THE ASYMMETRIC STOCK MESSAGE BOARD EFFECT

[1]XING LU, [2]YING ZHANG

[1,2]Indiana University South Bend, Fairfield University
E-mail: [1]Lu35@iusb.edu, [2]yzhang1@fairfield.edu

Abstract – As the first research to address the asymmetric effect of message board discussion, this paper has three key findings. First, message board discussion has a significant and positive impact on small stocks' abnormal returns, regardless of news arrival. More interestingly, this asymmetric effect becomes stronger with small stocks on no news days. For large stocks, there is a small abnormal return on news day and no abnormal return on no news day. Finally, online talks help predict the presence of material news regardless of firm size, although a stronger predictive power on news occurrence is found for small stocks. Our results are robust and suggesting online talks' impact on stock market is significant and asymmetric.

Index Terms— Stock Message Board, Online Trading, Investor Sentiment, Computational Linguistics.

I. INTRODUCTION

Online message board has been one of the most dominant venues for investors to share and discuss their thoughts and opinions on the stock market and investment, and thus has a considerable impact in helping investors make decisions on their own investment. Barber and Odean (2001) argue that the wide availability of rich information is perhaps the biggest benefit that the Internet can offer for investors. Therefore, online message boards have profound influence on investors' decision-making in general. It has become obvious that stock message boards have changed and will continue to impact investors' trading decisions (Litan and Rivlin, 2001). Since most investors are trading on Internet today, they also tend to seek information online. Among free online information sources, stock message boards are perfect for retail investors to learn how other people think about specific stocks and the market as a whole (Zhang et al., 2012).

Due to stock message boards' unique benefit of fast information dissemination among individual investors, not only the number of stock message boards has significantly increased, user participation on these boards has also become much more active (Loh and Ong, 1998; Teich, et al. 1999; and Hong, 2000). It should be of both practitioners' and researchers' interest to study the stock message boards and better understand investors' behaviors as well as the consequent impacts on the market. Meanwhile, policymakers are always interested in tracking the activities and trends on the stock message boards in order to prevent securities fraud and to protect investors' interests.

To practitioners, participating in online discussions and receiving feedbacks from others may improve their trading skills (Lu and Mizrach 2011). Learning from other experienced investors and tracking their performance is also beneficial to inexperienced investors. Breaking financial market news and stock related headlines are often heavily discussed and analyzed by forum participants on the message boards. Additionally, rumors are also circulated on boards before confirmed news. Van Bommel (2003) and Spiegel, et al. (2010). Anterweiler and Frank (2004) show that online talk is not just noise but exerts substantial influences on security trading activities. Therefore, trading strategies based on stock message board information are logically and practically feasible. For example, Gu, et al. (2006) suggest a trading strategy that involves buying stocks with low sentiments while selling stocks with high sentiments.

To researchers, studying stock message board activities helps us understand whether messages posted online contain valuable information not available in other traditional media. Although more recent studies largely agree upon online talk is not just noise (see Antweiler and Frank 2004 among others), researchers hold contrary arguments and have inconsistent findings on whether message board activities can predict future stock returns.

Tumarkin and Whitelaw (2001) study Internet service stocks and find that on days with abnormally high message posting activity, changes in online investor mood are correlated with contemporaneous abnormal industry-adjusted returns and trading volume but do not predict future returns and volume. Similar findings are reported by Tumarkin (2002). Antweiler and Frank (2004) report that a positive shock on message board predicts negative returns on the next day but the effect is not economically significant. Das et al. (2005) report that sentiment extracted from Motley Fool message board does not apparently predict returns. Koski et al. (2008) suggest that stock message board information increases stock return volatility but does not contribute to predicting stock returns in any fundamental way.

Recent scholars who document significant message board effect tend to adopt message-board-generated sample to avoid self-selection bias and most of these samples contain small cap stocks. Gu et al. (2006) argue that there are informed posters on stock message boards whose information is not fully incorporated into market prices and therefore message board sentiments can predict future stock returns. Sabherwal et al. (2011) study 64 heavily discussed fringe stocks and report that when there is no market news, online traders' sentiment causes positive contemporaneous return and negatively predicts the next day return. In addition, both Clarkson et al. (2006) and Bettman et al. (2011) report that online "rumors" significantly predict abnormal returns and trading volumes. Oh and Sheng (2011) focus on top 10 discussed stocks and reveal that micro-blog sentiments have predictive power on simple and market-adjusted stock returns.

In this study, we define the stock message board effect as whether information content in online stock message board can influence contemporaneous and future stock trading activities. We extend previous literature and discern the true effect of posting activities on trading behaviors by controlling for firm size (small vs. large) and news appearance (with vs. without). Due to data limitation and/or imperfect model design, prior studies had no success in splitting their sample by firm size and/or news arrival. To fill this void, we conduct a comprehensive study to examine how posting activities measured by sentiment, disagreement, number of messages, and average message length influenced stock return, volatility, volume, and spread. In addition, to avoid aforementioned self-selection bias, we focus on message-board-generated sample stocks that have been actively discussed by online posters. Das et al. (2005) suggest that there is a close relationship between message board sentiment and news sentiment. Antweiler and Frank (2004) also show unusual message board postings predict the presence of news stories up to two days later. Therefore, we also investigate the difference in stock message board effect due to the presence of market news.

We document three important findings. First, event study reveals that there is a significant and positive market reaction on the heavy online discussion days (event days) for small stocks regardless of news arrival. Investors' reaction was skewed to optimistic sentiment in general. For large stocks, as contrast, a smaller (no) market reaction is reported with (without) news arrival on the discussion day. Second, with lead-lag robust regressions, we find that when news arrived, posting activities affected trading for all stocks, especially on the discussion days, with a larger impact on small stocks. Without news, however, online talks only influenced security trading for small stocks but had no impact on large stocks indicating an asymmetric stock message board effect. Finally, we show that posting activities helped predict the presence of news stories one or two days later with a stronger news predictive power for small stocks. We conduct ten robustness tests and our results hold. We conclude that online talks generate significant impacts on stock trading exploiting an information asymmetry.

The remainder of this study is organized as follows. Section 2 reviews the data. The methodologies are provided in Section 3. Section 4 presents empirical results and Section 5 offers discussion on robustness tests. The paper is concluded in Section 6.

II. DATA

Our data is downloaded from WallStreetPit message board, the largest online chat room managed through Thelion.com, using a web crawler (Visual Basic software) written by the authors. This data source has been widely cited in the literature, such as Leinweber (2001), Depken and Zhang (2010), Sabherwal, et al. (2008, 2011), Zhang, et al. (2012), and Klein, et al. (2012) among others. Different from other online message boards, such as Yahoo! Finance and Raging Bull, WallStreetPit is a chat-room-like message trading board which lists messages reverse chronologically with the latest post on the top. In a different structure, Yahoo! Finance and Raging Bull, commonly named "Stock-specific board", list chat messages under each stock's separate board. As a consequence, one is required to input the stock's symbol as the first step on the Yahoo! Finance or Raging Bull to access this stock's independent message board. Thus, to read a stock's messages on Yahoo! Finance, one must know a-priori the existence of a specific stock which gives rise to self-selected samples, used in Antweiler and Frank (2004) and Das et al. (2005) for examples. In fact, online traders who search for timely hot stocks do not necessarily pre-know the existence of a particular stock, see Sabherwal et al. (2011) and Zhang et al. (2012).

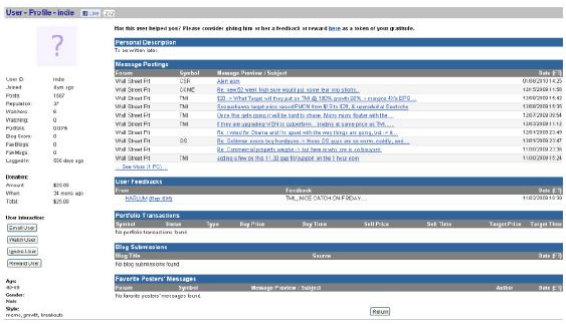


Figure 1 Screenshot of Messages

Figure 1 displays the contemporary structure of WallStreetPit message board. From left to right, there

are seven columns which are message number, stock symbol, message preview, sentiment, number of recommendations, poster username, and posting time. All messages are listed reverse chronologically according to their posting date and time shown in the last "Date (ET)" column. The latest message is placed on the top of the screen. Each message is assigned a message number shown in the "Msg #" column. As of July 2, 2013, WallStreetPit has accumulated over 2.3 million posts. If a message is referring to a stock, the ticker symbol inputted by the poster is displayed in the "Symbol" column. It is possible that a message has no referring stock so that its "Symbol" is empty (see Msg # 1296049 for an example). The first several words (80 characters) of the message are presented with a hyperlink in the "Message Preview" column so readers can grasp its keystone at first glance without further click into its separate page. Of course, a user can click on the message preview hyperlink to further read the complete content of the message. When writing a new message, user can choose one of the seven sentiments (Strong Buy, Buy, Hold, Scalp, Sell, Strong Sell, and Short) for the "Sentiment" column. Thelion.com allows peer reviewing using "Rec," which indicates how many other users have recommended a message. A "-" will be given in the absence of any endorsement.

The advantage of using WallStreetPit is threefold: First, one can directly discover a stock in a conversation chain shown in Figure 1 without prior knowledge of the stock. Therefore, WallStreetPit makes it a perfect venue to study the message board effect without using self-selected sample. Message-board-generated sample is based on the level of actual online discussion by all posters. Second, WallStreetPit is moderated by the administrators which tremendously reduce noises, Delort, et al. (2012). The message board's administrators use the moderation process to minimize spam, advertisement, profanity, defamatory, empty posts, etc. Third, WallStreetPit provides a reputation system with which readers can influence a poster's reputation in a pecuniary fashion. This reputation system can prevent posters from registering multiple screen names and reduce adverse selection problem documented by Depken and Zhang (2010). With a pecuniary reputation system, WallStreetPit might reduce stock manipulation and increase the incentives to offer accurate information (Gregg and Scott, 2006 and Depken and Zhang, 2010), thereby alleviating the identification problem stated in Vilardo (2004). With such pecuniary reputation system, WallStreetPit displays each online poster with a credit score to reflect his/her reputation in this financial community. Note that each poster's credit score has been constantly updated on a regular basis. Importantly, our software downloaded each poster's credit score at the time when the message was posted.

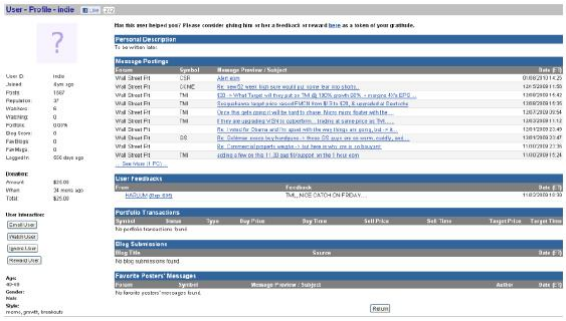


Figure 2 Screenshot of a Poster's Profile

For example, Figure 2 shows the user profile of "indie". On the left-bottom corner, we can see that "indie" is a middle age man whose trading style includes momentum, growth, and price breakouts. We further learn that "indie" joined the WallStreetPit community four years ago by the time we collected his information and has posted 1,687 messages with a 37 total credit score proxy for his reputation. "indie" was attracting 6 watchers, but put no other user in his "watching" list. Although "indie" became a member for four years, his last visit was 556 days ago by the time we collected this information. On the right-hand side of the profile, we find the latest messages posted by "indie". Followings are five steps that we purify our data for statistical analyses.

STEP 1—collect raw data: Our sample covers a six-year period, the longest in the literature, from January 1, 2006 to December 31, 2011 (message number from #880210 to #2221264) totaling 1,058,366 valid messages on WallStreetPit. These messages are downloaded daily through our web crawler. We consider all valid messages posted from 4:00:01 p.m. ET from the previous trading day t-1 to 4:00:00 p.m. ET on the trading day t as the number of messages for day t. Each message's content, including the message number, message length, posting time, poster ID, whether it is an original posting or subsequent reply, the total real-time credit score attributed to the poster, the length of poster's membership, the total number of postings by the poster, number of followers, and what stock the message referred to, are saved in a database while the body of the message is saved in a separate plain text file which is used as either training or testing data sample for the computational linguistics method—text classifier. Among these raw messages, only 16.26% are with a self-disclosed sentiment, which implies a majority of messages posted on the forum are with no explicit sentiment. Finally, 33.01% of all messages are replies and 66.99% are original postings.

STEP 2—use meaningful data: Tumarkin and Whitelaw (2001) and Tumarkin (2002) suggest that only stocks with a sufficient number of messages are meaningful and should be included in a message board study because significant noises and errors are introduced by including stocks with a small bulletin board following. Thus, to ensure meaningful and sufficient posting activity, we impose the following six criteria:

A. Following Sabherwal et al. (2011), each stock must have at least 30 messages (either first-post or subsequent reply messages) during the event day;
B. Each stock must have at least five different posters (alias) in order to prevent hyping by one or two dominant posters;
C. Minimum average trading price (average of opening and closing prices) is set to $1 to avoid excess transaction costs;
D. We remove all the over-the-counter bulletin board (OB) or pink sheet (PK) stocks;
E. Funds such as QQQQ, GLD, USO, and SLV are excluded;
F. Some stocks occur more than once in our sample. Following Sabherwal et al. (2011), we consider those multiple occurrences as isolated events if there is a gap of at least two trading days between any two occurrences. Otherwise, only the first appearance for a stock with consecutive appearances is counted. For instance, if t0 is the first appearance, subsequent appearances in day t+1 and t+2 do not count as events. This criteria implies that we require at least two quite days (with no news) prior to any event day. If t+3 and t+4 do not have any news, then news event on t+5 reappearance is considered as a new event.

STEP 3—News control: To identify market material news, we follow Demers and Lewellen (2003) and Antweiler and Frank (2004, 2006) and use Lexis-Nexis Academic to collect media news related to a specific company from "All News (English)" source which covers major news providers, such as the Wall Street Journal, the New York Times, the Washington Post, the Thomson Reuters One, the PR Newswire, and many others. In accordance with Sabherwal et al. (2011), we define material news as corporate fundamentals-related news which includes news such as unusual corporate events (e.g. stock splits, mergers, divestitures, significant changes in management, downsizing, recognition of key officials, FDA approvals, material news regarding competitors and/or industry, introduction of new product or service, and competitor company's material news, etc.), earnings announcements, analyst recommendations, etc. As contrast, technical news such as daily winner/loser alerts, moving average alerts, and abnormal trading volume alerts is not considered as substantial news. We impose our seventh criterion on our sample: if at least one piece of substantial material news is recorded on event day t0 and no other material news is announced on day t-2 and t-1, our news dummy variable is set to be 1, 0 if there is no news over three days from t-2 to t0. Although there are only 28 events that had consecutive subsequent material news after t0, we don't restrict news occurrence on subsequent day t+1 to t+2. Our final sample consists of 1558 events which coincide with all aforementioned seven criteria, accounted for 1216 distinct stocks. Among the 1558 events, 952 are with news while 606 are not.

STEP 4—determine firm size: We obtain daily trading data from the Center for Research in Security Prices (CRSP). To determine firm size, we match the market value of each stock on the discussion (event) day t0 with the average firm cutoff size in the same month from Professor Kenneth R. French's data source. In particular, any stock whose market value is smaller than Lo 30 cutoff number in the same month is considered as a small cap stock. The rest of stocks in Med 40 and Hi 30 are considered as mid-large stocks. Among the 1216 distinct stocks 702 are small stocks, 159 are large stocks, and 355 are mid cap stocks, Our sample is tilted to small stocks so that we combine mid and large stocks (159+355=514) as a comparable counterpart group named large stocks. On the one hand, among the 952 events with news, small (large) cap stocks occupied 635 (317) events. On the other hand, among the 606 events without news, small (large) cap stocks account for 471 (135) events. It is clear that there were more small-stock-related events in either with or without news groups.

STEP 5—intraday trading data: We obtain intraday trading data from the Trade and Quote database (TAQ) database. We use the standard procedures to filter the TAQ data. Specifically, we exclude the following data on any trading day t: trades and quotes outside of the regular trading hours (9:30:00 a.m. to 4:00:00 p.m. ET): opening and closing trades and quotes; regional trades and quotes; trades marked as irregular; negative bid or ask quotes; and spreads that are negative or in excess of 40% of the bid-ask midpoint. For each trading day, stock messages and news are collected from 4:00:01 p.m. ET on the previous day to 4:00:00 p.m. ET today. Intraday trading activities trades and quotes are from 9:30:00 a.m. ET to 4:00:00 p.m.ET on each trading day.

III. METHODOLOGY

3.1. Event Study

We employ a standard event study methodology to measure the impact online talks have on stock return and volume separated by size and news. For estimating the market model, we choose a 255-day

pre-event estimation window starting on t-300, if possible, and ending on t-46. We require three days as the minimum length of the estimation window. The abnormal return for stock j on day t is computed as:

$$A_{jt} = R_{jt} - (\hat{\alpha}_j + \hat{\beta}_j R_{mt}) \quad (1)$$

where the coefficients and are Scholes-Williams (1977) estimates of and , is the return on stock j on day t and is the return on an value-weighted CRSP index, with dividend reinvested, on day t. The average abnormal return (AAR) on day t is the sample mean:

$$AAR_t = \frac{\sum_{j}^{N} A_{jt}}{N} \quad (2)$$

where N is the number of sample stocks on event day t. Furthermore, the cumulative average abnormal return (CAAR) on a portfolio starting from day t1and ending on day t2 is as follows:

$$CAAR_{t_1,t_2} = \frac{1}{N}\sum_{j=1}^{N}\sum_{t_1}^{t_2} A_{jt} \quad (3)$$

We conduct the Patell (1976) Z-test to examine the null hypothesis that the average abnormal stock return equals zero assuming asymptotically unit normal. In addition, abnormal trading volume is often used to assess the information content. Average abnormal volume (AAV) is computed in a similar manner as shown above, using the daily NYSE-AMEX log-transformed value-weighted volume as the benchmark. All volumes are also log-transformed for individual stocks.

3.2. Computational Linguistics

The fact that online posters may or may not voluntarily disclose sentiment means that not all messages include a self-disclosed sentiment. Only 16.26% of our sample messages have self-disclosed sentiments. This is unfortunate because the sentiment index should exhibit minimal bias and ideally should include all messages, both with and without self-disclosed sentiment, Antweiler and Frank (2004). To handle non-self-disclosed sentiment messages and create an effective aggregate sentiment index in the message board study, researchers often rely on a text classifier. In finance-related literature, Antweiler and Frank (2004) adopt the Naïve Bayesian (NB, hereafter) classifier and report similar results when using a Support Vector Machine (SVM, hereafter) classifier. Takahashi, et al. (2007) also use the NB classifier to analyze the relation between stock price returns and market headline news in text format. Similar work is conducted by Antweiler and Frank (2006) and Schumaker and Chen (2006, 2008) who apply both NB and SVM text classifiers to examine stock markets' reaction to financial news. Bettman et al. (2011) use both NB and SVM for testing message board takeover rumors on the stock market while Sabherwal et al. (2011) employ NB to study online talks with no fundamental news. Sprenger and Welpe (2012) use NB to study the information content of stock on tweeter micro-blog. Thomas and Sycara (2002) and Zhang and Swanson (2010) apply Maximum entropy (ME, hereafter) in their stock message board studies while Das and Chen (2003, 2007) and Gu, et al. (2007) use their own text classifiers.

Although many researchers prefer the NB classifier, it is based on a probability model that incorporates strong independence assumptions that often have no bearing in reality, see Joachimes (1997) and Antweiler and Frank (2004) for details of NB. Hence it is naive and problematic (Friedman 1997). Such a severe assumption often generates systemic problems that reduce the accuracy of the probability estimates (Rennie et al. 2003). Recent research has found ME, which makes no conditional independence assumptions, to be a promising technique for text classification. Particularly, Nigam et al. (1999) find that ME is competitive with, and two out of three times better than, NB on experimental results. The superiority of ME over other text classifiers is widely documented in Ratnaparkhi (1998), Raychaudhuri, et al. (2002), Macskassy, et al. (2003) and Li, et al. (2005) among others.

In fact, the choice of the text classification algorithm is a somewhat ad hoc decision and must be resolved first. In nonfinancial fields, different classifiers have been found to be effective on different features of data. However, stock message board applications are unique. Unlike articles in newspapers or reports from analysts, stock messages are style free, short, elliptical, and in a dialogue-like format (Admati and Pfleiderer 2001), that is, in a form of fragmentary utterances that are a common phenomenon in spoken language, Sabherwal et al. (2011) and Zhang et al. (2012). Many studies, such as Berger et al. (1996), Choi et al. (1999) and Pettibone and Pon-Barry (2003), report that ME has consistently exhibited high performance for conversational discourse messages compared with any other examined classifiers. Since forum messages exist in the form of fragmentary utterances which is a common phenomenon in spoken language, ME should be a promising classifier with a high degree of accuracy potential. Based on empirical results, Bender et al. (2003) and Fernandez et al. (2005) support this view. In the finance literature, using stock message board data, Zhang et al. (2012) conclude that ME is a superior classifier among other popular classifiers including NB and SVM. All in all, we adopt ME as our exclusive text classifier. For more details of ME algorithm, see Berger et al. (1996), Lau

et al. (1993), Rosenfeld (1996), Ratnaparkhi (1997, 1998) and Zhang et al. (2012).

Following Antweiler and Frank (2004, 2006) and Zhang et al. (2012) among others, we use software called RAINBOW (McCallum, 1996) to build our ME text classifier based on training datasets (messages with pre-known self-reported sentiments) and then use it to evaluate messages without a self-reported sentiment. To reduce human bias, we choose only self-reported sentiment messages as training dataset, see Zhang et al. (2012). When building the ME text classifier, we restrict each of our training datasets to exactly 1,000 messages per class in order to avoid the class imbalance problem. We do not include reply messages in the training datasets in order to keep the training datasets as less noisy as possible, Zhang et al. (2012). ME classifier's average accuracy is presented in Table 1.

Table 1 Classifier Average Accuracy and Comparison

Panel A. An Example of Confusion Matrix

	Evaluated Optimistic	Evaluated Neutral	Evaluated Pessimistic	Classifier Accuracy
Actual 1000 Optimistic	**800**	100	100	80.00%
Actual 1000 Neutral	50	**900**	50	90.00%
Actual 1000 Pessimistic	100	200	**700**	70.00%

Panel B. Comparison of Average Accuracy

	In-Sample Test	Equal Out-of-Sample Test
Average Accuracy	73.11%***	67.28%***
Standard Error	0.88	0.81

Notes: Panel A exhibits an example of the confusion matrix on adjusted/combined sentiment types. When evaluating classifier accuracy, we do not further separate Buy from Strong Buy because they belong to the same directional sentiment. Thus, sum of the numbers across Buy and Strong Buy represents optimistic sentiment, sum of the numbers across Sell and Strong Sell represents pessimistic sentiment, and sum of the Hold numbers represents Neutral sentiment. As training data, we randomly select a sample of 1000 messages with self-disclosed sentiment per sentiment class. Panel B reports the average accuracy with standard error per classifier and the Chi-square statistics between ME and the random classifier. For the In-Sample Test, we repeat *RAINBOW* thirty times and report the average accuracy with *RAINBOW* standard error. For the Equal Out-of-Sample test for each text classifier, we randomly select 1000 messages in each class, repeat the evaluation thirty times and report the calculated standard errors.

Panel A first demonstrates an example of confusion matrix proposed by Kohavi and Provost (1998) for calculating the classifier's accuracy. Because each training message comes with a self-disclosed sentiment by the poster, we are able to determine its true sentiment. The confusion matrix is a tableau that presents a cross-classification of true sentiment versus evaluated sentiment type. The confusion matrix shown in Panel A has three rows and three columns. Each of the rows shows the pooled sentiment (Optimistic=Strong Buy + Buy, Neutral=Hold + Scalp, or Pessimistic=Short + Strong Sell + Sell) self-disclosed by the poster. WallStreetPit provides "Scalp" which is classified as a neutral sentiment in the same way as "Hold". It is shown that confusion matrix accuracy measure is based on pooled sentiments which share the same directional opinion (for instance, Strong buy and Buy both signal optimistic tone). The columns detail how many of these messages are classified by the ME classifier in each of three pooled sentiment categories: Optimistic, Neutral, or Pessimistic. The greater the weight of the diagonal of the confusion matrix, the more accurate of the ME classifier. Panel A shows an example of 80% accuracy.

We compare ME classifier accuracy with a random classifier which implies no classification ability. Because there are three pooled sentiment classes (optimistic, neutral, and pessimistic), the random classifier would result in a 33.33% accuracy. We use a standard Chi-square x2 test to compare ME with a random classifier as follow, (Das and Chen, 2007):

$$x^2(4) = \frac{1}{9}\sum_{i=1}^{9}\frac{(MaxEntropy - Random)^2}{Random}$$

with degrees of freedom = 4 (4)

where MaxEntropy represents the superior Maximum Entropy classifier accuracy in percentage, and Random represents the inferior random classifier with 33% accuracy. If ME is significantly better than the random classifier, the superior model has statistically higher classification ability. Panel B shows that ME is an effective text classifier evidenced by significant high accuracy in both in-sample (73.11%) and equal out-of-sample (67.28) tests. Our ME algorithm performs extremely well given the relatively small number of misclassifications. As a comparison, Das and Chen (2007) compare eight different algorithms and find that the out-of-sample accuracy is less than 40% on average. Thus, a 67.28% out-of-sample accuracy is outstanding. In addition, a detailed comparison between ME and NB is analyzed by Zhang, Swanson and Prombutr (2012) in which they show that ME performs the best either in- or out-of-sample among other seven text classifiers based on stock message board data.

When evaluating messages without a self-reported sentiment, ME classifier returns a probability P for each of the predetermined sentiment category. According to the additional rule for probability that either class A or B will occur is simply the sum of the unconditional probabilities for the message, P(A or B) = P(A) + P(B), see Zhang et al. (2012). When coding each sentiment, we follow Tumarkin and Whitelaw (2001) and Tumarkin (2002) and code sentiment scores as -2 for Strong Sell, -1 for Sell, 0 for Hold, +1 for Buy, and +2 for Strong Buy. We combine Short and Strong Sell as -2 given their similar semantics. We also combine Hold and Scalp as 0 given their neutral sentiment. We assign the sentiment score defined as to the evaluated message m associated with stock i follows:

$$x_m^i \begin{cases} = 2, \ if\ P(\text{Strong Buy or Buy}) > P(\text{Strong Sell or Sell or Hold})\ and\ P(\text{Strong Buy}) > P(\text{Buy}) \\ = 1, \ if\ P(\text{Strong Buy or Buy}) > P(\text{Strong Sell or Sell or Hold})\ and\ P(\text{Strong Buy}) < P(\text{Buy}) \\ = 0, \ if\ P(\text{Strong Buy or Buy}) = P(\text{Strong Sell or Sell or Short}) \geq P(\text{Hold}) \\ = 0, \ if\ P(\text{Hold}) \geq P(\text{Strong Buy or Buy or Strong Sell or Sell}) \\ = 0, \ if\ P(\text{Hold}) = P(\text{Strong Buy or Buy}) > P(\text{Strong Sell or Sell or Short}) \\ = 0, \ if\ P(\text{Hold}) = P(\text{Strong Sell or Sell}) > P(\text{Strong Buy or Buy}) \\ = -1, \ if\ P(\text{Strong Sell or Sell}) > P(\text{Strong Buy or Buy or Hold})\ and\ P(\text{Sell}) > P(\text{Strong Sell}) \\ = -2, \ if\ P(\text{Strong Sell or Sell}) > P(\text{Strong Buy or Buy or Hold})\ and\ P(\text{Sell}) < P(\text{Strong Sell}) \end{cases} \quad (5)$$

Therefore, each messages without a self-reported sentiment will be given a sentiment score by the ME classifier from -2 to 2. If a message has a self-reported sentiment, we use its original sentiment. With each message is quantified with a sentiment score, we are able to calculate an aggregate sentiment index and disagreement index for stock i on day t among all posters.

3. 3. Sentiment and Disagreement Measure

WallStreetPit has a pecuniary reputation system for its posters, with the credit score being a proxy for the poster's reputation and influence in the online community. We are able to construct a credit-weighted sentiment index, as suggested in some other studies (Gu et al., 2006; Cook and Lu, 2009). Many previous studies emphasize the importance of incorporating posters' credibility into the message bullishness effect analysis, Cook and Lu (2009) and Depken and Zhang (2010) among others. Our formula for the sentiment index is based on Antweiler and Frank (2004) sentiment index but differs in two ways: 1. our index is credit-weighted, with the log of a poster's reputation being the poster's weight; 2. Following Cook and Lu (2009), we include "Hold" sentiment messages in the formula.

Each message has a sentiment score ranges from Strong Sell (-2) to Strong Buy (+2), which is either provided by its original poster or assigned by our ME text classifier. Our sentiment index starts with Antweiler and Frank (2004) bullishness ratio:

$$B_t^i = \frac{M_t^{i,BUY} - M_t^{i,SELL}}{M_t^{i,BUY} + M_t^{i,SELL}} \quad (6)$$

In Antweiler and Frank (2004) study, about 78.8% of their messages are classified as "Hold" which are subsequently excluded from their study. In accordance with Cook and Lu (2009)'s argument, we suggest that "Hold" messages should not be excluded in the sentiment formula even if "Hold" is coded as 0. Consider the following two examples: case one is with 100 number of Buy sentiments while case two is with only 1 Buy and 99 number of Hold sentiments. Following equation (6) stated in Antweiler and Frank (2004), these two cases have the equivalent bullishness ratio of 1 even though the latter is unequivocally less optimistic than the former one. We content in this paper that purging "Hold" messages bias the sentiment index. Therefore, our modified bullishness ratio for stock i on day t is as follow:

$$B_t^i = \frac{M_t^{i,O} - M_t^{i,P}}{M_t^{i,O} + M_t^{i,P} + M_t^{i,N}} \quad (7)$$

where is the number of messages with sentiment Buy and Strong Buy, is the number of messages with sentiment Sell, Strong Sell and Short, and is the number of messages with sentiment Hold and Scalp. Since Antweiler and Frank (2004)'s message sentiment is either 1 or -1 with an equal weight allocation, M in equation (6) is equal to the number of messages. However, our sentiment score in message m on day t, ranges from -2 to +2, with a log of poster k's credit score, as the message weight for m on day t. Thus, our modified bullishness ratio, , with a modified numerator from equation (7) becomes:

$$MB_t^i = \frac{\sum_{m=1}^{M_t^{i,O}} LCS_m^{k,t} * x_m^{O,t} + \sum_{m=1}^{M_t^{i,P}} LCS_m^{k,t} * x_m^{P,t}}{M_t^{i,O} + M_t^{i,P} + M_t^{i,N}} \quad (8)$$

where is the modified bullishness ratio for stock i on day t. Note that given none negative credit score and pessimistic sentiment . Besides, =0 is omitted in the numerator because all neutral (Hold or Scalp) sentiment scores are set to be 0. The numerator represents the total number of messages related to stock i on day t, which accounts for neutral "Hold" information. Antweiler and Frank (2004) construct their final sentiment index by taking into account the number of traders expressing a particular sentiment. Because they assume each message is written by a different poster, the number of messages is the number of posters. Thus, their final sentiment index becomes , where is the total number of messages (or posters) for stock i on day t, making sure that a stock with no message, (=0) is given a zero score . In this study, we relax this assumption by using the actual number of different posters in that a more realistic case would be one poster can leave more than one messages about stock i on day t. Because we require each stock must have at least five different posters in order to prevent hyping by one or two dominant posters, our final sentiment index for stock i on day t is

$$SENT_t^i = MB_t^i * \log(NPosters_t^i) \quad (9)$$

where is the number of unique posters who discuss stock i on day t. We have . A higher implies a stronger optimistic sentiment considering number of posters, vice versa.

Antweiler and Frank (2004) further propose a reduced form of standard deviation of sentiments as an agreement proxy where is from equation (6). In their setup, sentiment is either 1 or -1 and they assume an equal weight for each message. Because our sentiment score ranges from -2 to +2, our disagreement index is a credit-weighted standard deviation of the sentiment scores for the stock i on day t. Different from Antweiler and Frank (2004)'s formula, neutral sentiment Hold and Scalp are included. We first calculate each message m's weight related to stock i on day t:

$$w_m^i = \frac{LCS_m^k}{\sum_{m=1}^{M_t^{i,A}} LCS_m^k} \quad (10)$$

where is all the number of messages related to stock i on day t. Our disagreement index is therefore:

$$DISG_t^i = \sqrt{\sum_{m=1}^{M_t^{i,A}} w_m^i * (x_m^i - \bar{x}_m^i)} \quad (11)$$

where mean sentiment .We have . A higher implies a higher disagreement with different opinions among posters, vice versa.

3.4. Lead-lag Regressions

Different from prior studies, we investigate the stock message board effect by controlling for both firm size (small vs. large) and news arrival (with vs. without). Particularly, we study the relations between four market variables (raw return—RETN, intra-day volatility—VOLT, turnover ratio—TURN, and standardized bid-ask spread—SPRD) and four message board variables (credit-weighted sentiment index—SENT, credit-weighted disagreement index—DISG, the log of number of messages—NMSG, and the average log of message length—ALEN). RETN is the raw return, but we control for the market return as one of the exogenous variables in the model following Antweiler and Frank (2004)'s model. VOLT is the standardized volume-weighted intraday price volatility. TURN is the proportion of small sized trading volume because retail investors who trade in smaller sizes are the ones likely to be influenced by message board activities, Sabherwal et al. (2011). Accordingly, we use the ratio of dollar trading volume in small-sized trades (trades of $3,750 or less) over total dollar trading volume as the turnover ratio. SPRD is the standardized bid-ask spread which is the ratio of the stock's average bid-ask spread and mean ask price on a particular day.

Many prior studies support the view that posting activities influence both current and future market trading activities. Wysocki (1999) suggests that messages posted on stock message boards influence "public opinion" about a firm. Antweiler and Frank (2004) suggest that the causality for volatility is more from message boards to the market. Gu et al. (2006) argue that message board sentiments drive future stock returns. Sabherwal et al. (2011) find that when there is no market news, online traders' sentiment causes positive contemporaneous return and negatively predicts the next day return. Both Clarkson et al. (2006) and Bettman et al. (2011) find that online rumors cause significant abnormal returns and trading volumes. Oh and Sheng (2011) reveal that stock micro-blog sentiments have predictive power for simple and market-adjusted stock returns. Bird and Thapa (2010) find that with unscheduled company news, contemporaneous message board activity causes positive return drifts and high volume of message board activity increases trading volatility. Other than the conventional metrics including number of messages, sentiment and disagreement, we propose an additional posting measure—the average log of message length (ALEN), as a proxy for average amount of information in the messages. Accordingly, we set one of the four market variables as endogenous Y variable at a time while message board variables as explanatory exogenous X variables. The contemporaneous model is as follow:

$$Y_t = B_0 + B_1 X_t + B_2 SIZE + B_3 INTERACT_t + B_4 CONTROL_t + B_5 Y_{t-1} + \varepsilon_t \quad (12)$$

where the dependent variable is one of the four market variables (RETN, VOLT, VOLM, or SPRD) in a separate test at a time. The explanatory variable is a vector of corresponding four message board variables on day t [SENTt, DISGt, NMSGt, and ALENt] and B1 is the vector of coefficients on message board variables. Size is a dummy variable where SIZE=1 if stock i is a small cap, =0 otherwise. B2 is the coefficient on the SIZE variable. The vector INTERACT is the multiplication between dummy variable SIZE and each of the four components in the vector .B3 contains a vector of coefficients on these four interaction terms. CONTROL contains a set of control variables: controls for market return which is the raw return on the CRSP value-weighted index and controls for a day being the first trading day after a weekend or holiday when people are generally hectic. Since trading volume is likely to be positively correlated with volatility (Jones et al. 1994) and negatively correlated with bid-ask spread (Amihud and Mendelson, 1986), we include , the log of the number of trades of the stock, as an additional control variable in both the volatility and spread regressions. It could be the case that retail investors have a preference on NASDAQ-listed stocks (Sabherwal et al. 2011), we include a dummy variable LIST, =1 if NASDAQ-listed stocks, =0 otherwise. B4 contains a vector of coefficients on control variables. is the one-day lagged of the dependent variable in order to control for 1-day autocorrelation. B5 is the coefficient on the lagged dependent variable.

For a one-day or two-day prediction model, we follow Antweiler and Frank (2004)'s time sequencing tests and lead the dependent variable and the control variables by one or two days relative to the independent variables in the above contemporaneous equation (12). In total, we conduct 24 separate lead-lag regressions among which 12 (4 x 3) are under no news scheme while the other 12 (4 x 3) regressions are conducted when there was news. All lead-lag regressions are controlled for heteroscedasticity using White (1980) approach. Robust t-test statistics and adjusted R-squares are used in these regressions.

3.5. News Prediction

Antweiler and Frank (2004) show that message board activities are not only associated with same-day Wall Street Journal news, but active message board postings predict the presence of news stories up to two days later. Since the sample stocks studied in Antweiler and Frank (2004) are in DIA and XLK funds, which are primarily large stocks, they do not separate the news prediction effect of stock message board by firm size. With rich data, we study the predictive power of message board posting variables on future news occurrence by splitting into small and large stock pools. We design a Probit model for news prediction as follow:

$$NEWS_0 = \alpha + \beta_1 X_t + \beta_2 SIZE + \beta_3 INTERACT_t + \beta_4 CONTROL_t + u_t \quad (13)$$

where the binary dependent variable takes a value of one (zero) if a stock i is with (without) a material market news on the event day t0 (952 news vs. 606 no news). We fix t0 for NEWS0 variable while lag the exogenous variables by one or two trading days. X is a vector including four message posting variables: SENT, DISG, NMSG, and ALEN. SIZE is a binary dummy variable as defined above. INTERACT is a vector of interaction terms between SIZE and X. CONTROL contains two dummy variables: LIST and NWHK. Fang and Peress (2009) report that media coverage generally is larger for NYSE stocks than for NASDAQ stocks, we impose the LIST dummy variable in the model. There could be a skewed probability of media coverage after a weekend or holiday, we add the NWHK as an additional control variable.

IV. EMPIRICAL RESULTS

4.1 Summary Statistics

Based on the 1,058,366 messages posted on WallStreetPit stock message board from January 1, 2006 to December 31, 2011.Figure 3-A shows that among all messages with self-disclosed sentiments, "Strong Buy" dominated. In addition, optimistic sentiments outweighed pessimistic ones imply a bullish tone on the WallStreetPit message board on average. Of all the messages, most of them were posted during trading hours (9:30 a.m. to 4:00 p.m. EST) and few were posted off-hours evidenced in Figure 3-B. As expected, Figure 3-C shows that posting activity during weekdays dramatically exceeded that of during weekends. There is little difference among weekdays although Wednesday is the peak of posting activity. Figure 3-D further discloses that there is no significant difference in terms of posting activity across months although October is shown to have the highest posting volume. These posting patterns are largely in line with previous studies using Yahoo! Finance, Raging Bull, and other stock message boards. Such similarities indicate online posters follow similar patterns when posting messages across different message boards and over different time periods.

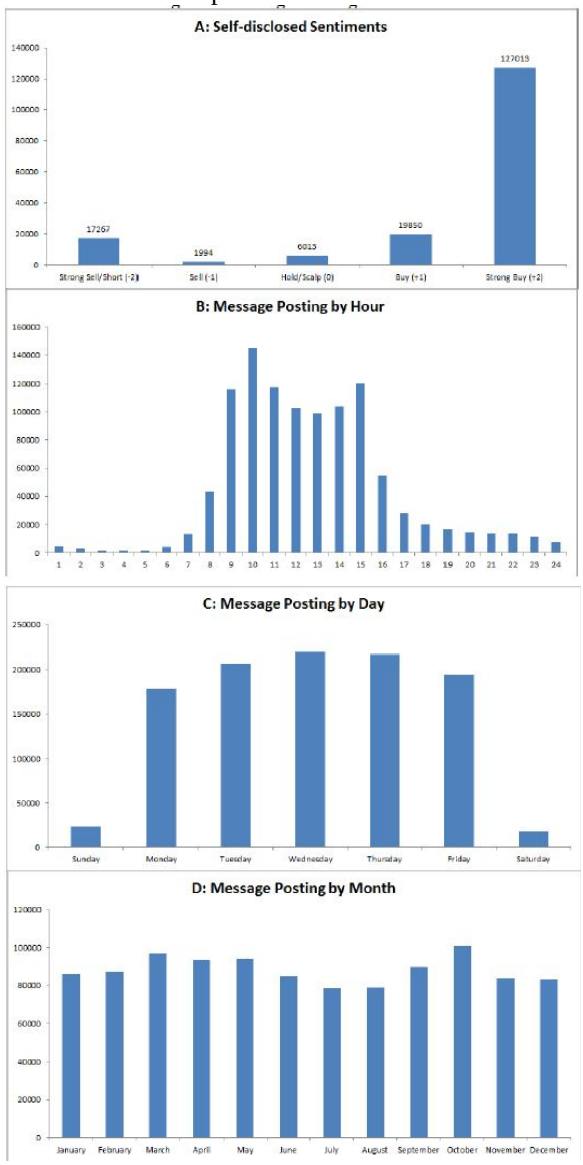


Figure 3 Message Posting Patterns

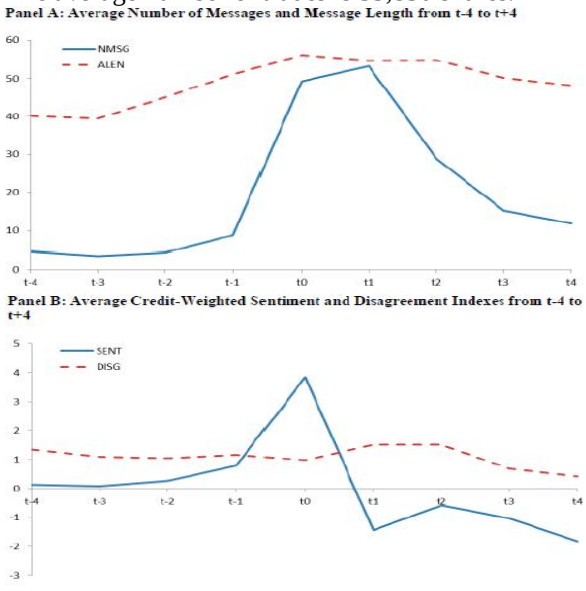
Table 2 reveals the summary statistics of 76,701 messages account for our final 1558 events. There were 49.23 messages per event on average with the minimum of 30 as our criteria. The average message length measured in number of characters is 57.15 which imply that dialogue-like messages are commonly concise on WallStreetPit message board. The average number of posters is 13.23, which are more than double the minimum requirement of 5 posters. As expected and consistent with Figure 1-A, the average bullishness index measured in equation (7) is 0.72 signaling a bullish tone on the board. There were cases when all messages are in bullish sentiment evidenced by the maximum bullishness index equals to 1.00. Furthermore, sentiment index, which ranges from negative infinite to positive infinite, has an average of 3.84. Consistent with Figure 3-A, it

expresses an optimistic tone among posters. Disagreement index, which is similar to a standard deviation and must be a non-negative term, shows an average of 0.98. This result shows certain level of incongruous opinions among posters so that the higher the disagreement index the fewer consensuses among posters. Thus, 0.98 disagreement index implies certain level of disagreement among posters.

Table 2 Summary Statistics of 1558 on the discussion day t0

Message board variables	*Mean*	*Std. Dev.*	*Minimum*	*Maximum*
Number of messages	49.23	16.11	30.00	179.00
Message length (characters)	56.15	24.69	1.00	228.62
Number of posters	13.23	5.15	5.00	47.00
Bullishness index [-1,1]	0.72	0.29	-0.88	1.00
Sentiment index [-∞, +∞]	3.84	2.02	-8.71	7.55
Disagreement index [0, +∞]	0.98	0.17	0.26	2.22
Stock market variables				
Stock trading price	20.07	59.83	1.00	639.62
Stock raw return (%)	19.40%	53.24%	-83.89%	825.00%
CRSP value-weighted return (%)	0.05%	1.61%	-8.98%	6.91%
S&P 500 index return (%)	0.03%	1.60%	-9.04%	7.08%
Volume-weighted volatility	0.41	0.33	0.01	3.13
NASDAQ listing	0.67	0.47	0.00	1.00
Trading volume (in millions)	33.80	113.00	100	1,870.00
Number of trades (in thousands)	35.85	67.19	3.00	716.58
Small trades (≤$3,750) proportion	0.69	0.12	0.02	0.99
Medium trades proportion	0.17	0.09	0.05	0.60
Large trades (≥$30,000) proportion	0.14	0.04	0.00	0.28

The bottom part of Table 2 reveals stock trading variables associated with the 1558 events. The average trading price is $20.07 with the minimum $1.00 as the criteria and the maximum $639.62 per share. Raw stock return has an average of 19.40%, which coincides with the average optimistic tone among posters. As a comparison, we see that average CRSP index return and S&P500 index return on these event days are not significantly different from zero. The average volume-weighted volatility is 0.41 and 67% of the events were associated with NASDAQ-listed stocks. Total trading volume is averaging at 33.80 million shares among which 69% were small trades, 17% were medium trades, and 14% were large trades. The average number of trades is 35,850 shares.



Notes:Average posting statistics are based on 76,701 observations account for 1,558 events. NMessages represents the average number of messages and NPosters represents the average number of posters. Sentiment is the average credit-weighted sentiment index and Disagreement is the average credit-weighted disagreement index. NMessages ∈ [30, +∞] on day t, the event day, and NMessages ∈ [0, +∞] on other days. NPosters ∈ [5, +∞] on day t and NPosters ∈ [0, +∞] on other days; Sentiment ∈ [-∞, +∞]; and Disagreement ∈ [0, +∞].

Figure 4

Focusing on our final sample of 1558 events, Figure 4 plots summary statistics for four major posting variables, NMSG, ALEN, SENT, and DISG, from day t-4 to t+4 with t0 be the discussion event day. Summary statistics on day t0 for these four variables match with the numbers presented in Table 2. In Panel A of Figure 4, we see that there was a prominent increase in number of postings and message length from t-4 to t0. NMSG peaked on day t1 while ALEN diminished after day t0. In Panel B, sentiment also increased before t0 and tapered afterwards while disagreement kept the same level on or before the discussion day t0. Interestingly, DISG was lower on t0 compared with preceding days from t-4 to t-1, which indicates a possible herding and concordant tone among traders on the discussion day t0. DISG rose higher after t0 and peaked on t2.

4.2 Event Study

In Table 3, we report the results of daily event study for abnormal and cumulative abnormal returns and volumes from day t−4 to t+4, with t0 be the event day when stocks were heavily discussed among at least five posters. The extended return and volume event study results over a wider (t-30, t+30) window are shown pictorially in Figures 5 and 6, respectively. As described above, we demarcate our sample in two dimensions: news arrival and firm size. Horizontally, Panel A and B (Panel C and D) cover events with (without) news arrival. Vertically, Panel A and C (Panel B and D) cover events with small (large) cap stocks.

Table 3 Event Study

Panel A. Small Size with News (591 Observations)

Event days	*AAR(%)*	*CAAR(%)*	*AAV(%)*	*CAAV(%)*
t-4	1.76[a]	-0.20	0.15	-1.97
t-3	0.32	0.12	1.56	-0.41
t-2	2.00[b]	2.12	5.33[c]	4.92
t-1	2.13[b]	4.25	9.38[c]	14.30
t0	8.75[c]	13.00	11.64[c]	25.94
t+1	3.39[c]	16.39	10.03[c]	35.97
t+2	2.18[b]	18.57	1.95[a]	37.92
t+3	-1.43[a]	17.14	1.11	39.03
t+4	-0.96	16.18	-0.09	38.94

Panel B. Large Size with News (361 Observations)

Event days	*AAR(%)*	*CAAR(%)*	*AAV(%)*	*CAAV(%)*
t-4	0.62	4.55	-1.36	1.68
t-3	0.39	4.94	-1.76	-0.08
t-2	0.55	5.49	1.26	1.18
t-1	1.42[a]	6.91	5.32[c]	6.50
t0	3.44[b]	10.35	10.09[c]	16.59
t+1	1.18[a]	11.53	13.86[c]	30.45
t+2	0.99	12.52	9.65[c]	40.10
t+3	-0.53	11.99	5.92[c]	46.02
t+4	-0.66	11.33	3.69[b]	49.71

Panel C. Small Size with No News (493 Observations)

Event days	*AAR(%)*	*CAAR(%)*	*AAV(%)*	*CAAV(%)*
t-4	1.38	2.78	356.30[c]	2858.08
t-3	1.22	4.00	514.64[c]	3372.72
t-2	0.47	4.47	526.72[c]	3899.44
t-1	0.98	5.45	401.15[c]	4300.59
t0	12.62[c]	18.07	913.03[c]	5213.62
t+1	-1.98[b]	16.09	400.28[c]	5613.90
t+2	-1.70[b]	14.39	621.77[c]	6235.67
t+3	-1.32[a]	13.07	515.25[c]	6750.92
t+4	-0.31	12.76	333.01[c]	7083.93

Panel D. Large Size with No News(113 Observations)

Event days	*AAR(%)*	*CAAR(%)*	*AAV(%)*	*CAAV(%)*
t-4	0.03	1.15	-0.09	14.71
t-3	0.98	2.13	0.63	15.34
t-2	-1.08	1.05	0.75	16.09
t-1	1.15	2.2	0.76	16.85
t0	0.79	2.99	0.43	17.28
t+1	-0.55	2.44	-1.00	16.28
t+2	1.16[a]	3.6	-0.13	16.15
t+3	-0.41	3.19	1.56[a]	17.71
t+4	0.85	4.04	0.30	18.01

Notes: This table shows the event study results for four groups: Small Size with News (Panel A), Large Size with News (Panel B), Small Size with No News (Panel C), and Large Size with No News (Panel D). Superscript a, b, and c denote statistical significance at the 0.10, 0.05 and 0.01, levels, respectively. CAAR and CAAV on t-4 is actually CAAR (-30, -4).

On the one hand, comparing results in Panel A with Panel B of Table 3, we find that on the event day t0, AAR is significantly higher (8.75%) for small stocks than large stocks (1.42%). Similarly, AAV are also significant for these two samples with a higher jump with small stocks (11.64%) than large cap ones (10.09%). Although not tabulated, CAAR (-30, 0) are

both insignificant (13.00% for small cap and 10.35% for large cap, respectively) for these two samples. In addition, CAAV (-30, 0) are both insignificant (25.94% for small cap and 16.59% for large cap, respectively). Collectively, when material news arrived, we see that there was a significant spike on the event date with heavy online discussion. But there was a larger and more significant market reaction to news for small stocks. A clear view of this pattern can be seen from the top two graphs in Figure 5.

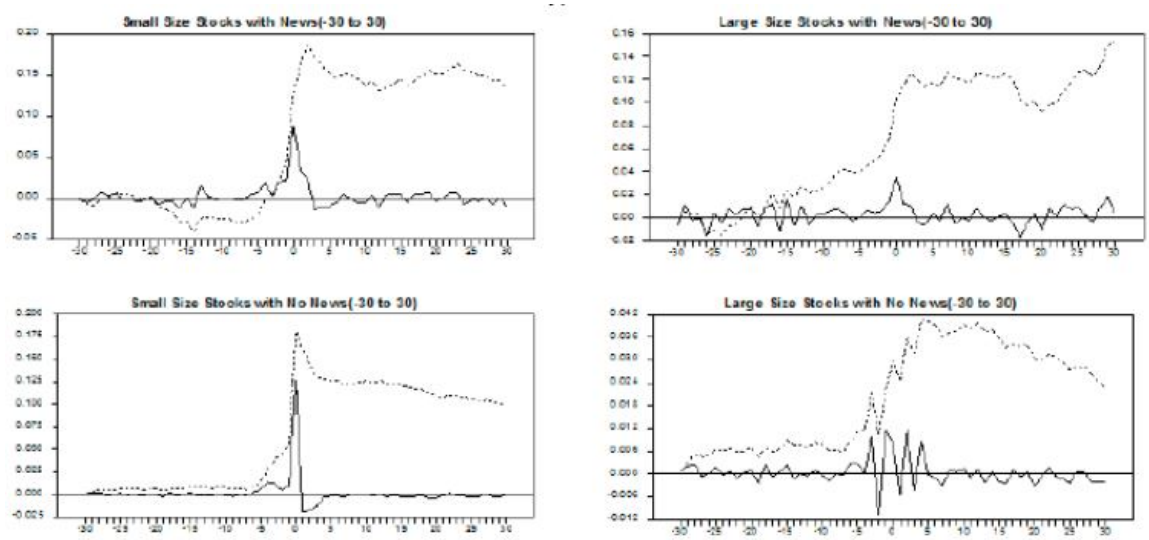


Figure 5

On the other hand, comparing results in Panel C with Panel D of Table 3, we observe an asymmetric market reaction to online discussion between small and large stock samples when there was no news presented. Particularly, we find that there was a significant AAR (12.62%) jump for small stocks while no significant market reaction was detected for large stocks (0.79%). Besides, AAV presented in Panel C shows a significant 913.03% jump on the event day t0 for small stocks while an inactive AAV (0.43%) for large stocks found in Panel D. CAAVs are in the similar pattern across Panel C and D. Graphical demonstration of AAV and CAAV over a wider window from t-30 to t+30 are reported in the bottom two graphs in Figure 6.

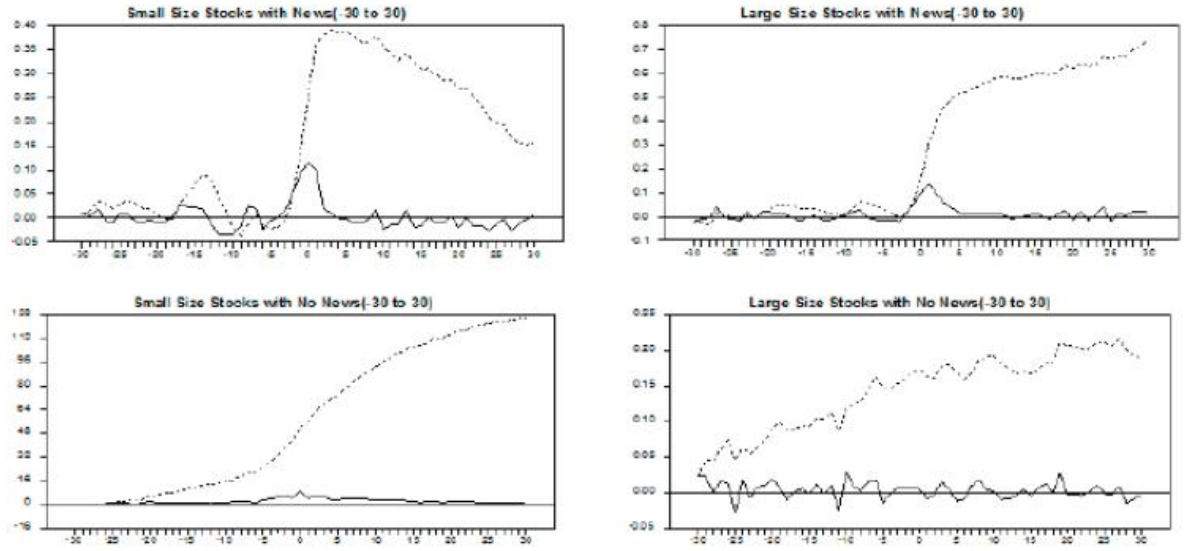


Figure 6 Volume Event Study

In sum, event study reveals that there was a significantly large and positive market reaction on the discussion day for small stocks regardless of news arrival although there was an even larger market response to online talks (12.62% versus 8.75%) under no news circumstance. For large stocks, however, a relative smaller (no) market reaction is detected with (without) news arrival on the discussion day.

4.3 Robust Lead-lag Regression

In this section, we investigate comprehensively how online posting activities (sentiment, disagreement, number of messages, and message length) affect stock

trading (return, volatility, turnover, and bid-ask spread). We first present the pairwise correlation among these eight primary parameters in Table 4.

Table 4 Pairwise Correlation

	SENT	*DISG*	*NMSG*	*ALEN*	*RETN*	*VOLT*	*TURN*	*SPRD*
SENT	1.0000							
DISG	-0.0129	1.0000						
NMSG	0.0378	0.0658	1.0000					
ALEN	-0.0728	-0.0741	-0.0818	1.0000				
RETN	0.1287	0.041	-0.001	-0.0117	1.0000			
VOLT	0.0332	-0.008	-0.1055	-0.0505	0.1100	1.0000		
TURN	0.1280	0.0649	-0.0349	0.0365	0.0820	0.1411	1.0000	
SPRD	-0.0013	-0.0758	0.0358	0.0031	0.1488	-0.1208	-0.187	1.0000

Notes: raw return—*RETN*, intra-day volatility—*VOLT*, turnover ratio—*TURN*, and standardized bid-ask spread—*SPRD*; credit-weighted sentiment index—*SENT*, credit-weighted disagreement index—*DISG*, the log of number of messages—*NMSG*, and the average log of message length—*ALEN*.

Due to different design of our parameters, from equation (9) and (11), we expect SENT is uncorrelated with DISG. The correlation between them is insignificant at -0.0129. Thus, multicollinearity seems not to be a concern in a regression model containing both variables. NMSG measures the number of posted messages while SENT measures opinions among posters, they measure different aspects of online talks. ALEN measures the amount of information in the message, which is shown to be uncorrelated with either SENT (-0.0728) or NMSG (-0.0818).

Regarding stock trading variables, except SPRD (-0.0013), we find that RETN (0.1287), VOLT (0.0332), and TURN (0.1280) all positively correlate with SENT. Besides, RETN (0.0410) and TURN (0.0649) positively correlate with DISG while VOLT (-0.0008) and SPRD (-0.0758) negatively correlate with it. Except SPRD (0.0358), all other trading variables negatively correlate with NMSG. TURN (0.0365) and SPRD (0.0031) positively correlate with ALEN but RETN (-0.0117) and VOLT (-0.0505) negatively correlate with it. Among the four trading variables themselves, other than VOLT and SPRD (-0.1208) as well as TURN and SPRD (-0.1870) whose correlation are negative, all other correlations are positive.

Table 5 Lead-lag Regressions with News Presence (952 Observations)

	Contemporaneous Y_t (t=0)				*One-day prediction Y_t(t=1)*				*Two-day prediction Y_t(t=2)*			
Variables	$RETN_t$	$VOLT_t$	$TURN_t$	$SPRD_t$	$RETN_t$	$VOLT_t$	$TURN_t$	$SPRD_t$	$RETN_t$	$VOLT_t$	$TURN_t$	$SPRD_t$
Intercept	0.4846	-1.3795	-1.8325	-0.1063	-1.2425[a]	-1.1323	-2.357	-0.0372	0.4234	-2.5715	2.7303	0.0475
$SENT_0$	0.0173[b]	0.0330[a]	-0.0143	-0.0008[a]	0.0093[a]	0.1095[c]	-0.0019	-0.0002	0.0087	-0.0002	0.0049	0.0002
$DISG_0$	-0.019	0.0106	0.1030[c]	-0.0015[b]	-0.0173	0.0026	-0.0053	0.0034	0.0097	-0.0013	-0.0252	-0.0007
$NMSG_0$	0.1611[c]	0.1160[c]	0.0320[a]	-0.0010[b]	0.0784[a]	0.0212[a]	0.0126	-0.0039	-0.0075	-0.0016	-0.008	0.0018
$ALEN_0$	-0.0015	0.0004	-0.0028	0.0001	-0.0197	-0.0002	-0.0086	0	-0.0001	0	-0.0001	0.0003
SIZE	0.0398[b]	0.3127[b]	2.8512[c]	0.0003	0.0188[a]	0.1740[a]	1.0858[a]	-0.001	-0.0317[a]	0.0956	-0.3641	0.0003[a]
$SIZE*SENT_0$	0.0234[c]	0.1104[b]	0.1172[a]	-0.0001	0.0161[c]	0.0148	-0.0033	-0.0004	0.0004	-0.0002	-0.0025	-0.0002
$SIZE*DISG_0$	-0.0789[c]	0.004	0.1315[c]	-0.0022	-0.0023[a]	0.0031	-0.0108	-0.0006	-0.0039	-0.0013	0.0013	-0.0008
$SIZE*NMSG_0$	0.0316	-0.0301	0.0719[c]	-0.0055[b]	0.0813[b]	0.0166	0.0056	-0.0014	-0.0001	0.0011	0.0002	0.001
$SIZE*ALEN_0$	-0.009	-0.0018	-0.0009	-0.0006	0	-0.0091	-0.0066	-0.0002	0	0.0006	0.003	-0.001
$MKTR_t$	3.5801[c]	2.1991[c]	0.3806[c]	-0.0039[b]	2.3514[c]	0.5406[b]	0.1974[a]	-0.0278[c]	1.0640[c]	0.0493	0.2081[b]	-0.018
$NWKH_t$	0.033	-0.1539	-0.0662	-0.0056[b]	0.0994	0.4054	0.6195[a]	-0.0017	-0.0343	0.3865[a]	0.5104[b]	-0.0012
$TRAD_t$		0.0739[b]		-0.0025[c]		0.0588[b]		-0.0021[c]		0.0064[a]		-0.0010[c]
LIST			0.0582				0.0047				0.007	
$RETN_{t-1}$	-0.0023				0.0086				0.0004			
$VOLT_{t-1}$		1.3229[c]				0.3143[c]				0.7816[c]		
$TURN_{t-1}$			1.0938[c]				0.5538[c]				0.4013[c]	
$SPRD_{t-1}$				0.3254[c]				0.6621[c]				0.9022[c]
Adj-R	0.26	0.77	0.48	0.78	0.26	0.93	0.63	0.94	0.16	0.85	0.83	0.92

Notes: Superscript a, b, and c denote statistical significance at the 0.10, 0.05 and 0.01, levels, respectively.

We first run the lead-lag regressions with events under news influence. We present contemporaneous regression with posting and trading variables occurred on the same day. We then conduct one-day predictive regression by leading the endogenous variable and control variables by one day while keeping the rest of exogenous variables unchanged on day t0. Two-day predictive regression is then conducted in a similar way. Of all the lead-lag regressions, we impose a dummy variable SIZE to control for stock market cap. Thus, variables included in vector Xt represent the message board effect on large stocks while interaction terms in vector INTERACT capture the effect on small stocks. Focusing on large stocks, we find that sentiment (SENTt0=0.0173) and postings (NMSGt0=0.1611) both significantly and positively influence the same day stock return after controlling for market return, trading day dummy, and lagged endogenous variable. On the contrary, disagreement (DISGt0=-0.0190) and message length (ALENt0=-0.0015) both have negative impact on return although they are insignificant. A positive and significant coefficient on the dummy variable SIZE (0.0398) implies that small stocks on average have a higher return than that of the large stocks on the discussion days. The interaction terms reflect the effect on RETN by online talks for small stocks. First, SENT has a 0.0407 positive and significant effect on RETN evidence by the sum of coefficients on SENTt0 and SIZE*SENTt0 (0.0173+0.0234=0.0407). Clearly,

with the arrival of market news on the discussion days, sentiments among online posters tend to influence small stocks' returns more significantly than large stocks (0.0407 versus 0.0173). For small stocks, disagreement (-0.0789-0.0190=-0.0979) has a stronger negative impact, postings (0.0316+0.1611=0.1927) has a more positive impact, and message length (-0.0090-0.0015=-0.0105) has a more negative impact on RETN, respectively. Collectively, online talks entailed larger and more significant stock returns for small than large stocks when market news arrival. MKTR (3.5801) as a control variable is significantly positive on the discussion day with news arrival. This is expected in that previous evidence shows that online traders tend to hold long positions, acting bullishly, and talk more when the market is generally rising. Return (RETNt-1=-0.0023) in the previous day has no significant prediction suggests return series is stationary with no substantial serial autocorrelation.

For large stocks, sentiment (SENTt0=0.0330) and postings (NMSGt0=0.1160) both have significant and positive explanatory power on the contemporaneous volume-weighted volatility while disagreement (DISGt0=0.0106) and message length (ALENt0=0.0004) have not. On average, small stocks are more volatile than large ones evidenced by the significant coefficient on SIZE (SIZE=0.3127). For small stocks, except SENT (SENTt0=0.1104+0.0330=0.1434) which exerts a stronger and more significantly positive impact on volatility, other three posting variables do not significantly affect trading volatility. Volatility tend to be higher on the discussion day with higher market return (MKTRt0=2.1991). Number of trades which is imposed as a control variable is positively significant (TRADt0=0.0739). Previous day volatility is also significant (VOLTt-1=1.3229) indicates an autocorrelation exists in this variable. As a result, autocorrelation of volatility largely contributes to the high adjusted R-square (0.77). Simply put, sentiment and disagreement (postings and message length) entailed larger (smaller) volatility when market news presented.

For turnover ratio, we find that higher disagreement (DISGt0=0.1030) and more postings (NMSGt0=0.0320) induce a higher small size trading volume for large stocks while sentiment (SENTt0=-0.0143) and message length (ALENt0=-0.0028) have no significant impact. For small stocks, disagreement (0.1315+0.1030=0.2345) and postings (0.0719+0.0320=0.1039) impacted the turnover ratio with a stronger positive degree. However, higher online sentiment (0.1172-0.0143=0.1029) among small stocks induced more small-sized trades. In addition, when the broad market was in green (MKTRt0=0.3806), there were more small size trades. Turnover ratio is proven to have a significant serial autocorrelation (TURNt-1=1.0938) which accounts for a high adjusted R-square (0.48) in this regression.

Concentrating on the spread regression, except message length (ALENt0=0.0001), all other posting variables significantly reduced standardized bid-ask spread for large stocks. There is no distinct difference between small and large stocks in terms of average standardized spread. For small stocks, only message postings had a significant negative impact on spread (SIZE*NMSG=-0.0055) so that the aggregate effect is -0.0055-0.0010=-0.0065. When the entire market was up, spread became narrower (MKTRt0=-0.0039) on average. We also find that spread was narrower on the first trading day after a weekend or holiday (NWKHt0=-0.0056) indicating higher liquidity. As expected, more trades induced narrower spread (TRADt0=-0.0025) and spread possesses a noticeable serial autocorrelation (SPRDt-1=0.3254) which largely contribute to the large adjusted R-square (0.78).

Next, we conduct a predictive regression by leading the endogenous and control variables by one day and keeping the exogenous variables unchanged. For large stocks under news occurrence, both sentiment index (SENTt0=0.0093) and message postings (NMSGt0=0.0784) continue to extend significant impact on stock returns on the next day which are consistent with the results in contemporaneous regression. Small stocks had higher returns than their counterpart on day t+1 on average (SIZE=0.0188). For small stocks, all posting variables had similar effects on the next day stock returns as reported in the contemporaneous regression. For the impact on small stocks trading volatility, we find similar effects as reported in the contemporaneous regression. For turnover ratio, however, disagreement (DISGt0=-0.0053) and message postings (NMSGt0=0.0126) no long carried their significant influences to the next day for large stock sample. The next day turnover ratio was still higher with small stocks on average (SIZE=1.0858). In addition, message posting activities did not significantly predict the next day turnover ratio for small stocks. Lastly, regardless stock size, posting activities on the discussion day did not affect the standardized bid-ask spread in the next one or two days appreciably. Although message posting activities predicted certain trading variables, such as return and volatility, their effects on stock trading on the next day fade. Further evidence can be found with the two-day prediction regression results in which we find no significant impact from the four posting variables on stock trading on day t+2, regardless of stock size. In sum, we

find posting activities dramatically affected trading for all stocks when news occurred, especially on the discussion day, with a larger impact on small stocks. However, the online talks influence lasted at most one trading days and disappeared on the second subsequent day.

Table 6 Lead-lag Regressions without News Presence (606 Obs)

	Contemporaneous Y_t (t=0)				*One-day prediction Y_t (t=1)*				*Two-day prediction Y_t (t=2)*			
Variables	$RETN_t$	$VOLT_t$	$TURN_t$	$SPRD_t$	$RETN_t$	$VOLT_t$	$TURN_t$	$SPRD_t$	$RETN_t$	$VOLT_t$	$TURN_t$	$SPRD_t$
Intercept	-0.115	-3.4373	-4.6222	0.0594	-0.012	1.2874	1.6013	0.0819	0.1428	-0.3702	1.3355	-0.0377
$SENT_0$	-0.0052	-0.019	-0.0511	0.0005	-0.0095	0.0094	-0.0212	0.0001	-0.0004	-0.0025	0.0069	-0.0002
$DISG_0$	0.0716	-0.0756	-0.0059	-0.0007	-0.0099	0.0445	-0.021	-0.0011	0.0124	0.006	0.105	0.0002
$NMSG_0$	0.028	0.0238	0.09	-0.0013	-0.0333	0.0089	-0.0286	-0.0005	-0.0251	-0.0622	-0.0459	0.002
$ALEN_0$	0.088	-0.0207	-0.1863	0.0001	0.0393	0.0416	0.0572	-0.0049	-0.0233	0.0041	0.205	0.0015
SIZE	0.1089[c]	0.6355[c]	1.7000[b]	0.0029[a]	-0.0094[a]	0.3930[a]	0.3248[b]	-0.0166[a]	-0.0218[b]	0.1216	0.3243[b]	-0.01
$SIZE*SENT_0$	0.0379[c]	0.0653[a]	0.1202[b]	-0.0016[b]	-0.0100[a]	0.0138[a]	0.1309[c]	-0.0007	-0.0162[a]	0.0155	0.0055	-0.0001
$SIZE*DISG_0$	-0.2211[b]	-0.0382	0.1403[a]	-0.0013[a]	-0.1316[a]	-0.0405	0.1188[b]	-0.0003	0.0017	0.0048	0.0892[c]	-0.0002
$SIZE*NMSG_0$	0.0367	-0.1562	0.1523[c]	-0.0027[a]	-0.0066	-0.0189	0.0565	-0.0006	-0.0148	-0.0556	0.0018	-0.0003
$SIZE*ALEN_0$	-0.0967[a]	0.0003	-0.002	0.0001	-0.0049	-0.0002	-0.0061	0	-0.0008	0.0061	0.004	0
$MKTR_t$	1.9918[c]	1.0174[b]	1.4723[c]	-0.0063[c]	-0.1255	-0.4324	-0.4322	-0.0107[c]	0.0658	0.1872	-0.3608	0.0007
$NWCH_t$	-0.0102	-0.0494	0.0538	-0.0027[a]	-0.0481	0.0086	-0.0623	-0.0013[a]	-0.002	-0.0183	-0.0377	-0.0018
$TRAD_t$		0.0370[b]		-0.0037[c]		0.0104[a]		-0.0022[c]		0.009		-0.0002[b]
LIST			0.0972[a]				0.0279				0.0527[a]	
$RETN_{t-1}$	-0.0057				0.0013				-0.0035			
$VOLT_{t-1}$		0.2932[b]				0.5070[c]				0.3760[c]		
$TURN_{t-1}$			1.4885[c]				0.4156[c]				0.6623[c]	
$SPRD_{t-1}$				0.2371[c]				0.7349[c]				0.9361[c]
Adj-R	0.15	0.39	0.51	0.7	0.04	0.63	0.47	0.83	0.1	0.46	0.6	0.84

Notes: Superscript a, b, and c denote statistical significance at the 0.10, 0.05 and 0.01, levels, respectively.

In Table 6, we run lead-lag regressions with events under no news circumstance. We present the contemporaneous results followed by predictive model results. As a dramatic difference from contemporaneous test results presented in Table 5, we find that online talks extended no significant impact on any of the four trading variables for large stocks. This implies that message board discussion itself, if not companions with market news, can be ignored by investors for large stocks. On average, when no news arrived, small stocks had higher return (0.1089), larger volatility (0.6355), higher turnover ratio (1.7000), and wider spread (0.0029) than large stocks evidenced by the dummy SIZE variable. Importantly, online talks still noticeably affected small stock trading on discussion days. When there was a high sentiment among posters, we experienced a higher return (0.0379-0.0052=0.0327), a larger volatility (0.0653-0.0190=0.0463), a higher turnover ratio (0.1202-0.0511=0.0691), and a narrower spread (-0.0016+0.0005=-0.0011). Although disagreement did not significantly affect contemporaneous volatility (SIZE*DISGt0=-0.0382) for small stocks, the higher the disagreement among posters, the smaller the return (-0.2211-0.0052=-0.2263), the higher the turnover ratio (0.0403-0.0059=0.0344), and the narrower the bid-ask spread (-0.0013-0.0007=-0.0020). For small stocks with no news arrival, more message postings induced more small-sized trades (SIZE*NMSGt0=0.1523) and narrower bid-ask spread (SIZE*NMSGt0=-0.0027). As we know that messages posted on stock forums are generally short and dialogue-like, we find that lengthier messages actually impacted stock return negatively (SIZE*ALENt0=-0.0967). For small stocks, when sentiments were more bullish, we experienced a higher intraday volatility. This effect is similar to what is shown in Table 5. Combined, we conclude that higher sentiment induced more contemporaneous and future volatility regardless of size and news arrival. We largely consent the ascertainment by Koski, et al. (2008) that message board postings cause current and future volatility.

With no news, both sentiment and disagreement positively and significantly affected small stocks turnover ratio in the contemporaneous (SIZE*SENTt0=0.1202; SIZE*DISGt0=0.1403) and one-day prediction (SIZE*SENTt0=0.1309; SIZE*DISGt0=0.1188) models. All posting variables, except the average message length, had an inverse relationship (SIZE*SENTt0=-0.0016; SIZE*DISGt0=-0.0013; SIZE*NMSGt0=-0.0027) with the bid-ask spread contemporaneously. However, no predictive power on the bid-ask spread by message posting activities is found for the following one or two days. Side-by-side comparison shows that the explanatory power of posting variables on trading activities tapered from day t0 to t+2. In sum, without news, online talks only influenced trading activities for small stocks but not for large stocks.

4.4 News Prediction Probit Model

Prior studies argue rumors circulated on the stock message boards foreshadow future news announcement, see Van Bommel (2003) and Spiegel, et al. (2010). Antweiler and Frank (2004) specifically show unusual message board postings predicted the presence of news stories up to two days later for stocks in DIA and XLK. However, Antweiler and Frank

(2004) do not separate news prediction test into small and large groups due to data limitation. To fill this void, we design a Probit model to investigate the predictive power of online talks on upcoming material news by controlling for firm size. Due to our sample restriction (Step 3 in section 3), we exclude any event that had news occurred on t-2 or t-1 but we don't restrict news occurrence on subsequent day t+1 and t+2. Also due to insufficient observations when news events took place on day t+1 and t+2, our Probit models are indeed testing how preceding posting activities on day t-1 and day t-2 predicted the news occurrence on the discussion day t0. In our models, we focus on the average marginal effects (AMEs) of each exogenous variable including interaction term on the probability of news present in the next one or two days. With the AMEs, a marginal effect is computed for each case, and then all the computed effects are averaged. We prefer AMEs over the conventional marginal effects at the means (MEMS) because AMEs provide a better and comprehensive representation of how changes in message posting variables affect the probability of future news presents P (News = 1).

Table 7 Message Board Activities Predict News

	t=-1		*t=-2*	
	Coeff.	*AMEs*	*Coeff.*	*AMEs*
Intercept	-0.7903[c]		-0.6210[c]	
$SENT_t$	0.0351	0.0098	0.0029	0.0017
$DISG_t$	0.0145	0.0046	0.0056	0.0009
$NMSG_t$	0.0814[b]	0.0270[b]	0.0451[a]	0.0148[a]
$ALEN_t$	0.0311	0.0108	0.019	0.0106
SIZE	0.1217[b]	0.0682[b]	0.0965[b]	0.0511[b]
$SIZE*SENT_t$	0.0528[c]	0.0154[c]	0.0216[c]	0.0077[c]
$SIZE*DISG_t$	-0.5468[b]	-0.1262[b]	-0.2224[b]	-0.0664[b]
$SIZE*NMSG_t$	0.1042[c]	0.0610[c]	0.0817[c]	-0.0231[c]
$SIZE*ALEN_t$	0.3603[b]	0.0881[b]	0.1413[a]	0.0331[a]
LIST	0.0066[b]	0.0013[b]	0.0005[c]	0.0002[c]
$NWHK_t$	0.0013	0.0007	0.0003	0.0001
Pseudo R^2	0.08		0.07	

Notes: Dependent variable is a binary variable =1 if news, =0 otherwise. Dependent variable is fixed on day t0. Among the 952 events with news, small (large) cap stocks occupied 635 (317) events. Among the 606 events with no news, small (large) cap stocks account for 471 (135) events. Superscript a, b, and c denote statistical significance at the 0.10, 0.05 and 0.01, levels, respectively.

We find that more message postings on day t-1 (NMSGt-1=0.0814, AMEst-1=0.0270) and t-2 (NMSGt-2=0.0451, AMEst-2=0.0148) significantly increased the odds of news occurrence on day t0 for large stocks. The predictive power of message postings is apparently stronger with one day closer to the news present (0.0270 versus 0.0148). Other than the number of messages, all other three message board variables did not predict the chance of news occurrence for large stocks. This is consistent with Antweiler and Frank (2004) who focus on large stocks in DIA and XLK sample.

As a stark comparison, however, we find that four posting variables one or two days earlier all had significant predictive power on news occurrence on the discussion day for small stocks. First, we find a significant coefficient on the dummy variable SIZE which is unsurprised due to more news events with small stocks (635 events) than large ones (317 events). Second, using interaction terms, we find that higher sentiment, more message postings, and lengthier message on t-1 and t-2 significantly increase the odds of a material news occurrence pertinent to small stocks on day t0. Online posters, who are mostly retail investors, are generally bullish due to their commonly adopted long positions, Odean (1999). Thus, rumor circulated online with higher positive sentiment is a premonition of forthcoming market news for small stocks. Rumor may contain fact-based information which is unconfirmed by the time it is posted. As expected, the larger number of messages posted online the higher the chance of a shortly followed news occurrence for small stocks. Significantly positive message length coefficient implies that posters tended to spread more detail and lengthy information on the message board when there was expecting news. This could also be an evidence of information leaking on the stock message board given the fact that many current employees of a firm post online, Felton and Kim (2002) and Kimmel (2004). Differently, disagreement had a negative impact on the odds of future news occurrence. Sabherwal et al. (2011) states that an online stock message board can serve as a herding venue to temporarily drive stock prices up and down. Thus, our finding on the relationship between disagreement and future news occurrence implies that a higher level of herding behavior proxy by a smaller disagreement level betokening a higher chance of upcoming news for small stocks. All in all, posting activities helped predict the presence of upcoming news in the next one or two days with a stronger news predictive power for small stocks while only posting volume predicts future news appearance for large stocks.

V. ROBUSTNESS CHECKS

In this section, we discuss ten robustness tests employed in this study. To enhance rigor of our study, we have conducted the following robustness tests:

1. In our event study, to avoid the dependence on normality of return distributions as in the Patell parametric test, we also conduct the Cowan (1992) Z-test, a nonparametric generalized sign test. Cowan (1992) Z-statistics are largely consistently with that of Patell (1976) Z-statistics. To conserve space, we only report the significant level using Patell (1976) Z-statistics.

2. We acknowledge that "Lionmaster" who is the creator of Thelion.com and a frequent poster on WallStreetPit has the highest credit score. We believe that it is likely this poster's credit score has been inflated. In this robustness test, we consider "Lionmaster" as an outlier to reduce credit bias. Our findings are consistent without "Lionmaster" in the sample.

3. We split mid and large stocks in two different groups and test them separately using above regressions. We find no notable difference between them which support the merger of mid and large stocks in a unity group.

4. The recent global financial crisis may have an impact on the frequency and voluntary of message postings and information disclosure. To incorporate such impact, we impose a dummy variable to discern any pattern shift in message posting activities pre and post the market-wide financial crisis. We first consider December 31, 2007, the National Bureau of Economic Research (NBER) cutoff date of US bull and bear equity markets during our sample period. We then consider another cutoff date October 9, 2007, the peak of US equity market (S&P 500 Index was 1,565.15 and Dow Jones Industrial Average was 14,164.53) prior to the recent global financial crisis. Results are consistent with the imposed dummy variable and our findings are unchanged.

5. In addition to messages collected during the 24-hour session from 4:00:01 p.m. on the previous day to 4:00:00 p.m. on the next day, we consider using messages posted only during the trading-hour session (9:30:00 a.m. to 4:00:00 p.m. on the same day) to synchronize with trading variables. It has been documented that online traders seldom post in off-hours (Antweiler and Frank, 2004). Although using trading-hour session messages slightly reduces sample observations, our findings are unchanged when using trading-hour session messages.

6. To capture any drift that might occur after the event day, we use a post-event 255-day estimation window instead of the pre-event window. As a standard procedure, we require three days as the minimum length of the estimation window. Please see MacKinlay (1997) and Cowan et al. (1990) for a justification of using a post-event window. To also capture any drift that might occur before the event day, we conduct the event study using a more comprehensive pooled estimation window, including the pre-event 255 days and the post-event 255 days, a total of 510 days. We run this model over pre, post and pooled estimation windows. These additional tests do not change the results appreciably.

7. For each regression model, we test for multicollinearity by calculating the un-centered variance inflation factors (VIFs) for the independent variables. We find that the VIFs do not indicate the presence of multicollinearity. Further, although sentiment index is not functionally correlated with disagreement index, we orthogonalize the two indexes and include only the residual as a proxy for disagreement. Turnover ratio and volume-weighted volatility is also shown to be positively correlated in Table 4 (0.1411), we orthogonalize these two variables and include only the residual as a proxy for volatility. We reach consistent findings.

8. The posting of a message does not imply an actual trade at the same time. Also, some online traders might use limit orders rather than market orders. Some of these limit orders may be reflected only in bid-ask prices but not in the actual trade price. Therefore, as an alternative to volume-weighted volatility of actual trade price, we use the standard deviation of midpoints of bid-ask quotes as a measure of volatility. We find consistent results. In addition, we use the realized volatility as proposed in Andersen et al. (2001) based on 15-minute intraday returns, our results hold.

9. Antweiler and Frank (2004) control for the first trading day after a weekend or a holiday. Also, previous studies find that people post more messages on Fridays. In addition, every third Friday is an option expiry day, which may induce more online talks. We include dummies for these days as control variables. Finally, since some investors prefer trading options over stocks, we include a dummy variable to control whether a stock has options. We find that the inclusion of these dummy variables does not qualitatively alter our results.

10. In addition, we have nine industry categories in our sample. For each stock i, we control for its industry effects using eight industry dummy variables, to , corresponding to basic materials (bm), services (se), consumer goods (cg), financial (fi), healthcare (hl), industrial goods (ig), technology (te), and utility industries (ut) with Conglomerates as the reference industry. Media coverage could also be biased toward some industries and we have stocks distributed

into nine industries. We control for industry effects in all the lead-lag regressions and news prediction regressions. We find that our results are robust to the inclusion of these industry dummy variables.

CONCLUSIONS

With rich data collected from WallStreetPit stock message board from January 1, 2006 to December 31, 2011, we are able to investigate, to what extent, information content of a stock message board can influence contemporaneous and future stock trading. Specifically, we look at how message board sentiments, disagreement, message volume, and average message length impact raw stock return, intraday volume-weighted volatility, small trade turnover ratio, and standardized bid-ask spread. Previous studies suggest that online talk matters and has significant impact on contemporaneous and future stock trading behaviors. However, no study so far has ever distinguished stock message board effect on different sized stocks. Besides, a good control over news arrival is lacking in prior studies. To fill such voids, we conduct a comprehensive study to examine the asymmetric stock message board effect accounting for news arrival and firm size. We come to three major findings in this study.

First, event study reveals that there is a significantly large and positive market reaction on the discussion day for small stocks regardless of news arrival. For large stocks, a smaller (no) market reaction is detected with (without) news arrival on the discussion day. Second, with lead-lag regressions, we find that when news arrived, posting activities affected trading for all stocks, especially on the discussion day, with a larger impact on small stocks. However, the stock message board effect lasted at most one trading days and disappeared on the second subsequent day. Without news, however, online talks only influenced trading for small stocks but had no impact on large stocks. Third, we show that posting activities helped predict the presence of news stories regardless of size with a stronger predictive power on news occurrence for small stocks.

It is of interest to investigate the causality between message board posting activities and stock trading behaviors. Granger causality test would serve such purpose should a sufficient time series of message board data become available. A better understanding of how and when message board posters place their actual trade orders could also be beneficial to academicians and practitioners. We exclude penny stocks, such as the over-the-counter bulletin board (OB) or pink sheet (PK) stocks, in this study. Online traders might have a high trading preference on low-priced stocks. We also exclude ETFs, mutual funds, and other funds from this study. ETF are gaining their popularity among investors, online traders might also have a high trading preference on them. Studying the stock message board effect on penny stocks and/or funds trading activities will be also beneficial to academicians and practitioners. Finally, understanding how online posters collaborate and interact to promote or manipulate a stock is crucial to stock message board literature. These proposed extensions remain promising areas of future research.

REFERENCES

[1] Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001) The Distribution of Realized Stock Return Volatility, Journal of Financial Economics, 61: 43-76.

[2] Abbasi, A., Chen, H. and Salem, A. (2008) Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums, ACM Transactions on Information Systems, 26: article 12.

[3] Admati, A.R. and Pfleiderer, P.C. (2001) Disclosing Information on the Internet: Is It Noise or Is It News?, Technical report, Graduate School of Business, Stanford University.

[4] Amihud Y. and Mendelson, H. (1986) Asset Pricing and the Bid-ask Spread, Journal of Financial Economics, 17: 223–249.

[5] Argamon, S., Whitelaw, C., Chase, P., Hota, S.R., Garg, N. and Levitan, S. (2007) Stylistic text classification using functional lexical features, Journal of the American Society for Information Science and Technology, 58: 802-822.

[6] Antweiler, W. and Frank, M.Z. (2004) Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards, Journal of Finance, 59: 1259-1295.

[7] Antweiler, W. and Frank, M.Z. (2006) Do US Stock Markets Typically Overreact to Corporate News Stories? Working paper, University of Minnesota.

[8] Barber, B. and T. Odean. (2001) The Internet and the Investor. Journal of Economic Perspectives, 15: 41-54.

[9] Berger, A.L., Della Pietra, V.J. and Della Pietra, S.A. (1996) A maximum entropy approach to natural language processing, Computational Linguistics, 22(1):39-71

[10] Bettman, J., Hallett, A. and Sault, S. (2011). Rumortrage: Can investors profit on takeover rumors on Internet stock message boards? Working paper, Australian National University.

[11] Bird R. and K. Thapa (2010) Stock Message Board and Investors' Reaction to Company News. Working paper, University of Technology Sydney.

[12] Burges, C.J.C. (1998) A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 2: 121-167

[13] Choi, W.S., Cho, J.M. and Seo, J. (1999) Analysis system of speech acts and discourse structures using maximum entropy model, Proceedings of the 37th conference on Association for Computational Linguistics, Morgan Kaufman, San Francisco, 230-237

[14] Clarkson, P.M., Joyce, D. and Tutticci I. (2006) Market reaction to takeover rumour in Internet Discussion Sites, Accounting & Finance, 46: 31-52.

[15] Cook, D.O. and Lu, X. (2009) Information Matters, Noises Don't: Internet Message Boards Affect Stock Returns, Working Paper, University of Alabama.

[16] Cowan, A.R. (1992) Nonparametric Event Study Tests. Review of Quantitative Finance and Accounting, 2: 343–58.

[17] Cowan, A.R., N. Nayar and A.K. Singh (1990) Stock Returns Before and After Calls of Convertible Bonds. Journal of Financial and Quantitative Analysis, 25: 549–54.

[18] Daniel, K. and S. Titman (1997), 'Evidence on the Characteristics of Cross Sectional

[19] Das, S.R. and Chen, M.Y. (2003) Yahoo! for Amazon: Sentiment Parsing from Small Talk on the Web, Proceedings of

the 8th Asia Pacific Finance Association annual conference, Working paper.
[20] Das, S.R. and Chen, M.Y. (2007) Yahoo! For Amazon: Sentiment extraction from small talk on the Web, Management Science, 53 (9): 1375-1388.
[21] Das, S.R., Martinez-Jerez, A. Fa. De. and Tufano, P. (2005) eInformation: A Clinical Study of Investor Discussion and Sentiment, Financial Management, 34(3):103-137
[22] Das, S.R. and Sisk, J. (2005) Financial Communities, Journal of Portfolio Management, 31: 112-123
[23] Delort, J., Arunasalam, B., Leung H. and M. Milosavljevic. 2011. The impact of manipulation in internet stock message boards. International Journal of Banking and Finance, 8: 1-18
[24] Demers, E. and K. Lewellen (2003) The marketing role of IPOs: evidence from internet stocks. Journal of Financial Economics, 68: 413–437
[25] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society, B.39, 1:1-38.
[26] Depken, C.A. and Zhang, Y. (2010) Adverse Selection and Reputation in a World of Cheap Talk, Quarterly Review of Economics and Finance, 50: 548-558.
[27] Dumais, S., Platt, J., Heckerman, D. and Sahami, M. (1998) Inductive Learning Algorithms and Representations for Text Categorization, In Proceedings of the Seventh of ACM-CIKM, 148-155.
[28] Estabrooks, A., Jo, T. and Japkowicz, N. (2004) A Multiple Resampling Method for Learning from Imbalances Data Sets, Computational Intelligence, 20:18-36.
[29] Fang L. and Peress, J. (2009) Media Coverage and the Cross-section of Stock Returns, Journal of Finance, 64: 2023-2052.
[30] Felton, J. and J. Kim. 2002. ''Warnings from the Enron message board.'', Journal of Investing, 11: 29-52.
[31] Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M. and Haussler D. (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data, Bioinformatics, 16:906-914.
[32] Gregg, D.G. and Scott, J.E. (2006) The role of reputation systems in reducing on-line auction fraud, International Journal of Electronic Commerce, 10: 95–120.
[33] Gu, B., Konana, P., Liu, A., Rajagopalan B., and Ghosh, J. (2006) Predictive Value of Stock Message Board Sentiments, working paper, the University of Texas at Austin.
[34] Gu, B., Konana, P., Liu, A., Rajagopalan B., and Chen, HWM (2007) Competition among virtual communities and user valuation: The case of investing-related communities, Information Systems Research, 18, 68-85.
[35] Hirschey M., Richardson, V.J., and Scholz, S. (2000) How "foolish" are internet investors. Financial Analyst Journal, 56:62–69.
[36] Hong, S-J (2000) Information-Processing Costs in Online Stock Trading, Electronic Markets, 10(3): 192 – 196.
[37] Japkowicz, N. and Stephen, S. (2002) The class imbalance problem: A systematic study, Intelligent Data Analysis Journal, 6(5), 429-450
[38] Joachims, T. (1997) A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization. In Proceedings of ICML-97, 14th International Conference on Machine Learning (Nashville, US, 1997), 143–151.
[39] Jones, C., Kaul, G., Lipson, M. (1994) Transactions, volume and volatility. Review of Financial Studies, 7: 631-652.
[40] Karpoff, J.M. (1986) A Theory of Trading Volume, Journal of Finance, 41: 1069-1087.
[41] Kimmel, A. J. 2004. "Rumors and the Financial Marketplace." Journal of Behavioral Finance, 5: 134-141.
[42] Klein, L., Dalko, V. and Wang, M. (2012) Regulating Competition in Stock Markets: Antitrust Measures to Promote Fairness and Transparency through Investor Protection and Crisis Prevention, published by John Wiley & Sons, Inc.
[43] Kohavi, R. and Provost, F. (1998) Glossary of Term, Machine Learning, 30:271-274.
[44] Koski J.L., Rice E.M. and Tarhouni, A. (2008) Day Trading and Volatility: Evidence from Message Board Postings in 2002 versus 1999, working paper, University of Washington.
[45] Kullback, S. and Leibler, R.A. (1951) On information and sufficiency, Annals of Mathematical Statistics, 22:79-86.
[46] Kumar A. and Lee, CMC. (2003) Individual Investor Sentiment and Comovement in Small Stock Returns, working paper, Cornell University Department of Economics.
[47] Leinweber D. J. (2001). Internet Information for Trading-Cost Control. AIMR Conference Proceedings, 2: 52-61.
[48] Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In Machine Learning: ECML-98, Tenth European Conference on Machine Learning, 4-15.
[49] Li, X. and Shi, Z. Innovating web page classification through reducing noise. Journal of Computer Science and Technology, 17(1), 9-17
[50] Litan, R. and A. Rivlin. (2001) Projecting the Economic Impact of the Internet. American Economic Review, 91: 313-317
[51] Loh, L. and Ong, Y-S. (1998) The adoption of Internet-based stock trading: a conceptual framework and empirical results. Journal of Information Technology, 13(2): 81-94.
[52] Lu J. and B. Mizrach. 2011 "Strategic Interaction in a Stock Trading Chat Room." Working paper, Rutgers University.
[53] MacKinlay, A. C. (1997) Event Studies in Economics and Finance. Journal of Economic Literature, 35: 13–39.
[54] McCallum, A.K. (1996) Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering, http://www.cs.cmu.edu/~mccallum/bow.
[55] Maron, M.E. and Kuhns, J.L. (1960) On Relevance, Probabilistic Indexing and Information Retrieval, Journal of the Association for Computing Machinery, 7:216-244.
[56] Mizrach,B. and Weerts, S. (2009) Experts online: An analysis of trading activity in a public Internet chat room, Journal of Economic Behavior and Organization, 70: 266-281.
[57] Nigam, K., Lafferty, J. and McCallum, A.K. (1999) Using Maximum Entropy for Text Classification, In Proceedings of IJCAI99 Workshop on Machine Learning for Information Filtering, 61-67.
[58] Nigam, K., and Hurst, M. (2004) Towards a Robust Metric of Opinion. In Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text, 598-603.
[59] Odean, T. 1999. "Do investors trade too much?" American Economic Review, 89: 1279-1298.
[60] Olson C.F. (1995) Probabilistic indexing for object recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(5): 518-521
[61] Orriols-Puig, A. and Bernad´o-Mansilla, E. (2005) The Class Imbalance Problem in Learning Classifier Systems: A Preliminary Study. In F. R. et al., editor, Genetic and Evolutionary Computation Conference (GECCO2005) workshop program, pages 74–78, Washington, D.C., USA, 2005. ACM Press.
[62] Pang, B. and Lee, L. (2005) Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, In Proceedings of the Annual Meeting of the Association for Computational Linguistics, 115-124
[63] Pang, B., Lee, L., and Vaithyanathain, S. (2002) Thumbs Up? Sentiment Classification using Machine Learning Techniques, In proceedings of the Empirical Methods in Natural Language Processing, 10: 79 – 86.
[64] Pettibone, J. and Pon-Barry, H. (2003) A Maximum Entropy Approach to Recognizing Discourse Relations in Spoken Language, Working paper, The Stanford Natural Language Processing Group. June 6.
[65] Ratnaparkhi, A. (1997) A simple introduction to maximum entropy models for natural language processing, Technical Report 97-08, Institute for Research in Cognitive Science, University of Pennsylvania.
[66] Ratnaparkhi, A. (1998) Maximum Entropy Models for Natural Language Ambiguity Resolution, PhD thesis, University of Pennsylvania
[67] Raychaudhuri, S., Chang, J.T., Sutphin, P.D. and Altman, R.B. (2002) Associating Genes with Gene Ontology Codes Using a Maximum Entropy Analysis of Biomedical Literature, Genome Research, 12:203-214
[68] Sabherwal, S., Sarkar, K.S and Zhang, Y. (2008) Online-talk, Does It Matter? Managerial Finance, 34(6) 423-436.

[69] Sebastiani, F. (2002) Machine Learning in Automated Text Categorization. ACM Computing Surveys, 34: 1-47.
[70] Schumaker, R.P. and Chen, H. (2006) Textual Analysis of Stock Market Prediction Using Financial News Articles. Americas Conference on Information Systems, Acapulco, Mexico.
[71] Schumaker, R.P. and Chen, H. (2008) Textual Analysis of Stock Market Prediction Using Breaking Financial New: The AZFinText System. Association for Computing Machinery Transactions on Information Systems, 27(2).
[72] Spiegel, U., T. Tavor, and Templeman J. (2010) The Effects of Rumors on Financial Market Efficiency. Applied Economics Letters, 17: 1461-1464.
[73] Sprenger, Timm O., Andranik Tumasjan, Philipp G. Sandner, Isabell M. Welpe (2013) Tweets and Trades: the Information Content of Stock Microblogs, European Financial Management, Forthcoming.
[74] Takahashi, S., Takahashi, M., Takahashi, H. and Tsuda, K. (2007) Analysis of the Relation Between Stock Price Returns and Headline News Using Text Categorization, Lecture Notes In Artificial Intelligence, 4693/2007, 1339 – 1345.
[75] Teich, J., Wallenius, H. and Wallenius J. (1999) World-Wide-Web technology in support of negotiation and communication, International Journal of Technology Management, 17 (1-2):223-239.
[76] Thomas, J.D. and Sycara, K. (2002) Integrating Genetic Algorithms and Text Learning for Financial Prediction. Genetic and Evolutionary Computation Conference (GECCO). 2002. Las Vegas, NV.
[77] Tumarkin, R. and Whitelaw, R.F. (2001) News or Noise? Internet Message Board Activity and Stock Prices, Financial Analysts Journal, 57(3), 41-51.
[78] Tumarkin, R (2002)—"Internet Message Board Activity and Market Efficiency: A Case Study of the Internet Service Sector using RagingBull.com", Financial Markets Institutions and Instruments, 11, 313-335.
[79] Turney, P. D. and Littman, M. L. (2003) Measuring praise and criticism: inference of semantic orientation from association, ACM Transactions on Information Systems, 21(4): 315-346.
[80] Van Bommel, J. (2003) Rumors. Journal of Finance, 58: 1499–1519.
[81] Vapnik, V. (1995) The Nature of Statistical Learning Theory, Springer-Verlag: New York.
[82] Vilardo, M.F. (2004) Online impersonation in securities scams. IEEE Security and Privacy, 2(3):82-85.
[83] White, H.L. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica , 48: 817-838.
[84] Wiebe, J., Wilson, T., Bruce, R., Bell, M. and Martin, M. (2004) Learning subjective language, Computational Linguistics, 30(3): 277-308.
[85] Wilson, T., Wiebe, J., and Hoffman, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), 347-354.
[86] Yi, L., Liu, B. and Li, X. (2003) Eliminating noisy information in Web pages for data mining, Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, D.C. Session: Research track, 296 – 305.
[87] Wysocki, P.D. (1999) Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards, working paper, University of Michigan.
[88] Zhang, Y. and Swanson, P.E. (2010) Are Day Traders Bias Free?—Evidence from Internet Stock Message Boards, Journal of Economics and Finance, 34:96–112.
[89] Zheng, R., Li, J., Huang, Z., and Chen, H. (2006) A Framework for Authorship Analysis of Online Messages: Writing-style Features and Techniques, Journal of the American Society for Information Science and Technology, 57(3): 378-393.

★★★