A Penny Is Worth a Thousand? Investigating the Relationship Between Social Media and Penny Stocks

KAR KEI LO and MICHAEL CHAU, The University of Hong Kong

Increasingly more investors are seeking information from social media to help make investment decisions. Considering that information on penny stocks is often less reported in traditional media, investors may rely more on social media to obtain such information for investment advice. Although previous research has shown that stock opinions in traditional media is a possible predictor of stock returns, no previous research has considered the effect of the stock opinions in social media on these stocks in terms of future stock performance and the moderation effect of penny stocks. In this research, we studied the relationship between social media and the financial performance of penny stocks. We used the net proportion of positive words in stock articles in social media to help predict the future stock performance for penny stocks. The moderation effect of penny stocks on the net fraction of positive words was found to be significant in short terms, revealing a stronger relationship between social media and stock performance at lower price and market capitalization (MC) levels. Based on the findings, we proposed simple strategies utilizing social media and our measure. The results of our applications will be of interest to individual and institutional investors, shareholders, and regulators.

CCS Concepts: • **Information systems** → **Web mining**; • **Computing methodologies** → **Natural language processing**;

Additional Key Words and Phrases: Text analytics, social media, stock opinions, penny stocks

ACM Reference format:
Kar Kei Lo and Michael Chau. 2019. A Penny Is Worth a Thousand? Investigating the Relationship Between Social Media and Penny Stocks. *ACM Trans. Manage. Inf. Syst.* 9, 4, Article 14 (February 2019), 35 pages.
https://doi.org/10.1145/3309704

1 INTRODUCTION

The use of social media in expressing stock opinions is becoming more prevalent. In contrast to traditional financial news articles that may occur in a sporadic manner for a particular stock, online stock forums serve as a convenient way to exchange ideas and news in real time on potential stocks. In addition to using traditional methods, such as technical and fundamental analyses, more and more investors rely on stock opinions in social media for investment advice. Cogent Research

This project was funded by the Hong Kong Research Grants Council through the General Research Fund (project no. 17502715B).

Authors' addresses: K. K. Lo, Faculty of Business and Economics, The University of Hong Kong, Pokfulam Road, Hong Kong; email: kkthomas@connect.hku.hk; M. Chau, Faculty of Business and Economics, The University of Hong Kong, Pokfulam Road, Hong Kong; email: mchau@business.hku.hk.

2158-656X/2019/02-ART14 $15.00
https://doi.org/10.1145/3309704

showed that 70% of the affluent investors in the United States referenced the social media to make investment decision,[1] reflecting the importance of obtaining financial views through this means.

There is no universal definition of penny stocks; however, *penny stock* generally refers to a security issued by a very small company that trades at less than $5 per share according to the U.S. Securities and Exchange Commission (SEC).[2] Although there is relatively higher risk, penny stocks may be very profitable and some investors often advocate trading such stocks for potential gains. For instance, financial analyst Leeds (2011) mentioned,

> *Penny stocks are big business, even though they sound small. Tens of millions of people trade them, and while some win and some lose, none can argue the place of small stocks on the markets.*

This shows that penny stocks may have potential value despite their small share prices. As pointed out by Kumar (2009), some individual investors may prefer stocks with lottery features involving low-priced stocks. To seek for potential gains, investors may search for more information in social media that offers them investment advice. In the event that relevant information is obtained, the future returns from penny stocks can be very profitable. As a result, penny stocks can be especially interesting in social media research.

Although previous literature has shown that stock opinions in social media may affect future share prices (Antweiler and Frank 2004; Chen et al. 2014; Das and Chen 2007), no literature has investigated such impacts specifically on penny stocks in the social media context. In addition, no previous research has investigated the moderation effect of penny stocks at different price and MC levels on such impacts or developed a potential trading strategy on such stocks by making use of the information from social media. Penny stocks can be profitable, yet the opportunity cannot be easily explored without applying appropriate analytics techniques on the social media data.

In this article, we investigate the effects of online stock opinions on the returns of penny stocks. It is also interesting to study how the moderation effect varies at different price levels of penny stocks. We use a measure based on the net fraction of positive words to help predict the future performance of penny stocks in the short term. The moderation effect of penny stocks on the net fraction of positive words was found to be significant, revealing a stronger relationship between social media and stock performance at lower price and MC levels. Penny stocks also had a reverse moderation effect in longer periods. Penny stocks with smaller MC levels could also have higher returns compared to those with higher MC levels in short terms. The results of our applications will be of interest to individual and institutional investors, shareholders, and regulators.

The rest of the article is organized as follows. We first discuss the relevant literature and some background of our study in Sections 2 and 3. We then explain the methodology of our study and the details of our data in Section 4. The main results and analyses involving our findings and predictions are then presented in Section 5. Next, further discussions on the findings and limitations of our study will be given in Section 6. We then conclude the article and provide some suggestions on future research work in Section 7.

2 LITERATURE REVIEW

2.1 Social Media and the Stock Market

According to the strong form of market efficiency hypothesis proposed by Malkiel and Fama (1970), stock prices should reflect all information, including public and private data. Li et al. (2015)

[1]https://www.businesswire.com/news/home/20130222005037/en/Cogent-Research-Trending---Social-Media-Fuels; accessed February 18, 2019.

[2]https://www.sec.gov/fast-answers/answerspennyhtm.html; accessed February 18, 2019.

showed that leveraging market information in the market surveillance task can be important. Nevertheless, given that information is costly, price cannot reflect the information immediately when it is available (Grossman and Stiglitz 1980). This can also be the reason brokerage analysts' recommendations and unofficial forecasts of earnings per share among traders and investors can have investment values given the market's inefficiency (Bagnoli et al. 1999; Womack 1996). Some studies also showed that the market does not follow a random walk and that stock prices can be predicted to some extent (Butler and Malaikah 1992; Kavussanos and Dockery 2001). As the financial market is often not strongly efficient, using appropriate information from social media to help make investment decisions is becoming more common.

Social media can have an influential role in information dissemination. Capturing the "wisdom of the crowd," social media content is updated at an unprecedented speed and provides first-hand information to investors ahead of other sources (Luo et al. 2013). Luo et al. (2013) showed that social media–based metrics can be a significant leading indicator of firm value, whereas Chen et al. (2014) found that stock opinions in social media can help predict stock returns at an aggregate level.

Social media can also reveal useful information for business (Chau and Xu 2012). In the marketing context, Chen and Xie (2008) found that users' opinions in online consumer reviews can help consumers identify their desirable products. In searching for the desirable social media, both posting volume and quality of the online opinions are important factors for potential investors in investing-related communities (Gu et al. 2007). Another study by Aggarwal et al. (2012) showed that electronic word-of-mouth (WOM) from popular blogs can help ventures get higher financing amounts. Consumer WOM and web visits could also contribute to firm value, and external WOM sources can have a significant impact on retailer's sales (Gu et al. 2012; Luo and Zhang 2013). Online WOM can also develop trust in the consumer, and reviewers who disclose identity-descriptive information can be more helpful (Awad and Ragowsky 2008; Forman et al. 2008). These studies suggested that the WOM effect in social media can be significant and have real impact.

Several studies have investigated the effects of traditional/social media and related Internet resources on the stock market. Table 1 summarizes some of these studies.

From Table 1, it can be seen that future stock returns can be predicted at an aggregate level for the tech-sector and large-cap firms by applying text analytics techniques on stock discussion forums (Antweiler and Frank 2004; Chen et al. 2011; 2014; Das and Chen 2007). Price directions can also be predicted via news articles or news sections (Schumaker and Chen 2009; Schumaker et al. 2012; Tetlock 2007; Tetlock et al. 2008). However, some studies showed that media coverage of company news can affect stock prices that would eventually reflect information gradually (Engelberg and Parsons 2011; Gurun and Butler 2012; Solomon 2012).

Making use of the stock opinions on a social media website called *Seeking Alpha* (SA), Chen et al. (2014) used the proportion of negative words in articles and comments and suggested that such a fraction can predict future abnormal returns. Similarly, Tetlock et al. (2008) made use of negative words in news articles to forecast firm earnings. Proportions of optimistic and pessimistic words in earnings press releases could also signal future firm performance or could reflect the tone of accounting reports (Davis et al. 2012; Li 2008). Text classifier scores were used in some studies to help predict future stock returns (Antweiler and Frank 2004; Sabherwal et al. 2011; Zhang and Swanson 2010). Using Twitter feeds, Bollen et al. (2011) found that changes in public mood state could be tracked using simple text processing techniques. Li et al. (2018) showed out that sentiment of messages was positively affected with contemporaneous daily abnormal stock returns, whereas Stieglitz and Dang-Xuan (2013) discovered that emotionally charged Twitter messages were retweeted more often compared to neutral ones. Hill and Ready-Campbell (2011) also found that the online crowd could, on average, perform better than the S&P 500 in terms of both overall return and risk-adjusted returns. However, Ho et al. (2017) found that models with only social

Table 1. Some Related Literature on Financial Data Analytics on Traditional/Social Media and Related Internet Resources

Type of Media	Media Used	Stocks Involved	Related Findings	Reference
Stock discussion forums	• Yahoo! message board • Raging Bull	45 large-cap firms that together made up the DJIA and XLK[3] in 2000	• The volume of messages has a significant positive effect on volatility of stocks. Its effect on stock returns is negatively significant but economically small. • Larger disagreement on one day results in fewer trades on the next day.	Antweiler and Frank (2004)
News section	• *Wall Street Journal*'s "Abreast of the Market"	From 1984 to 1999	• Downward pressure on market prices is induced for high values of media pessimism. • Unusually high or low pessimism results in a high market trading volume.	Tetlock (2007)
Stock discussion forums	• Yahoo! message board	24 tech-sector stocks in MSH[4] from July to August 2001	• Sentiment in 24 tech-sector stocks can predict returns in the aggregate stock level, but the effect is weak for the individual stock level. • In an aggregate stock level, changes in volume of messages have significant effects on changes in volatility and stock level. In contrast, disagreement in stock messages does not explain volatility.	Das and Chen (2007)
News articles	• *Wall Street Journal* • Dow Jones News Service	>350,000 articles for 980 S&P 500 firms from 1980 to 2004	• Low firm earnings can be forecast using negative words in the financial press. • There is a slight delay in the incorporation of the information embedded in negative words in the stock market. • For stories focused on fundamentals, the earnings and return predictability from negative words is largest.	Tetlock et al. (2008)
News articles	• Yahoo! Finance	2,809 articles from October 26, 2005 to November 28, 2005	• Applying a textual-based processing system, financial news articles partitioned by similar sector groupings are most predictable in measures of closeness, predicted directional accuracy, and simulated trading return.	Schumaker and Chen (2009)
Online social media	• Twitter publics	A time series of daily DJIA closing values from Yahoo! Finance from February 28, 2008 to December 20, 2008	• The accuracy of DJIA predictions can be significantly improved by the inclusion of a specific public mood dimension but not others.	Bollen et al. (2011)
Stock discussion forums and news articles	• Seeking Alpha • *Wall Street Journal*	>7,000 companies from 2005 to 2012	• Opinions from social media can predict future stock returns and earnings surprises. • Sentiment from discussion forums can have a larger impact on stock returns than views in news articles.	Chen et al. (2011, 2014)
News articles	• Various local and national newspapers	Large discount brokerage database from 1991 to 1996	• Local media coverage strongly predicts local trading after controlling for earnings, investor, and newspaper characteristics. • Local trading is strongly related to the timing of local reporting, which is a challenge to nonmedia explanations.	Engelberg and Parsons (2011)
Chat room–like Message Board	• TheLion.com	64 stocks from July 18, 2005 to July 18, 2006	• Message board sentiment can help predict trading-related activities.	Sabherwal et al. (2011)
News articles	• Yahoo! Finance	2,802 articles from October 26, 2005 to November 28, 2005	• Prediction of price direction is easier for subjective news articles compared to objective news articles. • Subjective articles have a 3.3% return using a simple trading engine.	Schumaker et al. (2012)
News articles and Website	• Lexis-Nexis and Wikipedia	375 Wikipedia entries of public companies	• Information aggregation on Wikipedia about public firms could moderate the timing of management disclosure of earnings disappointments and investors' negative reaction to bad news.	Xu and Zhang (2013)
News articles	• Yahoo! Finance	6 firms from October 30, 2012 to January 31, 2013	• The relationships between measures extracted from the forums and subsequent daily firm stock returns are examined to be statistically significant. • Measures at the stockholder group level better explain and predict daily stock returns than aggregate forum-level information.	Zimbra et al. (2015)

(Continued)

[3]DJIA and XLK refer to Dow Jones Industrial Average and Dow Jones Internet Commerce Index, respectively.
[4]MSH refers to the Morgan Stanley High-Tech index.

Table 1. Continued

Type of Media	Media Used	Stocks Involved	Related Findings	Reference
Searching website	• Yahoo! Finance	1,619 firms from September 15, 2011 to December 31, 2013	• Cosearch intensity across supply chain partners could help determine cross-return predictability. • Using returns of supply chain partners with low coattention, the simulated trading strategy could generate a significant and positive return above the market returns.	Agarwal et al. (2017)
Searching website	• Yahoo! Finance	2,900 firms from September 15, 2011 to January 14, 2013	• Search cluster–based habitats could reveal aggregate investment preferences and were more granular than fundamental-based habitats. • Search-based habitats could improve the predictability of return of related stocks.	Leung et al. (2017)
Online social media	• Weibo.com	40 Chinese stocks from October 8, 2013 to March 31, 2014	• There was significant evidence to support the intercorrelated relationship between the social attention and future abnormal return. • There was a higher explaining ability from the social attention to price shocks than the impact of historical abnormal returns on social attention.	Xiao et al. (2017)
News articles	• Yahoo! Finance	45 stocks drawn from two stock indices, namely DJIA and XLK from January 2009 to June 2009 and from June 2011 to June 2012	• The relationship between social media sentiments and stock returns is time varying. • Models with only social media sentiments and market returns perform at least as well as models that include Fama-French and Momentum factors. • There are significant correlations between stocks.	Ho et al. (2017)
Stock microblog messages	• Twitter publics	S&P 100 firms over a period of 7 months	• The sentiment of messages is positively affected with contemporaneous daily abnormal stock returns. • Message volume predicts 15-minute follow-up returns, trading volume, and volatility. • Disagreement in microblog messages positively influences stock features, both in interday and intraday analysis.	Li et al. (2018)

media sentiments and market returns performed as least as well as models that included momentum factors. By performing stakeholder analyses of firm-related web forums, Zimbra et al. (2015) identified distinctive groups of forum participants with shared characteristics expressed in discussion and evaluated their specific opinions and interests in the firm. Using English and Arabic forum postings, Abbasi et al. (2008) applied sentiment classification methodologies and developed an entropy-weighted genetic algorithm for efficient feature selection.

In addition, Xu and Zhang (2013) made use of news articles and Wikipedia and found that information aggregation on Wikipedia about public firms could moderate the timing of management disclosure of earnings disappointments and investors' negative reaction to bad news. Using Yahoo! Finance, Agarwal et al. (2017) showed that cosearch intensity across supply chain partners could help determine cross-return predictability, whereas Leung et al. (2017) pointed out that search cluster–based habitats could reveal aggregate investment preferences and that search-based habitats could improve the predictability of return of related stocks. However, Xiao et al. (2017) found that there was significant evidence to support the intercorrelated relationship between social attention and future abnormal return. Zimbra et al. (2018) conducted a benchmark evaluation of 28 top academic and commercial systems in tweet sentiment classification across five distinctive datasets, and the results revealed that the performances of the systems were lackluster overall.

Although the preceding reviewed studies used some measures of sentiment to predict future returns, the use of sentiment might not always succeed in the prediction. By analyzing the text in investment newsletters, Metrick (1999) found that there was no abnormal short-run performance persistence. Tumarkin and Whitelaw (2001) pointed out that message board activities did not predict industry-adjusted returns or abnormal trading volume for stocks in the Internet service sector. A study by Dewally (2003) showed that there was no evidence that new information was exchanged in forums. In addition, Das et al. (2005) found that sentiment from message boards did not predict returns apparently, but instead the returns drove sentiment.

2.2 Penny Stocks and Social Media

As discussed earlier, some investors may prefer low-priced stocks with lottery features (Kumar 2009). Nofsinger and Varma (2014) found that trading penny stocks was positively related to investor preference for lottery-like stocks. Penny stocks may yield considerable profits given their small prices. In an initial public offering (IPO) context, penny stocks seem more profitable in the short run compared to ordinary stocks (Bradley et al. 2006). Penny stocks can also be targeted by manipulators (Bouraoui et al. 2013; Hanke and Hauser 2008; Hu et al. 2010), implying that such stocks may have a higher risk given their small prices. They can be easily be manipulated, and stock prices were found to rise throughout the manipulation period (Aggarwal and Wu 2006).

Penny stocks can be speculative investments (Beatty and Kadiyala 2003). Urbański et al. (2014) showed that speculative stocks are mostly penny stocks, whereas Konku et al. (2012) showed that penny stocks generally are profitable investments in the short term but very poor investments beyond 1 year. Keasler and McNeil (2010) documented that there were significant market reactions to the stock recommendations particularly for the small capitalization stocks. Koulakiotis et al. (2016) pointed out that available information was first incorporated into large stocks and then absorbed by small-sized stocks, whereas Harris and Pisedtasalasai (2006) showed that the returns and volatilities of smaller stocks have less impact on the future dynamics of large stocks. In addition, Ap Gwilym et al. (2014) found that contemporaneous returns increase with the Google search volume of penny stocks, showing that individual investors' speculative demands can be referenced from penny stocks' searching demand. This finding is consistent with the claim that individual investors are net buyers of attention-grabbing stocks (Barber and Odean 2008). Investors therefore may look for profitable opportunity in penny stocks that have drawn attention in the media. Some investors may even treat certain stocks as a kind of lottery, implying that they may take advantage of a small loss for a small chance of higher gains (Barberis and Huang 2008; Kumar 2009; Markowitz 1952; Shefrin and Statman 2000; Statman 2002).

As information on penny stocks is often less reported in traditional media, investors may search for more information on penny stocks in social media to explore potential investment returns. Because of market inefficiency, obtaining relevant information in social media could be very important for predicting the future price movements of penny stocks. As a result, penny stocks can be a particularly interesting topic in social media research. The effect of stock opinions in the context of social media on the performance of penny stocks is a research gap in the literature. To the best of our knowledge, our study is the first to investigate the relationship between social media and penny stocks.

3 OUR STUDY

In this article, we study the impact of sentiment derived from social media on stock returns of penny stocks. Different from previous research works, such as Chen et al. (2014), our work specifically investigated both positive and negative opinions on penny stocks that are of potential interest to individual investors, institutional investors, shareholders, and regulators. The moderation effect on the net fraction of positive words at different price and MC levels was studied, which to the best of our knowledge had not been performed in any previous research. Considering that most penny stocks involve small companies containing relatively less information on its business outlook and financial situation, social media serve as important data sources for investors to seek investment advice. Social media are also good means for investors to discuss with each other on the future financial performance of penny stocks. Given their low prices, penny stocks may have greater fluctuations in prices. Despite that prices of penny stocks may be manipulated, WOM in the forum can be a useful reference for potential investors, and short-term potential gains may be explored through analyzing the sentiment levels of stock-related information in the forum.

As discussed in our review, negative sentiment has been shown to have an impact on stock performance. As our study specifically investigates penny stocks, potential investors may be interested in both the positive and negative sentiments in the online articles. Unlike institutional investors, individual investors normally buy stocks instead of short-sell. Investors may buy and hold the low-priced stocks and be relatively more interested in the positive content of the articles. It therefore is important to consider both the positive and negative words by seeing whether penny stocks will be affected by the net fraction of positive words in stock discussion forums. Despite being a simple measure, the use of the net fraction of positive words can cater to articles with both positive and negative comments, which are particularly essential to predict future returns of penny stocks.

The major research questions are as follows:

(1) How does the measure of sentiment level in social media help predict returns on penny stocks?
(2) How would the moderation effect differ for penny stocks at different price and MC levels?
(3) Can some simple strategies using our constructed measure be formulated to gain potential profits?

4 METHODOLOGY AND DATA

4.1 Methodology

Our research method involves four steps: developing programs, consolidating the data, running regressions, and making predictions. Our method is similar to those reported in previous studies (e.g., Chen et al. 2014). These steps facilitate both the data collection and data analyses process. In case some additional data or analyses are needed, the relevant step can be repeated. The details are discussed as follows.

Step 1: Develop Programs to Collect Articles and Financial Data

Considering that a large amount of data are involved in both social media (e.g., SA) and financial data (e.g., Compustat), we first selected the potential required fields and data in the first step. Java programs were written to extract the necessary data and store them in the server to facilitate further processing.

Step 2: Consolidate the Data and Perform Calculations

Textual analyses. In the second step, the data extracted were consolidated and textual analyses were conducted. Using the positive and negative word list by Loughran and McDonald (2011) for financial markets, textual analyses were performed to compute the proportions of positive and negative words in the articles and comments. Attention was also paid to the negation cases involving words such as *not, aren't*, and *isn't*. In such cases, the opposite sentiment would be used when calculating the proportions.

Let $Pos_{i,t}$ ($Neg_{i,t}$) be the average fraction of positive (negative) words across all articles published on the selected social media about company i on day t; we construct a measure

$$NPos_{i,t} \equiv Pos_{i,t} - Neg_{i,t},$$

which we call the *net fraction of positive words*[5] across all articles published on social media about company i on day t. We will investigate how it affects the returns of penny stocks in the regression analyses.

[5]For a robustness check, alternative definitions were considered and will be discussed in Section 5.2.

Defining penny stocks. We define a *penny stock* as one with price \$$P$ prior to the publication of the social media article being less than \$$x$ (cut-off price, where $x = 5$ according to the SEC). It should be noted that as the social media platform we use (SA) has an extra scrutiny on stocks trading at less than \$1, there is a limitation on the coverage for such stocks. However, there are higher credentials on investors' confidence in the presence of such scrutiny. Because stock prices may fluctuate around \$$x$ (i.e., P may be larger than or less than x at different time points), we first use the following method to define whether a stock is a penny stock:

(I) Method to define penny stock (Definition 1):

Let an article's posting time = s, current date = t, OP_t = opening price on date t, CP_t = closing price on date t, P = chosen price, and $Pen1(x)$ = indicator variable for penny stock with a cut-off price of \$$x$ based on this definition.

<u>Case 1: s is on a trading day t</u>

(1) If s is before opening of the stock market, set $P = CP_{t'}$, where t' is the closest prior trading date.
(2) If s is on or after opening of the stock market but before closing, set $P = OP_t$.
(3) If s is on or after closing of the stock market, set $P = CP_t$.

<u>Case 2: s is not on a trading day t</u>

Set $P = CP_{t'}$, where t' is the closest prior trading date.
If P in the preceding Cases 1 or 2 is $< x$, set $Pen1(x) = 1$ (penny stock).
Else, set $Pen1(x) = 0$ (non-penny stock).

The reason for using the price before the publication of social media articles is to ensure that all penny stocks are defined based on the information prior to the release of the articles. This facilitates our analyses to see if there are price changes of the penny stocks after the expression of views in the articles.

For a robustness check, we also considered an alternative definition of penny stocks using the average of the previous 30 trading days' closing prices.

(II) Method to define penny stock (Definition 2):

Let the article's posting time = s, current date = t, P = chosen price, and $Pen2(x)$ = indicator variable for penny stock with a cut-off price of \$$x$ based on this definition.

<u>Case 1: s is on a trading day t</u>

(1) If s is before closing of the stock market, set P = average of past 30 trading days' closing prices (excluding day t).
(2) If s is on or after closing of the stock market, set P = average of past 30 trading days' closing prices (including day t).

<u>Case 2: s is not on a trading day t</u>

Set P = average of past 30 trading days' closing prices.
If P in Cases 1 or 2 is $< x$, set $Pen2(x) = 1$ (penny stock).
Else, set $Pen2(x) = 0$ (non-penny stock).

As in Chen et al. (2014), we grouped the articles on a firm-day level based on the articles' publication date. In other words, the observations are grouped by both the firm and day. For instance, if there are three articles commenting on the company Google on the same date, we will treat these articles as one observation and the proportion of positive/negative words will be averaged across the same firm on the same firm day. As the observations are grouped in a firm-day level, the definitions of penny stocks for the articles on the same day may differ depending on whether the article was posted before or after closing of the stock market (although we also checked that the number of such cases is very small). In a firm-day level, we constructed the variables *IPen1*(x) and *IPen2*(x) as follows:

$$
\begin{aligned}
IPen1(x) &= 1(\text{penny stock}) && \text{if } Pen1(x) = 1\ \forall \text{ articles on the same publication date} \\
&= 0(\text{non-penny stock}) && \text{otherwise.} \\
IPen2(x) &= 1(\text{penny stock}) && \text{if } Pen2(x) = 1\ \forall \text{ articles on the same publication date} \\
&= 0(\text{non-penny stock}) && \text{otherwise.}
\end{aligned}
$$

In this way, the definition of penny stocks used in the regressions in the firm-day level would be consistent. Both methods of defining the penny stocks have their own merits. The first definition considers the stock price of the penny stocks on a single day and thus reflects the latest status of the stock. The second definition considers the average of 30 prior closing prices. This will avoid the penny stock status of a stock changing too quickly even when its price fluctuates around $5. As will be discussed in Section 5, including both methods can serve as a robustness check and they show similar results.

Formation of matching portfolios. A value-weighted matching portfolio consisting of similar size, book-to-market ratio, and past 1-year return was formed for each stock in our study as a comparison benchmark, in a way similar to the work by Daniel et al. (1997). Specifically, the stocks listed in NYSE, AMEX, and NASDAQ were placed into 125 portfolios. Abnormal returns, defined as raw returns of the stock minus the returns of its matching portfolios (Chen et al. 2014), were calculated. The idea is to see whether the stock mentioned in the articles can outperform its matching portfolio with similar characteristics. The abnormal return is used as a dependent variable in the regression analyses.

Calculations and derivations of variables for regression analyses. Calculations of variables and other controls (e.g., S&P 500 returns, volatility of stocks, and analysts' upgrades/downgrades) needed for running regression analyses were also performed.

Step 3: Run Regressions and Related Analyses

We investigated how the effects of stock opinions affect raw and abnormal returns on penny stocks using regressions in the third step. Fixed year/month effects together with fixed sector effects were included in the regressions as controls. As will be discussed in detail in Section 5, related regressions were run and graphs were plotted to study the direct and moderation effects. The graphs would then serve to investigate whether there are any strengthening effects when price levels of penny stocks vary.

Robustness checks were performed to check whether our results remained valid using the two definitions of penny stocks discussed earlier. Noting the results from regressions was an important step: Special attention was paid to choose the desired dependent variables for regressions and to see if more variables were needed as controls. If additional data were required to investigate the particular effects, the first and second steps were repeated to produce the required input data for regressions.

Step 4

In the final step, we tested whether we can make use of our findings to perform some predictions. Specifically, can penny stocks be really profitable and earn abnormal returns based on social media information? We used the values of our measure *NPos* to perform some simple predictions. Using

Table 2. List of Major Variables Used in the Analyses

Variable	Descriptions
$Pos_{i,t}$	Average fraction of positive words across all articles published on SA about company i on day t
$Neg_{i,t}$	Average fraction of negative words across all articles published on SA about company i on day t
$NPos_{i,t}$	Average net fraction of positive words across all articles published on SA about company i on day t
$PosCom_{i,t}$	Average fraction of positive words across all SA comments to the SA article about company i from days t to t+1
$NegCom_{i,t}$	Average fraction of negative words across all SA comments to the SA article about company i from days t to t+1
$NPosCom_{i,t}$	Average net fraction of positive words across all SA comments to the SA article about company i from days t to t+1
$IPen1(x)$	Indicator variable for penny stock using a threshold of $\$x$ according to Definition 1 (i.e., treated as penny stock if $Pen1(x) = 1$ ∀ SA articles on the same publication date); $IPen1(x) = 1$ for penny stock. and $IPen1(x) = 0$ if otherwise
$IPen2(x)$	Indicator variable for penny stock using a threshold of $\$x$ according to Definition 2 (i.e., treated as penny stock if $Pen2(x) = 1$ ∀ SA articles on the same publication date); $IPen2(x) = 1$ for penny stock, and $IPen2(x) = 0$ if otherwise
$Ret_{i,t,t+x}$	Holding period return of company i from days t to t+x
$ARet_{i,t,t+x}$	Abnormal return of company i from days t to t+x
$SP5Ret_{t,t+x}$	Holding period return of S&P 500 index from days t to t+x
$UpAnalyst_{i,t}$	Number of analysts upgrading company i on day t
$DnAnalyst_{i,t}$	Number of analysts downgrading company i on day t
$L\_MC_{i,t}$	Indicator for large value of MC of company i on day t; MC (in millions) is defined as price multiplied by the number of shares of company i on day t; $L\_MC_{i,t} = 1$ for companies with MC above the median, and $L\_MC_{i,t} = 0$ if otherwise
$Btm_{i,t}$	Book-to-market value ratio of company i at day t
$Vol_{t,t+x}$	Volatility (standard deviation) of the returns of company i from days t to t+x

simple strategies, the returns for buying penny versus non-penny stocks in both the short and long term would be investigated. The performance of penny stocks using the measure *NPos* would also be compared to that using only negative words, *Neg*.

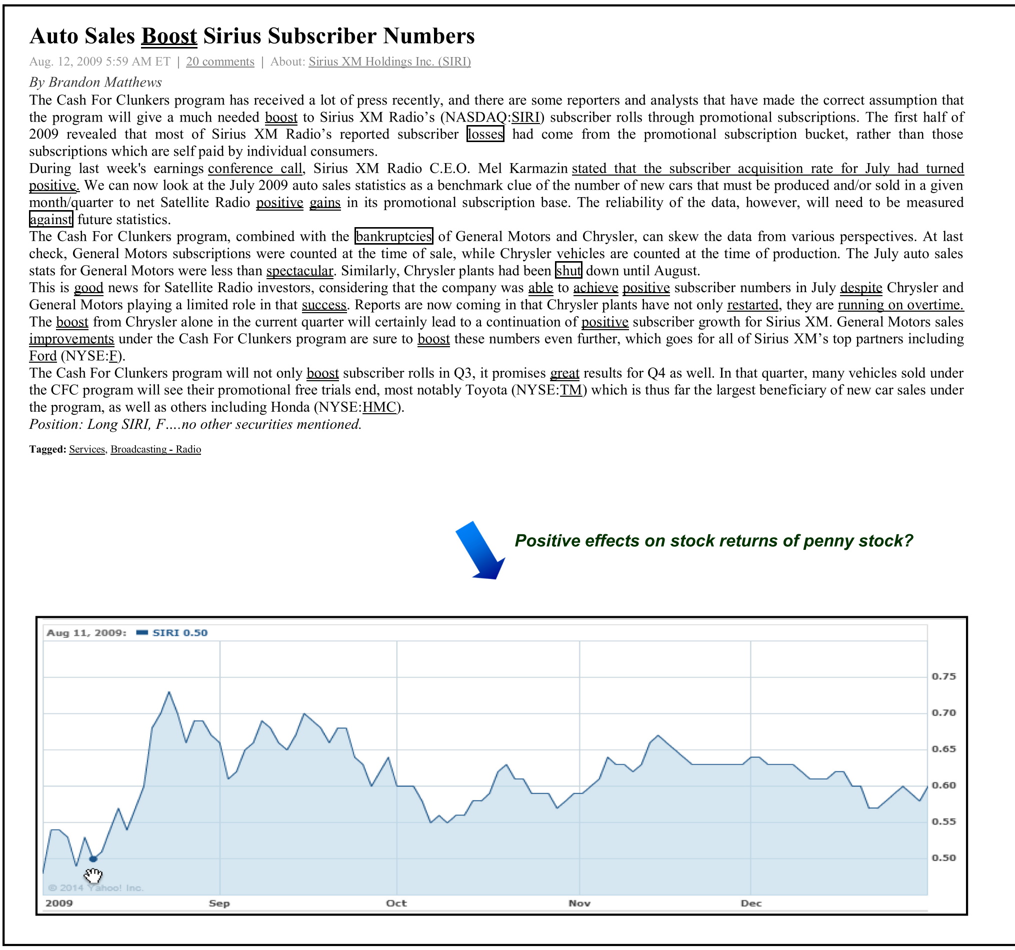
4.2 Data

The sources for the data used in our study consist of two categories:

(1) Social media data from SA (involving contents of the articles and comments, stock tickers, dates of publication, stock sectors, article information, author information, etc.)
(2) Financial data (including opening and closing prices of stocks, MC and book-to-market ratios of firms, S&P 500 indices prices, analysts' upgrade and downgrade information, etc.).

The list of major variables used in the analyses is shown in Table 2.

Social media data. To study the effect of stock opinions in social media, particularly on penny stocks, we need to choose an appropriate data source with stock-related views that can possibly affect stock performance. We chose SA as our data source for stock opinions because it has broad



Auto Sales Boost Sirius Subscriber Numbers

Aug. 12, 2009 5:59 AM ET | 20 comments | About: Sirius XM Holdings Inc. (SIRI)

By Brandon Matthews

The Cash For Clunkers program has received a lot of press recently, and there are some reporters and analysts that have made the correct assumption that the program will give a much needed boost to Sirius XM Radio's (NASDAQ:SIRI) subscriber rolls through promotional subscriptions. The first half of 2009 revealed that most of Sirius XM Radio's reported subscriber losses had come from the promotional subscription bucket, rather than those subscriptions which are self paid by individual consumers.

During last week's earnings conference call, Sirius XM Radio C.E.O. Mel Karmazin stated that the subscriber acquisition rate for July had turned positive. We can now look at the July 2009 auto sales statistics as a benchmark clue of the number of new cars that must be produced and/or sold in a given month/quarter to net Satellite Radio positive gains in its promotional subscription base. The reliability of the data, however, will need to be measured against future statistics.

The Cash For Clunkers program, combined with the bankruptcies of General Motors and Chrysler, can skew the data from various perspectives. At last check, General Motors subscriptions were counted at the time of sale, while Chrysler vehicles are counted at the time of production. The July auto sales stats for General Motors were less than spectacular. Similarly, Chrysler plants had been shut down until August.

This is good news for Satellite Radio investors, considering that the company was able to achieve positive subscriber numbers in July despite Chrysler and General Motors playing a limited role in that success. Reports are now coming in that Chrysler plants have not only restarted, they are running on overtime. The boost from Chrysler alone in the current quarter will certainly lead to a continuation of positive subscriber growth for Sirius XM. General Motors sales improvements under the Cash For Clunkers program are sure to boost these numbers even further, which goes for all of Sirius XM's top partners including Ford (NYSE:F).

The Cash For Clunkers program will not only boost subscriber rolls in Q3, it promises great results for Q4 as well. In that quarter, many vehicles sold under the CFC program will see their promotional free trials end, most notably Toyota (NYSE:TM) which is thus far the largest beneficiary of new car sales under the program, as well as others including Honda (NYSE:HMC).

Position: Long SIRI, F....no other securities mentioned.

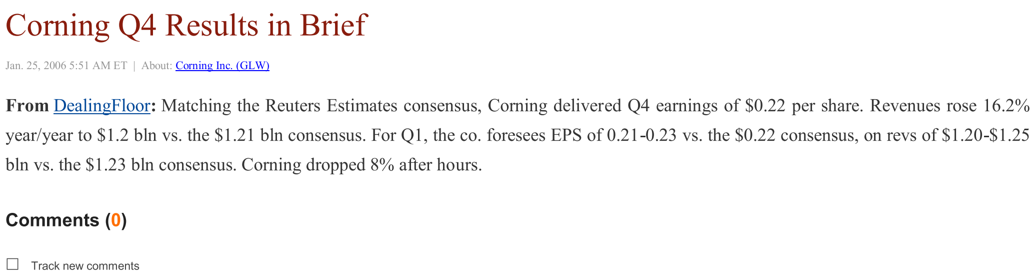
Tagged: Services, Broadcasting - Radio

Fig. 1. An SA article on a penny stock (SIRI) and its stock price chart from Yahoo! Finance.

coverage of stocks, including more than 3,000 small and medium capitalization stocks. It has higher credibility because the editor's picks and the numbers of followers for the authors is included. The articles that the authors submit must interest the readers (e.g., whether the article helps a fundamentally oriented investor decide whether to buy or sell the stock in question, whether the article provides meaningful information about the company) and conform to a standard of rigor and clarity. The author must agree in writing to a disclosure standard. SA is also used in some previous studies to analyze the effect of stock opinions (Chen et al. 2011, 2014; Fotak 2008).

In addition, SA contributors can earn $10 per 1,000 page views and a guaranteed minimum of $500 when an article is classified as "Top Idea". Flexible bonuses are also provided to authors for outstanding contributions. As a result, authors have the incentive to produce high-quality articles to build their reputation. It is also convenient for investors because daily email alerts and SA apps are available for Android, iPhone, and iPad. SA therefore is a suitable platform for our research, which demands credible stock information, particularly for penny stocks.

In SA, stock tickers relevant to an article are listed under the article title. In the example shown in Figure 1, only the stock SIRI is included; we refer to this as a *single-ticker article*. As seen in the "Tagged" section at the bottom of the passage, SA also provides us a convenient way to classify the


Corning Q4 Results in Brief

Jan. 25, 2006 5:51 AM ET | About: Corning Inc. (GLW)

From DealingFloor**:** Matching the Reuters Estimates consensus, Corning delivered Q4 earnings of $0.22 per share. Revenues rose 16.2% year/year to $1.2 bln vs. the $1.21 bln consensus. For Q1, the co. foresees EPS of 0.21-0.23 vs. the $0.22 consensus, on revs of $1.20-$1.25 bln vs. the $1.23 bln consensus. Corning dropped 8% after hours.

Comments (0)

Track new comments

Fig. 2. An SA article excluded from our analyses.

sector of the stock as "Services." In this article, we have double-underlined the positive words and boxed the negative words. According to the textual analysis result of this article, the proportion of positive words far exceeds that of negative ones, reflecting an overall positive view of the penny stock. Comments on the article can also be posted by the readers if they wish.

By looking at the stock price chart, the price prior to publication of the SA article was $0.50 on August 11, 2009. Over just a few weeks after the release of the SA article, the stock price went up to more than $0.70, although it dropped in price later in longer duration. It seems that penny stocks may be profitable given an optimistic view in stock prices as reflected in the article. Nevertheless, this is only one particular sample and cannot be conclusive. More rigid analyses are needed, as detailed in Section 5.

Articles posted on SA from January 2005 to February 2014 were included in our study. We considered stocks listed in NASDAQ, NYSE, and AMEX and excluded funds and indices. To avoid very short, uninteresting, and "advertisement" articles, we only included articles with at least one comment from users posted on the same or next day. This ensured that articles with no comments or no attention from investors, which in other words did not constitute in the price reflection process, were excluded in our analyses. An example of such an excluded article is shown in Figure 2. In addition, Figure 1 shows an SA article on a penny stock together with its stock price chart in Yahoo! Finance.

As shown in Figure 2, the article did not have any comments and possibly did not draw much attention from the readers.[6] The article itself was also quite short and did not involve any personal views. It simply quoted from "Dealing Floor" and acted like an "advertisement" post. This article was deleted from our observations.

Articles in SA are tagged with one or more stock tickers. We only considered single-ticker articles, as textual analyses would be difficult to perform if two or more stocks are mentioned in the same article. In addition, we required that the stocks selected must have data for all the control variables needed in the regressions (excluding analysts' upgrades and downgrades, which will be treated as 0 if missing). In total, we have 51,561 firm-day observations covering 2,872 tickers. The summary statistics on sentiment in SA articles are shown in Table 3.

SA categorizes the stocks into 9 sectors: Basic Materials, Technology, Consumer, Financial, Healthcare, Industrial, Services, Utilities, and Conglomerates. The proportions of each sector in the observations are listed in Table 4.

From Table 4, most of the articles are in technology (e.g., Apple, Yahoo) and services (e.g., Amazon, eBay) sectors with proportions of 23.7% and 20.2%, respectively. To cater to the differential

[6]One may believe that there might exist some articles that could draw attention from readers, yet they were difficult to be commented. We have checked that for popular articles, there were often at least several comments, such as "I agree," and "Well-written," even though it might not be easy to comment on them. We therefore assume that such articles with no comments were possibly those that did not draw much attention from readers.

Table 3. Summary Statistics of Sentiment Level in SA Articles on a Firm-Day Level

Variable	Mean	Median	Standard Deviation
$Pos_{i,t}$	0.013	0.012	0.007
$Neg_{i,t}$	0.013	0.012	0.009
$NPos_{i,t}$	−0.001	0.000	0.012
$PosCom_{i,t}$	0.015	0.011	0.018
$NegCom_{i,t}$	0.014	0.012	0.015
$NPosCom_{i,t}$	0.000	0.000	0.025

Table 4. Proportion of Sectors of the Observations on a Firm-Day Level

Basic Materials	16.1%
Technology	23.7%
Consumer	9.4%
Financial	12.1%
Healthcare	11.1%
Industrial	4.4%
Services	20.2%
Utilities	1.5%
Conglomerates	1.3%

effect of various sectors on stock returns, fixed effects on sectors were included in the regression analyses.

Financial data. Stock price–related data (e.g., opening and closing prices of stocks, MC and book-to-market ratios of firms, S&P 500 indices prices) were obtained from Compustat, whereas analysts' related data (e.g., dates for analysts' upgrades and downgrades, ratings given by analysts) were retrieved from the Institutional Brokers' Estimate System. Using a cut-off price of \$2.50, \$5 (the SEC's definition), and \$7.50, the number of observations treated as penny and non-penny stocks in a firm-day level is listed in Table 5.

Penny stocks constitute about 12% of our observations using a cut-off price of \$5. The proportion remains robust using either Definition 1 or Definition 2. With the cut-off price defined as \$2.50 and \$7.50, the proportions of penny stocks change to around 5% and 17%, respectively. Although penny stocks constitute around 12% in the SEC's definition, we are interested in investigating whether these stocks can in fact earn abnormal profits compared to non-penny ones. As described earlier in Table 2, several financial variables are used in our study. The dependent variables for the regressions are abnormal returns and raw returns of the stocks. A raw return for a stock i (i.e., $Ret_{i,\,t,\,t+x}$) is the holding period return from trading days t to $t+x$, whereas an abnormal return (i.e., $ARet_{i,\,t,\,t+x}$) is the difference between a raw return and the corresponding return of the stock's matching portfolio.

5 MAIN RESULTS AND ANALYSES

5.1 Summary Statistics

The summary statistics of major stock-related variables on a firm-day level are listed in Table 6.

Table 5. Number and Proportions of Penny and Non-Penny Stocks on a Firm-Day Level

	Definition 1		Definition 2	
Stocks	No. of Observations	Proportion	No. of Observations	Proportion
Cut-off price = $2.50				
Penny stocks	2,737	5.3%	2,752	5.3%
Non-penny stocks	48,824	94.7%	48,809	94.7%
Cut-off price = $5 (the SEC's definition)				
Penny stocks	6,343	12.3%	6,307	12.2%
Non-penny stocks	45,218	87.7%	45,254	87.8%
Cut-off price = $7.50				
Penny stocks	8,830	17.1%	8,790	17.0%
Non-penny stocks	42,731	82.9%	42,771	83.0%
Total	51,561	100%	51,561	100%

Table 6. Summary Statistics of Major Stock-Related Variables on a Firm-Day Level

Variable	Mean	Median	Standard Deviation
Dependent variables			
$Ret_{i,t,t+1}$	0.001	0.000	0.038
$Ret_{i,t,t+2}$	0.002	0.001	0.053
$Ret_{i,t,t+3}$	0.002	0.001	0.065
$Ret_{i,t,t+5}$	0.003	0.002	0.082
$Ret_{i,t,t+10}$	0.006	0.004	0.114
$Ret_{i,t,t+15}$	0.009	0.006	0.138
$Ret_{i,t,t+21}$	0.012	0.008	0.162
$Ret_{i,t,t+42}$	0.025	0.017	0.232
$Ret_{i,t,t+63}$	0.035	0.024	0.279
$ARet_{i,t,t+1}$	0.000	0.000	0.034
$ARet_{i,t,t+2}$	0.000	0.000	0.047
$ARet_{i,t,t+3}$	0.001	0.000	0.057
$ARet_{i,t,t+5}$	0.001	−0.001	0.071
$ARet_{i,t,t+10}$	0.001	0.000	0.097
$ARet_{i,t,t+15}$	0.002	−0.001	0.120
$ARet_{i,t,t+21}$	0.002	−0.001	0.140
$ARet_{i,t,t+42}$	0.004	−0.002	0.198
$ARet_{i,t,t+63}$	0.005	−0.006	0.239
Control variables			
$L\_MC_{i,t}$	0.499	0.000	0.500
$Btm_{i,t}$	0.654	0.382	2.429
$Ret_{i,t\text{-}252,t}$	0.180	0.059	0.794
$IPen1(5)$	0.123	0.000	0.328
$Vol_{t\text{-}21,t}$	0.052	0.036	0.067
$UpAnalyst_{i,t}$	0.007	0.000	0.096
$DnAnalyst_{i,t}$	0.008	0.000	0.110
$SP5Ret_{i,t,t+1}$	0.001	0.001	0.012

Table 7. Correlation Matrix of Major Stock-Related Variables on a Firm-Day Level

	$ARet_{i,t,t+1}$	$NPos_{i,t}$	$IPen1(5)$	$NPosCom_{i,t}$	$UpAnalyst_{i,t}$	$DnAnalyst_{i,t}$	$SP5Ret_{t,t+x}$	$ARet_{i,t-1,t}$	$Vol_{t-21,t}$
$ARet_{i,t,t+1}$	1.000								
$NPos_{i,t}$	0.024	1.000							
$IPen1(5)$	0.033	−0.042	1.000						
$NPosCom_{i,t}$	0.009	0.170	−0.001	1.000					
$UpAnalyst_{i,t}$	−0.000	0.001	0.005	0.002	1.000				
$DnAnalyst_{i,t}$	−0.000	−0.005	0.006	−0.001	0.124	1.000			
$SP5Ret_{t,t+x}$	0.010	0.012	0.007	−0.002	−0.009	−0.002	1.000		
$ARet_{i,t-1,t}$	0.032	0.067	0.037	0.021	0.008	−0.008	0.003	1.000	
$Vol_{t-21,t}$	0.001	−0.115	0.227	−0.045	−0.002	−0.004	0.009	0.116	1.000

The correlation matrix for the dependent variable $ARet_{i,\ t,\ t+1}$ and other independent variables on a firm-day level is listed in Table 7.

It seems difficult to earn profits in the stock market because average raw returns are less than 4%, whereas the abnormal returns are even near to 0% for holding periods ranging from 1 to 63 trading days. If stock opinions in forums serve as useful clues in predicting future returns, they can be informative particularly to penny stocks given their small investing prices. Our main results and analyses are discussed in the following.

5.2 Effects of Stock Opinions on Stock Returns for Penny Stocks

We first investigated how stock opinions affect the returns on penny stocks. We ran regressions using abnormal returns as follows:

$$ARet_{i,t,t+x} = \beta_0 + \beta_1 NPos_{i,t} + \beta_2 NPos_{i,t} \times IPen1(5) + \beta_3 IPen1(5) + \beta_4 NPosCom_{i,t} + \gamma C + \varepsilon_{i,t},$$

where C is a vector of control variables including the analysts' upgrades/downgrades, lagged abnormal returns, volatility of stock returns, and S&P 500 index returns. Year and month fixed effects, as well as sector fixed effects, were also included as controls. In our analyses on holding period returns, we considered only trading days and assumed that each month has 21 trading days. Both the holding period abnormal returns for shorter periods from 1 to 15 days ($x = 1$, 2, 3, 5, 10, and 15) and longer periods from 1 to 3 months ($x = 21$, 42, and 63) were studied. Heteroskedasticity robust standard errors were used in the regressions. The results are shown in Table 8.

From Table 8, it can be seen that the coefficient of $NPos_{i,t}$ is highly significant for nearly all regressions, although the significance becomes relatively smaller in longer periods. This shows that our measure $NPos_{i,t}$ can help predict future abnormal returns after controlling for various factors such as analysts' upgrades, volatility of stock returns, and S&P 500 index returns. The use of both negative and positive words is important because although negative voices may draw more attention than positive ones in general (Luo 2007), some investors may buy stocks with low prices to gamble for potential gains (Kumar 2009). As a result, such investors will be interested in both the positive and negative news for gaining considerable profits with a small amount of investment. Considering that information on penny stocks is relatively less available in traditional media, both positive and negative stock opinions can be influential in affecting future stock prices in the short period.

It is important to note that the coefficient of $IPen1(5)$ remains significant for all regressions. This is an interesting finding. One possible explanation is that since the trading volume of penny stocks is low, some reactions on an SA article by a few investors may already make an impact on the stock price, whereas such reactions would be less prominent or even negligible for non-penny stocks since the trading volume is high. As a result, this implies that stock forums do have a role

Table 8. Regression of Abnormal Returns (Using *IPen1*(5))

Variables	(1) $ARet_{i,t,t+1}$	(2) $ARet_{i,t,t+2}$	(3) $ARet_{i,t,t+3}$	(4) $ARet_{i,t,t+5}$	(5) $ARet_{i,t,t+10}$	(6) $ARet_{i,t,t+15}$	(7) $ARet_{i,t,t+21}$	(8) $ARet_{i,t,t+42}$	(9) $ARet_{i,t,t+63}$
$NPos_{i,t}$	0.0652***	0.0861***	0.0971***	0.120***	0.136***	0.158***	0.175***	0.175**	0.145*
	(4.856)	(4.692)	(4.363)	(4.458)	(3.712)	(3.545)	(3.498)	(2.498)	(1.673)
$NPos_{i,t}$×*IPen1*(5)	0.121*	0.0938	0.0162	−0.144	−0.139	−0.723**	−0.975***	−1.850***	−1.185*
	(1.648)	(0.899)	(0.129)	(−0.745)	(−0.567)	(−2.281)	(−2.617)	(−3.408)	(−1.944)
IPen1(5)	0.00355***	0.00553***	0.00525***	0.00647***	0.00996***	0.0137***	0.0152***	0.0250***	0.0402***
	(4.902)	(5.265)	(4.235)	(4.105)	(4.518)	(4.873)	(4.833)	(5.911)	(7.797)
$NPosCom_{i,t}$	0.00667	0.00493	0.00297	0.00319	−0.00220	−0.00578	−0.00331	0.0120	−0.0170
	(1.185)	(0.650)	(0.316)	(0.276)	(−0.145)	(−0.283)	(−0.153)	(0.396)	(−0.475)
$UpAnalyst_{i,t}$	0.000243	0.00106	−0.000585	−0.00115	0.00120	0.00503	0.00505	0.000211	0.00392
	(0.156)	(0.548)	(−0.259)	(−0.320)	(0.199)	(0.672)	(0.670)	(0.0212)	(0.357)
$DnAnalyst_{i,t}$	8.91e-05	4.60e-06	0.000870	−0.00156	−0.00312	−0.00598*	−0.00435	−0.00925	−0.00687
	(0.0862)	(0.00353)	(0.502)	(−0.783)	(−1.069)	(−1.668)	(−1.022)	(−1.537)	(−0.863)
$UpAnalyst_{i,t-1}$	−0.00135	−0.00277	−0.00247	−0.000698	−0.00348	−0.000982	0.000735	0.00387	0.00434
	(-0.861)	(−1.256)	(−1.016)	(−0.206)	(−0.710)	(−0.185)	(0.117)	(0.427)	(0.406)
$DnAnalyst_{i,t-1}$	−0.000412	−0.000433	−0.00193	−0.00248	−0.00288	−0.00136	−0.000550	−0.000620	0.0139*
	(-0.335)	(−0.222)	(−0.815)	(−0.973)	(−0.815)	(−0.344)	(−0.123)	(−0.0982)	(1.655)
$UpAnalyst_{i,t-2}$	−0.00236*	−0.000743	0.00134	0.00462	0.00234	0.00210	0.00709	0.0138	0.0136
	(-1.699)	(−0.392)	(0.486)	(1.142)	(0.384)	(0.317)	(0.951)	(1.452)	(1.180)
$DnAnalyst_{i,t-2}$	−0.000148	−0.00242	−0.00465**	−0.00595**	−0.00501	−0.00791*	−0.00857*	−0.00956	−0.00662
	(-0.124)	(−1.347)	(−2.351)	(−2.457)	(−1.565)	(−1.906)	(−1.684)	(−1.575)	(−0.667)
$SP5Ret_{t,t+x}$	0.0292	0.0315	0.0562***	0.0554***	0.0896***	0.0728***	0.0667***	0.104***	0.103***
	(1.261)	(1.473)	(2.838)	(2.810)	(3.798)	(3.825)	(3.564)	(4.477)	(3.839)
$ARet_{i,t-1,t}$	0.0165	0.0115	0.0122	0.0113	−0.0260	−0.0476	−0.0236	−0.0458	−0.0422
	(1.534)	(1.011)	(0.830)	(0.658)	(−1.121)	(−1.478)	(−0.925)	(−1.180)	(−0.946)
$ARet_{i,t-2,t-1}$	−0.0199**	−0.0425***	−0.0492***	−0.0528***	−0.0932***	−0.122***	−0.0833***	−0.101**	−0.116***
	(-2.425)	(−3.698)	(−3.886)	(−3.222)	(−3.932)	(−3.368)	(−2.627)	(−2.378)	(−2.578)
$ARet_{i,t-3,t-2}$	0.00465	−0.00193	−0.00928	−0.0252	−0.0254	−0.0428	−0.0281	−0.0319	−0.0302
	(0.567)	(−0.140)	(−0.572)	(−1.614)	(−1.073)	(−1.454)	(−0.786)	(−0.656)	(−0.535)
$ARet_{i,t-63,t-3}$	−0.00269***	−0.00280**	−0.00298	−0.00592***	−0.0103***	−0.0119***	−0.00751**	−0.0187***	−0.0196***
	(−2.759)	(−2.085)	(−1.623)	(−2.914)	(−3.904)	(−3.780)	(−2.051)	(−3.785)	(−3.387)
$Vol_{t-21,t}$	0.00157	−0.00653	−0.00330	−0.00624	0.00684	0.0197	−0.0123	0.0168	0.0164
	(0.261)	(−0.801)	(−0.319)	(−0.548)	(0.422)	(0.885)	(−0.537)	(0.625)	(0.466)
Year & month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.00616	0.0142	0.0193	0.0118	0.0306*	0.0239	0.0254	0.0704**	0.00596
	(0.763)	(0.568)	(0.594)	(0.660)	(1.879)	(1.310)	(0.661)	(2.059)	(0.125)
Observations	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561
Adjusted R^2	0.004	0.005	0.004	0.005	0.009	0.011	0.010	0.017	0.020

Robust *t*-statistics are shown in parentheses.
$^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

not only in reflecting views on the stock market but also in helping investors select penny stocks for potential investments.

Nevertheless, we also need to consider the moderation effect of penny stocks on the net fraction of positive words as reflected in the coefficient of $NPos_{i,t} \times IPen1(5)$. Interestingly, penny stocks have a positive moderation effect for shorter terms of 1 to 3 days but a significant negative moderation effect in longer terms. Considering that investors can rely on social media to explore profits for penny stocks, low-priced stocks with more positive views may be preferred by more investors and their prices may consequently rise in a short term. Nevertheless, as pointed out by Zhang et al. (2012), potential gains may only last for a short term and price reversions can exist later after the stocks have been heavily discussed on a stock message board. Price reversions may also happen because the price momentum effect can eventually reverse (Jegadeesh and Titman 1993; Lee and Swaminathan 2000). With the aim of exploring potential gains, investors for penny stocks rationally prefer to acquire such gains in a short term if an opportunity arises. Such an opportunity may last for only short period considering that the potential investor may capture this valuable time in a short duration (e.g., one trading day only). Stock forums therefore can provide instructive guidelines for penny stocks yielding more profitable returns in a short term compared to non-penny ones.

It is interesting to note that although the positive moderation effect only lasts for one trading day, the negative moderation effect lasts for a relatively long period (from t+15 to t+63) with high coefficient values. One possible explanation is that when the stock is mentioned on social media or when there is a news event with the company, investors may overreact and the stock price overshoots. After the potential investment opportunity of the attention-seeking penny stocks has been captured by the investors, it may become "ignored" or even "dumped" and their price may reverse due to the reversion of the price momentum effect. Another possible explanation is that other noises were captured by the abnormal return, as different events might happen over the longer period.

Robustness check: Definition of NPos. Instead of solely using one definition of *NPos*, we also performed regressions using three other definitions of *NPos* (denoted as *NPosA, NPosB*, and *NPosC*) as follows:

$$NPosA \equiv 2Pos_{i,t} - Neg_{i,t}$$
$$NPosB \equiv Pos_{i,t} - 2Neg_{i,t}$$
$$NPosC \equiv (Pos_{i,t} - Neg_{i,t})/(Pos_{i,t} + Neg_{i,t})$$

with the corresponding

$$NPosComA \equiv 2PosCom_{i,t} - NegCom_{i,t}$$
$$NPosComB \equiv PosCom_{i,t} - 2NegCom_{i,t}$$
$$NPosComC \equiv (PosCom_{i,t} - NegCom_{i,t})/(PosCom_{i,t} + NegCom_{i,t}).$$

As our study specifically investigates penny stocks, potential investors may be interested in both the positive and negative sentiments in the online articles. Nevertheless, it would be interesting to investigate whether there are any difference in results if the positive or negative sentiments are given higher weighting. As a demonstration, either the positive or negative sentiment is given a double-weighting of 2 and alternative definitions for *NPos* and *NPosCom* are used. The results were found to be similar, and for the sake of brevity, the results are included in Tables A.1 through A.3 in the Appendix.

Robustness check: Definition of penny stocks. For a robustness check, we also performed regressions using Definition 2 of penny stocks as follows:

$$ARet_{i,t,t+x} = \beta_0 + \beta_1 NPos_{i,t} + \beta_2 NPos_{i,t} \times IPen2(5) + \beta_3 IPen2(5) + \beta_4 NPosCom_{i,t} + \gamma C + \varepsilon_{i,t},$$

where C is a vector of control variables including the analysts' upgrades/downgrades, lagged abnormal returns, volatility of stock returns, and S&P 500 index returns. Year and month fixed effects, as well as sector fixed effects, were also included as controls. The results are shown in Table 9, which are very similar to those in Table 8.

Robustness check: Different cut-off prices as definition. We have been using \$5 as the cut-off in our definition of penny stocks. For a robustness check, we compared the coefficients of $NPos_{i,t} \times$

Table 9. Regression of Abnormal Returns (Using *IPen2*(5))

Variables	(1) $ARet_{i,t,t+1}$	(2) $ARet_{i,t,t+2}$	(3) $ARet_{i,t,t+3}$	(4) $ARet_{i,t,t+5}$	(5) $ARet_{i,t,t+10}$	(6) $ARet_{i,t,t+15}$	(7) $ARet_{i,t,t+21}$	(8) $ARet_{i,t,t+42}$	(9) $ARet_{i,t,t+63}$
$NPos_{i,t}$	0.0614***	0.0828***	0.0945***	0.124***	0.149***	0.171***	0.199***	0.194***	0.171*
	(4.609)	(4.544)	(4.296)	(4.601)	(4.108)	(3.786)	(3.946)	(2.743)	(1.946)
$NPos_{i,t} \times IPen2(5)$	0.152**	0.115	0.0295	−0.196	−0.285	−0.876***	−1.231***	−2.094***	−1.512**
	(2.028)	(1.094)	(0.234)	(−0.992)	(−1.125)	(−2.685)	(−3.201)	(−3.734)	(−2.411)
$IPen2(5)$	0.00308***	0.00466***	0.00409***	0.00567***	0.00844***	0.0127***	0.0150***	0.0246***	0.0375***
	(4.221)	(4.455)	(3.337)	(3.584)	(3.788)	(4.298)	(4.648)	(5.698)	(7.159)
$NPosCom_{i,t}$	0.00669	0.00500	0.00311	0.00326	−0.00203	−0.00570	−0.00338	0.0119	−0.0170
	(1.188)	(0.659)	(0.330)	(0.283)	(−0.134)	(−0.279)	(−0.157)	(0.393)	(−0.473)
$UpAnalyst_{i,t}$	0.000165	0.000958	−0.000645	−0.00123	0.00106	0.00490	0.00488	−2.20e-07	0.00334
	(0.107)	(0.503)	(−0.287)	(−0.346)	(0.178)	(0.661)	(0.655)	(−2.26e-05)	(0.312)
$DnAnalyst_{i,t}$	0.000146	8.17e-05	0.000941	−0.00151	−0.00304	−0.00595*	−0.00438	−0.00929	−0.00667
	(0.141)	(0.0629)	(0.544)	(−0.762)	(−1.045)	(−1.662)	(−1.027)	(−1.546)	(−0.844)
$UpAnalyst_{i,t-1}$	−0.00143	−0.00287	−0.00255	−0.000803	−0.00361	−0.00122	0.000467	0.00338	0.00361
	(−0.915)	(−1.309)	(−1.058)	(−0.238)	(−0.746)	(−0.233)	(0.0748)	(0.379)	(0.344)
$DnAnalyst_{i,t-1}$	−0.000401	−0.000413	−0.00190	−0.00246	−0.00284	−0.00132	−0.000515	−0.000568	0.0140*
	(−0.327)	(−0.212)	(−0.807)	(−0.964)	(−0.804)	(−0.334)	(−0.116)	(−0.0906)	(1.689)
$UpAnalyst_{i,t-2}$	−0.00241*	−0.000808	0.00130	0.00452	0.00219	0.00182	0.00671	0.0132	0.0128
	(−1.763)	(−0.435)	(0.476)	(1.131)	(0.364)	(0.279)	(0.914)	(1.418)	(1.137)
$DnAnalyst_{i,t-2}$	−0.000148	−0.00240	−0.00462**	−0.00589**	−0.00491	−0.00773*	−0.00835*	−0.00916	−0.00622
	(−0.124)	(−1.343)	(−2.344)	(−2.448)	(−1.543)	(−1.877)	(−1.648)	(−1.535)	(−0.636)
$SP5Ret_{t,t+x}$	0.0291	0.0316	0.0564***	0.0555***	0.0896***	0.0724***	0.0660***	0.104***	0.103***
	(1.260)	(1.480)	(2.847)	(2.817)	(3.799)	(3.811)	(3.539)	(4.496)	(3.847)
$ARet_{i,t-1,t}$	0.0160	0.0108	0.0115	0.0105	−0.0270	−0.0495	−0.0256	−0.0497	−0.0475
	(1.482)	(0.949)	(0.789)	(0.615)	(−1.165)	(−1.533)	(−1.004)	(−1.280)	(−1.068)
$ARet_{i,t-2,t-1}$	−0.0204**	−0.0434***	−0.0501***	−0.0542***	−0.0951***	−0.125***	−0.0878***	−0.109**	−0.125***
	(−2.486)	(−3.776)	(−3.958)	(−3.295)	(−3.996)	(−3.450)	(−2.764)	(−2.556)	(−2.784)
$ARet_{i,t-3,t-2}$	0.00408	−0.00289	−0.0102	−0.0266*	−0.0274	−0.0461	−0.0320	−0.0387	−0.0391
	(0.498)	(−0.209)	(−0.629)	(−1.701)	(−1.154)	(−1.562)	(−0.899)	(−0.800)	(−0.693)
$ARet_{i,t-63,t-3}$	−0.00283***	−0.00302**	−0.00321*	−0.00617***	−0.0107***	−0.0124***	−0.00802**	−0.0197***	−0.0211***
	(−2.880)	(−2.241)	(−1.748)	(−3.044)	(−4.027)	(−3.936)	(−2.196)	(−3.996)	(−3.650)
$Vol_{t-21,t}$	0.00204	−0.00562	−0.00196	−0.00540	0.00807	0.0212	−0.0119	0.0186	0.0195
	(0.338)	(−0.685)	(−0.189)	(−0.474)	(0.497)	(0.954)	(−0.520)	(0.690)	(0.551)
Year & month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.00611	0.0142	0.0192	0.0118	0.0306*	0.0241	0.0258	0.0710**	0.00644
	(0.756)	(0.565)	(0.590)	(0.660)	(1.877)	(1.320)	(0.674)	(2.079)	(0.134)
Observations	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561
Adjusted R^2	0.004	0.004	0.004	0.005	0.008	0.011	0.010	0.017	0.020

Robust t-statistics are shown in parentheses.
$^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

IPen1(x) and ran the abnormal return regressions for $x = 2.5, 3, 3.5, \ldots, 7, 7.5$. The results of the coefficients are shown in Table 10 and Figure 3.

Table 10 and Figure 3 give consistent results as expected. We also repeat the analyses using Definition 2 of penny stock. Results are shown in Table A.4 and Figure A.1 in the Appendix, revealing that our findings are robust.

Robustness check: Raw returns. As a comparison, we also ran regressions using raw returns as follows:

$$Ret_{i,t,t+x} = \beta_0 + \beta_1 NPos_{i,t} + \beta_2 NPos_{i,t} \times IPen1(5) + \beta_3 IPen1(5) + \beta_4 NPosCom_{i,t} + \gamma C' + \varepsilon_{i,t},$$

Table 10. Regression Coefficients of $NPos_{i,t} \times IPen1(x)$ in Abnormal Return Regression for $x = 2.5, 3, 3.5, \ldots, 7, 7.5$

Variables	(1) $ARet_{i,t,t+1}$	(2) $ARet_{i,t,t+2}$	(3) $ARet_{i,t,t+3}$	(4) $ARet_{i,t,t+5}$	(5) $ARet_{i,t,t+10}$	(6) $ARet_{i,t,t+15}$	(7) $ARet_{i,t,t+21}$	(8) $ARet_{i,t,t+42}$	(9) $ARet_{i,t,t+63}$
$NPos_{i,t} \times IPen1(2.5)$	0.195*	0.147	−0.0427	−0.521	−0.881**	−2.024***	−2.577***	−4.082***	−3.407***
	(1.804)	(0.871)	(−0.213)	(−1.488)	(−1.985)	(−3.520)	(−3.739)	(−3.987)	(−2.996)
$NPos_{i,t} \times IPen1(3)$	0.155	0.135	0.00241	−0.408	−0.627*	−1.537***	−2.022***	−3.246***	−2.482***
	(1.410)	(0.883)	(0.0134)	(−1.340)	(−1.651)	(−3.130)	(−3.455)	(−3.752)	(−2.588)
$NPos_{i,t} \times IPen1(3.5)$	0.169*	0.146	−0.00683	−0.334	−0.513	−1.370***	−1.747***	−2.937***	−2.045**
	(1.755)	(1.097)	(−0.0434)	(−1.274)	(−1.552)	(−3.220)	(−3.460)	(−3.944)	(−2.468)
$NPos_{i,t} \times IPen1(4)$	0.101	0.101	−0.0239	−0.272	−0.364	−1.084***	−1.397***	−2.400***	−1.472**
	(1.165)	(0.841)	(−0.167)	(−1.171)	(−1.236)	(−2.837)	(−3.108)	(−3.629)	(−1.998)
$NPos_{i,t} \times IPen1(4.5)$	0.0903	0.0903	−0.00431	−0.216	−0.289	−0.936***	−1.223***	−2.204***	−1.491**
	(1.156)	(0.829)	(−0.0330)	(−1.051)	(−1.109)	(−2.745)	(−3.043)	(−3.739)	(−2.267)
$NPos_{i,t} \times IPen1(5)$	0.121*	0.0938	0.0162	−0.144	−0.139	−0.723**	−0.975***	−1.850***	−1.185*
	(1.648)	(0.899)	(0.129)	(−0.745)	(−0.567)	(−2.281)	(−2.617)	(−3.408)	(−1.944)
$NPos_{i,t} \times IPen1(5.5)$	0.146**	0.138	0.0714	−0.0687	−0.108	−0.660**	−0.852**	−1.636***	−1.033*
	(2.095)	(1.395)	(0.602)	(−0.382)	(−0.473)	(−2.231)	(−2.461)	(−3.259)	(−1.827)
$NPos_{i,t} \times IPen1(6)$	0.153**	0.185*	0.112	−0.0213	−0.0333	−0.531*	−0.758**	−1.532***	−0.981*
	(2.274)	(1.954)	(0.973)	(−0.125)	(−0.154)	(−1.909)	(−2.325)	(−3.255)	(−1.844)
$NPos_{i,t} \times IPen1(6.5)$	0.171***	0.209**	0.138	0.0284	0.0135	−0.415	−0.601**	−1.364***	−0.863*
	(2.675)	(2.337)	(1.274)	(0.177)	(0.0666)	(−1.590)	(−1.963)	(−3.085)	(−1.716)
$NPos_{i,t} \times IPen1(7)$	0.166***	0.191**	0.111	0.0309	0.000820	−0.366	−0.568*	−1.232***	−0.799*
	(2.715)	(2.231)	(1.060)	(0.201)	(0.00421)	(−1.466)	(−1.943)	(−2.931)	(−1.672)
$NPos_{i,t} \times IPen1(7.5)$	0.157***	0.194**	0.125	0.0620	0.0851	−0.270	−0.455*	−1.094***	−0.774*
	(2.725)	(2.406)	(1.267)	(0.429)	(0.464)	(−1.152)	(−1.653)	(−2.759)	(−1.719)

Robust t-statistics are shown in parentheses.
$^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

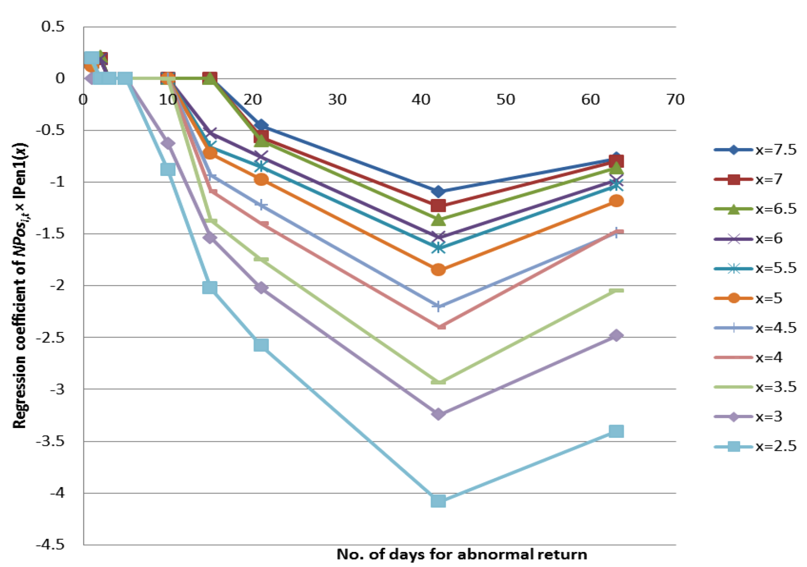


Fig. 3. Moderation effect of penny stocks for different cut-off prices (Definition 1).

Table 11. Regression of Raw Returns (Using *IPen1*(5))

Variables	(1) $Ret_{i,t,t+1}$	(2) $Ret_{i,t,t+2}$	(3) $Ret_{i,t,t+3}$	(4) $Ret_{i,t,t+5}$	(5) $Ret_{i,t,t+10}$	(6) $Ret_{i,t,t+15}$	(7) $Ret_{i,t,t+21}$	(8) $Ret_{i,t,t+42}$	(9) $Ret_{i,t,t+63}$
$NPos_{i,t}$	0.0668***	0.0937***	0.108***	0.136***	0.170***	0.197***	0.201***	0.227***	0.175*
	(4.683)	(4.737)	(4.507)	(4.760)	(4.289)	(4.153)	(3.786)	(2.993)	(1.870)
$NPos_{i,t}$×*IPen1*(5)	0.116	0.0968	0.00233	−0.151	−0.253	−1.055***	−1.357***	−2.405***	−1.888***
	(1.478)	(0.876)	(0.0175)	(−0.740)	(−0.930)	(−3.006)	(−3.260)	(−3.903)	(−2.736)
IPen1(5)	0.00365***	0.00617***	0.00610***	0.00741***	0.0123***	0.0182***	0.0209***	0.0372***	0.0572***
	(4.773)	(5.578)	(4.678)	(4.508)	(5.155)	(5.676)	(6.083)	(7.851)	(10.07)
$NPosCom_{i,t}$	0.00785	0.00353	0.000577	−0.00357	−0.00269	−0.000125	−0.00558	0.0174	−0.00470
	(1.323)	(0.441)	(0.0584)	(−0.295)	(−0.168)	(−0.00575)	(−0.244)	(0.542)	(−0.123)
$L\_MC_{i,t}$	−0.000185	−0.000316	−0.000445	−0.000235	0.00151*	0.00254**	0.00245*	0.00457***	0.00661***
	(−0.568)	(−0.718)	(−0.813)	(−0.351)	(1.670)	(2.410)	(1.950)	(2.582)	(3.015)
$Btm_{i,t}$	0.000399**	0.000724**	0.000903**	0.00182**	0.00464**	0.00496**	0.00572***	0.0103***	0.0109***
	(2.258)	(2.013)	(2.033)	(2.020)	(2.250)	(2.312)	(2.710)	(2.717)	(2.886)
$Ret_{i,t-252,t}$	−8.36e-06	0.000380	0.000874*	0.000993	0.00331***	0.00468***	0.00500***	0.00961***	0.0110***
	(−0.0280)	(0.919)	(1.713)	(1.532)	(3.360)	(4.181)	(3.858)	(5.170)	(5.085)
$UpAnalyst_{i,t}$	0.000163	0.000685	−0.000943	−0.00250	−0.000729	0.00172	0.00126	−0.00271	−0.00151
	(0.106)	(0.351)	(−0.403)	(−0.644)	(−0.115)	(0.208)	(0.148)	(−0.248)	(−0.127)
$DnAnalyst_{i,t}$	−1.50e-05	0.000467	0.00114	−0.00147	−0.00474	−0.00962**	−0.00932*	−0.0171**	−0.0174*
	(−0.0135)	(0.323)	(0.618)	(−0.677)	(−1.428)	(−2.315)	(−1.891)	(−2.422)	(−1.913)
$UpAnalyst_{i,t-1}$	−0.00101	−0.00221	−0.00251	−3.92e-05	−0.00230	−0.000201	−0.000391	0.00246	0.00650
	(−0.558)	(−0.926)	(−0.935)	(−0.0105)	(−0.434)	(−0.0336)	(−0.0555)	(0.252)	(0.585)
$DnAnalyst_{i,t-1}$	−0.000177	−0.000681	−0.000840	−0.00153	−0.00287	−0.00220	−0.00264	−0.00384	0.00849
	(−0.132)	(−0.320)	(−0.328)	(−0.571)	(−0.774)	(−0.535)	(−0.506)	(−0.533)	(0.910)
$UpAnalyst_{i,t-2}$	−0.00223	0.000163	0.00296	0.00651	0.00343	0.00356	0.00720	0.0130	0.0165
	(−1.583)	(0.0812)	(1.054)	(1.544)	(0.526)	(0.478)	(0.844)	(1.243)	(1.398)
$DnAnalyst_{i,t-2}$	0.000562	−0.00128	−0.00254	−0.00408	−0.00555	−0.0111**	−0.0128**	−0.0129*	−0.0111
	(0.415)	(−0.600)	(−1.073)	(−1.447)	(−1.581)	(−2.433)	(−2.108)	(−1.809)	(−1.087)
$SP5Ret_{t,t+x}$	1.209***	1.246***	1.295***	1.313***	1.355***	1.320***	1.333***	1.399***	1.391***
	(47.58)	(53.52)	(59.18)	(59.06)	(51.26)	(62.95)	(64.03)	(52.68)	(47.84)
$Ret_{i,t-1,t}$	0.0229**	0.0245**	0.0253*	0.0286*	−0.0149	−0.0350	0.00441	−0.0170	−0.0252
	(2.341)	(2.286)	(1.834)	(1.941)	(−0.606)	(−0.958)	(0.160)	(−0.473)	(−0.556)
$Ret_{i,t-2,t-1}$	−0.0126	−0.0345***	−0.0433***	−0.0440***	−0.0905***	−0.119***	−0.0764**	−0.0923**	−0.121***
	(−1.535)	(−2.934)	(−3.415)	(−2.752)	(−3.905)	(−3.429)	(−2.483)	(−2.229)	(−2.741)
$Ret_{i,t-3,t-2}$	−0.000248	−0.00742	−0.0156	−0.0324**	−0.0398*	−0.0491*	−0.0274	−0.0495	−0.0665
	(−0.0304)	(−0.572)	(−1.005)	(−2.025)	(−1.699)	(−1.748)	(−0.844)	(−1.038)	(−1.186)
$Ret_{i,t-63,t-3}$	−0.00263***	−0.00340**	−0.00448**	−0.00689***	−0.0142***	−0.0188***	−0.0156***	−0.0339***	−0.0361***
	(−2.776)	(−2.494)	(−2.371)	(−3.224)	(−4.822)	(−5.068)	(−3.840)	(−5.985)	(−5.604)
$Vol_{t-21,t}$	0.00295	−0.00339	0.00395	0.00525	0.0319*	0.0566**	0.0305	0.0839**	0.0910**
	(0.454)	(−0.375)	(0.346)	(0.407)	(1.697)	(2.128)	(1.115)	(2.512)	(2.155)
Year & month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.00462	0.0151	0.0214	0.0109	0.0371*	0.0334	0.0330	0.0712*	0.0123
	(0.522)	(0.550)	(0.616)	(0.493)	(1.864)	(1.565)	(0.830)	(1.945)	(0.224)
Observations	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561
Adjusted R^2	0.147	0.143	0.152	0.154	0.166	0.164	0.163	0.184	0.189

Robust t-statistics are shown in parentheses.
$^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

where C' is a vector of control variables including the analysts' upgrades/downgrades, lagged returns, volatility of stock returns, and S&P 500 index returns. Considering that the dependent variable is now the raw return, additional controls on the large MC indicator, book-to-market ratio, and past 1-year return are added. Year and month fixed effects, as well as sector fixed effects, were also included as controls. As before, we considered the holding period returns for both shorter periods from 1 to 15 days ($x = 1, 2, 3, 5, 10$, and 15) and longer periods from 1 to 3 months ($x = 21$, 42, and 63). The results are shown in Table 11.

Comparing the results to those using the abnormal returns, our findings remain robust. One observation is that the values of adjusted R^2 using raw returns are higher than those using abnormal ones, which is expected because raw returns are relatively much easier to predict and result in a better fitting.

For a robustness check, we performed the analysis on raw returns using Definition 2 of penny stocks:

$$Ret_{i,t,t+x} = \beta_0 + \beta_1 NPos_{i,t} + \beta_2 NPos_{i,t} \times IPen2(5) + \beta_3 IPen2(5) + \beta_4 NPosCom_{i,t} + \gamma C' + \varepsilon_{i,t},$$

where C' is a vector of control variables including the analysts' upgrades/downgrades, lagged returns, volatility of stock returns, S&P 500 index returns, large MC indicator, book-to-market ratio, and past 1-year return. Year and month fixed effects as well as sector fixed effects are also included as controls. Results are shown in Table A.5 in the Appendix and are similar to those in Table 11.

5.3 Moderation Effect of Penny Stocks Using Different MC Levels as Definition

In addition to trading prices, another way to identify "small stocks" is to look at the MC of the company. A smaller MC means that the stock's overall equity and liquidity on the market are smaller and trading the stocks of the company is a riskier investment. It therefore is worth investigating the following: Would the effects observed for penny stocks be higher for penny stocks with a smaller MC—that is, stocks with both a low trading price and low MC?

To investigate the effect of penny stocks with low MC levels, we generalize the definition of penny stocks as follows:

$$\begin{aligned} IPen1(x,y) &= 1 \text{ (penny stock with low MC)} \\ &\qquad \text{if } Pen1(x) = 1 \text{ and MC} < y \ \forall \text{ SA articles on the same publication date} \\ &= 0 \text{ otherwise.} \end{aligned}$$

Specifically, we analyze the moderation effect by looking at the coefficient of $NPos_{i,t} \times IPen1(x, y)$ for $x = 5$ and $y = 50, 100, \ldots, 250, 300$ (in millions).[7] The range from \$50 million to \$300 million is chosen because these stocks are referred to as microcap stocks.[8] The regression results are listed in Table 12.

From the results, we found that the magnitudes for the coefficients on the interaction term are generally higher when the MC level is reduced. This shows that penny stocks with smaller MC levels generally have a higher moderation effect compared to those with higher levels.

5.4 Prediction of Stock Price Using *NPos* Values

We found that stock opinions in social media have impacts on penny stock prices in the short term and a reverse effect in the long term. It is interesting to investigate the following: Can we make use of this to perform some predictions? Would investors making use of stock opinions from social media to buy penny stock earn a profit on average?

In the final analyses, we attempted to apply our findings and make use of our measure *NPos* to perform some predictions. We also tried to benchmark our measure *NPos* against the *Neg* measure.

[7]We have also considered the effect using Definition 2—that is, $IPen2(x, y)$, and for x with a value other than 5. The results were found to be similar. The reason we focused on microcap stocks (but not those with MC levels lower than \$50 million) was that stocks with *very* low MC levels were relatively less available in the SA forum.

[8]http://www.investopedia.com/terms/m/microcapstock.asp?layout=infini&v=3A; accessed January 2, 2016.

Table 12. Regression Coefficients of $NPos_{i,t} \times IPen1(5, y)$ in Abnormal Return Regression for $y = 50, 100, 150, \ldots, 250, 300$ (in Millions)

Variables	(1) $ARet_{i,t,t+1}$	(2) $ARet_{i,t,t+2}$	(3) $ARet_{i,t,t+3}$	(4) $ARet_{i,t,t+5}$	(5) $ARet_{i,t,t+10}$	(6) $ARet_{i,t,t+15}$	(7) $ARet_{i,t,t+21}$	(8) $ARet_{i,t,t+42}$	(9) $ARet_{i,t,t+63}$
$NPos_{i,t} \times IPen1(5, 50)$	0.786**	1.206**	0.817	−0.640	−0.393	−0.910	−2.979	−4.166	−1.531
	(2.423)	(2.511)	(1.551)	(−0.457)	(−0.296)	(−0.672)	(−1.163)	(−1.151)	(−0.459)
$NPos_{i,t} \times IPen1(5, 100)$	0.583***	1.023***	0.726**	0.217	0.677	−0.123	−0.951	−2.323	−1.822
	(3.076)	(3.759)	(2.326)	(0.304)	(0.941)	(−0.144)	(−0.713)	(−1.174)	(−0.919)
$NPos_{i,t} \times IPen1(5, 150)$	0.459***	0.724***	0.468*	0.152	0.555	−0.122	−0.641	−1.383	−0.653
	(2.895)	(3.217)	(1.836)	(0.273)	(0.972)	(−0.180)	(−0.615)	(−0.908)	(−0.422)
$NPos_{i,t} \times IPen1(5, 200)$	0.475***	0.686***	0.485**	0.278	0.526	−0.0546	−0.362	−1.510	−1.655
	(3.384)	(3.431)	(2.164)	(0.617)	(1.086)	(−0.0948)	(−0.422)	(−1.180)	(−1.228)
$NPos_{i,t} \times IPen1(5, 250)$	0.350**	0.492**	0.344	0.159	0.417	0.0320	−0.362	−1.779	−1.626
	(2.287)	(2.493)	(1.556)	(0.380)	(0.928)	(0.0603)	(−0.466)	(−1.526)	(−1.331)
$NPos_{i,t} \times IPen1(5, 300)$	0.329**	0.466**	0.337	0.124	0.158	−0.262	−0.612	−1.861*	−1.876*
	(2.342)	(2.535)	(1.631)	(0.328)	(0.380)	(−0.525)	(−0.843)	(−1.755)	(−1.670)

Robust t-statistics are shown in parentheses.
$^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Considering that our samples range from January 2005 to February 2014, we used the data from 2005 to 2012 to find the 99th percentile of the *NPos* values for penny stocks. Given the large number of financial prices available in penny stocks, the choice of a high 99th percentile value is to avoid a large amount of investment capital (because the wealth of most individual investors is limited). Such a high percentile of net fraction of positive words also increases the confidence of the investor for expecting a positive return in the future.

The 99th percentile of *NPos* is applied to the data from January 2013 to February 2014, as the testing data, using simple strategies.

Simple strategies.

Using the *NPos* measure (Strategy 1)

- Let k be the 99th percentile of the daily *NPos* values in a firm-day level for penny stocks from January 2005 to December 2012 on all trading days.
- Buy a penny stock i on date t (or the earliest future trading day if t is not a trading date) if *NPos* (average across all SA articles) for penny stock i on date $t > k$, where $t \in$ {January 1, 2013 to February 28, 2014}.

We then check whether there are any positive returns in the short versus long period and compare the results for both penny and non-penny stocks using the threshold proportion k in the period from January 2013 to February 2014. The results using the SEC's cut-off price of \$5 and Definitions 1 and 2 to define penny stocks are shown in Table 13.

Using both definitions, our measure *NPos* is able to predict positive returns for penny stocks in the short term. By using stock opinion information in *NPos*, penny stocks, on average, yield higher returns compared to non-penny stocks for a 1- or 2-day return. Price fluctuations and reversions also happen in which positive or negative returns may result in a longer term.

It may be interesting to look at individual penny stocks to see how sentiments on social media influence their return. For example, among the portfolio of penny stocks in the long position, the penny stock with the highest positive sentiment yielded an abnormal return of 3.6% on the first trading day after publication of the article. It then had an abnormal return of 4.1% on the second day but then only had a 0% return on the third day. Although this is only a particular example for

Table 13. Average Stock Returns Using Strategy 1 by *NPos*

	Definition 1			*Definition 2*		
	Penny Stock	Non-Penny Stock (*for comparison*)		Penny Stock	Non-Penny Stock (*for comparison*)	
Variable	Mean	Mean	Difference	Mean	Mean	Difference
$Ret_{i,t,t+1}$	0.016	0.001	0.015	0.014	0.001	0.012
$Ret_{i,t,t+2}$	0.014	0.004	0.010	0.011	0.004	0.007
$Ret_{i,t,t+3}$	0.002	0.004	−0.002	0.000	0.004	−0.004
$Ret_{i,t,t+5}$	−0.003	0.006	−0.009	0.000	0.005	−0.005
$Ret_{i,t,t+10}$	0.007	0.013	−0.005	0.016	0.012	0.003
$Ret_{i,t,t+15}$	0.027	0.013	0.014	0.031	0.012	0.019
$Ret_{i,t,t+21}$	0.044	0.015	0.029	0.047	0.015	0.032
$Ret_{i,t,t+42}$	0.052	0.039	0.013	0.055	0.039	0.016
$Ret_{i,t,t+63}$	−0.003	0.052	−0.055	0.016	0.051	−0.036

a simple illustration, this observation is consistent with our findings that penny stocks could yield profitable results in a short term.

Using the *Neg* measure (Strategy 2)

We have found that our measure *NPos* can predict positive returns for penny stocks in the short term. It is interesting to investigate the following: Would the performance of penny stocks in the strategy using the measure *NPos* (Strategy 1) be better than that using only negative words, *Neg* (Strategy 2)?

Strategy 2 is similar to Strategy 1, but this time we anticipate a drop in stock price if the fraction of negative words in the SA article is sufficiently low and we therefore should *short* such stocks. The strategy using *Neg* only is as follows:

—Let k' be the 99th percentile of the daily *Neg* values in a firm-day level for penny stocks from January 2005 to December 2012 on all trading days.
—Short a penny stock i on date t (or the earliest future trading day if t is not a trading date) if *Neg* (average across all SA articles) for penny stock i on date $t > k'$, where $t \in$ {January 1, 2013 to February 28, 2014}.

We then compare the results for penny stocks for Strategies 1 and 2 in the period from January 2013 to February 2014. The results using the SEC's cut-off price of $5 and Definitions 1 and 2 to define penny stocks are shown in Table 14.

From the results, strategies using *NPos* mostly result in positive returns, whereas that using *Neg* mainly results in negative returns. This shows that both *NPos* and *Neg* are able to capture stocks with strong and weak performances, respectively. In other words, our methodology of predicting stock returns using positive and negative sentiments can be effective. We proxy the relative performance of the measure *NPos* against that of *Neg* by the difference between the return using *NPos* and the return using *Neg*. Consistent with our previous findings, penny stocks can yield, on average, positive returns in the short term, and our measure *NPos* is able to perform better in the short run. Similarly, price fluctuations and reversions can happen in the longer term.

Using the *NPos* measure to long and short penny stocks (Strategy 3)

We have found that our measure *NPos* can be useful in predicting future stock performance. In this section, we attempt to use only the *NPos* measure to long and short penny stocks and explore if there are potential gains as follows:

Table 14. Average Stock Returns for Penny Stocks Using Strategy 2 by *Neg*

	Definition 1			*Definition 2*		
	NPos	*Neg*	Comparison	*NPos*	*Neg*	Comparison
Variable	Mean (1)	Mean (2)	(3)=(1)−[−(2)]	Mean (4)	Mean (5)	(6)=(4)−[−(5)]
$Ret_{i,t,t+1}$	0.016	−0.006	0.010	0.014	−0.006	0.008
$Ret_{i,t,t+2}$	0.014	0.004	0.018	0.011	0.004	0.015
$Ret_{i,t,t+3}$	0.002	−0.012	−0.010	0.000	−0.012	−0.012
$Ret_{i,t,t+5}$	−0.003	0.009	0.006	0.000	0.009	0.009
$Ret_{i,t,t+10}$	0.007	−0.052	−0.045	0.016	−0.052	−0.037
$Ret_{i,t,t+15}$	0.027	−0.064	−0.037	0.031	−0.064	−0.032
$Ret_{i,t,t+21}$	0.044	−0.068	−0.023	0.047	−0.068	−0.021
$Ret_{i,t,t+42}$	0.052	−0.160	−0.108	0.055	−0.160	−0.105
$Ret_{i,t,t+63}$	−0.003	−0.093	−0.096	0.016	−0.093	−0.078

Table 15. Average Stock Returns for Penny Stocks Using Strategy 3 by the 99th and 1st Percentiles of *NPos*

	Definition 1			*Definition 2*		
	99th Percentile	1st Percentile	Overall Performance	99th Percentile	1st Percentile	Overall Performance
Variable	Mean (1)	Mean (2)	(3)=(1)+[−(2)]	Mean (4)	Mean (5)	(6)=(4)+[−(5)]
$Ret_{i,t,t+1}$	0.016	−0.006	0.022	0.014	−0.006	0.020
$Ret_{i,t,t+2}$	0.014	0.004	0.010	0.011	0.004	0.008
$Ret_{i,t,t+3}$	0.002	−0.012	0.013	0.000	−0.012	0.012
$Ret_{i,t,t+5}$	−0.003	0.009	−0.012	0.000	0.009	−0.009
$Ret_{i,t,t+10}$	0.007	−0.052	0.060	0.016	−0.052	0.068
$Ret_{i,t,t+15}$	0.027	−0.064	0.090	0.031	−0.064	0.095
$Ret_{i,t,t+21}$	0.044	−0.068	0.112	0.047	−0.068	0.114
$Ret_{i,t,t+42}$	0.052	−0.160	0.212	0.055	−0.160	0.214
$Ret_{i,t,t+63}$	−0.003	−0.093	0.090	0.016	−0.093	0.109

—Let k be the 99th percentile (representing a relatively strong sentiment) of the daily *NPos* values in a firm-day level for penny stocks from January 2005 to December 2012 on all trading days.
—Let k'' be the 1st percentile (representing a relatively weak performance) of the daily *NPos* values in a firm-day level for penny stocks from January 2005 to December 2012 on all trading days.
—Long a penny stock i on date t (or the earliest future trading day if t is not a trading date) if *NPos* (average across all SA articles) for penny stock i on date $t > k$.
—Short a penny stock i on date t (or the earliest future trading day if t is not a trading date) if *NPos* (average across all SA articles) for penny stock i on date $t < k''$, where $t \in$ {January 1, 2013 to February 28, 2014}.

The results using the SEC's cut-off price of $5 and Definitions 1 and 2 to define penny stocks are shown in Table 15.

Table 16. Average Stock Returns for Penny Stocks Using Strategy 4 by the 99th and 1st Percentiles of *NPos*

	Definition 1 (Using All Penny Stocks)			*Definition 1 (Using Penny Stocks With MC Smaller Than 300 Million)*		
	99th Percentile	1st Percentile	Overall Performance	99th Percentile	1st Percentile	Overall Performance
	Mean	Mean		Mean	Mean	
Variable	(1)	(2)	(3)=(1)+[−(2)]	(4)	(5)	(6)=(4)+[−(5)]
$Ret_{i,t,t+1}$	0.016	−0.006	0.022	0.057	0.009	0.048
$Ret_{i,t,t+2}$	0.014	0.004	0.010	0.044	0.022	0.022
$Ret_{i,t,t+3}$	0.002	−0.012	0.013	−0.015	0.032	−0.046
$Ret_{i,t,t+5}$	−0.003	0.009	−0.012	−0.089	0.023	−0.113
$Ret_{i,t,t+10}$	0.007	−0.052	0.060	−0.152	−0.070	−0.082
$Ret_{i,t,t+15}$	0.027	−0.064	0.090	−0.152	−0.055	−0.096
$Ret_{i,t,t+21}$	0.044	−0.068	0.112	−0.078	−0.057	−0.021
$Ret_{i,t,t+42}$	0.052	−0.160	0.212	−0.091	−0.139	0.048
$Ret_{i,t,t+63}$	−0.003	−0.093	0.090	−0.192	−0.123	−0.068

From the results, returns using the 99th percentile of *NPos* mostly result in positive returns, whereas that using the 1st percentile mainly results in negative returns.[9] As comparisons, we proxy the overall performance by taking the sum of returns (i.e., returns from long position of stocks plus those from short position of stocks). It is found that the overall performance is mostly positive but with some price reversions. The results are robust using two definitions of penny stocks.

Using the *NPos* measure to long and short penny stocks with MC levels less than \$300 million (Strategy 4)

As discussed, penny stocks with lower MC levels can have a higher moderation effect. The process is similar to Strategy 3, but this time we consider only penny stocks with MC levels less than \$300 million in Strategy 4.[10] The results for such penny stocks with low MC levels using the SEC's cut-off price of \$5 and Definition 1 are shown in Table 16. The original results from Strategy 3 using *all* penny stocks are also listed for comparisons.

From Table 16, it can be seen that penny stocks with a smaller MC have higher returns in short terms of 1 to 2 days. It should be noted that these effects can appear in a very short term. This is rational because such information on small-sized companies is not readily available in traditional media. Consequently, the related information in the social media for such stocks may be reflected in the stock price quickly. Price reversions can also happen in the longer term that are consistent with our previous results.

Owing to the dataset's constraint, we did not conduct further out-of-sample analyses for the proposed strategies. Readers should be alerted that the four strategies are presented here for illustration purpose, and cautions need to be taken when interpreting the results.

[9]Interestingly, given our low percentile at the 1% level for the *NPos* measure, the stocks involved are mostly those with mainly bad sentiment with extremely weak positive content. Consequently, the returns using a very low level for the *NPos* measures could be highly similar to those using a very high level for the *Neg* measure.

[10]We have also considered other low MC levels (e.g., lower than \$50 million), and the results were found to be similar.

6 DISCUSSIONS ON FINDINGS AND CONTRIBUTIONS

Social media and financial data constitute valuable data sources for investors. Our data analytics give us insight that the efficient market hypothesis (Malkiel and Fama 1970) may not hold and stock price may not fully incorporate all available information instantly. Our results have addressed our three major questions stated in Section 3.

6.1 Sentiment Level in Social Media and Returns on Penny Stocks

Our measure on *NPos* and findings based on the analytics offer a new perspective to understand the relationship between information in social media and financial implications, particularly to the effects in penny stocks. Our study has demonstrated that data analytics can be applied on stock opinions in the social media and financial dataset to make predictions about the presumably risky penny stocks. Social media can generate a massive amount of useful information in real time, and financial data appear in a variety of forms. By applying textual analyses, we have used the net proportion of positive words in stock articles that can cater to both positive and negative views on stocks. Our findings suggested that social media information can predict the general future performance of penny stocks particularly in shorter terms.

6.2 Moderation Effect for Penny Stocks at Different Price and MC Levels

To the best of our knowledge, we are the first to investigate the relationship between social media and the financial impact on penny stocks. We also analyzed the moderation effect of penny stocks at different price and MC levels, and our results showed that penny stocks with lower levels can have stronger effects. Penny stocks have shown a reverse moderation effect in a longer period, and such an effect is consistent with a price reversion of stock prices in finance.

6.3 Simple Strategies Using Our Constructed Measure

Our simple strategies also showed that penny stocks can earn positive returns and with higher profits relative to non-penny ones in the short term. Our findings are consistent with Konku et al. (2012) stating that penny stocks generally are profitable investments in the short term but very poor investments beyond 1 year. Our findings are also consistent with another finding in the IPO context that penny stocks seem more profitable in the short run compared to ordinary stocks (Bradley et al. 2006). All of our findings remained robust when different definitions of penny stocks were used.

6.4 Contributions in Terms of Theoretical and Practical Implications

Our results are important in terms of both theoretical and practical implications in the information systems discipline. In terms of theoretical implications, our findings showed that penny stocks with smaller price and MC levels can constitute to a potential gain in a short term by utilizing the information in social media. Such a gain can reverse in the longer term and is consistent with price reversions. We therefore found that information diffusion occurs rapidly in social media among penny stocks. This could probably be due to the lack of relevant information on penny stocks reported in traditional media, making investors rely more on social media to obtain such information for investment advice. Considering that the opportunity may need to be captured quickly for potential profits, the effect of social media sentiment on the price of penny stocks may be reflected only in a short term. We also showed that both positive and negative opinions in social media can be essential in affecting future stock performance. Based on the results, we benchmarked the strategy using *NPos* against that using only the *Neg* measure on social media and found that the former measure could perform better for penny stocks in the short run.

In terms of practical implications, our study can be of valuable reference to individual investors, institutional investors, shareholders, and regulators. First, individual investors may hope to have a large capital gain given a small investment. Our study showed that social media can be a useful information source for investment opinions for penny stocks, as such information is difficult to obtain in traditional media. Second, given that institutional investors may trade securities in relatively large shares and they are concerned in the potential larger risks faced in the securities concerned, they may consider trading strategies on options by making use of both the positive and negative sentiment in the social media. Third, considering that shareholders are concerned with the firm value, they may be interested to know how the information in social media can affect the stock price and, in turn, the MC of the firm. However, given that information in social media can be important for penny stocks and that such stocks can be subject to manipulations, regulators may pay attention to the information in social media to see if any manipulated article pertaining to a penny stock is found. In the event that there are articles suspected to cause a manipulation in stock markets, regulators may decide on some appropriate regulations to ensure the proper function of the stock market.

7 LIMITATIONS AND FUTURE RESEARCH

7.1 Limitations of the Study

There are several limitations to our research. First, textual analyses are difficult to handle when multiple stocks are mentioned in the same article. We therefore choose to focus only on single-ticker articles. Even when single-ticker articles are chosen and our measure has already considered both positive and negative views, sometimes sentiment is not easy to analyze accurately when the sentence structure is complicated.

Second, SA has an extra scrutiny on stocks trading at less than $1 or with an MC below $100 million. Considering that SA authors will be cautious in posting about stocks that are less than $1, our sample size for very cheap types of penny stocks is inevitably affected. Nevertheless, from another point of view, this extra scrutiny gives higher credentials on investors' confidence on the view of SA's articles.

In addition, the stock behavior of penny stocks in our findings is a general phenomenon, and care should be taken in the simple strategy because stock prices can fluctuate when the market condition is unstable. In an extreme case of a financial turmoil, extra care is needed because penny stocks generally are riskier than non-penny ones.

7.2 Future Research

There are a few directions for our future research. One possible direction is to further seek a proprietary dataset from stock forum websites and to investigate other potential moderators with financial implications. Another avenue for research is to compare how the designs of different social media characteristics affect the prediction accuracy of future stock performance. For instance, different features of websites and degrees of investors' interaction can probably have different effects on financial information dissemination, which may affect future stock returns.

APPENDIX

Table A.1. Regression of Abnormal Returns (Using $IPen1(5)$) [Using $NPosA \equiv 2\ Pos_{i,t} - Neg_{i,t}$ and $NPosComA \equiv 2\ PosCom_{i,t} - NegCom_{i,t}$]

Variables	(1) $ARet_{i,t,t+1}$	(2) $ARet_{i,t,t+2}$	(3) $ARet_{i,t,t+3}$	(4) $ARet_{i,t,t+5}$	(5) $ARet_{i,t,t+10}$	(6) $ARet_{i,t,t+15}$	(7) $ARet_{i,t,t+21}$	(8) $ARet_{i,t,t+42}$	(9) $ARet_{i,t,t+63}$
$NPosA_{i,t}$	0.0411***	0.0557***	0.0634***	0.0785***	0.0893***	0.109***	0.121***	0.126***	0.121**
	(4.587)	(4.574)	(4.299)	(4.383)	(3.664)	(3.696)	(3.614)	(2.642)	(2.053)
$NPosA_{i,t} \times IPen1(5)$	0.0972*	0.0874	0.0318	−0.0143	−9.61e-05	−0.374*	−0.553**	−1.123***	−0.739*
	(1.931)	(1.221)	(0.372)	(−0.104)	(-0.000582)	(−1.776)	(−2.159)	(−3.096)	(−1.795)
$IPen1(5)$	0.00232**	0.00446***	0.00490***	0.00687***	0.0102***	0.0189***	0.0227***	0.0400***	0.0500***
	(2.420)	(3.229)	(2.912)	(2.717)	(3.151)	(4.579)	(4.541)	(5.691)	(6.328)
$NPosComA_{i,t}$	0.00202	−0.000448	−0.00344	−0.00352	−0.00493	−0.00783	−0.00615	−0.00324	−0.0209
	(0.612)	(−0.0989)	(−0.602)	(−0.508)	(−0.551)	(−0.668)	(−0.476)	(−0.176)	(−0.970)
$UpAnalyst_{i,t}$	0.000213	0.00103	−0.000608	−0.00115	0.00119	0.00508	0.00513	0.000402	0.00402
	(0.136)	(0.531)	(−0.268)	(−0.320)	(0.198)	(0.678)	(0.679)	(0.0403)	(0.366)
$DnAnalyst_{i,t}$	8.39e-05	−2.35e-06	0.000862	−0.00157	−0.00312	−0.00598*	−0.00435	−0.00924	−0.00687
	(0.0811)	(−0.00180)	(0.498)	(−0.787)	(−1.073)	(−1.669)	(−1.022)	(−1.535)	(−0.862)
$UpAnalyst_{i,t-1}$	−0.00137	−0.00279	−0.00250	−0.000737	−0.00352	−0.00105	0.000661	0.00375	0.00430
	(−0.868)	(−1.262)	(−1.026)	(−0.217)	(−0.718)	(−0.198)	(0.105)	(0.415)	(0.402)
$DnAnalyst_{i,t-1}$	−0.000454	−0.000487	−0.00199	−0.00255	−0.00296	−0.00142	−0.000605	−0.000626	0.0138*
	(−0.369)	(−0.250)	(−0.840)	(−0.998)	(−0.836)	(−0.360)	(−0.136)	(−0.0993)	(1.652)
$UpAnalyst_{i,t-2}$	−0.00239*	−0.000772	0.00132	0.00462	0.00233	0.00215	0.00717	0.0140	0.0138
	(−1.718)	(−0.408)	(0.478)	(1.140)	(0.382)	(0.324)	(0.956)	(1.464)	(1.189)
$DnAnalyst_{i,t-2}$	−0.000134	−0.00240	−0.00463**	−0.00594**	−0.00500	−0.00794*	−0.00861*	−0.00963	−0.00668
	(−0.112)	(−1.339)	(−2.343)	(−2.456)	(−1.566)	(−1.913)	(−1.691)	(−1.587)	(−0.674)
$SP5Ret_{t,t+x}$	0.0292	0.0314	0.0562***	0.0552***	0.0897***	0.0729***	0.0670***	0.104***	0.103***
	(1.261)	(1.472)	(2.836)	(2.804)	(3.803)	(3.827)	(3.570)	(4.480)	(3.845)
$ARet_{i,t-1,t}$	0.0168	0.0118	0.0124	0.0111	−0.0263	−0.0487	−0.0249	−0.0481	−0.0439
	(1.565)	(1.038)	(0.850)	(0.645)	(−1.132)	(−1.507)	(−0.970)	(−1.230)	(−0.978)
$ARet_{i,t-2,t-1}$	−0.0197**	−0.0423***	−0.0490***	−0.0529***	−0.0933***	−0.122***	−0.0839***	−0.103**	−0.117***
	(−2.402)	(−3.686)	(−3.871)	(−3.230)	(−3.939)	(−3.387)	(−2.647)	(−2.404)	(−2.598)
$ARet_{i,t-3,t-2}$	0.00491	−0.00167	−0.00902	−0.0252	−0.0255	−0.0432	−0.0286	−0.0330	−0.0311
	(0.598)	(−0.121)	(−0.556)	(−1.614)	(−1.075)	(−1.471)	(−0.803)	(−0.680)	(−0.552)
$ARet_{i,t-63,t-3}$	−0.00261***	−0.00272**	−0.00289	−0.00592***	−0.0103***	−0.0120***	−0.00767**	−0.0191***	−0.0199***
	(−2.686)	(−2.027)	(−1.581)	(−2.927)	(−3.915)	(−3.829)	(−2.100)	(−3.856)	(−3.442)
$Vol_{t-21,t}$	0.00104	−0.00708	−0.00385	−0.00621	0.00693	0.0208	−0.0110	0.0194	0.0183
	(0.173)	(−0.871)	(−0.374)	(−0.548)	(0.429)	(0.935)	(−0.480)	(0.719)	(0.520)
Year & month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.00570	0.0137	0.0187	0.0109	0.0297*	0.0228	0.0241	0.0691**	0.00490
	(0.708)	(0.545)	(0.575)	(0.616)	(1.825)	(1.245)	(0.627)	(2.017)	(0.103)
Observations	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561
Adjusted R^2	0.004	0.005	0.004	0.005	0.009	0.011	0.010	0.017	0.020

Robust t-statistics are shown in parentheses.
***$p < .01$, **$p < .05$, *$p < .1$.

Table A.2. Regression of Abnormal Returns (Using *IPen1*(5)) [Using $NPosB \equiv Pos_{i,t} - 2\ Neg_{i,t}$ and $NPosComB \equiv PosCom_{i,t} - 2\ NegCom_{i,t}$]

Variables	(1) $ARet_{i,t,t+1}$	(2) $ARet_{i,t,t+2}$	(3) $ARet_{i,t,t+3}$	(4) $ARet_{i,t,t+5}$	(5) $ARet_{i,t,t+10}$	(6) $ARet_{i,t,t+15}$	(7) $ARet_{i,t,t+21}$	(8) $ARet_{i,t,t+42}$	(9) $ARet_{i,t,t+63}$
$NPosB_{i,t}$	0.0381***	0.0491***	0.0547***	0.0674***	0.0765***	0.0863***	0.0953***	0.0923**	0.0633
	(4.698)	(4.433)	(4.078)	(4.169)	(3.487)	(3.183)	(3.176)	(2.206)	(1.219)
$NPosB_{i,t} \times IPen1(5)$	0.0586	0.0376	−0.00454	−0.138	−0.142	−0.490***	−0.628***	−1.141***	−0.718**
	(1.348)	(0.607)	(−0.0608)	(−1.262)	(−0.978)	(−2.577)	(−2.885)	(−3.518)	(−1.963)
$IPen1(5)$	0.00424***	0.00594***	0.00513***	0.00453**	0.00796***	0.00729**	0.00701*	0.0103**	0.0310***
	(4.584)	(4.478)	(3.317)	(2.354)	(2.972)	(2.050)	(1.853)	(1.981)	(4.673)
$NPosComB_{i,t}$	0.00677*	0.00749	0.00848	0.00892	0.00348	0.00212	0.00340	0.0201	0.00268
	(1.735)	(1.444)	(1.344)	(1.128)	(0.331)	(0.146)	(0.229)	(0.981)	(0.109)
$UpAnalyst_{i,t}$	0.000271	0.00108	−0.000560	−0.00115	0.00120	0.00498	0.00500	8.42e-05	0.00385
	(0.174)	(0.563)	(−0.248)	(−0.320)	(0.200)	(0.667)	(0.663)	(0.00845)	(0.350)
$DnAnalyst_{i,t}$	8.97e-05	4.88e-06	0.000869	−0.00156	−0.00312	−0.00599*	−0.00436	−0.00928	−0.00689
	(0.0868)	(0.00375)	(0.502)	(−0.784)	(−1.070)	(−1.672)	(−1.026)	(−1.542)	(−0.865)
$UpAnalyst_{i,t-1}$	−0.00136	−0.00278	−0.00248	−0.000699	−0.00348	−0.000975	0.000744	0.00390	0.00432
	(−0.868)	(−1.262)	(−1.021)	(−0.206)	(−0.710)	(−0.184)	(0.118)	(0.430)	(0.403)
$DnAnalyst_{i,t-1}$	−0.000384	−0.000401	−0.00190	−0.00245	−0.00285	−0.00135	−0.000543	−0.000673	0.0138*
	(−0.312)	(−0.205)	(−0.802)	(−0.961)	(−0.803)	(−0.340)	(−0.121)	(−0.107)	(1.650)
$UpAnalyst_{i,t-2}$	−0.00235*	−0.000735	0.00134	0.00460	0.00231	0.00202	0.00700	0.0137	0.0135
	(−1.690)	(−0.388)	(0.486)	(1.136)	(0.381)	(0.306)	(0.942)	(1.437)	(1.169)
$DnAnalyst_{i,t-2}$	−0.000143	−0.00241	−0.00464**	−0.00593**	−0.00498	−0.00786*	−0.00851*	−0.00947	−0.00655
	(−0.119)	(−1.341)	(−2.347)	(−2.446)	(−1.555)	(−1.896)	(−1.675)	(−1.561)	(−0.660)
$SP5Ret_{t,t+x}$	0.0293	0.0317	0.0563***	0.0556***	0.0896***	0.0728***	0.0667***	0.103***	0.103***
	(1.268)	(1.483)	(2.846)	(2.821)	(3.799)	(3.827)	(3.565)	(4.479)	(3.839)
$ARet_{i,t-1,t}$	0.0165	0.0116	0.0122	0.0118	−0.0255	−0.0467	−0.0226	−0.0445	−0.0412
	(1.534)	(1.016)	(0.832)	(0.687)	(−1.098)	(−1.453)	(−0.891)	(−1.153)	(−0.925)
$ARet_{i,t-2,t-1}$	−0.0199**	−0.0425***	−0.0492***	−0.0525***	−0.0928***	−0.121***	−0.0828***	−0.101**	−0.115**
	(−2.421)	(−3.691)	(−3.885)	(−3.202)	(−3.915)	(−3.355)	(−2.615)	(−2.371)	(−2.567)
$ARet_{i,t-3,t-2}$	0.00461	−0.00195	−0.00934	−0.0250	−0.0252	−0.0423	−0.0276	−0.0313	−0.0296
	(0.562)	(−0.142)	(−0.576)	(−1.602)	(−1.062)	(−1.437)	(−0.771)	(−0.643)	(−0.524)
$ARet_{i,t-63,t-3}$	−0.00271***	−0.00281**	−0.00300	−0.00587***	−0.0102***	−0.0117***	−0.00738**	−0.0186***	−0.0195***
	(−2.769)	(−2.088)	(−1.630)	(−2.885)	(−3.874)	(−3.738)	(−2.015)	(−3.756)	(−3.351)
$Vol_{t-21,t}$	0.00165	−0.00650	−0.00321	−0.00669	0.00624	0.0186	−0.0134	0.0154	0.0152
	(0.274)	(−0.795)	(−0.310)	(−0.588)	(0.386)	(0.837)	(−0.584)	(0.572)	(0.430)
Year & month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.00658	0.0148	0.0199	0.0125	0.0314*	0.0248	0.0264	0.0716**	0.00659
	(0.814)	(0.589)	(0.613)	(0.701)	(1.925)	(1.359)	(0.688)	(2.093)	(0.138)
Observations	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561
Adjusted R^2	0.004	0.005	0.004	0.005	0.009	0.012	0.010	0.017	0.020

Robust *t*-statistics are shown in parentheses.
$^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Table A.3. Regression of Abnormal Returns (Using *IPen1*(5)) [Using $NPosC \equiv (Pos_{i,t} - Neg_{i,t})/(Pos_{i,t} + Neg_{i,t})$ and $NPosComC \equiv (PosCom_{i,t} - NegCom_{i,t})/(PosCom_{i,t} + NegCom_{i,t})$]

Variables	(1) $ARet_{i,t,t+1}$	(2) $ARet_{i,t,t+2}$	(3) $ARet_{i,t,t+3}$	(4) $ARet_{i,t,t+5}$	(5) $ARet_{i,t,t+10}$	(6) $ARet_{i,t,t+15}$	(7) $ARet_{i,t,t+21}$	(8) $ARet_{i,t,t+42}$	(9) $ARet_{i,t,t+63}$
$NPosC_{i,t}$	0.00184***	0.00230***	0.00247***	0.00331***	0.00351***	0.00419***	0.00519***	0.00640***	0.00572**
	(4.998)	(4.621)	(4.138)	(4.504)	(3.461)	(3.356)	(3.684)	(3.227)	(2.356)
$NPosC_{i,t} \times IPen1(5)$	0.00245	0.00265	0.000519	−0.00316	−0.00216	−0.00872	−0.0159	−0.0350**	−0.0162
	(1.142)	(0.874)	(0.147)	(−0.545)	(−0.330)	(−1.106)	(−1.511)	(−2.444)	(−0.993)
$IPen1(5)$	0.00336***	0.00553***	0.00519***	0.00661***	0.0103***	0.0150***	0.0160***	0.0261***	0.0423***
	(4.423)	(5.101)	(4.088)	(3.946)	(4.399)	(5.008)	(4.790)	(5.810)	(7.815)
$NPosComC_{i,t}$	0.000427	0.000447	0.000550	0.000444	0.000327	8.25e-05	6.29e-05	−0.000619	−0.00131
	(1.645)	(1.258)	(1.310)	(0.810)	(0.447)	(0.0950)	(0.0623)	(−0.438)	(−0.758)
$UpAnalyst_{i,t}$	2.13e-05	0.000799	−0.00101	−0.00188	0.00163	0.00553	0.00573	0.00127	0.00694
	(0.0130)	(0.397)	(−0.424)	(−0.498)	(0.257)	(0.705)	(0.731)	(0.124)	(0.621)
$DnAnalyst_{i,t}$	8.41e-06	−9.50e-05	0.00106	−0.00173	−0.00354	−0.00623*	−0.00397	−0.00535	−0.00282
	(0.00784)	(−0.0727)	(0.604)	(−0.844)	(−1.153)	(−1.649)	(−0.893)	(−0.904)	(−0.358)
$UpAnalyst_{i,t-1}$	−0.00156	−0.00240	−0.00241	−0.00116	−0.00523	−0.00303	−0.000428	0.00266	0.00279
	(−0.959)	(−1.049)	(−0.954)	(−0.315)	(−0.996)	(−0.536)	(−0.0642)	(0.284)	(0.252)
$DnAnalyst_{i,t-1}$	−0.000490	−0.000257	−0.00215	−0.00254	−0.00183	−0.000101	−0.000813	−0.00252	0.0111
	(−0.395)	(−0.132)	(−0.886)	(−0.949)	(−0.495)	(−0.0250)	(−0.180)	(−0.386)	(1.231)
$UpAnalyst_{i,t-2}$	−0.00235	−0.000327	0.00165	0.00549	0.00391	0.00406	0.00930	0.0176*	0.0167
	(−1.621)	(−0.167)	(0.572)	(1.297)	(0.614)	(0.581)	(1.184)	(1.768)	(1.373)
$DnAnalyst_{i,t-2}$	0.000363	−0.00215	−0.00438**	−0.00583**	−0.00509	−0.00816*	−0.00938*	−0.00904	−0.00672
	(0.301)	(−1.168)	(−2.143)	(−2.320)	(−1.522)	(−1.863)	(−1.749)	(−1.425)	(−0.648)
$SP5Ret_{t,t+x}$	0.0257	0.0272	0.0582***	0.0648***	0.0981***	0.0738***	0.0684***	0.101***	0.106***
	(1.024)	(1.182)	(2.743)	(3.011)	(3.776)	(3.582)	(3.373)	(4.183)	(3.592)
$ARet_{i,t-1,t}$	0.0230*	0.00808	0.00438	0.000550	−0.0438*	−0.0661***	−0.0189	−0.0200	−0.0311
	(1.697)	(0.602)	(0.294)	(0.0326)	(−1.929)	(−2.748)	(−0.681)	(−0.497)	(−0.770)
$ARet_{i,t-2,t-1}$	−0.0194**	−0.0398***	−0.0450***	−0.0503***	−0.0906***	−0.120***	−0.0741**	−0.0927**	−0.106**
	(−2.251)	(−3.296)	(−3.438)	(−2.922)	(−3.623)	(−3.158)	(−2.236)	(−2.120)	(−2.327)
$ARet_{i,t-3,t-2}$	−0.00126	−0.00478	−0.0106	−0.0339**	−0.0259	−0.0399	−0.0136	−0.0131	−0.00287
	(−0.128)	(−0.332)	(−0.634)	(−2.152)	(−0.949)	(−1.170)	(−0.345)	(−0.241)	(−0.0435)
$ARet_{i,t-63,t-3}$	−0.00243***	−0.00311**	−0.00317*	−0.00634***	−0.0110***	−0.0119***	−0.00799**	−0.0197***	−0.0206***
	(−2.598)	(−2.381)	(−1.726)	(−3.057)	(−4.011)	(−3.687)	(−2.141)	(−3.896)	(−3.463)
$Vol_{t-21,t}$	0.00139	−0.0109	−0.00897	−0.0114	0.00385	0.00746	−0.0200	0.0131	0.000841
	(0.215)	(−1.293)	(−0.877)	(−0.990)	(0.227)	(0.340)	(−0.900)	(0.471)	(0.0235)
Year & month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.00263	−0.0123**	−0.0168**	−0.00567	0.0235*	0.0378**	0.0747***	0.0789*	−0.00177
	(−0.601)	(−2.142)	(−2.091)	(−0.475)	(1.761)	(2.035)	(2.800)	(1.859)	(−0.0485)
Observations	46,225	46,225	46,225	46,225	46,225	46,225	46,225	46,225	46,225
Adjusted R^2	0.005	0.005	0.004	0.006	0.009	0.012	0.010	0.017	0.021

Robust t-statistics are shown in parentheses.
$^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Table A.4. Regression Coefficients of $NPos_{i,t} \times IPen2(x)$ in Abnormal Return Regression for $x = 2.5, 3, 3.5, \ldots, 7, 7.5$

Variables	(1) $ARet_{i,t,t+1}$	(2) $ARet_{i,t,t+2}$	(3) $ARet_{i,t,t+3}$	(4) $ARet_{i,t,t+5}$	(5) $ARet_{i,t,t+10}$	(6) $ARet_{i,t,t+15}$	(7) $ARet_{i,t,t+21}$	(8) $ARet_{i,t,t+42}$	(9) $ARet_{i,t,t+63}$
$NPos_{i,t} \times IPen2(2.5)$	0.157	0.178	0.0359	−0.550	−0.724	−1.556***	−1.982***	−3.274***	−2.438**
	(1.440)	(1.060)	(0.178)	(−1.543)	(−1.614)	(−2.781)	(−2.889)	(−3.319)	(−2.299)
$NPos_{i,t} \times IPen2(3)$	0.188*	0.156	0.0241	−0.446	−0.530	−1.465***	−1.903***	−3.181***	−2.561***
	(1.938)	(1.060)	(0.136)	(−1.447)	(−1.366)	(−2.901)	(−3.158)	(−3.555)	(−2.594)
$NPos_{i,t} \times IPen2(3.5)$	0.135	0.0822	−0.0543	−0.435	−0.526	−1.311***	−1.714***	−2.865***	−2.020**
	(1.397)	(0.613)	(−0.338)	(−1.621)	(−1.551)	(−2.993)	(−3.304)	(−3.753)	(−2.377)
$NPos_{i,t} \times IPen2(4)$	0.132	0.0911	−0.0392	−0.362	−0.529*	−1.202***	−1.588***	−2.574***	−1.835**
	(1.514)	(0.763)	(−0.274)	(−1.542)	(−1.778)	(−3.117)	(−3.483)	(−3.853)	(−2.452)
$NPos_{i,t} \times IPen2(4.5)$	0.159**	0.101	−0.00293	−0.266	−0.413	−0.995***	−1.359***	−2.186***	−1.528**
	(1.988)	(0.906)	(−0.0218)	(−1.251)	(−1.520)	(−2.849)	(−3.298)	(−3.638)	(−2.273)
$NPos_{i,t} \times IPen2(5)$	0.152**	0.115	0.0295	−0.196	−0.285	−0.876***	−1.231***	−2.094***	−1.512**
	(2.028)	(1.094)	(0.234)	(−0.992)	(−1.125)	(−2.685)	(−3.201)	(−3.734)	(−2.411)
$NPos_{i,t} \times IPen2(5.5)$	0.144**	0.117	0.0247	−0.181	−0.269	−0.832***	−1.117***	−1.876***	−1.201**
	(2.073)	(1.191)	(0.208)	(−0.983)	(−1.141)	(−2.740)	(−3.116)	(−3.612)	(−2.057)
$NPos_{i,t} \times IPen2(6)$	0.149**	0.112	0.0303	−0.153	−0.221	−0.729**	−1.046***	−1.713***	−1.232**
	(2.258)	(1.201)	(0.271)	(−0.887)	(−1.005)	(−2.549)	(−3.119)	(−3.516)	(−2.242)
$NPos_{i,t} \times IPen2(6.5)$	0.168***	0.156*	0.0862	−0.0602	−0.121	−0.573**	−0.884***	−1.647***	−1.142**
	(2.677)	(1.755)	(0.808)	(−0.367)	(−0.579)	(−2.112)	(−2.777)	(−3.553)	(−2.187)
$NPos_{i,t} \times IPen2(7)$	0.165***	0.176**	0.123	0.0143	−0.0359	−0.438*	−0.721**	−1.483***	−1.052**
	(2.767)	(2.075)	(1.202)	(0.0922)	(−0.182)	(−1.714)	(−2.411)	(−3.414)	(−2.145)
$NPos_{i,t} \times IPen2(7.5)$	0.180***	0.206**	0.180*	0.117	0.0602	−0.308	−0.603**	−1.415***	−1.056**
	(3.167)	(2.552)	(1.851)	(0.794)	(0.322)	(−1.274)	(−2.132)	(−3.442)	(−2.269)

Robust t-statistics are shown in parentheses.
$^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

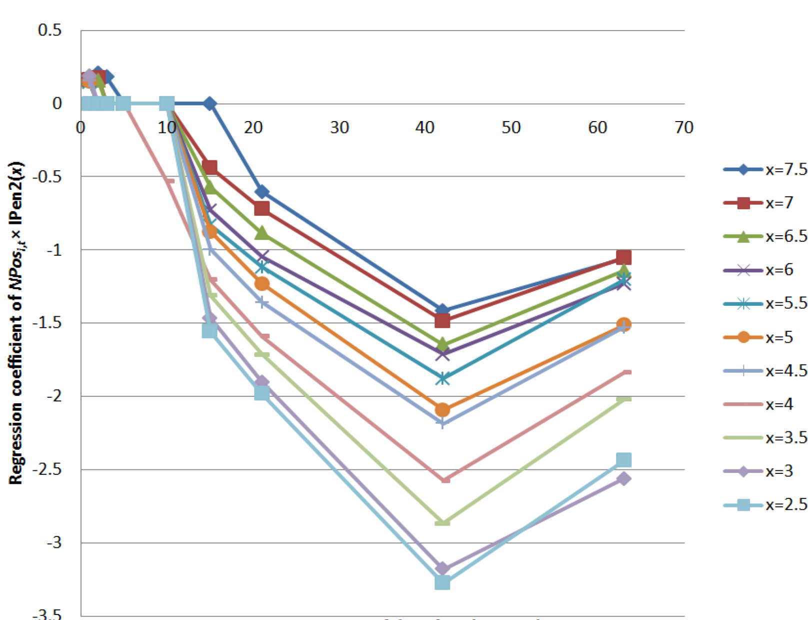


Fig. A.1. Moderation effect of penny stocks for different cut-off prices (Definition 2).

Table A.5. Regression of Raw Returns (Using *IPen2*(5))

Variables	(1) $Ret_{i,t,t+1}$	(2) $Ret_{i,t,t+2}$	(3) $Ret_{i,t,t+3}$	(4) $Ret_{i,t,t+5}$	(5) $Ret_{i,t,t+10}$	(6) $Ret_{i,t,t+15}$	(7) $Ret_{i,t,t+21}$	(8) $Ret_{i,t,t+42}$	(9) $Ret_{i,t,t+63}$
$NPos_{i,t}$	0.0617***	0.0882***	0.102***	0.136***	0.178***	0.204***	0.223***	0.235***	0.193**
	(4.371)	(4.509)	(4.312)	(4.758)	(4.488)	(4.261)	(4.164)	(3.057)	(2.030)
$NPos_{i,t} \times IPen2(5)$	0.160**	0.140	0.0519	−0.168	−0.353	−1.164***	−1.611***	−2.571***	−2.170***
	(2.020)	(1.266)	(0.391)	(−0.811)	(−1.258)	(−3.186)	(−3.759)	(−4.086)	(−3.085)
IPen2(5)	0.00301***	0.00496***	0.00454***	0.00628***	0.0109***	0.0173***	0.0211***	0.0370***	0.0543***
	(3.934)	(4.511)	(3.522)	(3.814)	(4.496)	(5.130)	(5.986)	(7.688)	(9.451)
$NPosCom_{i,t}$	0.00789	0.00363	0.000745	−0.00346	−0.00258	−6.77e-05	−0.00569	0.0173	−0.00468
	(1.329)	(0.453)	(0.0754)	(−0.287)	(−0.161)	(−0.00312)	(−0.249)	(0.539)	(−0.122)
$L\_MC_{i,t}$	−0.000292	−0.000520	−0.000709	−0.000438	0.00125	0.00231**	0.00238*	0.00436**	0.00593***
	(−0.910)	(−1.194)	(−1.299)	(−0.653)	(1.374)	(2.190)	(1.900)	(2.468)	(2.716)
$Btm_{i,t}$	0.000412**	0.000741**	0.000924**	0.00183**	0.00464**	0.00496**	0.00570***	0.0103***	0.0109***
	(2.277)	(2.030)	(2.049)	(2.026)	(2.251)	(2.313)	(2.712)	(2.722)	(2.891)
$Ret_{i,t-252,t}$	−4.32e-05	0.000323	0.000808	0.000949	0.00326***	0.00465***	0.00502***	0.00959***	0.0109***
	(−0.144)	(0.782)	(1.580)	(1.459)	(3.303)	(4.137)	(3.862)	(5.141)	(5.006)
$UpAnalyst_{i,t}$	9.57e-05	0.000597	−0.000987	−0.00258	−0.000890	0.00155	0.00101	−0.00310	−0.00232
	(0.0627)	(0.310)	(−0.425)	(−0.669)	(−0.142)	(0.190)	(0.121)	(−0.292)	(−0.202)
$DnAnalyst_{i,t}$	5.26e-05	0.000569	0.00124	−0.00139	−0.00464	−0.00957**	−0.00933*	−0.0171**	−0.0170*
	(0.0474)	(0.396)	(0.676)	(−0.645)	(−1.404)	(−2.312)	(−1.896)	(−2.430)	(−1.902)
$UpAnalyst_{i,t-1}$	−0.00108	−0.00231	−0.00260	−0.000159	−0.00250	−0.000551	−0.000808	0.00166	0.00539
	(−0.601)	(−0.976)	(−0.976)	(−0.0432)	(−0.478)	(−0.0937)	(−0.116)	(0.174)	(0.497)
$DnAnalyst_{i,t-1}$	−0.000160	−0.000649	−0.000801	−0.00150	−0.00282	−0.00216	−0.00261	−0.00379	0.00863
	(−0.120)	(−0.306)	(−0.314)	(−0.560)	(−0.765)	(−0.529)	(−0.508)	(−0.534)	(0.944)
$UpAnalyst_{i,t-2}$	−0.00226	0.000113	0.00294	0.00642	0.00323	0.00318	0.00666	0.0121	0.0153
	(−1.636)	(0.0576)	(1.059)	(1.541)	(0.502)	(0.434)	(0.795)	(1.190)	(1.341)
$DnAnalyst_{i,t-2}$	0.000570	−0.00125	−0.00249	−0.00400	−0.00543	−0.0109**	−0.0125**	−0.0124*	−0.0105
	(0.422)	(−0.586)	(−1.053)	(−1.430)	(−1.559)	(−2.410)	(−2.080)	(−1.780)	(−1.054)
$SP5Ret_{t,t+x}$	1.209***	1.245***	1.294***	1.312***	1.354***	1.319***	1.332***	1.399***	1.390***
	(47.60)	(53.53)	(59.18)	(59.06)	(51.25)	(63.10)	(64.31)	(52.61)	(47.78)
$Ret_{i,t-1,t}$	0.0225**	0.0238**	0.0247*	0.0278*	−0.0162	−0.0375	0.00143	−0.0228	−0.0327
	(2.289)	(2.233)	(1.797)	(1.894)	(−0.660)	(−1.025)	(0.0520)	(−0.633)	(−0.722)
$Ret_{i,t-2,\ t-1}$	−0.0130	−0.0353***	−0.0442***	−0.0455***	−0.0930***	−0.124***	−0.0824***	−0.103**	−0.134***
	(−1.585)	(−3.005)	(−3.486)	(−2.835)	(−3.983)	(−3.529)	(−2.674)	(−2.481)	(−3.032)
$Ret_{i,t-3,\ t-2}$	−0.000709	−0.00829	−0.0165	−0.0338**	−0.0422*	−0.0536*	−0.0329	−0.0592	−0.0790
	(−0.0873)	(−0.636)	(−1.060)	(−2.113)	(−1.798)	(−1.895)	(−1.017)	(−1.245)	(−1.408)
$Ret_{i,t-63,\ t-3}$	−0.00272***	−0.00355***	−0.00465**	−0.00712***	−0.0146***	−0.0195***	−0.0164***	−0.0354***	−0.0380***
	(−2.854)	(−2.600)	(−2.450)	(−3.328)	(−4.900)	(−5.164)	(−3.993)	(−6.140)	(−5.797)
$Vol_{t-21,\ t}$	0.00335	−0.00258	0.00518	0.00612	0.0328*	0.0581**	0.0311	0.0861**	0.0939**
	(0.514)	(−0.284)	(0.451)	(0.473)	(1.735)	(2.183)	(1.135)	(2.556)	(2.209)
Year & month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.00465	0.0152	0.0215	0.0110	0.0375*	0.0339	0.0338	0.0723**	0.0139
	(0.525)	(0.553)	(0.618)	(0.500)	(1.878)	(1.592)	(0.852)	(1.979)	(0.253)
Observations	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561	51,561
Adjusted R^2	0.147	0.143	0.152	0.154	0.166	0.164	0.163	0.185	0.189

Robust t-statistics are shown in parentheses.
$^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

ACKNOWLEDGMENTS

We would like to thank the associate editor and two anonymous reviewers for their constructive comments on the article. We would also like to show our gratitude to Reno Ha for his help in writing programs for retrieving the relevant data from various data sources.

REFERENCES

A. Abbasi, H. Chen, and A. Salem. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems* 26, 3 (2008), 1–34.

A. Agarwal, A. C. M. Leung, P. Konana, and A. Kumar. 2017. Cosearch attention and stock return predictability in supply chains. *Information Systems Research* 28, 2 (2017), 265–288.

R. Aggarwal, R. Gopal, A. Gupta, and H. Singh. 2012. Putting money where the mouths are: The relation between venture financing and electronic Word-of-Mouth. *Information Systems Research* 23, 3 (2012), 976–992.

R. K. Aggarwal and G. Wu. 2006. Stock market manipulations. *Journal of Business* 79, 4 (2006), 1915–1953.

W. Antweiler and M. Z. Frank. 2004. Is all that talk just noise? The information content of Internet stock message boards. *Journal of Finance* 59, 3 (2004), 1259–1294.

O. Ap Gwilym, A. Kita, and Q. Wang. 2014. Speculate against speculative demand. *International Review of Financial Analysis* 34 (2014), 212–221.

N. F. Awad and A. Ragowsky. 2008. Establishing trust in electronic commerce through online word of mouth: An examination across genders. *Journal of Management Information Systems* 24, 4 (2008), 101–121.

M. Bagnoli, M. D. Beneish, and S. G. Watts. 1999. Whisper forecasts of quarterly earnings per share. *Journal of Accounting and Economics* 28, 1 (1999), 27–50.

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

N. Barberis and M. Huang. 2008. Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review* 98, 5 (2008), 2066–2100.

R. Beatty and P. Kadiyala. 2003. Impact of the Penny Stock Reform Act of 1990 on the initial public offering market. *Journal of Law and Economics* 46, 2 (2003), 517–541.

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

T. Bouraoui, M. Mechanaoui, and B. Bahli. 2013. Stock spams: Another kind of stock prices manipulation. *Journal of Applied Business Research* 29, 1 (2013), 79–90.

D. J. Bradley, J. J. W. Cooney, S. D. Dolvin, and B. D. Jordan. 2006. Penny stock IPOs. *Financial Management* 35, 1 (2006), 5–29.

K. C. Butler and S. J. Malaikah. 1992. Efficiency and inefficiency in thinly traded stock markets: Kuwait and Saudi Arabia. *Journal of Banking and Finance* 16, 1 (1992), 197–210.

M. Chau and J. Xu. 2012. Business intelligence in blogs: Understanding consumer interactions and communities. *MIS Quarterly* 36, 4 (2012), 1189–1216.

H. Chen, P. De, Y. Hu, and B. Hwang. 2011. Sentiment revealed in social media and its effect on the stock market. In *Proceedings of the 2011 IEEE Statistical Signal Processing Workshop (SSP'11)*. 25–28.

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

Y. Chen and J. Xie. 2008. Online consumer review: Word-of-Mouth as a new element of marketing communication mix. *Management Science* 54, 3 (2008), 477–491.

K. Daniel, M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52, 3 (1997), 1035–1058.

S. Das, A. Martinez-Jerez, and P. Tufano. 2005. eInformation: A clinical study of investor discussion and sentiment. *Financial Management* 34, 3 (2005), 103–137.

S. R. Das and M. Y. Chen. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science* 53, 9 (2007), 1375–1388.

A. K. Davis, J. M. Piger, and L. M. Sedor. 2012. Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research* 29, 3 (2012), 845–868.

M. Dewally. 2003. Internet investment advice: Investing with a rock of salt. *Financial Analysts Journal* 59, 4 (2003), 65–77.

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

C. Forman, A. Ghose, and B. Wiesenfeld. 2008. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research* 19, 3 (2008), 291–313.

V. Fotak. 2008. The Impact of Blog Recommendations on Security Prices and Trading Volumes. Available at https://ssrn.com/abstract=1089868.

S. J. Grossman and J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70, 3 (1980), 393–408.

B. Gu, P. Konana, B. Rajagopalan, and H. W. M. Chen. 2007. Competition among virtual communities and user valuation: The case of investing-related communities. *Information Systems Research* 18, 1 (2007), 68–85.

B. Gu, J. Park, and P. Konana. 2012. Research Note—The impact of external Word-of-Mouth sources on retailer sales of high-involvement products. *Information Systems Research* 23, 1 (2012), 182–196.

U. G. Gurun and A. W. Butler. 2012. Don't believe the hype: Local media slant, local advertising, and firm value. *Journal of Finance* 67, 2 (2012), 561–598.

M. Hanke and F. Hauser. 2008. On the effects of stock spam e-mails. *Journal of Financial Markets* 11, 1 (2008), 57–83.

R. D. F. Harris and A. Pisedtasalasai. 2006. Return and volatility spillovers between large and small stocks in the UK. *Journal of Business Finance and Accounting* 33, 9–10 (2006), 1556–1571.

S. Hill and N. Ready-Campbell. 2011. Expert stock picker: The wisdom of (experts in) crowds. *International Journal of Electronic Commerce* 15, 3 (2011), 73–102.

C.-S. Ho, P. Damien, B. Gu, and P. Konana. 2017. The time-varying nature of social media sentiments in modeling stock returns. *Decision Support Systems* 101 (2017), 69–81.

B. Hu, T. McInish, and L. Zeng. 2010. Gambling in penny stocks: The case of stock spam e-mails. *International Journal of Cyber Criminology* 4, 1–2 (2010), 610–629.

N. Jegadeesh and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48, 1 (1993), 65–91.

M. G. Kavussanos and E. Dockery. 2001. A multivariate test for stock market efficiency: The case of ASE. *Applied Financial Economics* 11, 5 (2001), 573–579.

T. Keasler and C. McNeil. 2010. *Mad Money* stock recommendations: Market reaction and performance. *Journal of Economics & Finance* 34, 1 (2010), 1–22.

D. Konku, V. Bhargava, and D. K. Malhotra. 2012. Long-run performance of penny stock IPOs. *Journal of Wealth Management* 15, 1 (2012), 104–121.

A. Koulakiotis, V. Babalos, and N. Papasyriopoulos. 2016. Financial crisis, liquidity and dynamic linkages between large and small stocks: Evidence from the Athens Stock Exchange. *Journal of International Financial Markets, Institutions and Money* 40 (2016), 46–62.

A. Kumar. 2009. Who gambles in the stock market? *Journal of Finance* 64, 4 (2009), 1889–1933.

C. M. C. Lee and B. Swaminathan. 2000. Price momentum and trading volume. *Journal of Finance* 55, 5 (2000), 2017–2069.

P. Leeds. 2011. *Invest in Penny Stocks: A Guide to Profitable Trading*. Wiley.

A. C. M. Leung, A. Agarwal, P. Konana, and A. Kumar. 2017. Network analysis of search dynamics: The case of stock habitats. *Management Science* 63, 8 (2017), 2667–2687.

F. Li. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45, 2 (2008), 221–247.

T. Li, D. Jan Van, and P. J. Van Rees. 2018. More than just noise? Examining the information content of stock microblogs on financial markets. *Journal of Information Technology* 33, 1 (2018), 50–69.

X. Li, S. X. Sun, K. Chen, T. Fung, and H. Wang. 2015. Design theory for market surveillance systems. *Journal of Management Information Systems* 32, 2 (2015), 278–313.

T. Loughran and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66, 1 (2011), 35–65.

X. Luo. 2007. Consumer negative voice and firm-idiosyncratic stock returns. *Journal of Marketing* 71, 3 (2007), 75–88.

X. Luo and J. Zhang. 2013. How do consumer buzz and traffic in social media marketing predict the value of the firm? *Journal of Management Information Systems* 30, 2 (2013), 213–238.

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

B. G. Malkiel and E. F. Fama. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25, 2 (1970), 383–417.

H. Markowitz. 1952. The utility of wealth. *Journal of Political Economy* 60, 2 (1952), 151–158.

A. Metrick. 1999. Performance evaluation with transactions data: The stock selection of investment newsletters. *Journal of Finance* 54, 5 (1999), 1743–1775.

J. R. Nofsinger and A. Varma. 2014. Pound wise and penny foolish? OTC stock investor behavior. *Review of Behavioral Finance* 6, 1 (2014), 2–25.

P. C. Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62, 3 (2007), 1139–1168.

S. Sabherwal, S. K. Sarkar, and Y. Zhang. 2011. Do Internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news. *Journal of Business Finance and Accounting* 38, 9–10 (2011), 1209–1237.

R. P. Schumaker and H. Chen. 2009. A quantitative stock prediction system based on financial news. *Information Processing and Management* 45, 5 (2009), 571–583.

R. P. Schumaker, Y. Zhang, C.-N. Huang, and H. Chen. 2012. Evaluating sentiment in financial news articles. *Decision Support Systems* 53, 3 (2012), 458–464.

H. Shefrin and M. Statman. 2000. Behavioral portfolio theory. *Journal of Financial and Quantitative Analysis* 35, 2 (2000), 127–151.

D. H. Solomon. 2012. Selective publicity and stock prices. *Journal of Finance* 67, 2 (2012), 599–637.

M. Statman. 2002. Lottery players/stock traders. *Financial Analysts Journal* 58, 1 (2002), 14–21.

S. Stieglitz and L. Dang-Xuan. 2013. Emotions and information diffusion in social media-sentiment of microblogs and sharing behavior. *Journal of Management Information Systems* 29, 4 (2013), 217–248.

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

R. Tumarkin and R. F. Whitelaw. 2001. News or noise? Internet postings and stock prices. *Financial Analysts Journal* 57, 3 (2001), 41–51.

S. Urbański, P. Jawor, and K. Urbański. 2014. The impact of penny stocks on the pricing of companies listed on the Warsaw Stock Exchange in light of the CAPM. *Folia Oeconomica Stetinensia* 14, 2 (2014), 163.

K. L. Womack. 1996. Do brokerage analysts' recommendations have investment value? *Journal of Finance* 51, 1 (1996), 137–167.

K. Xiao, Q. Liu, C. Liu, and H. Xiong. 2017. Price stock detection with an influence-based model of social attention. *ACM Transactions on Information Systems* 9, 1 (2017), 1–21.

S. X. Xu and X. M. Zhang. 2013. Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction. *MIS Quarterly* 37, 4 (2013), 1043–1068.

Y. Zhang and P. E. Swanson. 2010. Are day traders bias free?—Evidence from Internet stock message boards. *Journal of Economics and Finance* 34, 1 (2010), 96–112.

Y. Zhang, P. E. Swanson, and W. Prombutr. 2012. Measuring effects on stock returns of sentiment indexes created from stock message boards. *Journal of Financial Research* 35, 1 (2012), 79–114.

D. Zimbra, A. Abbasi, D. Zeng, and H. Chen. 2018. The state-of-the-art in Twitter sentiment analysis: A review and benchmark evaluation. *ACM Transactions on Management Information Systems* 9, 2 (2018), 1–29.

D. Zimbra, H. Chen, and R. F. Lusch. 2015. Stakeholder analyses of firm-related web forums: Applications in stock return prediction. *ACM Transactions on Information Systems* 6, 1, (2015), Article 2, 38 pages.

Received May 2018; revised November 2018; accepted December 2018