Impact of Social Media and News Media on Financial Markets

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

Social Media has been a popular platform for individuals to share their opinions of financial markets. Meanwhile, journalists and analysts contribute to news media to express their analysis on financial securities. This research compares the effect of opinions expressed via social media and news media platforms on the subsequent market movement. We use text mining methods to capture the sentiments revealed on the most popular stock discussion form (HotCopper) and a popular news media platform Google Finance. We find that social media does not seem to replicate the information produced on news media. Investor sentiments from social media predict future stock returns at the individual firm level. We find no evidence that analysts' opinions on news media predict subsequent stock returns.

Introduction

Instead of only focusing on experts' recommendations, retail investors increasingly turn to other fellow investors when looking for recommendation for investment. With the emergence of social media and high volume of user-generated content, stock discussion forums have become one of the many sources of information for investors. Financial analysis is traditionally the domain of professional forecasters but now is increasingly performed and broadcasted by retail investors (Chen et al., 2014). However, it is still not clear whether activities of stock discussion forum could predict the market movement. For instance, Leung and Ton (2015) have shown that the number of board messages and message sentiment are significantly and positively related to the contemporaneous returns of underperforming small capitalization stocks. In comparison, Kim and Kim (2014) demonstrate that they find no evidence that investor sentiment forecasts future stock returns. In this research, we investigate whether messages about large capitalization stocks generated on the largest Australian stock discussion forum, HotCopper (HC), have any influence on the Australian stock market movement.

Even with the proliferation of social media, news media is still an important source for investors to get a peek of stock markets. Early work shows that the stock market responds negatively to news about higher real economic activity when the economy is strong (McQueen & Roley, 1993). 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 sentiment, 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 effect of news media on financial markets.

With the abundant research on the effect of social media and news media on stock markets, we don't know if social media outperforms news media in the stock market or the opposite. In this research, we compare the performance of social media and news media in predicting the market return of 46 large cap stocks in the Australian market. Traditional news media, such as newspapers or online news media, limits their influence with limited capability for readers to interact or further spread information. In comparison, social media users could search for specific information and spread information immediately (Westen, 2000; Rubin & Rubin, 2010). As users could spread value-relevant information in social media more quickly than news media, and social media users could receive feedback while most of the news media users could not, we posit that social media activity reflects more quickly into prices than the news media. Our results are consistent with this expectation.

We find a strong relationship between stock discussion forum sentiment and individual stock returns. More bullishness sentiment from stock discussion forum will contribute to higher stock returns, with 1 to 3 days holding time. We also find that sentiment from Google Finance fails to have enough significant influence on the stock returns, with 1 to 3 holding days. Thus, the effect of social media is revealed in stock returns more quickly than news media. We also find that the sentiments from social media and news media are not correlated. Social media does not seem to duplicate the information from news media. Instead, social media provides more market relevant, genuine information in a highly regulated market like the Australian market.

The remainder of this paper is structured as follows. In Section 2 the literature is reviewed. Section 3 describe the social media messages, news media articles and how we extract the data, as well as sample characteristics and how we construct the proxy variables for investor sentiment. Section 4 examines the effect of social media and news media on financial variables in contemporaneous and intertemporal regressions. Section 5 set forth our results and section 6 is conclusion.

Literature Review

This research studies the different effect of social media and news media on individual stock returns. Social media is a heated research topic with the proliferation of the internet and a huge amount of user-generated content. 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. Zhang and Swanson's (2010) findings indicate that day traders' sentiment is not bias-free, which instead contains a buy side bias because they would rather not speak against their own position. Zhang et al., (2012) apply eight text classifiers to stock message board data and find that their new sentiment index provides a significant "same-day positive and next-day negative" directional indicator. Chen et al. (2014) research on an equity review website and demonstrate that views expressed in both articles and commentaries predict future stock returns and earning surprises. 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 returns of small and large capitalization stocks reflect public available social media information quite quickly. However, there is also another research showing no evidence that investor sentiment forecasts future stock returns (Kim & Kim, 2014). In this research, we demonstrate whether social media has an effect on the market movements.

Some research also provides evidence why social media could move the market. 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 the propagation of information to a broad investor group (Fang & Peress, 2009). Much recent research has studied the media effect on stock markets. Barber and Loeffler (1993) analyze 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. Engelberg and Parsons (2011) find that local media coverage strongly forecasts local trading. Dougal et al. (2012) show that financial journalists have the potential to impact investor behavior, at least in a short time interval. 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. Solomon et al. (2014) show that investor relations firms generate more media coverage of positive press releases for their clients and increase announcement returns.

With the abundant research on social media and news media, not research has been conducted in the comparison of sentiment between these two media sources. Whether do they provide different insights for investors? Does social media only duplicate the information already disclosed in news media? 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. One approach is to use machine learning. 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. Zhang et al. (2012) contrast eight widely used text classifiers and come up a sentiment index, which is quite like the bullishness index. 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) propose to use NB and Support Vector Machine and compute bullishness and agreement indexes. In this research, we use Bullishness and Agreement index to gauge the sentiment of social media.

Another approach uses dictionaries to capture the sentiment from varied information sources. There are four widely used dictionaries, which are Henry (2008), Harvard's General Inquirer (GI), Diction, and Loughran and Mcdonald (2011). In the accounting and finance literature, it has been shown that the L&M list has two main advantages over the other three dictionaries. First, it is a relatively comprehensive dictionary, without any apparent missing negative or positive words. Second, it is created with financial communication in mind (Loughran & McDonald, 2015). Kearney and Liu (2014) demonstrate that "the L&M lists have become predominant in more recent studies". Specifically, many papers have used the L&M dictionary to gauge the tone of business communication (Dougal et al., 2012; Ferris et al., 2012; Gurun & Butler, 2012; Mayew & Venkatachalam, 2012; SOLOMON, 2012; Twedt & Rees, 2012; Garcia, 2013; Liu & McConnell, 2013; AHERN & SOSYURA, 2014; Chen et al., 2014; Hillert et al., 2014; Huang et al., 2014; Solomon et al., 2014). In this study, we utilize the L&M to capture the sentiment of 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, so, have been removed from our dataset (Tirunillai and Tellis, 2012). The firms in our sample, in general, are large. Table 1 shows the composition of 46 stocks.

Table 1. Summaries for the 46 companies from ASX 50

Industry	Number	Industry	Number
Financials	17	Health Care	2
Industrials	8	Utilities	2
Materials	8	Consumer Discretionary	1

Energy	4	Information Technology	1
Consumer Staples	2	Telecommunications Services	1

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. 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 2 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 "Buy" sentiment. Same is done for "Sell" sentiments. We do this as we only need "Buy" and "Sell" polarity, not valence, of sentiment in the bullishness index (equation 1). Messages with "Hold" sentiment are discarded in computing the bullishness index as suggested in the literature (Antweiler & Frank, 2004). Following this approach, among the messages with self-disclosed sentiment, "Buy" messages are 32.26% and "Sell" messages are 9.62%. We find that retail investors try to reveal "Buy" sentiment rather than "Sell" sentiment. This is consistent with the finding that investors try to use more positive words in messages (Boehme et al., 2009).

Table 2. 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 range of media sites. The summary of the collected messages and news articles is shown in Table 3.

Table 3. Summary for the collected messages from HC and GF

	HC Messages	GF Articles
Total# Messages (or Articles)	43375	65658
Avg.# Words per message	64.79	407.2
StDev.# Words per message	100.74	725.9

Computation of Investor Sentiment from Discussion Forum Data

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

$$Bullishness_{i,t} = \frac{M_{i,t}^{BUY} - M_{i,t}^{SELL}}{M_{i,t}} * \ln(1 + M_{i,t}) \tag{1}$$

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

In addition, we also compute an agreement index $Agreement_{i,t}$ (Antweiler & Frank, 2004) to measure the disagreement between the sentiments of messages. In the literature, there have been controversial meanings of "disagreement" on the trading volume. Harris and Raviv (1993) showed that "disagreement" could increase trading volume while Milgrom and Stokey (1982) demonstrated that "disagreement" gives rise to ‘no trade’ behavior in the financial markets. Antweiler and Frank (2004) proposed a proxy for capture disagreement among message posters, which is given by:

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

Basic Summary of the Sentiment Index

Table 4 shows the basic characteristics of the sentiment measures, among which $PosNews_{i,t}$ is percentage of positive words from all the news articles that appeared on Google Finance for stock i at time t, $NegNews_{i,t}$ is the percentage of negative words from Google Finance for stock i at time t, $Bullishness_{i,t}$ is the bullishness index from equation 1, $Agreement_{i,t}$ is agreement index from equation 2. $PosNews_{i,t}$ and $NegNews_{i,t}$ are calculated using L&M dictionary (Loughran & Mcdonald, 2011).

Table 4. Min, Max, Mean, Standard Deviations

	Min	*Max*	*Mean*	*Median*	*S.D.*
$PosNews_{i,t}$	0	0.400	0.011	0.009	0.021
$NegNews_{i,t}$	0	0.667	0.018	0.014	0.03
$Bullishness_{i,t}$	-2.773	3.754	0.648	0.693	1.002
$Agreement_{i,t}$	0	1	0.783	1.000	0.391

Table 5 presents the correlations among the sentiment measures.

Table 5. Correlations

	$PosNews_{i,t}$	$NegNews_{i,t}$	$Bullishness_{i,t}$	$Agreement_{i,t}$
$PosNews_{i,t}$	1.000	0.139	0.007	-0.056
$NegNews_{i,t}$	0.139	1.000	0.034	0.014
$Bullishness_{i,t}$	0.007	0.034	1.000	0.226
$Agreement_{i,t}$	-0.056	0.014	0.226	1.000

Methodology

Previous studies have focused on 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). Only a few have examined the relationship between information revealed on social media and news media. For example, Sabherwal et al. (2011) argue that news media coverage has an influence on the message boards posting activity, which indicates that message boards only replicate the information already revealed by news media. However, we show that social media does not replicate the information from news media by comparing the results of equation 4 and 5. Table 5 shows the low correlation between the sentiment from social media and news media.

In the literature, researchers have examined the contemporaneous regression with one holding day (Kim & Kim, 2014; Leung & Ton, 2015), or 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 analyze the sentiment from social media, together with news media, and shed light on the predictive power of these two sources of information on individual stock return with three different holding periods (one to three days). We control for market index and firms' characteristics.

$$\mathrm{Ret}_{i,t,t+2} = \ln(\frac{P_{t+2}}{P_{t-1}}) \quad \mathrm{Ret}_{i,t,t+1} = \ln(\frac{P_{t+1}}{P_{t-1}}) \quad \mathrm{Ret}_{i,t} = \ln(\frac{P_{t}}{P_{t-1}}) \tag{3}$$

$$\begin{aligned}\mathrm{Ret}_{i,t,t+2} &= \alpha + \beta_1 NegNews_{i,t} + \beta_2 PosNews_{i,t} + \beta_3 Log(MarketCap)_{i,t} \\ &+ \beta_4 Log(StockIndex)_{i,t} + \beta_5 Log(Messages)_{i,t} + \beta_6 \mathrm{Ret}_{i,t-1} + \varepsilon\end{aligned} \tag{4}$$

$$\begin{aligned}\mathrm{Ret}_{i,t,t+2} &= \alpha + \beta_1 Bullishness_{i,t} + \beta_2 Agreement_{i,t} + \beta_3 Log(MarketCap)_{i,t} \\ &+ \beta_4 Log(StockIndex)_{i,t} + \beta_5 Log(Messages)_{i,t} + \beta_6 \mathrm{Ret}_{i,t-1} + \varepsilon\end{aligned} \tag{5}$$

$Ret_{i,t,t+2}$ 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 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 in equation 3. P_t is the adjusted close price of stock i for day t. $Bullishness_{i,t}$ is the bullishness index as denoted in equation 1, $Agreement_{i,t}$ is the agreement index as denoted in equation 2, $NegNews_{i,t}$ is the percentage of negative words in GF, $PosNews_{i,t}$ is the percentage of positive words in GF, $Log(MarketCap)_{i,t}$ is the log of market capitalization, $Log(StockIndex)_{i,t}$ is the log of ASX 50 stock index, $Log(Messages)_{i,t}$ is the log of $(1+M_t)$, $Ret_{i,t-1}$ is the raw return on day $t-1$.

Equation 4 uses the L&M dictionary to gauge the sentiment from the news media. Equation 5 uses the bullishness and agreement indexes to capture the tone from social media. Both of them have used control variables for market capitalization. One-day lagged returns have been used in each model to control for possible autocorrelation (Sabherwal et al., 2011).

Results and Contribution

Table 6 shows the summary of results focusing on two regressions (equation 4 and 5). Coefficients are standardized.

We first focus on the results for equation 4. If there are no messages or news posted on any day for a stock, $NegNews_{i,t}$ is treated as neutral on that day (Hu & Tripathi, 2015). In this research, we discard the trading days with no messages posted. It is shown that raw returns are not significantly related to sentiment from news media. The reason could be that the L&M is designed and first used for 10-Ks (annual report). The median 10-K contains about 20,000 words (Loughran & Mcdonald, 2011), in comparison with the median GF article containing 257 words. The limited length of a news article makes it difficult to capture the sentiments by using the L&M dictionary. Previous studies have used L&M dictionary for structured data, such as 10-K, 10-Q (Feldman et al., 2010), columns of newspapers (Dougal, Engelberg, Garcia, & Parsons, 2012; Garcia, 2013), news articles (Liu & McConnell, 2013; Solomon et al., 2014), earnings press release (Huang et al., 2014), analyst report (Twedt & Rees, 2012), conference call (Matsumoto et al., 2011), and IPO prospectus (Ferris et al., 2012; Loughran & McDonald, 2013). We are the first to examine the applicability of L&M dictionary on unstructured text (online news media).

In equation 5, we use bullishness and agreement index to capture the sentiments of an online discussion forum. We demonstrate that raw returns are significantly positively related to $Bullishness_{i,t}$. This is consistent with previous result (Leung & Ton, 2015). Our results show that if the standard deviation (SD) of bullishness index increases by one, then the return will increase by 0.164 for the same day and 0.15 for two or 0.144 for three holding days. These results confirm that sentiment from social media is reflected in market price quite quickly.

Previous research has studied the impact of social media and news media in two separate streams. Chen et al. (2014) study social media and include news media as a control variable. Their data set of social media is an equity review website, whose articles are quite long, and thus appropriate for L&M dictionary. In comparison, many social media messages are quite short, which could not use the L&M approach. In this research, we get the bullishness index of the short messages from social media and test the correlation between the sentiment from social media and news media. To make our points stronger, Table 5 shows that the correlation between *$Bullishness_{i,t}$* and *$PosNews_{i,t}$* (0.007), and the correlation between *$Bullishness_{i,t}$* and *$NegNews_{i,t}$* (0.034) are quite small. Together with the result from table 6, we show that sentiment from social media is more significantly related to market movements than sentiment from news media.

Table 6 also puts *$Bullishness_{i,t}$*, *$PosNews_{i,t}$* and *$NegNews_{i,t}$* in the same regression showing that sentiment from social media leads the market even with the presence of sentiment from news media. Our research differs from Chen et al. (2014) in several ways. First, Chen et al. (2014) use the L&M dictionary to estimate sentiments on social media, while we use the bullishness index (based on author revealed sentiments) instead. Second, Chen et al. (2014) use experts generated articles on an equity review website, while we focus on retail investors contributed messages on a stock discussion forum. Different users join social media for different reasons and have different professional levels. Different type of social media has different type of contents. For instance, an equity research website attracts experts in the industry as this website pays its contributors, and a stock discussion forum like HotCopper, attracts retail investors as they could speak freely and get feedback. Third, Chen et al. (2014) use Dow Jones News Service, while we use Google Finance news service. Google Finance aggregated a wide coverage of news from many different news media platforms. But this high coverage of news could also dilute the sentiments. Forth, 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 & Ton, 2015). Thus, by only focusing on large cap stocks, we remove the bias of the different effect of social media on stocks with large and small caps.

Table 6. Summary of Results

	$Ret_{i,t}$			$Ret_{i,t,t+1}$			$Ret_{i,t,t+2}$		
$PosNews_{i,t}$	-0.021 (0.03)		-0.02 (0.029)	-0.046 (0.03)		-0.044 (0.029)	-0.026 (0.03)		-0.025 (0.029)
$NegNews_{i,t}$	-0.039 (0.03)		-0.045 (0.029)	-0.025 (0.03)		-0.031 (0.029)	0.011 (0.03)		0.006 (0.029)
$Bullishness_{i,t}$		0.164 *** (0.033)	0.166 *** (0.033)		0.15 *** (0.033)	0.152 *** (0.033)		0.144 *** (0.033)	0.145 *** (0.033)
$Agreement_{i,t}$		0.015 (0.038)	0.014 (0.038)		0.039 (0.038)	0.036 (0.038)		0.021 (0.038)	0.019 (0.038)
$Log(Market Cap)_{i,t}$	-0.005 (0.029)	-0.001 (0.029)	0.002 (0.029)	-0.01 (0.029)	-0.006 (0.029)	-0.004 (0.029)	-0.011 (0.029)	-0.006 (0.029)	-0.005 (0.029)
$Log(StockIndex)_{i,t}$	0.119 *** (0.03)	0.094 ** (0.03)	0.094 ** (0.03)	0.121 *** (0.03)	0.096 ** (0.03)	0.096 ** (0.03)	0.108 *** (0.03)	0.086 ** (0.03)	0.085 ** (0.03)
$Log(Messages)_{i,t}$	-0.034 (0.029)	-0.06 (0.038)	-0.06 (0.038)	-0.002 (0.029)	-0.012 (0.038)	-0.013 (0.038)	0.005 (0.029)	-0.013 (0.038)	-0.014 (0.038)
$Ret_{i,t-1}$	0.026 (0.03)	0.016 (0.029)	0.016 (0.029)	-0.017 (0.03)	-0.028 (0.029)	-0.027 (0.029)	-0.04 (0.03)	-0.05 (0.029)	-0.049 (0.029)

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

Conclusion

This research examines the impact of sentiments expressed on online discussion forums and in news media on market returns. We have collected data from an online discussion forum, HotCopper (HC), and investigated the impact of users' sentiments on this forum on the stocks listed in ASX 50. News articles related to these stocks were collected from Google Finance, and sentiments were estimated using L&M dictionary. Our study shows a significant effect of sentiments on an online discussion forum on financial market whereas the effect of new media wasn't significant.

Our results show that sentiments on online investor forums do not seem to mirror the ones from news media. Instead, the sentiments in these two platforms differ from each other. We conjecture that content generated on these platforms comes from different sources. For example, news articles are written by expert/journalists well versed with financial markets whereas the content on online discussion forum comes from ordinary investors.

As with any study, there are limitations in our research that provides opportunities for future research. First, we conduct sentiment analysis using two approaches in this research. Future studies could manually classify an adequate amount of news articles and use machine learning to classify news articles into different sentiment groups and then use the bullishness index for news media. Second, we only focus on the top 50 stocks in respect of market capitalization. Future research could try to include more small cap stocks with a longer sample time interval. Third, this is research conducted on the Australian market. Google Finance may not be the best portal for Australian investors to read news articles. Further research could come up with a more popular news media or try the American market and use Dow Jones News Service as the source for news media.

References

Ahern, K.R. & Sosyura, D.(2014). "Who Writes the News? Corporate Press Releases during Merger Negotiations". *The Journal of Finance*, 69(1), pp.241–291. Available at: http://doi.wiley.com/10.1111/jofi.12109 [Accessed January 9, 2015].

Antweiler, W. & 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.

Barber, B.M. & Loeffler, D.(1993). "The "Dartboard" Column: Second-Hand Information and Price Pressure". *The Journal of Financial and Quantitative Analysis*, 28(2), p.273. Available at: http://journals.cambridge.org.ezproxy.auckland.ac.nz/abstract_S0022109000008498 [Accessed April 30, 2016].

Boehme, R.D., Danielsen, B.R. & Sorescu, S.M.(2009). "Short-Sale Constraints, Differences of Opinion, and Overvaluation". *Journal of Financial and Quantitative Analysis*, 41(02), p.455. Available at: http://journals.cambridge.org/abstract_S0022109000002143 [Accessed April 29, 2016].

Busse, J.A. & Clifton Green, T.(2002). "Market efficiency in real time". *Journal of Financial Economics*, 65(3), pp.415–437. Available at: http://www.sciencedirect.com/science/article/pii/S0304405X02001484 [Accessed April 30, 2016].

Chen, H. et al.(2014). "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media". *Rev. Financ. Stud.*, 27, p.hhu001–. Available at: http://rfs.oxfordjournals.org/content/early/2014/02/04/rfs.hhu001.abstract.html?papetoc\nhttp://rfs.oxfordjournals.org/content/27/5/1367.abstract.html?etoc.

Das, S.R. & Chen, M.Y.(2007). "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web". *Management Science*, 53(9), pp.1375–1388. Available at: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1070.0704 [Accessed March 12, 2015].

Dougal, C. et al.(2012). "Journalists and the Stock Market". *Review of Financial Studies*, 25(3), pp.639–679. Available at: http://rfs.oxfordjournals.org/content/25/3/639.abstract [Accessed March 25, 2015].

Engelberg, J.E. & Parsons, C.A.(2011). "The Causal Impact of Media in Financial Markets". *The Journal of Finance*, 66(1), pp.67–97. Available at: http://doi.wiley.com/10.1111/j.1540-6261.2010.01626.x [Accessed April 30, 2016].

Fang, L. & Peress, J.(2009). "Media Coverage and the Cross-section of Stock Returns". *The Journal of Finance*, 64(5), pp.2023–2052. Available at: http://doi.wiley.com/10.1111/j.1540-6261.2009.01493.x [Accessed September 15, 2015].

Feldman, R. et al.(2010). "Management's tone change, post earnings announcement drift and accruals". *Review of Accounting Studies*, 15(4), pp.915–953.

Ferris, S.P., Hao, Q. & Liao, M.-Y.(2012). "The Effect of Issuer Conservatism on IPO Pricing and Performance". *Review of Finance*, 17(3), pp.993–1027. Available at:

http://rof.oxfordjournals.org/content/17/3/993.abstract [Accessed March 25, 2015].

Garcia, D.(2013). "Sentiment during Recessions". *The Journal of Finance*, 68(3), pp.1267–1300. Available at: http://doi.wiley.com/10.1111/jofi.12027 [Accessed March 25, 2015].

Gurun, U.G. & 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. Available at: http://doi.wiley.com/10.1111/j.1540-6261.2012.01725.x [Accessed February 6, 2015].

Harris, M. & Raviv, A.(1993). "Differences of Opinion Make a Horse Race". *Review of Financial Studies*, 6(3), pp.473–506. Available at: http://rfs.oxfordjournals.org/content/6/3/473.abstract [Accessed November 20, 2015].

Henry, E.(2008). "Are Investors Influenced By How Earnings Press Releases Are Written?". *Journal of Business Communication*, 45(4), pp.363–407. Available at: http://job.sagepub.com/content/45/4/363.short [Accessed March 25, 2015].

Hillert, A., Niessen-Ruenzi, A. & Ruenzi, S.(2014). "Mutual Fund Shareholder Letter Tone Do Investors Listen?". *SSRN Electronic Journal*. Available at: http://papers.ssrn.com/abstract=2524610 [Accessed March 25, 2015].

Hu, T. & Tripathi, A.K.(2015). "The Effect of Social Media on Market Liquidity". *SSRN Electronic Journal*. Available at: http://papers.ssrn.com/abstract=2661099 [Accessed November 25, 2015].

Huang, X., Teoh, S.H. & Zhang, Y.(2014). "Tone Management". *The Accounting Review*, 89(3), pp.1083–1113. Available at: http://aaajournals.org/doi/abs/10.2308/accr-50684 [Accessed March 25, 2015].

Huberman, G. & 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. Available at: http://doi.wiley.com/10.1111/0022-1082.00330 [Accessed April 30, 2016].

Kearney, C. & Liu, S.(2014). "Textual sentiment in finance: A survey of methods and models". *International Review of Financial Analysis*, 33, pp.171–185. Available at: http://www.sciencedirect.com/science/article/pii/S1057521914000295 [Accessed February 18, 2015].

Kim, S.-H. & Kim, D.(2014). "Investor sentiment from internet message postings and the predictability of stock returns". *Journal of Economic Behavior & Organization*, 107, pp.708–729. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167268114001206.

Leung, H. & 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), pp.37–55. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0378426615000175.

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*. Available at: http://papers.ssrn.com/abstract=1267235 [Accessed March 23, 2015].

Liu, B. & McConnell, J.J.(2013). "The role of the media in corporate governance: Do the media influence managers' capital allocation decisions?". *Journal of Financial Economics*, 110(1), pp.1–17. Available at: http://www.sciencedirect.com/science/article/pii/S0304405X13001761 [Accessed March 14, 2015].

Loughran, T. & McDonald, B.(2013). "IPO first-day returns, offer price revisions, volatility, and form S-1 language". *Journal of Financial Economics*, 109(2), pp.307–326. Available at: http://www.sciencedirect.com/science/article/pii/S0304405X13000603 [Accessed March 24, 2015].

Loughran, T. & McDonald, B.(2015). "Textual Analysis in Finance and Accounting: A Survey". *SSRN Electronic Journal*. Available at: http://papers.ssrn.com/abstract=2504147 [Accessed March 23, 2015].

Loughran, T. & Mcdonald, B.(2011). "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks". *The Journal of Finance*, 66(1), pp.35–65. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2010.01625.x/full [Accessed March 19, 2015].

Matsumoto, D., Pronk, M. & Roelofsen, E.(2011). "What Makes Conference Calls Useful? The Information Content of Managers' Presentations and Analysts' Discussion Sessions". *The Accounting Review*, 86(4), pp.1383–1414. Available at: http://aaajournals.org/doi/abs/10.2308/accr-10034 [Accessed March 25, 2015].

Mayew, W.J. & Venkatachalam, M.(2012). "The Power of Voice: Managerial Affective States and Future Firm Performance". *The Journal of Finance*, 67(1), pp.1–43. Available at:

http://doi.wiley.com/10.1111/j.1540-6261.2011.01705.x [Accessed March 23, 2015].

McQueen, G. & Roley, V.V.(1993). "Stock Prices, News, and Business Conditions". *Review of Financial Studies*, 6(3), pp.683–707. Available at: http://rfs.oxfordjournals.org.ezproxy.auckland.ac.nz/content/6/3/683.short [Accessed April 29, 2016].

Milgrom, P. & Stokey, N.(1982). "Information, trade and common knowledge". *Journal of Economic Theory*, 26(1), pp.17–27. Available at: http://www.sciencedirect.com/science/article/pii/0022053182900461 [Accessed October 17, 2015].

Rubin, A. & Rubin, E.(2010). "Informed Investors and the Internet". *Journal of Business Finance & Accounting*, 37(7-8), pp.841–865. Available at: http://doi.wiley.com/10.1111/j.1468-5957.2010.02187.x [Accessed May 5, 2015].

Sabherwal, S., Sarkar, S.K. & 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. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-83455212034&partnerID=tZOtx3y1 [Accessed February 11, 2015].

Solomon, D.H.(2012). "Selective Publicity and Stock Prices". *The Journal of Finance*, 67(2), pp.599–638. Available at: http://doi.wiley.com/10.1111/j.1540-6261.2012.01726.x [Accessed March 25, 2015].

Solomon, D.H., Soltes, E. & Sosyura, D.(2014). "Winners in the spotlight: Media coverage of fund holdings as a driver of flows". *Journal of Financial Economics*, 113(1), pp.53–72. Available at: http://www.sciencedirect.com/science/article/pii/S0304405X14000403 [Accessed March 25, 2015].

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.

Tetlock, P.C., Saar-Tsechansky, M. & MacSkassy, S.(2008). "More than words: Quantifying language to measure firms' fundamentals". *Journal of Finance*, 63(3), pp.1437–1467.

Tirunillai, S. & Tellis, G.J.(2012). "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance". *Marketing Science*, 31(November 2014), pp.198–215.

Tumarkin, R. & Whitelaw, R.F.(2001). "News or Noise? Internet Postings and Stock Prices". *Financial Analysts Journal*, 57(3), pp.41–51. Available at: http://www.cfapubs.org/doi/abs/10.2469/faj.v57.n3.2449 [Accessed May 14, 2015].

Twedt, B. & Rees, L.(2012). "Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports". *Journal of Accounting and Public Policy*, 31(1), pp.1–21. Available at: http://www.sciencedirect.com/science/article/pii/S0278425411001189 [Accessed March 25, 2015].

Westen, T.(2000). "E-Democracy: Ready or Not, Here It Comes". *National Civic Review*, 89(3), pp.217–228. Available at: http://doi.wiley.com/10.1002/ncr.89304 [Accessed May 5, 2015].

Zhang, Y. & Swanson, P.E.(2010). "Are day traders bias free?—evidence from internet stock message boards". *Journal of Economics and Finance*, 34(1), pp.96–112. Available at: http://link.springer.com/10.1007/s12197-008-9063-1 [Accessed May 14, 2015].

Zhang, Y., Swanson, P.E. & Prombutr, W.(2012). "MEASURING EFFECTS ON STOCK RETURNS OF SENTIMENT INDEXES CREATED FROM STOCK MESSAGE BOARDS". *Journal of Financial Research*, 35(1), pp.79–114. Available at: http://doi.wiley.com/10.1111/j.1475-6803.2011.01310.x [Accessed May 14, 2015].