a negative relationship with volume. These results both confirmed with the noise trader hypothesis and disagreement hypothesis.

[INSERT Table 4-B HERE]

Fig 1 depicts the monthly performance of the standardized RSCI and the return of the representative stock Shanghai Pudong Development Bank (600000.SS) during our sample period. The blue and red pillars represent the RSCI and return respectively. It shows a synchronicity between monthly RSCI and stock return.

[INSERT Figure 1 HERE]

4. Results

4.1 Online sentiment contagion and stock return

As we claimed in section 2.1, it is important to identify the variable as a proxy for sentiment contagion rather than information diffusion. We give the correlation of SCI and daily posts of each kind of sentiments in Table 5. We calculate posts as $P_{k,t}^{C} = \log(1 + N_{k,t}^{C})$, where C is the type of the sentiment. We also define the relative sentiment posts as $RP_{k,t} = \frac{P_{k,t}^{positive} - P_{k,t}^{negative}}{1 + P_{k,t}^{positive} + P_{k,t}^{negative}}$. The correlation coefficient showed a weak correlation between $SCI_{k,t}^{positive}$ and $P_{k,t}^{C}$ (-0.122) or $RP_{k,t}^{C}$ (-0.133). We can also draw a similar conclusion on $SCI_{k,t}^{negative}$. Moreover, the $RSCI_{k,t}$ also has a low relationship with $P_{k,t}^{C}$ (0.005 and 0.051) or $RP_{k,t}^{C}$ (-0.028).

[INSERT Table 5 HERE]

We adopt a autoregressive model to test whether the online sentiment contagion predicts returns. All variables in the model is estimated as 5 lags before trading day. The depended variable is stock return. The independent variables of the model defined as the SCI, prior return of stock k and volume. The control variables include market return, volatility, investor attention, Monday Effect and Weekend Effect. Following Tetlock (2007), we define the lag operator L5 as $L5(x_t) = [x_{t-1}\ x_{t-2}\ x_{t-3}\ x_{t-4}\ x_{t-5}]$. With the settings above, we build a regression model as:

$$R_{k,t} = \alpha + \beta_1 \cdot \mathrm{L5}(\mathrm{Contagion}_{k,t}) + \beta_2 \cdot \mathrm{L5}(R_{k,t}) + \beta_3 \cdot \mathrm{L5}(Volume_{k,t}) + \beta_4 \cdot \mathrm{L5}(Control^{a}_{k,t}) + \beta_4 \cdot Control^{b}_{k,t} + \varepsilon_{k,t} \quad (4.1)$$

Where $\mathrm{Contagion}_{\mathrm{k,t}}$ represents a certain kind of sentiment, i.e. $\mathrm{SCI}^{\mathrm{positive}}_{\mathrm{k,t}}$, $\mathrm{SCI}^{\mathrm{negative}}_{\mathrm{k,t}}$ or $\mathrm{RSCI}_{\mathrm{k,t}}$. $\mathrm{Control}^{\mathrm{a}}_{\mathrm{k,t}}$ represents the market return, volatility, investor attention, and $\mathrm{Control}^{\mathrm{b}}_{\mathrm{k,t}}$ represents the Monday Effect dummy and Weekend Effect dummy.

We also give a comparison on the predictability between the proxy of sentiment contagion we used and the number of posts on certain sentiment. We use $\mathrm{Posts}_{\mathrm{k,t}}$ to represents the number of posts on sentiment, i.e. $\mathrm{P}^{\mathrm{positive}}_{\mathrm{k,t}}$, $\mathrm{P}^{\mathrm{negative}}_{\mathrm{k,t}}$ or $\mathrm{RP}_{\mathrm{k,t}}$, and the regression gives as:

$$R_{k,t} = \alpha + \beta_1 \cdot \mathrm{L5}(Posts_{k,t}) + \beta_2 \cdot \mathrm{L5}(R_{k,t}) + \beta_3 \cdot \mathrm{L5}(Volume_{k,t}) + \beta_4 \cdot \mathrm{L5}(Control^{a}_{k,t}) + \beta_4 \cdot Control^{b}_{k,t} + \varepsilon_{k,t} \quad (4.2)$$

The regression result is given in Table 6. Panel A reports the results of model 4.1, while Panel B reports the results of model 4.2. Results in panel A cannot reject the hypothesis that the online sentiment cognation have predictability on stock return. For online sentiment cognation, we use $\mathrm{L5}(\mathrm{SCI}^{\mathrm{positive}}_{\mathrm{k,t}})$, $\mathrm{L5}(\mathrm{SCI}^{\mathrm{negative}}_{\mathrm{k,t}})$ and $\mathrm{L5}(\mathrm{RSCI}_{\mathrm{k,t}})$ to predict stock return respectively. We can find a significant positive predictability of $\mathrm{SCI}^{\mathrm{positive}}_{\mathrm{k,t-1}}$ (0.022, significant at 1% level), a significant negative predictability of $\mathrm{SCI}^{\mathrm{negative}}_{\mathrm{k,t-1}}$ (-0.028, significant at 1% level) and a significant positive predictability of $\mathrm{RSCI}_{\mathrm{k,t-1}}$ (0.184, significant at 1% level), which consistent with the noise trader hypothesis that the sentiment may affect noise trader and generate upward or downward pressure on stock price. Moreover, the results also suggested the proxy we used is sentiment rather than information. We can find a reversal appears both in $\mathrm{L5}(\mathrm{SCI}^{\mathrm{positive}}_{\mathrm{k,t}})$ and $\mathrm{L5}(\mathrm{SCI}^{\mathrm{negative}}_{\mathrm{k,t}})$. The coefficient of

$\mathrm{SCI}_{k,t-3}^{positive}$ and $\mathrm{SCI}_{k,t-4}^{positive}$ are significantly negative and the sum of the absolute value of coefficients is bigger than the coefficient of $\mathrm{SCI}_{k,t-1}^{positive}$ (0.026+0.016>0.022), which suggested the sentiment contagion only caused a temporary pressure on price in short term and following a fully reversal in the long term. Similar results can observed from the negative sentiment contagion (a reversal with 2 positive coefficient of 0.018, significant at 1% level) and relative sentiment contagion (a reversal with negative coefficient of -0.102, significant at 5% level, and negative coefficient of -0.076, significant at 5% level).

[INSERT Table 6 HERE]

Results in panel B suggested that the number of posts have less predictability on stock return. We use $\mathrm{L5}(\mathrm{P}_{k,t}^{positive})$, $\mathrm{L5}(\mathrm{P}_{k,t}^{negative})$ and $\mathrm{L5}(\mathrm{RP}_{k,t})$ to predict stock return respectively. For the $\mathrm{L5}(\mathrm{P}_{k,t}^{positive})$, we cannot find a significant positive predictability. Coefficients of $\mathrm{L5}(\mathrm{P}_{k,t}^{positive})$ are not significant. For the $\mathrm{L5}(\mathrm{P}_{k,t}^{negative})$, it has a predictability of $\mathrm{P}_{k,t-1}^{negative}$(-0.121, significant at 1% level). More importantly, We cannot find predictability on $\mathrm{L5}(\mathrm{RP}_{k,t})$, which is consistent with the results of Tumarkin and Whitelaw (2001), Antweiler and Frank (2004), Das and Chen (2007) and Kim and Kim (2014). Therefore, we suggested that the clicks might be a better proxy for measure sentiment contagion, and have more predictability than number of posts.

4.2 Online Sentiment Contagion and direction of trades

More evidence of online sentiment contagion affect investors might be find from trading orders. If the noise trader theory holds, investors tend to buy assets driven by positive sentiment contagion and sell assets driven by negative sentiment contagion. We try to find the predictability of online sentiment contagion on buying and selling orders of investors. We separate the trades data of stock k at trading day t into 3 parts: buy initiated order ($\mathrm{Order}_{k,t}^{buy}$), sell initiated order ($\mathrm{Order}_{k,t}^{sell}$) and order imbalance

($\mathrm{Order}_{k,t}^{imba}$), where we calculated $\mathrm{Order}_{k,t}^{imba} = \mathrm{Order}_{k,t}^{buy} - \mathrm{Order}_{k,t}^{sell}$.

We test hypothesis 2 and expect to find positive sentiment contagion can predict buy initiated order, negative sentiment contagion can predict sell initiated order, and relative sentiment contagion can predict order imbalance. In the model of this section, the depended variable is three kinds of orders. The independent variables defined as the SCI, prior each kind of orders, and prior stock returns. With the settings above, we build a regression as:

$$Order_{k,t}^{C} = \alpha + \beta_1 \cdot \mathrm{L5}(\mathrm{Contagion}_{k,t}) + \beta_2 \cdot \mathrm{L5}(Order_{k,t}^{C}) + \beta_3 \cdot \mathrm{L5}(R_{k,t}) + \beta_4 \cdot \mathrm{L5}(Control_{k,t}^{a}) + \beta_4 \cdot Control_{k,t}^{b} + \varepsilon_{k,t} \qquad (4.3)$$

Where $\mathrm{Order}_{k,t}^{C}$ represents a certain kind of orders, i.e. $\mathrm{Order}_{k,t}^{buy}$, $\mathrm{Order}_{k,t}^{sell}$ or $\mathrm{Order}_{k,t}^{imba}$. $\mathrm{Control}_{k,t}^{a}$ represents the trading volume, market return, volatility, investor attention, and $\mathrm{Control}_{k,t}^{b}$ represents the Monday Effect dummy and Weekend Effect dummy. Other variables are the same as in the previous model.

Moreover, we consider an trade-size classification to identified small investors' trading behavior. Following Lee and Ready (1991) and Mikhail, Walther and Willis (2007), we employ a cutoff method to get small investors' trades. The cutoff shield we used is 1,000,000 shares. If the value of one trade is smaller than 1,000,000 shares, we identified the trade is initiated by small or individual investor. We also report the results of trade-size classification in Table 7. More cutoff shields are tested in section 5.

[INSERT Table 7 HERE]

The regression result is given in Table 7. Panel A reports the results of model 4.3 with all trades data, while Panel B reports the results with the data which only contain the trades smaller than 1,000,000 shares. Results in Table 7 cannot reject the hypothesis that the online sentiment cognation have predictability on investors' trading behaviors. In panel A, We can find a significant positive predictability of all

$L5(SCI_{k,t}^{positive})$. The coefficient of $SCI_{k,t-1}^{positive}$ is positive (2118.04) and significant at 1% level. $SCI_{k,t-2}^{positive}$ to $SCI_{k,t-4}^{positive}$ also have positive coefficients, and all of which are significant at 5% level. The results showed that online positive sentiment contagion can affect investors and generate buy initiated orders. For the negative sentiment contagion, the significant predictability on the sell initiated orders exists in negative sentiment contagion with one day lag $SCI_{k,t-1}^{negative}$ (1215.45, significant at 1% level), two day lag $SCI_{k,t-2}^{negative}$(1033.12, significant at 1% level), and four day lag $SCI_{k,t-4}^{negative}$(777.87, significant at 5% level). However, for the relative sentiment contagion, $RSCI_{k,t-1}$ does not show a significant predictability, while $RSCI_{k,t-3}$ appears a negative predictability (-5350.87, significant at 1% level).

Results in panel B showed that online sentiment contagion can better predict individual investors' behavior. Both $L5(SCI_{k,t}^{positive})$ and $L5(SCI_{k,t}^{negative})$ showed similar predictability compare with results in panel A, which means the more positive / negative sentiment contagion, the more buy / sell orders submitted by individual investors. Moreover, for $L5(RSCI_{k,t})$, $RSCI_{k,t-1}$ positively predicts order imbalance (3337.23, significant at 5% level) and $RSCI_{k,t-3}$ negative predicts order imbalance (-3966.37, significant at 5% level).

4.3 Opposite sentiment and volume

Several hypothesis associated sentiment contagion with trading volume. The noise trader theory assumes that sentiment will have a temporary pressure on stock price, and accompany with high trading volume. The alternative hypothesis is disagreement hypothesis, which suggested that trading volume is induced by the opposing opinions on the future price held by investors in the market (Hirshleifer [1977], Diamond and Verrecchia [1981], Hong and Stein [2007]). Herding model also suggested that sentiment contagion generates optimistic and pessimistic investors of stock market. Besides hypothesis above, Tetlock (2007) find that media pessimistic sentiment negatively predict trading volume, which provide evidence that pessimism is a proxy of trading costs. We test these hypothesis by using sentiment contagion index and opposite sentiment contagion. Our regression model is built as follows:

$$Volume_{k,t} = \alpha + \beta_1 \cdot \mathrm{L5}\left(Contagion_{k,t}\right) + \beta_2 \cdot \mathrm{L5}\left(Volume_{k,t}\right) + \beta_3 \cdot \mathrm{L5}\left(R_{k,t}\right) + \beta_4 \cdot \mathrm{L5}\left(Control^{a}_{k,t}\right) + \beta_4 \cdot Control^{b}_{k,t} + \varepsilon_{k,t} \quad (4.4)$$

Where $\mathrm{Contagion}_{k,t}$ represents one kind of contagion in $\mathrm{SCI}^{\mathrm{positive}}_{k,t}$, $\mathrm{SCI}^{\mathrm{negative}}_{k,t}$ or $\mathrm{OSCI}_{k,t}$. $\mathrm{Control}^{a}_{k,t}$ represents the market return, volatility, investor attention, and $\mathrm{Control}^{b}_{k,t}$ represents the Monday Effect dummy and Weekend Effect dummy. The results are listed in the Panel A of Table 8.

Panel A suggested that neither of the noise trader hypothesis and disagreement hypothesis can be rejected according to the results. For positive or negative sentiment contagion, $\mathrm{SCI}^{\mathrm{positive}}_{k,t-1}$ and $\mathrm{SCI}^{\mathrm{negative}}_{k,t-1}$ both showed positive and highly significant predictability (0.033 and 0.027) on trading volume. Other lags of sentiment contagion also positively related with trading volume. These results are inconsistent with Tetlock (2007), while confirmed with the noise trader hypothesis. Three of the five lags of $\mathrm{OSCI}_{k,t}$ is positive related with trading volume. The first lag of $\mathrm{OSCI}_{k,t}$ is significantly predicts (3.472) trading volume, which implied that the opposite sentiment contagion induced the disagreement of investors and consist with disagreement hypothesis: the more opposite sentiment contagion, the more trading volume in stock market. However, the coefficient of the third lag of $\mathrm{OSCI}_{k,t}$ is negative (-1.521) and significant at 10% level.

[INSERT Table 8 HERE]

We provide more test to find out whether sentiment is associated with trading volume. We calculated trading volume of individual investors by using the classification method in 4.2, and to see if trading volume positively related with sentiment contagion. The regression is built as follows:

$$IdVolume_{k,t} = \alpha + \beta_1 \cdot \mathrm{L5}\left(Contagion_{k,t}\right) + \beta_2 \cdot \mathrm{L5}\left(IdVolume_{k,t}\right) + \beta_3 \cdot \mathrm{L5}\left(R_{k,t}\right) + \beta_4 \cdot \mathrm{L5}\left(Control^{a}_{k,t}\right) + \beta_4 \cdot Control^{b}_{k,t} + \varepsilon_{k,t} \qquad (4.5)$$

The results of model 4.5 listed in Panel B of Table 8. In panel B, we still can observed evidence to support the hypothesis that sentiment contagion predict trading volume. Trading volume has significantly positive predictability on the positive, negative and opposite sentiment contagion. The third lag of opposite sentiment contagion is positively related with individual trading volume, which is different in the regression of total trading volume.

4.4 What drives online sentiment contagion?

It is important to know where does sentiment contagion come from and why investors affected by the online messages. Social influence (Hong, Kubik and Stein [2004]; Brown et al. [2008]) studies emphasis the social norms and reputation are important for the influence happens. In our study, anonymous posters lacked incentive to increase reputation. Therefore, the contagion happened might be because of the trading environment. Feedback trading theory suggested that noise traders tend to buy assets when prices rise and sell assets when prices fall. Sentiment might be generated and spread among feedback traders. To test the feedback trader hypothesis that pervious stock price will predict investors' sentiment contagion process, we employ the model with depended variable as $\mathrm{SCI}^{\mathrm{positive}}_{\mathrm{k,t}}$, $\mathrm{SCI}^{\mathrm{negative}}_{\mathrm{k,t}}$ and $\mathrm{RSCI}_{\mathrm{k,t}}$. The independent variables defined as prior stock returns, prior each kind of sentiments, and prior volatility. The regression model is given as:

$$\mathrm{Contagion}_{k,t} = \alpha + \beta_1 \cdot \mathrm{L5}\left(R_{k,t}\right) + \beta_2 \cdot \mathrm{L5}\left(\mathrm{Contagion}_{k,t}\right) + \beta_3 \cdot \mathrm{L5}\left(V_{k,t}\right) + \beta_4 \cdot \mathrm{L5}\left(Control^{a}_{k,t}\right) + \beta_4 \cdot Control^{b}_{k,t} + \varepsilon_{k,t} \qquad (4.6)$$

In model 4.6, $\mathrm{Control}^{\mathrm{a}}_{\mathrm{k,t}}$ represents the trading volume, market return, investor attention, and $\mathrm{Control}^{\mathrm{b}}_{\mathrm{k,t}}$ represents the Monday Effect dummy and Weekend Effect dummy. Other variables are the same as in the previous model. Moreover, we consider using trading order to provide more evidence whether historical market status can affect online sentiment contagion process. We employ the model with independent variable of three kinds of orders ($\mathrm{Order}_{k,t}^{buy}$, $\mathrm{Order}_{k,t}^{sell}$ or $\mathrm{Order}_{k,t}^{imba}$) to replace the prior stock returns. The model is given in 4.7:

$$\text{Contagion}_{k,t} = \alpha + \beta_1 \cdot \mathrm{L5}\left(Order_{k,t}^{C}\right) + \beta_2 \cdot \mathrm{L5}\left(\text{Contagion}_{k,t}\right) + \beta_3 \mathrm{L5}\left(V_{k,t}\right) + \beta_4 \cdot \mathrm{L5}\left(Control_{k,t}^{a}\right) + \beta_4 \cdot Control_{k,t}^{b} + \varepsilon_{k,t} \quad (4.7)$$

[INSERT Table 9 HERE]

The regression result is given in Table 9. Panel A reports the results of model 4.6 with stock returns, while Panel B reports the results of model 4.7 with trading orders. We cannot reject the feedback trader hypothesis that the stock return and buy/sell orders can predict the online sentiment contagion process according to Table 9. In panel A, one to three lag returns have significant positive predictability on positive sentiment contagion (0.017, 0.013, 0.010, both significant at 1% level). The similar results can draw from the coefficient of negative sentiment contagion (0.002, 0.006, 0.005) and relative sentiment contagion (0.002, 0.001, 0.001). These results confirmed that the relationship between feedback traders and sentiment contagion. When prior stock return is big, positive sentiment contagion accelerated, and low stock return accelerates the contagion of negative sentiment. All of the lag sentiment contagion indexes are highly significant, which implied that sentiment contagion process persists at long term. Moreover, volatility cannot predict sentiment contagion.

Results in panel B also provide evidence that online sentiment contagion affected by historical market status. Prior buy initiated order also can accelerate positive sentiment contagion in 4 of the 5 lags. Sell initiated order with 4 lags can accelerated negative sentiment contagion. These results confirmed the robustness of conclusions in Panel A. However, our results find that order imbalance cannot predict relative sentiment contagion in 5 trading days.

4.5 Portfolio based on sentiment contagion

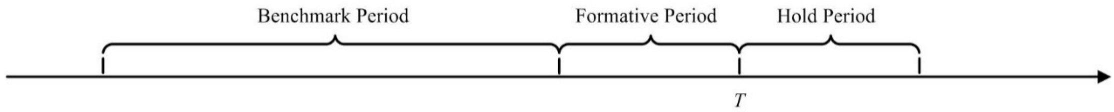
If online sentiment contagion can predict stock return, which implied that we can build a portfolio based on the SCI to gain the excess profit. Inspired by Fang and Peress (2009) who built a portfolio based on the media coverage and found that buying and holding the non-coverage stocks and selling out the high coverage stocks could gain an excess return even controlling the market risk. We use the SCI and RSCI to build the portfolio to study whether the sentiment-based portfolio gain a profit without systemic risk. It will be another piece of evidence that the sentiment contagion of online messages moves the stock price.

We built portfolio based on the Abnormal Sentiment Index (ASI), which is calculated by the abnormal SCI. We define three periods to get ASI and portfolio: the benchmark period, the forming period and the holding period, as shown in Fig 2. ASI is defined as the difference of the averages SCI between the forming period and the benchmark period:

$$ASI_{k,T}^{c} = \frac{1}{FL}\sum_{t\in(T-FL,T-1)} Contagion_{k,t} - \frac{1}{BL}\sum_{t\in(T-BL-FL,T-1)} Contagion_{k,t} \qquad (4.8)$$

Where $\mathrm{Contagion}_{k,t}$ represents a certain kind of sentiment, i.e. $\mathrm{SCI}_{k,t}^{positive}$, $\mathrm{SCI}_{k,t}^{negative}$ or $\mathrm{RSCI}_{k,t}$. T is the time point to calculate ASI. For an example of ASI with $\mathrm{RSCI}_{k,t}$, if the benchmark length (BL) is 4 weeks and the format length (FL) is 1 week, the $\mathrm{ASI}_{i,T}^{c}$ will be the average $\mathrm{RSCI}_{k,t}$ of last week minus the averages RSCI in the period of [-5, -2] weeks.

Fig 2



[INSERT Table 9 HERE]

We consider the average SCI in the benchmark period as a normal performance of the stock and the one in the forming period as its latest performance. The ASI shows abnormal fluctuation of the recent sentiment contagion. By comparing the ASI of each stocks, we selected the stocks with different levels of abnormal positive or negative sentiments to construct our portfolio.

We even divide 290 stocks into 5 portfolios based on their ASI, which are named as Max, Big, Median, Small and Min, and have 58 stocks in each portfolio. We calculate their size-weighted return as the portfolio return and use the following regression to compute alpha:

$$Portfolio\ Return_t = Alpha + \beta_1 Rm_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 CAR_{t-1} + \beta_5 CAR_{t-2,t-6} + \varepsilon_t \tag{4.9}$$

Where $\mathrm{Rm_t}$ is the daily return of the market. $\mathrm{SMB_t}$ is the daily difference of size-weighted portfolio return between the listed companies of small size and the big size. $\mathrm{HML_t}$ is the daily difference of size-weighted portfolio return between the listed companies of high book-to-market ratio and low book-to-market ratio. These factors are calculated following Fama and French (1993, 1996). $\mathrm{CAR_{t-1}}$ and $\mathrm{CAR_{t-2,t-6}}$ are the factors of momentum effect (Jegadeesh and Titman [1993]) We build $\mathrm{CAR_{t-1}}$ and $\mathrm{CAR_{t-2,t-6}}$ as the accumulative abnormal return in certain period following Tetlock, Saar-tsechansky and Macskassy (2008). With these control variables, the Alpha is calculated as the return excluding the market risk and the momentum effect.

[INSERT Table 10 HERE]

Table 10 gives the alpha of each portfolio. We made a grid search on the different BL, FL and HL to maximum the return of the portfolio in a short term holding period (less than 20 trading days). The result of best combination with highest returns in the three periods is given in table 10. The best BL, FL and HL are 22 trading days, 9 trading days and 9 trading days respectively. For each portfolio, table 10 reports its alpha, t-test of alpha and the r-square of the regression. The sixth column of table 10 shows the alpha by buying the maximum and selling out the minimum. It is important to note that we do not consider the transaction cost and the short-sale constraint. With these restrictions, our alpha is calculated as a theoretical value.

Table 10 shows that the portfolio based on online sentiment contagion is profitable. For RSCI, if we buy the max ASI portfolio and sell out the min ASI portfolio, we could get a significant abnormal return of 0.6190% in 9 trading days. Accompany with the ASI increase, the portfolio alpha increase significantly. For positive SCI, selling out the min ASI portfolio of positive SCI could gain a 0.2811% return in 9 trading days (with t-statistics 4.9227). However, buying the portfolio with max ASI of positive sentiment will gain a significance low alpha (-0.5509). Moreover, the arbitrage portfolio of positive SCI is unprofitable. For negative SCI, selling out the max ASI portfolio of negative SCI could gain a significantly 0.7247% return in holding periods, while the arbitrage portfolio of negative SCI can also gain 0.6253% return in holding period. These results suggested that the hypothesis that buying portfolio with higher ASI and selling portfolio with lower ASI can bring a significant excess profit cannot be rejected. We provide more results of the portfolios with long holding periods in robustness part.

5. Robustness tests

5.1 Sentiment or information?

For better test that whether the proxy based on number of clicks that we used is a measurement of sentiment contagion or information diffusion, we employ another method used by Hasbrouck (1991) to measure the impact of information on stock returns. Hasbrouck (1991) build a structural VAR model and calculate cumulative impulse response of the innovations to test whether the stock trades contains information. If the stock trades do not contain information, the impact of the trades on stock return should be zero in long term. Following Hasbrouck (1991), we can get our test model as follows:

$$\begin{pmatrix} R_{k,t} \\ SCI_{k,t}^{C} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \beta_{1,t-1} & \beta_{2,t-1} \\ \beta_{3,t-1} & \beta_{4,t-1} \end{pmatrix} \begin{pmatrix} R_{k,t-1} \\ SCI_{k,t-1}^{C} \end{pmatrix} + \cdots + \begin{pmatrix} \beta_{1,t-5} & \beta_{2,t-5} \\ \beta_{3,t-5} & \beta_{4,t-5} \end{pmatrix} \begin{pmatrix} R_{k,t-5} \\ SCI_{k,t-5}^{C} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \quad (5.1)$$

We calculated the expected cumulative $R_{k,t}$ revision conditional on $\varepsilon_{2,0}$, which capture the permanent sentiment contagion impact on stock price. The 10 lags cumulative revisions of one representative stocks is shown in figure 3. The statistics of the distribution is provided by Table 11.

[INSERT Figure 3 HERE]

[INSERT Table 11 HERE]

Figure 3 showed that, for stock 000001.SZ, the 10 lags cumulative revisions caused by the impact of positive or negative sentiment contagion is close to zero. The statistics also confirm the conclusion that the cumulative revisions are no significant differences from zero. In table 11, the mean of cumulative revision caused by the impact of positive sentiment contagion is zero, with a low skewness (0.0268). The Kurtosis is also 3.6008, which is close to 3. More importantly, we give a t-test on the cumulative revisions of 10 lags with t-statistics value equal to 0.39. Similar results can result from the statistics of impact from negative sentiment contagion. These results consist with the conclusion that the proxy based on number of clicks we used is a measurement of sentiment contagion, which cannot give the stock price a permanent impact.

5.2 More trading order classification

We test more trading order classification to find whether the conclusion in section 4.2 has robustness. In 4.2, we consider the cutoff shield as 500,000 CNY. In this section, we consider double cutoff shields: 500,000 CNY and 50,000 CNY. Prior studied (Lee and Ready [1991]; Mikhail, Walther and Willis [2007]) conclude that the informed trader tend to break up their trades to hide information advantage, therefore the medium-sized trades may have more information. We suggested the trade is bigger than 500,000 CNY is initiated by institutional investors, while trade is bigger than 50,000 CNY is initiated by individual investors. Our regression model is the same as in section 4.2.

[INSERT Table 12 HERE]

The results are shown in Table 12. Panel A provide the results of sentiment contagion predict the institutional investors' trading behavior. Panel B provide the results of sentiment contagion predict the individual investors' trading behavior. We can find results in Panel B are similar as the conclusion in section 4.2 that the sentiment contagion significantly predicts the trading behavior of individual investors. For the results in Panel A, only $\mathrm{SCI}_{k,t-5}^{positive}$ and $\mathrm{SCI}_{k,t-3}^{negative}$ can predict the trading behavior of institutional investors. None of the relative sentiment contagion factors can predict the trading activity of institutions. These result suggested implied that the online sentiment contagion more likely affect individual investors rather than institutional investors.

5.3 Portfolios with long holding periods

In section 4.5, we limited the holding period of portfolio as a short term. We also provide a results of long term (More than 20 trading days) holding period. In this section, we calculated ASI, portfolio return and control variables with weekly data. The result of best combination with highest returns in the three periods is given in Table 13. The best BL, FL and HL are 16 weeks, 4 weeks and 4 weeks respectively. Table 13 shows the portfolio alpha with different kinds of sentiment contagion, which are also profitable as in section 4.5. For RSCI, when ASI increase, the portfolio alpha increase significantly and the alpha of arbitrage portfolio is 1.0674% with a 9.3255 t-test value. Selling out the max ASI portfolio of negative SCI could gain a 1.2997% monthly return and the arbitrage strategy can get 1.1876% per month. Similar with the results in section 4.5, positive SCI is hard to bring profit. Buying the portfolio with max ASI will gain a significance low alpha(-1.0916 and the t-test value is 11.0330). For negative SCI, the alpha of arbitrage strategy on negative sentiment contagion is 1.1876%. Shorting the highest ASI portfolio brings 1.2997% abnormal return, even higher than the arbitrage strategy.

[INSERT Table 13 HERE]

6. Conclusion

This paper provides evidence that sentiment is contagious through social influence in stock market. First of all, this study makes a difference for "sentiment" and "information". By employ two filters to remove the message might contain macro / industry news and firm specific news, this paper find the number of clicks have better predictability on stock returns. The results are consistent with the noise trader model developed by De Long et al. (1990a). Moreover, this paper find that sentiment contagion can significantly predict trading orders, which find the way of how investors affect by others' sentiment through internet and what drives the sentiment contagion. The results are also consistent with the noise trading model and feedback trading theory. Additionally, this paper associates sentiment contagion with trading volume and find the results consistent both with noise trader theory and disagreement hypothesis. Finally, this study tests whether we can get a profit from sentiment contagion. A portfolio is built based on abnormal sentiment contagion and can gain an excess return without consideration of transaction costs.

Reference

[1] Antweiler, Werner, and Murray Z. Frank, 2004, Is all that talk just noise? The information content of internet stock message boards, The Journal of Finance 59, 1259-1294.

[2] Baker, Malcolm, Jeffrey Wurgler, and Yu Yuan, 2012, Global, local, and contagious investor sentiment, Journal of Financial Economics 104, 272-287.

[3] Banerjee, Abhijit V., 1992, A simple model of herd behavior, The Quarterly Journal of Economics 107, 797-817.

[4] Banerjee, Snehal, and Ilan Kremer, 2010, Disagreement and learning: Dynamic patterns of trade, The Journal of Finance 65, 1269-1302.

[5] Barber, Brad M., and Terrance 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, 785-818.

[6] Barsade, Sigal G., 2002, The ripple effect: Emotional contagion and its influence on group behavior, Administrative Science Quarterly 47, 644-675.

[7] Bollen, Johan, Huina Mao, and Xiaojun Zeng, 2011, Twitter mood predicts the stock market, Journal of Computational Science 2, 1-8.

[8] Brown, Gregory W., and Michael T. Cliff, 2004, Investor sentiment and the near-term stock market, Journal of Empirical Finance 11, 1-27.

[9] Da, Z. H. I., Joseph Engelberg, and Pengjie Gao, 2011, In search of attention, The Journal of Finance 66, 1461-1499.

[10] Das, Sanjiv R., and Mike Y. Chen, 2007, Yahoo! For amazon: Sentiment extraction from small talk on the web, Management Science 53, 1375-1388.

[11] De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Noise trader risk in financial markets, Journal of Political Economy 98, 703-738.

[12] Diamond, Douglas W., and Robert E. Verrecchia, 1981, Information aggregation in a noisy rational expectations economy, Journal of Financial Economics 9, 221-235.

[13] Dong, Zhendong, and Qiang Dong, 2003, Hownet - a hybrid language and knowledge resource, Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003 International Conference on.

[14] Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work*, The Journal of Finance 25, 383-417.

[15] Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, Journal of Financial Economics 33, 3-56.

[16] Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, The Journal of Finance 51, 55-84.

[17] Fang, Lily, and Joel Peress, 2009, Media coverage and the cross-section of stock returns, The Journal of Finance 64, 2023-2052.

[18] Foroni, Francesco, and Gün R. Semin, 2009, Language that puts you in touch with your bodily feelings: The multimodal responsiveness of affective expressions, Psychological Science 20, 974-980.

[19] Garfinkel, Jon A., and Jonathan Sokobin, 2006, Volume, opinion divergence, and returns: A study of post–earnings announcement drift, Journal of Accounting Research 44, 85-112.

[20] Gentzkow, Matthew, and Jesse M. Shapiro, 2010, What drives media slant? Evidence from u.S. Daily newspapers, Econometrica 78, 35-71.

[21] Hirshleifer, Jack, 1971, The private and social value of information and the reward to inventive activity, The American Economic Review 61, 561-574.

[22] Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2004, Social interaction and stock-market participation, The Journal of Finance 59, 137-163.

[23] Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2005, Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, The Journal of Finance 60, 2801-2824.

[24] Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, The Journal of Finance 54, 2143-2184.

[25] Ivković, Zoran, and Scott Weisbenner, 2007, Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices, Review of Financial Studies 20, 1327-1357.

[26] Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, The Journal of Finance 48, 65-91.

[27] Karpoff, Jonathan M., 1986, A theory of trading volume, The Journal of Finance 41, 1069-1087.

[28] Kim, Soon-Ho, and Dongcheol Kim, 2014, Investor sentiment from internet message postings and the predictability of stock returns, Journal of Economic Behavior & Organization 107, Part B, 708-729.

[29] Kirman, Alan, 1993, Ants, rationality, and recruitment, The Quarterly Journal of Economics 108, 137-156.

[30] Li, Xiu, Huimin Wang, and Xinwei Yan, 2015, Accurate recommendation based on opinion mining, in Hui Sun, Chin-Yu Yang, Chun-Wei Lin, Jeng-Shyang Pan, Vaclav Snasel, and Ajith Abraham, eds.: Genetic and evolutionary computing (Springer International Publishing).

[31] Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing, 2011, How social influence can undermine the wisdom of crowd effect, Proceedings of the National Academy of Sciences 108, 9020-9025.

[32] Lux, Thomas, 1995, Herd behaviour, bubbles and crashes, The Economic Journal 105, 881-896.

[33] Neumann, R., and F. Strack, 2000, "Mood contagion": The automatic transfer of mood between persons, Journal of personality and social psychology 79, 211-223.

[34] Shiller, Robert J., Stanley Fischer, and Benjamin M. Friedman, 1984, Stock prices and social dynamics, Brookings Papers on Economic Activity 1984, 457-510.

[35] Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, The Journal of Finance 62, 1139-1168.

[36] Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, The Journal of Finance 63, 1437-1467.

[37] Zouaoui, Mohamed, Geneviève Nouyrigat, and Francisca Beer, 2011, How does investor sentiment affect stock market crises? Evidence from panel data, Financial Review 46, 723-747.

Table 1. Samples of online messages in Eastmoney

Title	Date	User	Clicks
万科出现见顶信号！！	2014/7/23	浙江丽水股友	207
疯啦股指意欲何为	2014/7/23	北京股友	114
000001 明天又到 9.6	2014/7/23	天天好 9988	341
淫威之下，小散付出惨痛代价	2014/7/23	湖北孝感股友	220
青奥题才为啥直往下打。	2014/7/23	重庆股友	206
为啥世青奥会没行清	2014/7/23	重庆股友	114
为啥青奥题才没行清	2014/7/23	重庆股友	181
牢牢站上 9.88，否则减半[财力]	2014/7/23	月冷星清	236
打到股指期货	2014/7/23	安徽铜陵股友	258

Table 2. Summary statistics on stocks and online messages

	Return (%)	Market Return (%)	Capitalization (billion)	Book/Market (%)	Posts (thousand)	Clicks (million)
Mean	-0.033	-0.006	40.570	0.578	11.997	12.297
Median	-0.037	-0.002	15.500	0.533	11.034	11.508
Std.	2.132	1.097	119.840	0.296	3.079	5.021
Maximum	9.677	4.348	1483.860	1.731	21.360	25.371
Minimum	-8.567	-5.277	2.890	0.082	5.245	3.277

Table 3-A. In-sample classification by using Naïve Bayesian Learning

In-sample classification accuracy				
Manually Classify	%	Naïve Bayesian Learning Classification		
		Positive	Neutral	Negative
Positive	15.6	10.9	4.4	0.3
Neutral	62.6	1.6	58.8	2.2
Negative	21.8	0.1	6.1	15.7
5000 messages		12.6	69.2	18.2

Table 3-B. Out-sample classification by using Naïve Bayesian Learning

Out-sample classification accuracy				
Manually Classify	%	Naïve Bayesian Learning Classification		
		Positive	Neutral	Negative
Positive	12.5	7.1	5.0	0.4
Neutral	69.3	4.9	60.5	3.9
Negative	18.2	0.2	7.7	10.3
1000 messages		12.2	73.2	14.6

Table 3-C. Summary statistics on deleted messages by two filters

	Del/Sum	Del/SumofNewsday	Mean Del	Max Del	Min Del	Name/SumofNewsday	Name/Del
Posts	7.74%	40.52%	24.35	220.38	6.79	33.48%	54.00%
Clicks	10.96%	62.15%	36728.06	334733.67	10426.50	51.45%	64.14%

Table 4-A. Summary statistics on Sentiment Contagion Indexes

	$SCI_{k,t}^{positive}$	$SCI_{k,t}^{negative}$	$RSCI_{k,t}$	$OSCI_{k,t}$	Volume	$V_{k,t}$	$Att_{k,t}$
Mean	6.559	6.329	-0.013	0.987	18.921	0.022	19.388
Median	6.56	6.322	-0.011	0.964	14.334	0.021	19.302
Std.	0.604	0.514	0.142	0.015	15.888	0.005	1.264
Maximum	9.482	8.866	0.576	1.000	154.849	0.042	24.028
Minimum	4.697	4.852	-0.677	0.736	3.028	0.015	15.905

Table 4-B. Correlation for variables and control variables

	$SCI_{k,t}^{positive}$	$SCI_{k,t}^{negative}$	$RSCI_{k,t}$	$Att_{k,t}$	$OSCI_{k,t}$	$R_{k,t}$	Rm_t	$Volm_{k,t}$
$SCI_{k,t}^{negative}$	0.391							
$RSCI_{k,t}$	0.640	-0.453						
$Att_{k,t}$	0.807	0.795	0.118					
$OSCI_{k,t}$	0.294	-0.071	0.328	0.109				
$R_{k,t}$	0.041	-0.044	0.035	0.061	0.011			
Rm_t	0.025	0.014	0.012	0.031	0.011	0.597		
$Volm_{k,t}$	0.026	0.036	0.004	-0.043	-0.062	0.125	0.046	
$V_{k,t}$	0.050	0.014	0.036	0.025	-0.020	0.014	0.024	0.162

Figure 1. Monthly RSCI and stock return of 600000.SS

Table 5. Correlation for cognation proxies with clicks and posts

	$SCI_{k,t}^{positive}$	$SCI_{k,t}^{negative}$	$RSCI_{k,t}$	$P_{k,t}^{positive}$	$P_{k,t}^{negative}$
$SCI_{k,t}^{negative}$	0.391				
$RSCI_{k,t}$	0.640	-0.453			
$P_{k,t}^{positive}$	-0.122	-0.157	0.006		
$P_{k,t}^{negative}$	-0.133	-0.172	0.051	0.615	
$RP_{k,t}$	0.013	0.030	-0.028	0.436	-0.411

Table 6. Predictability of sentiment contagion and posts on stock return

Panel A: Sentiment Contagion				Panel B: Posts			
Variables	$R_{k,t}$			Variables	$R_{k,t}$		
$SCI_{k,t-1}^{positive}$	0.022^{c}			$P_{k,t-1}^{positive}$	0.033		
$SCI_{k,t-2}^{positive}$	-0.003			$P_{k,t-2}^{positive}$	-0.058		
$SCI_{k,t-3}^{positive}$	-0.026^{c}			$P_{k,t-3}^{positive}$	-0.032		
$SCI_{k,t-4}^{positive}$	-0.016^{b}			$P_{k,t-4}^{positive}$	-0.071		
$SCI_{k,t-5}^{positive}$	-0.010			$P_{k,t-5}^{positive}$	-0.032		
$SCI_{k,t-1}^{negative}$		-0.028^{c}		$P_{k,t-1}^{negative}$		-0.121^{c}	
$SCI_{k,t-2}^{negative}$		0.018^{b}		$P_{k,t-2}^{negative}$		-0.046	
$SCI_{k,t-3}^{negative}$		0.018^{b}		$P_{k,t-3}^{negative}$		-0.005	
$SCI_{k,t-4}^{negative}$		-0.005		$P_{k,t-4}^{negative}$		0.027	
$SCI_{k,t-5}^{negative}$		-0.010		$P_{k,t-5}^{negative}$		0.024	
$RSCI_{k,t-1}$			0.184^{c}	$RP_{k,t-1}$			0.157
$RSCI_{k,t-2}$			-0.044	$RP_{k,t-2}$			0.031
$RSCI_{k,t-3}$			-0.102^{b}	$RP_{k,t-3}$			0.001
$RSCI_{k,t-4}$			-0.076^{b}	$RP_{k,t-4}$			-0.018
$RSCI_{k,t-5}$			-0.045	$RP_{k,t-5}$			-0.024
$R_{k,t-1}$	0.013	0.013	0.013	$R_{k,t-1}$	0.013	0.008	0.010
$R_{k,t-2}$	-0.041^{c}	-0.042^{c}	-0.041^{c}	$R_{k,t-2}$	-0.041^{c}	-0.046^{c}	-0.044^{c}
$R_{k,t-3}$	-0.020	-0.022	-0.020	$R_{k,t-3}$	-0.020	-0.024	-0.022
$R_{k,t-4}$	-0.020	-0.021	-0.020	$R_{k,t-4}$	-0.018	-0.024	-0.021
$R_{k,t-5}$	-0.035^{c}	-0.036^{c}	-0.035^{c}	$R_{k,t-5}$	-0.034^{c}	-0.038^{c}	-0.036^{c}
$Volm_{k,t-1}$	0.028	0.029	0.028	$Volm_{k,t-1}$	0.028	0.032	0.032
$Volm_{k,t-2}$	0.036	0.037	0.036	$Volm_{k,t-2}$	0.035	0.039	0.038
$Volm_{k,t-3}$	0.023	0.024	0.023	$Volm_{k,t-3}$	0.022	0.024	0.024
$Volm_{k,t-4}$	-0.014	-0.014	-0.015	$Volm_{k,t-4}$	-0.016	-0.014	-0.014
$Volm_{k,t-5}$	0.078^{c}	0.078^{c}	0.077^{c}	$Volm_{k,t-5}$	0.077^{c}	0.079^{c}	0.079^{c}
Control Variables				Control Variables			
R^2	0.053	0.052	0.052	R^2	0.044	0.044	0.044

(Superscript c, b and a denote the coefficient is different to 0 under 99%, 95% and 90% significance level separately)

Table 7. Predictability of sentiment contagion on trading order

Panel A: Orders for all sample				Panel B: Orders <1,000,000 Shares			
Variables	$Order_t^{buy}$	$Order_t^{sell}$	$Order_t^{imba}$	Variables	$Order_t^{buy}$	$Order_t^{sell}$	$Order_t^{imba}$
$SCI_{k,t-1}^{positive}$	2118.04 [c]			$SCI_{k,t-1}^{positive}$	1733.20 [c]		
$SCI_{k,t-2}^{positive}$	867.64 [b]			$SCI_{k,t-2}^{positive}$	690.49 [b]		
$SCI_{k,t-3}^{positive}$	90.91			$SCI_{k,t-3}^{positive}$	78.54		
$SCI_{k,t-4}^{positive}$	660.09 [b]			$SCI_{k,t-4}^{positive}$	542.01 [b]		
$SCI_{k,t-5}^{positive}$	560.80			$SCI_{k,t-5}^{positive}$	418.37		
$SCI_{k,t-1}^{negative}$		1215.45 [c]		$SCI_{k,t-1}^{negative}$		923.46 [c]	
$SCI_{k,t-2}^{negative}$		1033.12 [c]		$SCI_{k,t-2}^{negative}$		845.29 [c]	
$SCI_{k,t-3}^{negative}$		605.24		$SCI_{k,t-3}^{negative}$		473.01	
$SCI_{k,t-4}^{negative}$		777.87 [b]		$SCI_{k,t-4}^{negative}$		606.47 [b]	
$SCI_{k,t-5}^{negative}$		531.33		$SCI_{k,t-5}^{negative}$		461.73	
$RSCI_{k,t-1}$			2494.34	$RSCI_{k,t-1}$			3337.23 [b]
$RSCI_{k,t-2}$			1486.92	$RSCI_{k,t-2}$			1397.49
$RSCI_{k,t-3}$			-5350.87 [c]	$RSCI_{k,t-3}$			-3966.37 [b]
$RSCI_{k,t-4}$			-531.99	$RSCI_{k,t-4}$			-244.16
$RSCI_{k,t-5}$			-2960.72	$RSCI_{k,t-5}$			-2227.92
$Order_{t-1}$	0.40 [c]	0.45 [c]	0.02	$Order_{t-1}$	0.41 [c]	0.45 [c]	0.02 [a]
$Order_{t-2}$	0.12 [c]	0.12 [c]	0.03 [b]	$Order_{t-2}$	0.12 [c]	0.12 [c]	0.03 [b]
$Order_{t-3}$	0.09 [c]	0.08 [c]	0.03 [b]	$Order_{t-3}$	0.09 [c]	0.08 [c]	0.03 [b]
$Order_{t-4}$	0.07 [c]	0.06 [c]	0.01	$Order_{t-4}$	0.07 [c]	0.06 [c]	0.01
$Order_{t-5}$	0.08 [c]	0.10 [c]	0.01	$Order_{t-5}$	0.08 [c]	0.10 [c]	0.01
$R_{k,t-1}$	-2248.22 [c]	1100.33	-2278.99 [c]	$R_{k,t-1}$	-1733.50 [c]	924.13 [a]	-1850.11 [c]
$R_{k,t-2}$	527.10	-868.70	948.16 [b]	$R_{k,t-2}$	419.58	-709.83	783.99 [b]
$R_{k,t-3}$	-996.34	-1193.74	502.38	$R_{k,t-3}$	-878.30	-1004.87 [a]	392.76
$R_{k,t-4}$	-51.42	115.57	0.36	$R_{k,t-4}$	-37.89	44.76	37.32
$R_{k,t-5}$	267.68	93.19	452.15	$R_{k,t-5}$	199.91	17.65	395.99
Control Variables				Control Variables			
R^2	0.48	0.48	0.07	R^2	0.49	0.49	0.07

(Superscript c, b and a denote the coefficient is different to 0 under 99%, 95% and 90% significance level separately)

Table 8. Relationship between sentiment contagion and volume

Panel A: Total volume				Panel B: Individual volume			
Variables	$Volm_{k,t}$			Variables	$IdVolm_{k,t}$		
$SCI^{positive}_{k,t-1}$	0.033 [c]			$SCI^{positive}_{k,t-1}$	0.030 [c]		
$SCI^{positive}_{k,t-2}$	0.012 [c]			$SCI^{positive}_{k,t-2}$	0.013 [c]		
$SCI^{positive}_{k,t-3}$	0.019 [c]			$SCI^{positive}_{k,t-3}$	0.015 [c]		
$SCI^{positive}_{k,t-4}$	0.015 [c]			$SCI^{positive}_{k,t-4}$	0.018 [c]		
$SCI^{positive}_{k,t-5}$	0.018			$SCI^{positive}_{k,t-5}$	0.021		
$SCI^{negative}_{k,t-1}$		0.027 [c]		$SCI^{negative}_{k,t-1}$		0.013 [b]	
$SCI^{negative}_{k,t-2}$		0.030 [c]		$SCI^{negative}_{k,t-2}$		0.023 [c]	
$SCI^{negative}_{k,t-3}$		0.026 [c]		$SCI^{negative}_{k,t-3}$		0.030 [c]	
$SCI^{negative}_{k,t-4}$		0.013 [b]		$SCI^{negative}_{k,t-4}$		0.015 [c]	
$SCI^{negative}_{k,t-5}$		0.024		$SCI^{negative}_{k,t-5}$		0.025	
$OSCI_{k,t-1}$			3.472 [c]	$OSCI_{k,t-1}$			5.338 [c]
$OSCI_{k,t-2}$			1.084	$OSCI_{k,t-2}$			1.884 [a]
$OSCI_{k,t-3}$			-1.521 [a]	$OSCI_{k,t-3}$			1.482
$OSCI_{k,t-4}$			-0.871	$OSCI_{k,t-4}$			1.607
$OSCI_{k,t-5}$			0.349	$OSCI_{k,t-5}$			0.625
$Volm_{k,t-1}$	0.273 [c]	0.274 [c]	0.274 [c]	$Volm_{k,t-1}$	0.265 [c]	0.264 [c]	0.265 [c]
$Volm_{k,t-2}$	0.117 [c]	0.118 [c]	0.119 [c]	$Volm_{k,t-2}$	0.116 [c]	0.117 [c]	0.116 [c]
$Volm_{k,t-3}$	0.058 [b]	0.059 [c]	0.060 [c]	$Volm_{k,t-3}$	0.057 [b]	0.057 [c]	0.058 [c]
$Volm_{k,t-4}$	0.044 [a]	0.044 [b]	0.045 [b]	$Volm_{k,t-4}$	0.042 [a]	0.043 [b]	0.043 [b]
$Volm_{k,t-5}$	0.015	0.016	0.017	$Volm_{k,t-5}$	0.013	0.014	0.014
$R_{k,t-1}$	0.023	0.022	0.023	$R_{k,t-1}$	0.029	0.030	0.030
$R_{k,t-2}$	0.013	0.012	0.012	$R_{k,t-2}$	0.009	0.010	0.009
$R_{k,t-3}$	-0.003	-0.004	-0.004	$R_{k,t-3}$	0.005	0.006	0.006
$R_{k,t-4}$	0.008	0.007	0.007	$R_{k,t-4}$	0.006	0.007	0.007
$R_{k,t-5}$	0.001	0.001	0.001	$R_{k,t-5}$	0.004	0.005	0.005
Control Variables				Control Variables			
R^2	0.305	0.304	0.304	R^2	0.306	0.305	0.304

(Superscript c, b and a denote the coefficient is different to 0 under 99%, 95% and 90% significance level separately)

Table 9. Regression of the prediction on sentiment contagion

Panel A: Regression with returns				Panel B: Regression with orders			
Variables	$SCI_{k,t}^{positive}$	$SCI_{k,t}^{negative}$	$RSCI_{k,t}$	Variables	$SCI_{k,t}^{positive}$	$SCI_{k,t}^{negative}$	$RSCI_{k,t}$
$R_{k,t-1}$	0.017^{c}	0.002^{b}	0.002^{c}	$Order_{t-1}$	$2.2\text{E-}06^{c}$	$1.5\text{E-}06^{c}$	8.4E-09
$R_{k,t-2}$	0.013^{c}	0.006^{c}	0.001^{c}	$Order_{t-2}$	$1.2\text{E-}06^{c}$	$4.5\text{E-}07^{c}$	-7.4E-09
$R_{k,t-3}$	0.010^{c}	0.005^{c}	0.001^{c}	$Order_{t-3}$	$8.0\text{E-}07^{c}$	$3.8\text{E-}07^{b}$	2.6E-08
$R_{k,t-4}$	0.003	0.001	0.000^{b}	$Order_{t-4}$	$4.1\text{E-}07^{b}$	$2.9\text{E-}07^{a}$	-2.0E-08
$R_{k,t-5}$	0.001	0.004	0.000	$Order_{t-5}$	3.3E-07	4.7E-07	3.4E-08
$SCI_{k,t-1}^{positive}$	0.121^{c}			$SCI_{k,t-1}^{positive}$	0.110^{c}		
$SCI_{k,t-2}^{positive}$	0.094^{c}			$SCI_{k,t-2}^{positive}$	0.085^{c}		
$SCI_{k,t-3}^{positive}$	0.079^{c}			$SCI_{k,t-3}^{positive}$	0.071^{c}		
$SCI_{k,t-4}^{positive}$	0.080^{c}			$SCI_{k,t-4}^{positive}$	0.072^{c}		
$SCI_{k,t-5}^{positive}$	0.074^{c}			$SCI_{k,t-5}^{positive}$	0.067^{c}		
$SCI_{k,t-1}^{negative}$		0.136^{c}		$SCI_{k,t-1}^{negative}$		0.127^{c}	
$SCI_{k,t-2}^{negative}$		0.095^{c}		$SCI_{k,t-2}^{negative}$		0.089^{c}	
$SCI_{k,t-3}^{negative}$		0.082^{c}		$SCI_{k,t-3}^{negative}$		0.076^{c}	
$SCI_{k,t-4}^{negative}$		0.071^{c}		$SCI_{k,t-4}^{negative}$		0.067^{c}	
$SCI_{k,t-5}^{negative}$		0.069^{c}		$SCI_{k,t-5}^{negative}$		0.065^{c}	
$RSCI_{k,t-1}$			0.066^{c}	$RSCI_{k,t-1}$			0.065
$RSCI_{k,t-2}$			0.051^{c}	$RSCI_{k,t-2}$			0.051
$RSCI_{k,t-3}$			0.042^{c}	$RSCI_{k,t-3}$			0.042
$RSCI_{k,t-4}$			0.042^{c}	$RSCI_{k,t-4}$			0.042
$RSCI_{k,t-5}$			0.038^{c}	$RSCI_{k,t-5}$			0.037
$V_{k,t-1}$	-7.737	-4.992	-0.322	$V_{k,t-1}$	-23.017^{b}	-7.483	-0.355
$V_{k,t-2}$	3.324	-1.925	0.449	$V_{k,t-2}$	5.406	-2.332	0.412
$V_{k,t-3}$	-1.035	1.141	-0.186	$V_{k,t-3}$	2.447	2.101	-0.181
$V_{k,t-4}$	1.143	-0.860	0.113	$V_{k,t-4}$	0.033	-2.556	0.101
$V_{k,t-5}$	7.131	4.732	0.203	$V_{k,t-5}$	10.637	7.824	0.270
Control Variables				Control Variables			
R^2	0.129	0.143	0.060	R^2	0.140	0.151	0.064

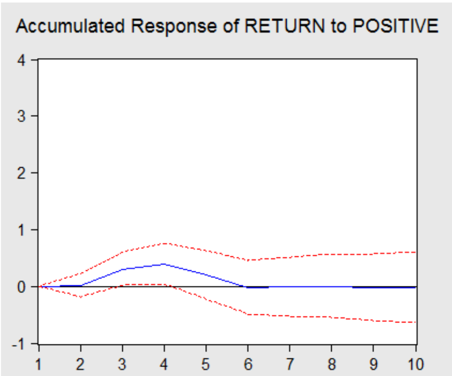
(Superscript c, b and a denote the coefficient is different to 0 under 99%, 95% and 90% significance level separately)

Table 10. Performance of portfolios based on sentiment contagion (Daily)

BL = 22 trading days, FL = 9 trading days, HL = 9 trading days							
		Min	Small	Median	Big	Max	Max-Min
$RSCI$	Alpha	-0.5368^c	-0.4168^c	-0.0702	-0.0595	0.0575	0.6190^c
		[7.5353]	[7.6479]	[1.8481]	[1.1776]	[0.9640]	[7.9415]
	R^2	0.8806	0.8936	0.9375	0.9076	0.9099	0.0942
$SCI^{positive}$	Alpha	-0.2811^c	-0.0605	-0.1672^c	0.0928	-0.5509^c	-0.1624
		[4.9227]	[1.2055]	[3.6959]	[1.8054]	[7.9452]	[1.5627]
	R^2	0.9077	0.9153	0.9228	0.9062	0.8690	0.0462
$SCI^{negative}$	Alpha	-0.0783	-0.2328^c	-0.0236	-0.0106	-0.7247^c	-0.6253^c
		[1.3733]	[4.7822]	[0.5459]	[0.2160]	[11.3649]	[7.6479]
	R^2	0.9081	0.9156	0.9228	0.9150	0.8857	0.0821

(Superscript c, b and a denote the coefficient is different to 0 under 99%, 95% and 90% significance level separately)

Figure 3. Statistics of cumulative revision of returns at 10 lags



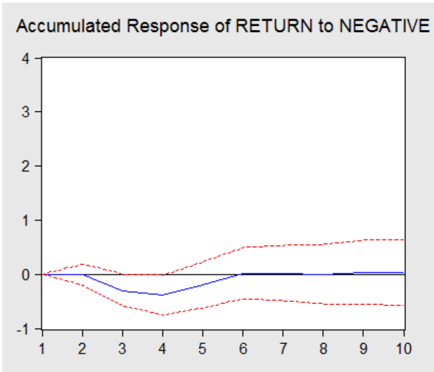


Table 11. Summary statistics of cumulative revision of returns at 10 lags

	Positive Sentiment Contagion	Positive Sentiment Contagion
Mean	0.0000	0.0000
Std	0.0100	0.0100
Max	0.0303	0.0299
Min	-0.0362	-0.0298
Kurtosis	3.6008	3.3959
Skewness	0.0268	0.0121
T-Stat	0.398	0.492

Table 12. Regressions on trading orders with more classification

Panel A: Orders >500,000 CNY				Panel B: Orders <50,000 CNY			
Variables	$Order_t^{buy}$	$Order_t^{sell}$	$Order_t^{imba}$	Variables	$Order_t^{buy}$	$Order_t^{sell}$	$Order_t^{imba}$
$SCI_{k,t-1}^{positive}$	416.74			$SCI_{k,t-1}^{positive}$	870.39 ^{c}		
$SCI_{k,t-2}^{positive}$	458.17			$SCI_{k,t-2}^{positive}$	695.23 ^{b}		
$SCI_{k,t-3}^{positive}$	464.21			$SCI_{k,t-3}^{positive}$	668.50 ^{b}		
$SCI_{k,t-4}^{positive}$	399.51			$SCI_{k,t-4}^{positive}$	604.61 ^{a}		
$SCI_{k,t-5}^{positive}$	527.76 ^{b}			$SCI_{k,t-5}^{positive}$	533.88 ^{a}		
$SCI_{k,t-1}^{negative}$		415.59		$SCI_{k,t-1}^{negative}$		2287.47 ^{c}	
$SCI_{k,t-2}^{negative}$		408.12		$SCI_{k,t-2}^{negative}$		1660.84 ^{a}	
$SCI_{k,t-3}^{negative}$		949.64 ^{c}		$SCI_{k,t-3}^{negative}$		-548.29	
$SCI_{k,t-4}^{negative}$		601.34		$SCI_{k,t-4}^{negative}$		677.88	
$SCI_{k,t-5}^{negative}$		602.43 ^{c}		$SCI_{k,t-5}^{negative}$		1444.87 ^{a}	
$RSCI_{k,t-1}$			-837.33	$RSCI_{k,t-1}$			2042.24 ^{b}
$RSCI_{k,t-2}$			-251.58	$RSCI_{k,t-2}$			2747.94 ^{b}
$RSCI_{k,t-3}$			837.29	$RSCI_{k,t-3}$			-3737.52 ^{c}
$RSCI_{k,t-4}$			566.38	$RSCI_{k,t-4}$			-3524.05 ^{c}
$RSCI_{k,t-5}$			931.12	$RSCI_{k,t-5}$			3347.42 ^{c}
$Order_{t-1}$	0.34 ^{c}	0.37 ^{c}	0.11 ^{c}	$Order_{t-1}$	0.45 ^{c}	0.52 ^{c}	0.19 ^{c}
$Order_{t-2}$	0.12 ^{b}	0.06 ^{a}	0.07 ^{a}	$Order_{t-2}$	0.16 ^{c}	0.13 ^{c}	0.18 ^{c}
$Order_{t-3}$	0.08 ^{a}	0.09 ^{a}	0.11 ^{b}	$Order_{t-3}$	0.10 ^{c}	0.09 ^{c}	0.14 ^{c}
$Order_{t-4}$	0.06 ^{a}	0.08 ^{a}	0.02	$Order_{t-4}$	0.11 ^{c}	0.08 ^{c}	0.13 ^{c}
$Order_{t-5}$	0.00	0.01	0.01	$Order_{t-5}$	0.10 ^{c}	0.11 ^{c}	0.05 ^{c}
$R_{k,t-1}$	1189.07 ^{c}	2777.59 ^{c}	2118.21 ^{c}	$R_{k,t-1}$	2546.73 ^{c}	3809.70 ^{c}	5659.85 ^{c}
$R_{k,t-2}$	-57.19	35.52	411.67	$R_{k,t-2}$	-73.08	269.21 ^{b}	2694.93 ^{c}
$R_{k,t-3}$	3.21	85.28	922.69	$R_{k,t-3}$	1.12	148.69	2316.55 ^{c}
$R_{k,t-4}$	-55.03	23.99	418.21	$R_{k,t-4}$	-364.18 ^{b}	19.52	1922.01 ^{c}
$R_{k,t-5}$	-198.47	-96.33	339.55	$R_{k,t-5}$	-961.05 ^{c}	-557.93 ^{c}	1858.56 ^{c}
Control Variables				Control Variables			
R^2	0.1532	0.1612	0.0392	R^2	0.4532	0.4612	0.1389

(Superscript c, b and a denote the coefficient is different to 0 under 99%, 95% and 90% significance level separately)

Table 13. Performance of portfolios based on sentiment contagion (Weekly)

BL = 16 week, FL = 4 week, HL = 4 week							
		Min	Small	Median	Big	Max	Max-Min
$RSCI$	Alpha	-0.9763[c]	-0.7939[c]	-0.2611[c]	-0.1723[a]	-0.0954	1.0674[c]
		[8.9407]	[9.9304]	[4.8778]	[2.2222]	[1.1408]	[9.3255]
	R^2	0.8859	0.9027	0.9461	0.9068	0.9247	0.1127
$SCI^{positive}$	Alpha	-0.3651[c]	-0.3262[c]	-0.3014[c]	0.1988[b]	-1.0916[c]	-0.0781
		[4.2846]	[4.3813]	[4.5673]	[2.8083]	[11.0330]	[0.5173]
	R^2	0.9056	0.9238	0.9316	0.9206	0.8910	0.0473
$SCI^{negative}$	Alpha	-0.0553	-0.5104[c]	-0.1783[b]	-0.0781	-1.2997[c]	-1.1876[c]
		[0.7286]	[6.6357]	[2.7436]	[1.0126]	[14.6936]	[10.5716]
	R^2	0.9289	0.9062	0.9266	0.9162	0.9036	0.1068

(Superscript c, b and a denote the coefficient is different to 0 under 99%, 95% and 90% significance level separately)