The Effect of Social and News Media Sentiments on Financial Markets

Regular Paper

Tianyou Hu
University of Auckland
Address
Owen G Glenn Building, 12 Grafton Road,
Auckland, New Zealand
t.hu@auckland.ac.nz

Arvind Tripathi
University of Auckland
Address
Owen G Glenn Building, 12 Grafton Road,
Auckland, New Zealand
a.tripathi@auckland.ac.nz

Abstract

As new technologies continue to push towards democratization of knowledge, various platforms have been vying to become the premiere source of knowledge in this knowledge driven economy. At one end, we have online discussion forums morphing into social media platforms, where retail investors flock to seek and share their opinions about financial markets, a classical example of a *self-sustainable knowledge factory*, where community generates and consumes the knowledge. On the other end, we have news media platforms, fighting to maintain their status as premier knowledge sources, touting domain experts as their contributors. This research compares the effect of these two knowledge sources on financial markets. We use text mining methods to capture the sentiments revealed on the most popular stock discussion forum (HotCopper) and a very popular news media platform (Google Finance) in Australia. We find that both of sentiments from social media and news media predict future stock returns at the individual firm level. But the effect of social media appears to stronger and long lasting compared to news media.

Introduction

Instead of only focusing on experts' recommendations, retail investors increasingly turn to other fellow investors when looking for recommendations for investment. With the emergence of social media and high volume of user-generated content, stock discussion forums have become a good source of information for investors. Traditionally, financial analysis has been the domain of professional forecasters but now is increasingly performed and broadcasted by retail investors (Chen et al. 2014). However, we are yet to understand, the value and effect of information generated by retail investors on stock discussion forums. While a few recent studies show significant positive relationships between number and sentiment of board messages and returns of underperforming small caps stocks (Leung and Ton, 2015), others find no evidence that investor sentiment forecasts future stock returns (Kim and Kim, 2014). In this research, we investigate whether messages about large capitalization stocks generated on HotCopper (HC) have any influence on the Australian stock market.

Even with the proliferation of social media, news media are still an important source for investors to get recommendations for stock markets. Tetlock (2007) collected data from Wall Street Journal and found that high media pessimism leads to downward pressure on market prices followed by a reversion to the fundamentals. Instead of showing the effect of media sentiments, Fang and Peress (2009) argue that stocks with no media coverage have higher returns than stocks with high media coverage. Therefore, we are yet to understand the effect of news media on financial markets.

With the abundant research on the effect of social media and news media on stock markets, we are yet to understand how these two different information sources stack-up in predicting market trends.

In this research, we compare the performance of social media and news media in predicting the market returns of 46 large cap stocks in the Australian market. Traditional news media, such as online news media, are limited in their influence with lack of information sharing and other tools available on social media platforms. In contrast, social media provides tools for sharing and searching which contributes to information diffusion (Westen, 2000; Rubin & Rubin, 2010).

We find a strong relationship between sentiments on stock discussion forum and individual stock returns. Bullish sentiments from stock discussion forum contribute to higher stock returns with 1 to 3 days holding time. We also find that bullish sentiment from Google Finance is followed by higher returns with one holding day. Thus, sentiments on both social and news media platforms influence market returns. We also find that the sentiments from social media have a stronger and longer effect than news media.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature, section 3 describe the social media messages, news media articles, data collection strategy and summary statistics of our dataset. Later we discuss our research model, methodology employed and results. We conclude with discussion of our results and direction for future research.

Literature Review

This paper compares the effect of social media and news media on individual stock returns. Advancement in technologies and applications in the last decade has contributed to more connected and intertwined World via a wide range of online applications. In particular, social media platforms have created a "web of communities" and has gained interest of researchers, businesses and policy makers. Tumarkin and Whitelaw (2001) suggest that message board opinions and stock returns are linked on days of abnormal board activity, with no evidence that opinion predicts future returns. Antweiler and Frank (2004) show that higher discussion forum posting volume will be followed by significant negative returns on the following day, but with small economic impact. Das and Chen (2007) argue that the combined high-tech sector sentiment is linked with high-tech sector index returns, but not for single stocks. Chen et al. (2014) focus on an equity review website (www.seekingalpha.com) and demonstrate that views expressed in expert articles and users comments predict future stock returns and earnings. Leung and Ton (2015) study a stock message board (HC) and find that the number of messages and message sentiment have an effect on the contemporaneous stock returns. To sum up, these studies show that sentiments on social media platforms affect returns of small and large capitalization stocks. However, there is also another research showing no evidence that investor sentiment forecasts future stock returns (Kim and Kim 2014). In this research, we compare the effect of social and new media platforms on market returns.

A few studies provide evidence why social media could influence the financial markets. For instance, Tumarkin and Whitelaw (2001) argue that company or sector professionals may want to disseminate value-relevant information on the internet, perhaps they have framed a long position in these stocks themselves. Boehme et al. (2009) show that online investors are more likely to disseminate information about stocks they are about to buy, instead of spreading false information to earn a profit because of the high cost of or the prohibitions on short selling.

Instead of only focusing on social media, more traditional news media such as newspaper, TV, and online news media have also been taking a leading role in propagating information to a broader investor group (Fang and Peress 2009b). Much recent research has studied the effect of media on stock markets. Barber and Loeffler (1993) analyse the Wall Street Journal column and observe average positive abnormal returns of 4 percent for the two days following the publication of the recommendation. Huberman and Regev (2001) study a Sunday New York Times article on a possible improvement of new cancer-curing drugs, which give rise to biotechnology stocks on the following Monday and in the three following weeks. Busse and Green (2002) focus on the Morning Call and Midday Call segments on CNBC TV and find that prices respond to reports within seconds of initial mention, with positive reports fully incorporated within one minute. Tetlock et al. (2008) find that more negative words in news focusing on specific firms predict low firm earnings. Dougal et al. (2012) show that financial journalists have the potential to impact investor behaviour, at least in a short term. Gurun and Butler (2012) demonstrate that local media uses fewer negative words when reporting local companies in comparison with reporting nonlocal companies. Abnormal positive local media slant is strongly linked with firm equity values.

With the abundant research on social media and news media, no research has been conducted in the comparison of sentiment between these two media sources. Whether do they provide different insights for investors? Does the high frequency of information dissemination and users' interaction through social media help it generate more wisdom of crowds, which could be counted on for investment? In this research, we shed light these concerns and help investors better understand social media, news media and the difference between them.

To understand the sentiment and opinions from media platforms, researchers have been trying different approaches. Different machine learning algorithms have been used to classify social media posts and generate bullishness and agreement indexes. One of the first papers using bullishness and agreement indexes is Antweiler and Frank (2004). Antweiler and Frank (2004) use Naïve Bayes (NB) Algorithm to classify posts and propose three ways to generate the bullishness index together with one way to generate the agreement index. Li (2008) uses NB to classify sentences from 10-K and 10-Q into different tone and content groups. Both of the following two research use the Bullishness index from Antweiler and Frank (2004). Kim and Kim (2014) use NB and compute bullishness using two methods. Hu and Tripathi (2015) use NB and Support Vector Machine then compute bullishness and agreement indexes. In this research, we use Bernolli Naïve Bayes to classify messages and articles and use Bullishness and Agreement index to gauge the sentiments of social media and news media.

Data

Basic Summary of the Data

HotCopper (HC) is the largest and most popular stock discussion platform in the Australasian region. A web crawler was used to download and store messages in a database. We focus on 46 companies from the ASX 50 index from January 2014 to March 2015. Four stocks in ASX 50 underwent identity change during the sample period and have been removed from our dataset (Tirunillai and Tellis, 2012). The firms in our sample, in general, are large capitalization stocks. Our dataset contains 43375 messages from HotCopper (http://hotcopper.com.au). We selected HotCopper for this study because all the messages have self-disclosed sentiments by the authors, which made it good for training machine learning classifiers. Downloaded messages contain author sentiments (“None”, “LT Buy”, “ST Buy” “Buy”, “Hold”, “Sell”, “ST Sell” and “LT Sell”), title, posting time, author, content, and ticker symbol of the firm. ST means short term while LT means long term. Table 1 presents the summary statistics of the posted messages. Among all the messages, 64.98% of them reveal sentiment explicitly. In this research, we combine “LT Buy”, “ST Buy”, and “Buy” and term them as “Bullish” sentiment. Same is done for “Bearish” sentiments. We do this because we only need “Bullish” and “Bearish” polarity, not valence of sentiment to compute bullishness index (equation 1). Messages with “Hold” sentiment are discarded in computing the bullishness index following the literature (Antweiler & Frank, 2004). Following this approach, among the messages with the self-disclosed sentiment, “Bullish” messages are 32.26% and “Bearish” messages are 9.62%. We conjecture that retail investors are more likely to reveal “Bullish” instead of “Bearish” sentiment. This observation is consistent with other studies that investors try to use more positive words in messages (Boehme et al. 2009).

We classify collected messages into “Bullish” and “Bearish” using Bernoulli Naïve Bayes (BNB) classifier. The training set is constructed using messages with self-disclosed sentiments.

Table 1. Summary Statistics

#Revealed Messages	%LT Buy	%ST Buy	%Buy	%Hold	%Sell	%ST Sell	%LT Sell	%None
43375	2.83%	0.54%	28.89%	23.10%	9.05%	0.45%	0.12%	35.02%

In this research, we have developed an agent to collect news articles from Google Finance (GF) based on stock tickers. Google Finance covers a wide range of media sites. The summary of the collected messages and news articles is shown in Table 2. We classify messages into “Bullish” and “Bearish” using BNB classifier, which performs best when compared with Multinomial Naïve

Bayes, Liner Support Vector Machine and Support Vector Machine with rdf kernel. To make the results comparable and consistent, we use BNB to classify messages from HC. The training set for news articles is downloaded from Date for Everyone (https://www.crowdflower.com/data-for-everyone/). Contributors viewed a news article and rated the positivity of the article on a scale 1-9 with 1 being negative and 9 being positive. We classified all the articles with score of 1, 2 as negative, and 7, 8, 9 as positive (there are no news with score 10). As a result, we have 746 news articles in the training set, with about equal number of positive and negative articles.

Table 2. Summary for the collected messages from HC and GF

	HC Messages	GF Articles
Total#Messages/Articles	43375	65658
Avg.# WordsPerMessage	64.79	407.2
StDev.#WordsPerMessage	100.74	725.9

Computation of Investor Sentiment from Discussion Forum Data

The standardised bullishness index $Bullishness_{i,t}$ (Antweiler & Frank, 2004) for stock i at time t can be calculated as following:

$$Bullishness_{i,t} = \frac{M_{i,t}^{Bullish} - M_{i,t}^{Bearish}}{M_{i,t}} * LN(1 + M_{i,t}) \tag{1}$$

$M_{i,t}^{Bullish}$ is the number of messages with "Bullish" sentiment, $M_{i,t}^{Bearish}$ is the number of messages with "Bearish" sentiment. $M_{i,t} = M_{i,t}^{Bullish} + M_{i,t}^{Bearish}$ is the total number of relevant messages.

We also compute an agreement index $Agreement_{i,t}$ (Antweiler and Frank 2004) to measure the disagreement between the sentiments of messages. Literature has shown somewhat controversial and conflicting effect of "disagreement" on the trading volume. For example, Harris and Raviv

(1993) showed that "disagreement" could increase trading volume while Milgrom and Stokey (1982) demonstrated that "disagreement" gives rise to 'no trade' behaviour in the financial markets. Antweiler and Frank (2004) proposed a proxy to capture disagreement among message posters, which is given by:

$$Agreement_{i,t} = 1 - \sqrt{1 - \left(\frac{M_{i,t}^{Bullish} - M_{i,t}^{Bearish}}{M_{i,t}}\right)^2} \in [0,1] \quad (2)$$

where 0 represents complete disagreement. They find that greater agreement on a given day is followed by more trades on the next day.

Basic Summary of the Sentiment Index

We report the summary statistics of the sentiment measures in Table 3, where $GF\_Bullishness_{i,t}$ is the bullishness index (equation 1) from all the news articles that appeared on Google Finance for stock i at day t, $GF\_Agreement_{i,t}$ is the agreement index (equation 2) from Google Finance for stock i at day t, $HC\_Bullishness_{i,t}$ is the bullishness index for HotCopper, $HC\_Agreement_{i,t}$ is agreement index for HotCopper.

Table 3. Min, Max, Mean, Standard Deviations

	Min	Max	Mean	Median	S.D.
$GF\_Bullishness_{i,t}$	-2.639	3.933	0.112	0	1.118
$GF\_Agreement_{i,t}$	0	1	0.604	1	0.454
$HC\_Bullishness_{i,t}$	-2.773	3.754	0.648	0.693	1.002
$HC\_Agreement_{i,t}$	0	1	0.783	1.000	0.391

Methodology

Previous studies have focused on a general relationship between social media and financial market activities (Antweiler & Frank, 2004; Das & Chen, 2007; Kim & Kim, 2014; Leung & Ton, 2015) or between news media and financial market variables (Tetlock, 2007; Tetlock et al., 2008; Chen et al., 2014). Researchers have employed the contemporaneous regression with one holding day (Kim & Kim, 2014; Leung & Ton, 2015), and one-day or two-days lead-lag (Antweiler & Frank, 2004; Chen et al., 2014; Leung & Ton, 2015). For holding days, Tetlock et al. (2008) used one holding day, which is a short holding time, while, others have used longer holding times, such as one month to 36 months holding time (Chen et al. 2014).

In this research, we examine how sentiments expressed on online stock discussion forum and news media affect individual stock return with three different holding periods (one, two or three days). We control for market index and firms' characteristics.

$$Ret_{i,t,t+2} = LN\left(\frac{P_{t+2}}{P_{t-1}}\right) \quad Ret_{i,t,t+1} = LN\left(\frac{P_{t+1}}{P_{t-1}}\right) \quad Ret_{i,t} = LN\left(\frac{P_{t}}{P_{t-1}}\right) \tag{3}$$

$$Ret_{i,t,t+2} = \alpha + \beta_1 GF\_Bullishness_{i,t} + \beta_2 GF\_Agreement_{i,t} + \beta_3 Log(MarketCap)_{i,t} + \beta_4 Log(StockIndex)_{i,t} + \beta_5 Ret_{i,t-1} + \varepsilon \tag{4}$$

$$Ret_{i,t,t+2} = \alpha + \beta_1 HC\_Bullishness_{i,t} + \beta_2 HC\_Agreement_{i,t} + \beta_3 Log(MarketCap)_{i,t} + \beta_4 Log(StockIndex)_{i,t} + \beta_5 Ret_{i,t-1} + \varepsilon \tag{5}$$

$$Ret_{i,t,t+2} = \alpha + \beta_1 HC\_Bullishness_{i,t} + \beta_2 HC\_Agreement_{i,t} + \beta_3 GF\_Bullishness_{i,t} + \beta_4 GF\_Agreement_{i,t} + \beta_5 Log(MarketCap)_{i,t} + \beta_6 Log(StockIndex)_{i,t} + \beta_7 Ret_{i,t-1} + \varepsilon \tag{6}$$

$Ret_{i,t,t+2}$ is the raw return of stock i with holding time of three days from day t to day $t+2$. We will also show results for $Ret_{i,t,t+1}$ (holding two days from day t to $t+1$) and $Ret_{i,t}$ (holding one day for day *t*). *t* represents the day on which the message appeared on HC (or GF) or the following trading day if the message/article was posted on a non-trading day. Previous research has used raw return, which is the natural logarithm of the last holding day's adjusted close price divided the adjusted close price on day t-1 (Kim & Kim, 2014; Leung & Ton, 2015). Raw returns are calculated as shown in equation 3. P_t is the adjusted close price of stock *i* on day *t*. $Log(MarketCap)_{i,t}$ is the log of market capitalization, $Log(StockIndex)_{i,t}$ is the log of ASX 50 stock index, $Ret_{i,t-1}$ is the raw return on day *t-1*.

Equation 4, 5 and 6 use the bullishness and agreement indexes to capture the sentiments from social media and news media. All of them have used control variables for market capitalization and stock index. One-day lagged returns have been used to control for possible autocorrelation (Sabherwal et al. 2011).

Results and Contribution

Table 4 shows the summary of results focusing on three regressions (equation 4, 5 and 6). Coefficients are standardised. In equation 4, we use bullishness and agreement index to capture the sentiments of news media. These results show that sentiments on news media platforms positively impact raw returns. Our results show that if the standard deviation (SD) of bullishness index increases by one, then the return will increase by 0.088 for the same day. Prior studies have shown that high media pessimism predicts downward pressure on market prices (Tetlock, 2007) and abnormal positive local media slant is strongly linked with positive increase in firm equity values (Gurun and Butler, 2012). Our findings are consistent with previous results.

In equation 5, we use bullishness and agreement index to capture the sentiments of social media. We demonstrate that raw returns are significantly positively related to the bullishness from HotCopper,which is consistent with previous result (Leung and Ton 2015). Our results show that if the standard deviation (SD) of bullishness index increases by one, then the return will increase by 0.127 for the same day and 0.163 for two or 0.157 for three holding days. These results confirm that sentiment from social media is reflected in market price quite quickly.

Equation 6 puts $GF\_Bullishness_{i,t}$, $HC\_Bullishness_{i,t}$ in the same regression showing that effect of sentiments on social media remains significant on market returns for at least 3 days, while the effect of sentiments on news media on market returns last only for one day. We conjecture that HC is a specialized investor forum where passionate investors contribute based on their own experience and market research and therefore the information quality and focus of the content is likely to be better compared to news articles on Google Finance which tend to discuss both pros and cons of the market and firms. Further, short messages on HC have clear positive and negative sentiments compared to news articles, which tend to be more balanced. Users on social media forums such as HC, are quick to respond and echo sentiments from community leaders quickly leading to bearish or bullish sentiments on the stock discussion forum (HotCopper).

Prior research (Chen et al., 2014) has shown effect of sentiments on social media platforms on market returns. Our research confirms and adds nuances to the extant literature. Focusing on articles written by experts on an equity review website (www.seekingalpha.com), Chen et al. (2014) established the impact of sentiments in expert articles on market returns. In this research, we extended that line of research to examine the impact of sentiments generated in retail investor forums on market returns. Retail investor forums, such as HotCopper used in this study, attract hobbyist and amateur investors and often compete with expert forums, such as SeekingAlpha.com, in terms of value of information. Note that experts are paid to contribute on equity research websites (e.g., seekingAlpha), whereas hobbyists provide information on HotCopper for free. Further, articles written by experts are structured and follow accepted financial jargons; therefore, sentiments of these articles can be analyzed using standard dictionary tools such as L&M dictionary. In contrast, short text messages posted on amateur investor forums that are used in this study, are unstructured, therefore require machine-learning approaches for sentiment analysis. Chen et al. (2014) use Dow Jones News Service, while we use Google Finance news service. Google Finance aggregates a wide coverage of news from many different news media platforms. Fourth, we use all large cap stocks, whereas Chen et al. (2014) use all stocks that have been discussed on their chosen social media. Stocks with large and small caps will be influenced by social media differently (Leung and Ton 2015). Thus, by only focusing on large cap stocks, we control for the heterogeneity emanating due to market cap size.

Conclusion

This research examines the impact of sentiments expressed in social media and news media on market returns. We have collected data from an online discussion forum, HotCopper (HC), and investigated the impact of users' sentiments expressed on this forum on the stocks listed in ASX 50. News articles related to these stocks were collected from Google Finance. Messages and articles are classified using Bernoulli Naïve Bayes classifier and sentiments are analyzed using the bullishness index. Our study shows a significant effect of sentiments from social media and news media on market returns. We find that sentiments from social media have a long lasting and stronger effect on market returns than the sentiments from news media.

Table 4. Summary of Results

	$Ret_{i,t}$			$Ret_{i,t,t+1}$			$Ret_{i,t,t+2}$		
$GF\_Bullishness_{i,t}$	0.088 * (0.039)		0.084 * (0.038)	0.040 (0.039)		0.035 (0.038)	0.059 (0.039)		0.054 (0.038)
$GF\_Agreement_{i,t}$	0.041 (0.038)		0.041 (0.038)	0.034 (0.039)		0.035 (0.038)	0.023 (0.039)		0.024 (0.038)
$HC\_Bullishness_{i,t}$		0.130 ** (0.040)	0.127 ** (0.040)		0.164 *** (0.040)	0.163 *** (0.040)		0.155 *** (0.040)	0.153 *** (0.040)
$HC\_Agreement_{i,t}$		0.0190 (0.039)	0.017 (0.039)		0.007 (0.039)	0.006 (0.039)		-0.010 (0.039)	-0.010 (0.039)
$Log(MarketCap)_{i,t}$	-0.051 (0.039)	-0.0036 (0.038)	-0.041 (0.038)	-0.042 (0.039)	-0.030 (0.038)	-0.032 (0.038)	-0.036 (0.039)	-0.024 (0.038)	-0.028 (0.039)
$Log(StockIndex)_{i,t}$	0.102 ** (0.039)	0.077 ** (0.039)	0.080 * (0.039)	0.115 ** (0.039)	0.087 * (0.039)	0.089 * (0.039)	0.102 ** (0.039)	0.075 . (0.039)	0.077 * (0.039)
$Ret_{i,t-1}$	0.079 * (0.039)	0.064 (0.039)	0.067 (0.039)	0.002 (0.039)	-0.014 (0.038)	-0.012 (0.039)	-0.035 (0.039)	-0.049 (0.039)	-0.047 (0.039)

Standard errors in parentheses; . $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

References

Antweiler, W., and Frank, M. Z. 2004. “Is all that talk just noise? The information content of Internet stock message boards,” *Journal of Finance* (59:3), pp. 1259–1294 (doi: 10.1111/j.1540-6261.2004.00662.x).

Barber, B. M., and Loeffler, D. 1993. “The ‘Dartboard’ Column: Second-Hand Information and Price Pressure,” *The Journal of Financial and Quantitative Analysis* (28:2), Cambridge University Press, p. 273 (doi: 10.2307/2331290).

Boehme, R. D., Danielsen, B. R., and Sorescu, S. M. 2009. “Short-Sale Constraints, Differences of Opinion, and Overvaluation," *Journal of Financial and Quantitative Analysis* (41:02), Cambridge University Press, p. 455 (doi: 10.1017/S0022109000002143).

Busse, J. A., and Clifton Green, T. 2002. "Market efficiency in real time," *Journal of Financial Economics* (65:3), pp. 415–437 (doi: 10.1016/S0304-405X(02)00148-4).

Chen, H., De, P., Hu, Y. Y. (Jeffrey), and Hwang, B.-H. B.-H. 2014. "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media," *Rev. Financ. Stud.* (27), p. hhu001– (doi: 10.1093/rfs/hhu001).

Das, S. R., and Chen, M. Y. 2007. "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science* (53:9), INFORMS, pp. 1375–1388 (available at http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1070.0704).

Dougal, C., Engelberg, J., Garcia, D., and Parsons, C. A. 2012. "Journalists and the Stock Market," *Review of Financial Studies* (25:3), pp. 639–679 (doi: 10.1093/rfs/hhr133).

Fang, L., and Peress, J. 2009a. "Media Coverage and the Cross-section of Stock Returns," *The Journal of Finance* (64:5), pp. 2023–2052 (doi: 10.1111/j.1540-6261.2009.01493.x).

Fang, L., and Peress, J. 2009b. "Media Coverage and the Cross-section of Stock Returns," *The Journal of Finance* (64:5), pp. 2023–2052 (doi: 10.1111/j.1540-6261.2009.01493.x).

Gurun, U. G., and Butler, A. W. 2012. "Don't Believe the Hype: Local Media Slant, Local Advertising, and Firm Value," *The Journal of Finance* (67:2), pp. 561–598 (doi: 10.1111/j.1540-6261.2012.01725.x).

Harris, M., and Raviv, A. 1993. "Differences of Opinion Make a Horse Race," *Review of Financial Studies* (6:3), pp. 473–506 (doi: 10.1093/rfs/5.3.473).

Hu, T., and Tripathi, A. K. 2015. "The Effect of Social Media on Market Liquidity," *SSRN Electronic Journal* (doi: 10.2139/ssrn.2661099).

Huberman, G., and Regev, T. 2001. "Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar," *The Journal of Finance* (56:1), pp. 387–396 (doi: 10.1111/0022-1082.00330).

Kim, S.-H., and Kim, D. 2014. "Investor sentiment from internet message postings and the predictability of stock returns," *Journal of Economic Behavior & Organization* (107), Elsevier B.V., pp. 708–729 (doi: 10.1016/j.jebo.2014.04.015).

Leung, H., and Ton, T. 2015. "The impact of internet stock message boards on cross-sectional returns of small-capitalization stocks," *Journal of Banking & Finance* (55:July 1998), Elsevier B.V., pp. 37–55 (doi: 10.1016/j.jbankfin.2015.01.009).

Li, F. 2008. "The Determinants and Information Content of the Forward-looking Statements in Corporate Filings - A Naive Bayesian Machine Learning Approach," *SSRN Electronic Journal* (doi: 10.2139/ssrn.1267235).

Milgrom, P., and Stokey, N. 1982. "Information, trade and common knowledge," *Journal of Economic Theory* (26:1), pp. 17–27 (doi: 10.1016/0022-0531(82)90046-1).

Rubin, A., and Rubin, E. 2010. "Informed Investors and the Internet," *Journal of Business Finance & Accounting* (37:7-8), pp. 841–865 (doi: 10.1111/j.1468-5957.2010.02187.x).

Sabherwal, S., Sarkar, S. K., and Zhang, Y. 2011. "Do Internet Stock Message Boards Influence Trading? Evidence from Heavily Discussed Stocks with No Fundamental News," *Journal of Business Finance & Accounting* (38:9-10), pp. 1209–1237 (doi: 10.1111/j.1468-5957.2011.02258.x).

Tetlock, P. C. 2007. “Giving content to investor sentiment: The role of media in the stock market,” *Journal of Finance* (62:3), pp. 1139–1168 (doi: 10.1111/j.1540-6261.2007.01232.x).

Tetlock, P. C., Saar-Tsechansky, M., and MacSkassy, S. 2008. “More than words: Quantifying language to measure firms’ fundamentals,” *Journal of Finance* (63:3), pp. 1437–1467 (doi: 10.1111/j.1540-6261.2008.01362.x).

Tirunillai, S., and Tellis, G. J. 2012. “Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance,” *Marketing Science* (31:November 2014), pp. 198–215 (doi: 10.1287/mksc.1110.0682).

Tumarkin, R., and Whitelaw, R. F. 2001. “News or Noise? Internet Postings and Stock Prices,” *Financial Analysts Journal* (57:3), Association for Investment Management and Research, pp. 41–51 (doi: 10.2469/faj.v57.n3.2449).

Westen, T. 2000. “E-Democracy: Ready or Not, Here It Comes,” *National Civic Review* (89:3), pp. 217–228 (doi: 10.1002/ncr.89304).