UNDERSTANDING THE VALUE OF FINANCIAL ADVICE ON SOCIAL MEDIA

Tianyou Hu

胡天游

A thesis submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN

INFORMATION SYSTEMS

The University of Auckland, 2021

ABSTRACT

This thesis includes three separate studies which examine different aspects of the value of financial advice on social media. In the first study, we examine the effect of the long or short positions of non-professional analysts (hereafter, NPAs) writing on the social media outlet Seeking Alpha (hereafter SA) on the direction of investor trading and subsequent stock returns. We find that NPA positions contribute directly to short-window (less than one week) order imbalances after the article's publication. We also find that purchasing stocks with the most favorable sentiment and short selling those with the least favorable sentiment, together with daily portfolio rebalancing, yield a positive abnormal gross return that is statistically, but not economically, significant. There is no evidence that the information on Seeking Alpha can be used to generate economically significant abnormal returns. In the second study, we study the effect of providing financial incentives to NPAs on SA on the quality of stock recommendations. We find that NPAs are more likely to join the premium partnership program on SA and receive monetary payments if they have joined SA for a longer time and contributed more articles. We show that financial incentives reduce the quality of free stock recommendations. NPAs react to financial incentives and put their best work out where it generates the most income. The quality of NPAs' long (short) stock position recommendations in fee-based articles after joining the premium partnership program is worse (better) than the quality of their long (short) stock position recommendations in free articles prior to joining the premium partnership program. This study contributes to the literature on the role of social media in financial markets, the role of sell-side analysts in financial markets, and the understanding of the role of financial incentives in influencing the quality of user-generated content provided by NPAs on social media. In the third study, we examine the performance of textual analysis methods on data collected from financial microblogging websites HotCopper (hereafter, HC) and StockTwits (hereafter, ST), which have been frequently used for sentiment

analysis and stock market return predictions. We show that machine-learning classifiers have better accuracy than the Loughran & McDonald (2011) dictionary when classifying short text from HC. When conducting sentiment analysis on short text from social media and examining the effect of social media sentiment on stock market abnormal returns, researchers should try to use a financial social media like ST rather than a more informal social media like Twitter.

ACKNOWLEDGEMENTS

First of all, I would like to thank my main supervisor, Professor Arvind Tripathi, and co-supervisor, Professor Henk Berkman for their invaluable advice, continuous support, and patience through my Ph.D. journey. I am very grateful to them for sharing thoughtful recommendations, providing insightful feedback, helping me to improve my writing, and inspiring me along the way. I have enjoyed their mentorship, and I am very fortunate to have such an excellent supervision team.

I would like to thank Professor Michael Myers for his guidance and support over the years. I would like to thank the faculty of the Department of ISOM, especially Professor Tava Olsen, Professor Ananth Srinivasan, Professor Cecil Chua, and Dr. Valery Pavlov for their kind support in different ways during my Ph.D. study. Many thanks also go to faculty administrative staff, Chrissy Bretherton, Elena Tang, and Brogan Work, who kindly assisted me in the program. I would like to thank my friends and colleagues for a cherished time spent together in the office and social settings.

I thank Business School for providing me with the Conference Fund to attend several conferences. I'm grateful for the comments I received from participants of several conferences and workshops, including the International Conference on Information Systems (ICIS), Workshop on Information Technologies and Systems (WITS), Workshop on e-Business (WeB).

Last but not least, my special thanks go to my parents, who have always understood my feelings, cheered me up, and had a firm belief in me. Without their understanding and encouragement in the past few years, it would be hard to complete my study. Thanks to my grandmother. Please stay healthy. I miss my family so much. Thanks for their most incredible love and support.

TABLE OF CONTENTS

ABSTRACT **2**

ACKNOWLEDGEMENTS **4**

TABLE OF CONTENTS **6**

LIST OF FIGURES **8**

LIST OF TABLES **9**

INTRODUCTION **11**

CHAPTER 1- DOES CROWDSOURCED INVESTMENT ADVICE WORK? **16**

Paper (I) Social Media Contributors as New Analysts: Social Media Sentiment and Stock Abnormal Returns. **16**

1.1 Introduction 16

1.2 Related Literature 20

1.3 Seeking Alpha, Sample, and Descriptive Statistics 27

1.4 The Effect of Seeking Alpha Sentiment on Investors’ Trading 34

1.5 Can Retail Investors Capture the Benefits of Retail Financial Advice? 40

1.6 Conclusion 60

CHAPTER 2- THE EFFECT OF FINANCIAL INCENTIVE ON THE QUALITY OF STOCK RECOMMENDATIONS ON SOCIAL MEDIA **61**

Paper (II) The Effect of Financial Incentive on the Quality of Stock Recommendations on Social Media **61**

2.1 Introduction 61

2.2 Related Literature 65

2.3 Data 71

2.4 What Drives the Authors’ Decision to Charge for Their Financial Advice? 76

2.5 Do financial incentives affect the quality of stock recommendations? 82

2.6 Conclusion 94

CHAPTER 3 - SOCIAL MEDIA SENTIMENT ANALYSIS, SHORT TEXT CLASSIFICATION **95**

Paper (III) The Performance Evaluation of Textual Analysis Tools in Financial Markets **95**

3.1.1 Introduction 95

3.1.2 Literature Review 97

3.1.3 Data 100

3.1.4 Methodology 100

3.1.5 Results and Contribution 103

3.1.6 Conclusion 108

Paper (IV) The Performance Evaluation of Machine Learning Classifiers on Financial Microblogging Platforms 110

3.2.1 Introduction 110

3.2.2 Literature Review 111

3.2.3 Data 113

3.2.4 Machine Learning Models 116

3.2.5 Results 120

3.2.6 Conclusion 122

CONCLUSION 123

APPENDICES 125

Appendix A: The Effect of Seeking Alpha Disclosed Positions on Investors' Trading 125

Appendix B: Market Prediction 128

Appendix C: Portfolio Construction Using Articles with Disclosed Positions 131

Appendix D: Asset Pricing Models 134

Appendix E: Effect of Financial Incentive on Stock Recommendation Quality using BHAR 137

Appendix F: Professional Analysts Recommendations 141

REFERENCES 146

LIST OF FIGURES

CHAPTER 1

Figure 1 Total Number of Articles by Year..33
Figure 2 Average Order Imbalance on Different Days....................................36
Figure 3 Holding Periods of Weekly Return 47
Figure 4 Holding Periods of Monthly Return..47

CHAPTER 2

Figure 1 Research Design 83

CHAPTER 3

Figure 1 10-Fold Validation on Dataset with 1,000 Posts 104
Figure 2 10-Fold Validation for Dataset S...105
Figure 3 10-Fold Validation for Different Size of Dataset 105
Figure 4 Accuracy for Machine Learning Classifiers with Different Number of Characters Taken into Consideration...107
Figure 5 Accuracy for L&M Dictionary with Different Sample Size 107

Figure 1 Learning Curve..121
Figure 2 Accuracy, Precision, Recall and F1 Measure 121

LIST OF TABLES

CHAPTER 1

Table 1 SA Subscription Plans 29

Table 2 Sample Attrition..32

Table 3 Descriptive Statistics of the SA Single-Ticker Articles, by Year..................34

Table 4 Seeking Alpha Article Sentiment and Direction of Investor Trading 39

Table 5 Descriptive Statistics of Percentage of Negative Words for the Single-Ticker Articles, by Year 41

Table 6 Sixteen Portfolio Groups with Different Articles Periods and Holding Periods 44

Table 7 Fama-French-Carhart Filtering of Portfolio Groups with 1 Day Holding Period 51

Table 8 Fama-French-Carhart Filtering of Portfolio Groups with Different Holding Periods 52

Table 9 Fama-French-Carhart Filtering of Portfolio Group d-d Formed Using Subsets of All Articles Generated by Different Slicing Methods...................................57

Table 10 Fama-French-Carhart Filtering of Portfolio Group d-d Formed Using Subsets of All Articles Ranked by Author Historical Consistency 59

CHAPTER 2

Table 1 Sample Attrition..75

Table 2 Sample Attrition..75

Table 3 Descriptive Statistics of Variables 79

Table 4 The Cox Proportional Hazards Model 80

Table 5 Descriptive Statistics of the Main Variables..86

Table 6 Effect of Financial Incentive on Stock Recommendation Quality (CAR with One Quarter Holding Period)...89

Table 7 Effect of Financial Incentive on Stock Recommendation Quality (CAR with One Month Holding Period) 91

Table 8 Effect of Financial Incentive on Stock Recommendation Quality (CAR with One Week Holding Period) 92

Table 9 Effect of Financial Incentive on Stock Recommendation Quality (CAR with One Day Holding Period) 93

CHAPTER 3

Table 1 Comparison of Accuracy 108

Table 1 Summary Statistics 116

INTRODUCTION

Professional financial analysts (hereafter, PAs) play an important role in the capital markets by presenting new information and analyzing information previously published (Asquith et al., 2005; Bradshaw et al., 2017; Brown et al., 2016; Womack, 1996). There is a proliferation of sharing information and data over the Internet and social media in the financial industry. Investors consume financial information and are affected by others' sentiments, feelings, and opinions in the process. Access to financial information has increased exponentially, leading to the growth of nonprofessional analysts (hereafter, NPAs) on social media (Chen et al., 2014a; M. S. Drake et al., 2017). Many NPAs contribute stock recommendations in articles on Seeking Alpha (hereafter SA), which is a crowd-sourced online equity research platform. The long form of the articles, which are similar in format to the sell-side analysts' reports, together with content reviewed by the SA editorial board for clarity, consistency, and effect, allows SA NPAs to share something valuable if they have access to such information (Chen et al., 2014a).

This thesis contributes to our understanding of stock recommendations on social media. It includes three studies; each of them investigating a separate but related aspect of stock recommendations on social media. Each study is presented as a standalone chapter.

The first study, presented in Chapter 1, examines two research questions. First, does SA article sentiment influence the direction of investor trading? We find that SA article sentiment contributes directly to short-window order imbalances after the article's publication. It appears that investors react to the SA article sentiment provided by NPAs.

Second, can investors capture positive abnormal returns (alphas) following NPAs' stock recommendations on SA? We use calendar time portfolio strategies (hereafter, CTIME) in this study. We form portfolios by buying stocks with the most favorable sentiment (lowest average percentage of negative words) and short selling those with the least favorable sentiment (highest average percentage of negative words) and examine different portfolio rebalancing windows. Among these studied portfolio groups with a 1-day holding period, the portfolio group considering articles on a daily basis yields a statistically but not economically significant positive abnormal return. The abnormal returns of weekly, monthly, or quarterly rebalancing are not significant.

This study further explores several additional tests to identify subsamples that might contain more valuable stock recommendations. We use subsamples of SA articles with different methods: 1) free and fee-based articles (website-level premium service), 2) articles with a high and a low number of comments, 3) articles contributed by NPAs who have individual NPA level premium services, and 4) articles contributed by NPAs who performed best in previous years. However, none of these portfolios generate significant abnormal returns.

This study makes two contributions that extend our understanding of the role of NPA stock recommendations in financial markets. First, we apply CTIME to account for the cross-sectional dependence of event firm abnormal returns that overlap in calendar time. Major corporate events cluster by industry over time, which leads to a positive cross-correlation of abnormal results. Event firm abnormal returns that overlap in calendar time result in cross-sectional dependence of abnormal results. This cross-sectional dependence is likely to make t-statistics overstated. CTIME is robust to the most serious statistical problems (Mitchell & Stafford, 2000). Second, rather than equal-weighted abnormal returns, we use value-weighted abnormal returns to better capture the economic significance of our findings. As a result, larger firms will be more heavily reflected in the portfolio return than smaller firms.

The second study, presented in Chapter 2, explores two research questions. In April 2015, SA launched Seeking Alpha Investor Marketplace (hereafter, Marketplace). Marketplace is an investment services platform provided by SA individual authors and author teams. The

Marketplace is an NPA-level subscription plan. First, we examine what drives NPAs' decision to start offering financial advice on Marketplace. The decision of NPAs to offer their services on Marketplace is mainly determined by the number of years the NPA has been active on SA (*NumYearsOnSA*) and the total number of articles contributed by the NPA per year (*NumArticles*). In terms of economic significance, we find that a one-year increase in *NumYearsOnSA* is associated with a 13.5% increase in the probability that an author will join Marketplace. One article increase in *NumArticles* is associated with a 1.6% increase in the probability that an author will join Marketplace. These findings suggest that an NPA is more likely to join Marketplace if he or she has contributed to SA for a longer period of time and has published more articles each year.

Second, how do financial incentives provided by social media platform owners affect the behavior of NPAs on social media? We use a research design with NPA fixed effects to control for time-invariant variation in skill levels of NPAs. The financial market context allows us to construct an objective measure of quality, i.e., abnormal returns of disclosed stock holding positions in SA articles. We compare the abnormal returns corresponding to NPAs' disclosed stock positions in free articles and Marketplace articles. Using three-month abnormal returns after the article's publication, we find that financial incentives have negative effects on the quality of free articles. When NPAs have joined Marketplace, the quality of their recommendations in Marketplace articles is better than the quality of their recommendations in free articles. We also conduct additional analyses using different holding periods. Our results still mostly hold with a one-month holding period but lose some significance with a one-week or one-day holding period. In addition, our results are robust to the use of different asset pricing models for calculating abnormal returns. A more detailed discussion of the results can be found in chapter 2.5.

We contribute to the literature on the role of social media in financial markets, particularly for equity research platforms such as SA. These studies show that NPAs contribute value-relevant information to the market in that their articles and SA comments predict future stock returns (Chen et al., 2014a); their articles lead to more informed retail trading (Farrell et al., 2020); improve liquidity during earnings announcements (Gomez et al., 2020); and their disclosed positions magnify investor responses to article tone (Campbell et al., 2019). Most of the prior studies have focused on the role of social media in financial markets without considering financial incentives. We build on the work of Chen et al. (2019), who explore the effect of providing financial incentives to NPAs on SA, as well as the implications for online investor communities. They show that financial payments from the platform[1] increase the number of articles but do not affect the quality of articles. Our paper extends Chen et al. (2019) by showing that financial incentives provided by social media platform owners affect the quality of stock recommendations in SA articles.

This study also contributes to the literature on the role of sell-side analysts in financial markets. Jame et al. (2016) show that crowdsourced earnings forecasts are incrementally useful to investors beyond those of sell-side analysts. Jame et al. (2017) show that crowdsourced earnings forecasts can discipline sell-side analysts and result in less biased forecasts. Each SA Marketplace NPA makes recommendations to SA's clients (Marketplace subscribers) and is paid based on the number of Marketplace subscribers, which makes Marketplace NPAs comparable to sell-side analysts. This research contributes to this literature by being the first study to examine the article quality in SA Marketplace and show that financial incentives influence the quality of NPAs' stock recommendations in Marketplace. This evidence is important because the amount of information available to investors via social media is likely

[1] SA launched a premium partnership program in January 2011. These "premium" articles are published only on SA and are not freely available elsewhere on the Internet. It is a website-level subscription.

to expand over time, while budgets and headcounts of sell-side equity research departments are likely to decline (Morris, 2017). The number of sell-side analysts has been steadily declining while the number of NPAs has been steadily increasing (M. S. Drake et al., 2020).

The third study, presented in Chapter 3, compares the performance of textual analysis tools on textual data collected from financial social media websites. Financial microblogging websites HotCopper (hereafter, HC) and StockTwits (hereafter, ST) have been frequently used by researchers for sentiment analysis and stock market return predictions. We demonstrate that machine learning classifiers have higher accuracy than Loughran & McDonald (2011) dictionary when classifying short text using data collected from HC. After careful feature engineering, we obtain 81.9% accuracy in classifying ST messages. In comparison, previous research on Twitter has a classification accuracy of 75.39%.

This study adds to our understanding of the performance of textual analysis tools in social media sentiment analysis. We show that machine learning classifiers will be preferred to dictionaries when classifying short text. When researchers want to conduct sentiment analysis on short text from social media and examine the effect of social media sentiment on stock market abnormal returns, they should try to use a financial social media like ST instead of more casual social media like Twitter.

CHAPTER 1 - Does Crowdsourced Investment Advice Work?

Paper (I) Social Media Contributors as New Analysts: Social Media Sentiment and Stock Abnormal Returns.

1.1 Introduction

Financial advisors prefer higher net worth investors due to the lower aggregate costs of engagement, which leaves small/retail investors disadvantaged in comparison with investors with large portfolios (West, 2012). In recent years, a variety of market platforms, from social media-driven investment communities, such as Stocktwits[2] to equity review websites, such as Seeking Alpha[3] (SA, hereafter), have started to provide low-cost, crowdsourced investment research. Crowdsourced financial analysis is more valuable to less sophisticated investors and mitigates the information asymmetry problem (Gomez et al., 2018).

In the last decade, SA has emerged as a popular crowdsourced investment research platform (Farrell et al., 2018). The crowdsourced research articles on SA are different from conventional Wall Street brokerage research in that SA covers more companies with low institutional ownership and higher breadth of ownership (Farrell et al., 2018). With more than 10 million registered users, it attracts 17 million unique visitors every month[4] seeking crowdsourced investment research contributed by thousands of contributors (*About Seeking Alpha*, 2020). Unlike social media stock discussion forums, content on SA is screened by the editors before publication to ensure quality. Further, contributors on SA, dubbed as nonprofessional analysts (NPAs, hereafter), are paid for their contributions based on the number of page views of their articles. These NPAs are required to disclose their stock holding positions in their research

[2] https://stocktwits.com/

[3] https://seekingalpha.com/

[4] https://en.wikipedia.org/wiki/Seeking_Alpha

articles (Campbell, DeAngelis, & Moon, 2019), possibly to signal SA's commitment to fairness and ethics and to avoid any conflict of interest.

Given the large readership of SA, it is important to understand if NPAs' advice on SA influences investors' trading behavior and whether NPAs' advice predicts future returns. The effect of NPAs' advice on investors' trading behavior can be gauged by examining order imbalances before and after the publication of SA articles, and the value of advice contained in these articles can be measured via abnormal returns. We focus on two research questions. First, do the financial positions of NPAs writing on SA influence the direction of investor trading? Second, can investors capture positive abnormal returns (alphas) following NPAs' stock recommendations on SA?

While scholars have examined the value of financial advice by NPAs to investors (Campbell et al., 2019; Chen et al., 2014a; Farrell et al., 2020), the research investigating the effect of this financial advice on investors' trading behavior is more limited. To the best of our knowledge, only one recent study (Farrell et al., 2020) examines retail trading around the publication of SA articles using ten half-hour intraday event windows and finds that retail trading is much higher following the publication of SA articles, but not before. However, one can argue that NPAs' advice, available on an online platform, is not restricted to retail investors and may influence the trading behavior of a wider range of investors. Note that NPAs on SA include fund managers, institutional investors[5], and highly qualified experts (Campbell et al., 2019) and therefore may also attract investors other than small retail investors, if not seeking advice, at least for understanding the trading behavior of the retail investors. Investors are strategic and factor in other market forces, even if irrational. Therefore, we conjecture that NPAs' advice may affect the trading behavior of all types of

[5] https://seekingalpha.com/page/become-a-seeking-alpha-contributor

investors and hence it is interesting to look at overall order imbalances instead of those related only to retail investors. We find that SA article sentiment contributes directly to short-window order imbalances after the article's publication. This evidence shows that investors react to the content of the articles contributed by NPAs.

We scrutinize the research articles published on SA to analyze the extent to which crowdsourced investment research articles can predict future stock abnormal returns. Our data includes more than two hundred thirteen thousand single ticker articles contributed by more than 10 thousand authors from 2005 to 2018. We form portfolios by purchasing the stocks with the most favorable sentiment and short-selling those with the least favorable sentiment, and examine different portfolio rebalancing windows.

We find positive and statistically significant alphas for portfolio groups with a 1-day holding period. Among these studied portfolio groups with a 1-day holding period, the portfolio group considering articles on a daily basis yields a positive abnormal gross return of 10 basis points per trading day. This positive abnormal return is statistically significant, but it is not economically significant after factoring in the bid-ask spread. The alphas are not significant for portfolios with a holding period longer than one day. We conclude that, on average, NPAs' stock analyses on SA do not contain valuable information.

We perform several additional tests in order to identify subsamples that might contain more valuable stock recommendations (NPAs' advice on SA). For example, we conduct tests comparing: 1) free and pro (website level premium) articles, 2) articles with high and low numbers of comments, 3) articles contributed by authors who have set up their own Marketplaces on SA (individual contributor level premium service), and 4) articles by authors who performed best in previous years. However, none of these portfolios generate significant abnormal returns.

Our study contributes to the literature on the value of information on social media (Campbell et al., 2019; Chen et al., 2014a; M. S. Drake et al., 2017; Jame et al., 2016; Tang, 2018). Using a similar dataset, Chen et al. (2014) have shown that stock opinions transmitted through articles (NPAs' advice) and comments on SA forecast future stock returns and earnings surprises. Further, Campbell et al. (2019) show that NPAs' stock holding positions magnify investor responses to both positive and negative tone of SA articles. Chen et al. (2014a) and Campbell et al. (2019) claim that certain contributors (NPAs) are more credible because of their established record of providing value-relevant analyses (Chen et al., 2014a) or their disclosed positions in the firm's stocks (Campbell et al., 2019).

To explain the difference between the results in the literature and our findings, we point out that, in contrast to previous work in this area, we use a calendar time portfolio strategy to account for the cross-sectional dependence of event firm abnormal returns that are overlapping in calendar time.[6] This approach is consistent with the advice in Fama (1998) to use the calendar-time portfolio approach (hereafter, CTIME). Mitchell & Stafford (2000) also argue the CTIME is robust for the most serious statistical problems. The first contribution of this study is that we apply CTIME with non-overlapping return periods on the data collected from SA to get robust test statistics. We have proposed a new approach to building the portfolio to assess the value offered by equity review platforms like SA. It is crucial to understand how to evaluate the value provided by such platforms. We also contribute to the literature by questioning whether the wisdom of crowds exists in a financial context.

The remainder of this paper is organized as follows. Section 2 conducts a literature review. Section 3 discusses data sources and descriptive statistics. In Section 4, we examine whether investors trade based on sentiment in SA articles using order imbalances. Section 5 examines

[6] Fama (1998) is a criticism of long term event studies that do not account for the fact that major corporate events cluster by industry over time, resulting in a positive cross-correlation of abnormal results and inflation of the test statistics.

the value of online financial advice using calendar time portfolio strategies with different return periods. Section 6 concludes.

1.2 Related Literature

Our study draws upon two streams of research, which include: 1) the value of investment advice and 2) the performance of retail investors.

1.2.1 The Value of Investment Advice

1.2.1.1 The Value of Crowdsourced Research

The desire to understand the value of signals that can predict share market returns has been a recurring theme in academic and practice literature. A large body of literature has examined the value of signals from conventional media in predicting stock market returns (Barber & Loeffler, 1993; Huberman & Regev, 2001; Busse & Clifton Green, 2002; Tetlock, 2007; J. Engelberg, 2008; Tetlock, Saar-Tsechansky, & MacSkassy, 2008; Fang & Peress, 2009; J. E. Engelberg & Parsons, 2011; Dougal, Engelberg, Garcia, & Parsons, 2012; Gurun & Butler, 2012; Solomon, 2012). Overall, the findings show that conventional media plays an important role in the origination and distribution of information.

Social media allows direct and immediate interaction between investors to complement information from traditional media sources, such as news media. The mechanism of information sharing does make a difference and is a critical component of WEB 2.0[7]. Although many recent studies have provided evidence that information on different types of social media has investment value (Jame, Johnston, Markov, & Wolfe, 2016; Tang, 2018; Campbell et al., 2019), there is little evidence that retail investors can obtain these benefits. On the contrary, some existing studies demonstrate that social media can reinforce the behavioral biases of retail

[7] Web 2.0 refers to websites that emphasize user-generated content, ease of use and participatory culture for end users.

investors and damage retail investors' performance (Ammann & Schaub, 2020; Cookson, Engelberg, & Mullins, 2020; Heimer, 2016).

Social Media Reinforce the Behavioral Biases of Retail Investors

We discuss the research on social media leading to behavioral biases of retail investors. The disposition effect (the tendency to sell winning assets and unwillingness to give up losing assets) is seen as a deviation from rational trading behavior. Heimer (2016) argues that social interaction contributes to the disposition effect of some traders. Cookson et al. (2020) find that users of StockTwits tend to follow others who agree with their sentiment (bullish/bearish) and build a personalized news source to support their original views, which are associated with poor ex-post returns.

Message Board

The consensus of early literature demonstrates that message board postings have a limited ability to forecast future price movements for individual stocks (Avery et al., 2015). Tumarkin and Whitelaw (2001) detect no association between message board activities and industry-adjusted returns or abnormal trading volume. Antweiler and Frank (2004) find a statistically significant, while economically meaningless association between internet message board activities and stock returns. Das and Chen (2007) argue there is no strong relationship from sentiment to stock prices on average across individual stocks. Message boards are proven to generate mostly noise.

Twitter

Multiple studies explore the link between Tweets and aggregate stock market movements. For example, Bollen, Mao, and Zeng (2011) find that aggregate sentiment resulting from sentiment analysis of daily Tweets can help predict the daily directional changes in the closing values of the Dow Jones Index. Mao, Wei, Wang, and Liu (2012) show that the daily volume of tweets that mention S&P 500 stocks is significantly correlated with the levels, changes, and absolute changes in the S&P 500 Index. Some other studies examine the association between Tweets and firm-level sales and earnings. Bartov, Faurel, & Mohanram (2018) argue that aggregate views from individual Tweets can forecast a company's subsequent quarterly earnings and announcement returns. Tang (2018) demonstrates that aggregate Twitter sentiment can predict firm-level future sales and sales growth. Tweets can predict aggregate stock market movements and individual stock movements.

Seeking Alpha

SA is a popular place where NPAs can share their financial securities analyses. It is ranked as the third most popular stock market news aggregation site after Google Finance and Yahoo! Finance (*Top Financial and Stock Market News Sites*, 2020). A few recent studies have examined the value of crowdsourced investment advice via SA (Campbell et al., 2019; Chen et al., 2014a). These studies aggregate SA articles from individual NPAs and claim that advice offered by NPAs predicts future stock abnormal returns. Chen et al. (2014) show that the long form of the articles, which are similar in format to the sell-side analysts' reports, together with content reviewed by the SA editorial board for clarity, consistency, and effect, gives SA NPAs the opportunity to share something valuable if they have access to such information. By analyzing research articles published on SA from 2005-2012, Chen et al. (2014) find that a 1% increase in the fraction of negative words results in a 0.379% decrease in three-month abnormal returns. Using SA articles published between 2005 and 2015, Campbell et al. (2019) discover that an NPA's disclosure of a long (short) position results in a two-day return of 0.4% (-1.2%), after controlling for the content of the article as well as news released at the same time. The authors claim that SA stands out from other social media platforms because the research contains solid and fine-tuned analyses, with the NPAs' disclosed stock holding positions.

There are some other crowdsourced research social media platforms, including Motley Fool CAPS, and Estimize.com. CAPS users contribute explicit forecasts on the future price of a specific stock, and CAPS is built with the purpose of facilitating the reputation of its contributors (Avery et al., 2015). Estimize.com only gives an earnings estimate without any detailed analyses (Jame et al., 2016). Another related set of studies involves online prediction markets, such as Intrade. These platforms host competitive forecasting markets for stock trading, which will pay off if a specific event occurs. Wolfers & Zitzewitz (2004) study the functioning of these prediction markets as one of the pioneering studies. Unlike Motley Fool CAPS, Estimaze, Intrade, or platforms like stock message boards and Twitter, which grant any user to write short messages (Tweets) without any quality assurance, SA research articles are long-form, with curated analysis, which has been peer-reviewed by SA editors for quality assurance.

Social Media Brings New Information to the Market

It is shown that social media brings new information to the market (Chen et al., 2014, Avery et al., 2015, Jame, Johnston, Markov, & Wolfe, 2016). For example, Jame et al. (2016) examine the value of crowdsourced earnings forecasts and show that Estimize.com forecasts are effective in predicting earnings. When there are a larger number of Estimize.com contributors, they find evidence of more accurate forecasts, which reflects the size of the crowd, increases the benefits of crowdsourcing. Drake, Thornock, & Twedt (2017) identify that articles published on SA advance price efficiency. Farrell et al. (2018) document the role of crowdsourced research as a source of information, especially for retail investors. Avery et al. (2015) form a portfolio that short sells stocks with an excessive number of negative picks and buys stocks with an excessive number of positive picks on Motley Fool, resulting in a yearly return of 12 percent. They show that crowdsourced earnings estimates are as accurate as those of professional analysts.

Methodology
Many studies have tried to correct for cross-sectional correlation in error terms. Chen et al. (2014) use panel regressions with standard errors clustered by firm and year-month. They use long holding periods from one month to three years, which introduces overlapping return periods. Campbell et al. (2019) use panel regressions with standard errors clustered by year-month. Major corporate events clustered by industry over time still contribute to the cross-correlation of abnormal results. Mitchell & Stafford (2000) show that overstated test statistics are likely to result from cross-sectional dependence of event firm abnormal returns that overlap in calendar time. They also argue that major corporate events cluster over time by industry, resulting in positive cross-relationships of abnormal returns that overstate test statistics. To mitigate the issue of inflated t-statistics due to these issues, we use CTIME (Fama, 1998) with non-overlapping return periods.

We use value-weighted abnormal returns instead of equal-weighted abnormal returns. Barber et al. (2001) show that 1) Equal weighting of returns results in severely exaggerated returns on portfolios; 2) Using a value weighting method helps us to better reflect the economic significance of our findings since the individual returns of larger companies are more heavily represented in the aggregate return than those of smaller firms.

1.2.1.2 The Role of Sell-side Analysts in Capital Markets
There are two types of PAs in general: buy-side analysts and sell-side analysts (hereafter, SSA). SA articles are relatively long and similar in format to those written by professional SSAs (Chen et al., 2014a).

SSAs have played an important role in capital markets for decades. Research of SSAs aids in establishing the market's expectations for earnings and stock price, supports trading recommendations, and gives investors important information on crucial stock investing debates. Their predictions and views are prominently reported in the business press and news media (Rees et al., 2015). Many studies examine the effect of their activities on markets, and show that their reports move markets (Beyer et al., 2010; Frankel et al., 2006; Gleason & Lee, 2003; Li et al., 2015). SSAs face several incentives, including pleasing management, generating trading commissions, and increasing investment banking transactions, which affect the objectivity of their recommendations and forecasts negatively (Jackson, 2005; Lin & McNichols, 1998; Mayew, 2008). Despite these conflicting interests, SSAs have been widely considered as the principal source of stock investment research for investors for years.

Regulatory changes and changes in the market's supply and demand for information have shifted the landscape of sell-side stock research over the past few years (M. Drake et al., 2020). Regulations limited equity research departments' ability to support and facilitate investment banking transactions for their brokerages, and as a result, many of the best analysts left the industry or moved to the buy-side (Guan et al., 2019). SSAs are now more concentrated on their efforts to generate revenue from their research via trading commissions (Groysberg & Healy, 2020; Kadan et al., 2008). For institutional clients with high commissions, SSAs devote more time to meeting their needs by providing them with more specialized and personalized services (Brown et al., 2015; Green et al., 2014). Budgets and headcounts of equity research departments have been continuously declining in recent years as a result of these changes (M. S. Drake et al., 2020; Groysberg & Healy, 2020).

Regulatory changes that affect the demand for and supply of sell-side research have contributed to the budgets cut for equity research. Meanwhile, alternative forms of equity research have emerged on social media, such as that contributed by NPAs. In comparison to the decreasing number of SSAs, the number of NPAs contributing on SA has increased in recent years (M. S. Drake et al., 2020). These NPAs provide value-relevant information at least on average, which we have shown in the section 1.2.1.1. A more detailed literature review on PAs can be found in Appendix F.

Each SA Marketplace NPA makes recommendations to SA's clients (Marketplace subscribers) and gets paid based on the number of Marketplace subscribers, which makes Marketplace NPAs comparable to SSAs. This research contributes to the literature on the role of paid analysts in financial markets by being the first study to examine the article quality in SA Marketplace.

1.2.2 The performance of retail investors

Early research shows that retail investors underperform. For example, Barber & Odean (2000) show that retail investors' relatively high turnover rates and poor performance can be explained by overconfidence. Hvidkjaer (2008) suggests that stocks preferred by retail investors seem to go through large and protracted underperformance. Barber & Odean (2013) show that retail investors (1) underperform standard benchmarks (e.g., low-cost index funds), (2) show the "disposition effect," (3) are significantly affected by limited attention and performance of past purchasing decisions (4) repeat behaviors that previously coincided with pleasure and avoid those that coincided with pain, and (5) invest in undiversified stock portfolios.

In comparison, more recent studies show that retail investors have the ability to choose stocks wisely as a group. Kaniel, Liu, Saar, & Titman (2012) is the first research to use US data to find evidence of informed retail investors trading around company events. Their interpretation of the results is: 1) more sophisticated retail investors are insiders who have privileged access to special information, 2) the aggregate information can be useful even though only a small percentage of the investors happen to come across some valuable information. Kelley & Tetlock (2017) show that retail short-sellers have unique information in the retail investor community, maybe because retail short-sellers have valuable information gleaned from geographic proximity to firms, social networks, and employment relationships. Boehmer et al. (2019) provide supportive evidence that "retail investors are better informed about firm-level news and are likely to have valuable private information." Farrell et al. (2018) indicate that SA research leads to more informed retail trading and plays a positive role in retail-investor decision-making.

1.3 Seeking Alpha, Sample, and Descriptive Statistics

This section discusses the data sources and sample selection. It also shows descriptive statistics. Our study uses data collected from SA articles, SA comments, and financial market data from the Center for Research in Security and Prices (CRSP), as well as abnormal returns and Fama-French factors both from WRDS (Wharton Research Data Services). The sample period is from January 2005 to December 2018.

1.3.1 Seeking Alpha Platform

SA is one of the largest crowdsourced investment research social media websites in the US. SA has "5.2 million unique visitors generating 40 million visits monthly and spending 7 minutes per visit on average" ("Who reads Seeking Alpha," 2020). There are in total more than 15,000 contributors who publish research articles on SA. Chen et al. (2014) demonstrate that the primary reasons for contributing to SA are: utility from attention and recognition, monetary rewards, feedback to rectify bad ideas, and convergence of market prices to what the authors claim to be the fundamental value. These motivations have been confirmed by SA author testimonials (*Author Testimonials*, 2020).

1.3.2 Seeking Alpha Subscription Plans

We collected research articles from SA, which were published before January 1, 2019. An editorial board reviews SA articles to ensure the highest quality standards. Starting from 2010, authors have to disclose their holding positions on the stock they write. We focus on articles (available at seekingalpha.com/article) instead of news (available at seekingalpha.com/news) or blog posts (not reviewed by the editorial team for publication). SA offers three types of website-level subscription services Basic, Premium, and PRO, as shown in Table 1.

Regarding the website level subscription, investors will subscribe to the SA website as a whole instead of subscribing to individual contributors. The “PRO” articles are exclusive to pro-subscribers for 24 hours before they are put into public for 30 days, after which the articles will be archived behind a paywall and only fully readable by PRO-subscribers (Campbell et al., 2019). Free articles are archived behind the paywall when they are more than ten days old. SA “PRO” subscription gives investors exclusive access to “PRO” articles and other archived articles, which are behind the paywall.

SA has another tier of subscription plan called “MARKETPLACE,” where you can subscribe to each Marketplace whose service suits you. One or more SA contributors can form a group, a.k.a SA Marketplace, to offer specific investment research and guidance. Services are led by individual authors and specific communities of investors with similar interests, focusing on a particular investment style and methods. It enables investors to obtain guidance and ideas that suit their needs (*SA Marketplace*, 2020). SA Marketplace subscription is different from the other three subscription plans as the investor will subscribe to each Marketplace instead of the whole website. What’s more, SA Marketplace “offers a curated, VIP experience: exclusive real-time investing or trading ideas, direct contact with the service’s leaders and a community chat room for in-depth discussion” (*Seeking Alpha’s Marketplace of Guided Investment Communities*, 2020). A detailed comparison of different subscription plans is shown in Table 1.

Table 1 SA Subscription Plans

SA has four tiers of subscription plans: BASIC, PREMIUM, PRO, and MARKETPLACE. The first three subscription plans are website-level subscription plans. One or a few more SA contributors can form a MARKETPLACE. As a result, a MARKETPLACE subscription plan is an individual-level subscription plan. The monthly subscription price and features of each subscription plan are listed in the table.

Plan	BASIC	PREMIUM	PRO	MARKET PLACE
Monthly Price if Billed Monthly	Free	$29.99 USD	$299.99 USD	$25 USD to $375 USD per market place
Monthly Price if Billed Annually	Free	$19.99 USD	$199.99 USD	$14.17 USD to $208.25 USD per market place
Level	Website	Website	Website	Individual
Features	Stock News & analysis alerts	**All BASIC features, plus:**	**All PREMIUM features, plus:**	Exclusive experience from each Marketplace
	Investing newsletters	Unlimited access: 1 million articles	Top Ideas	Exclusive real-time investing or trading ideas
	Follow authors and receive new article alerts	Seeking Alpha Author Ratings	PRO content & newsletters	Direct contact with the service's leaders
	Save articles	Seeking Alpha Author Performance	Short ideas portal	A community chat room for an in-depth discussion
	Comments	Quant Ratings	Idea screener/filter	
	Blogs & StockTalks	Dividend scores & forecasts	VIP Service	
		Ad-lite	No ads	

1.3.3 Seeking Alpha Data

In this research, we have collected all the free and "PRO" research articles published between January 2005 and December 2018. Overall, we have 771,039 SA articles in our sample.

SA appoints a unique article ID to each article. One or more tickers are tagged on each article before publication. Each SA article has two types of ticker groups: "About" and "Includes." Typically, "About" shows the tickers that the article is focusing on, while "Includes" demonstrates the tickers that the article mentions. Single-ticker articles concentrate on one stock, which makes it easier to extract the author's opinion of the firm (Chen et al., 2014a). In comparison, extracting the author's disparate views on each tagged stock is more difficult for multiple-ticker articles. In this research, we consider articles with a single ticker. We consider an article a single-ticker article when there is only one ticker in "About" and "Includes" tickers of each article. Excluding multiple-ticker articles reduces our sample by 491,129 articles. We have 279,910 single-ticker articles. We deleted 30,743 SA transcripts and 90 articles, which have less than 100 characters. As a result, we have 249,077 single-ticker articles, which is about one one-third of all articles published on SA during the sample period and is consistent with prior samples (Chen et al. 2014).

In general, disclosures follow the same basic format. The author incorporates a disclosure statement at the beginning or end of each article. For instance, an author may include "I am/we are long XXXX," "I/we have no positions in any stocks mentioned, and no plans to initiate any positions within the next 72 hours," or "I am/we are short XXXX." However, the positions are sometimes less clear because the author may disclose complex option holdings or multiple positions in different stocks (i.e., long XXXX and short YYYY). As a result, we use the following procedure to parse these disclosures.

We follow the literature (Campbell et al., 2019) in identifying the SA contributors' positions in their articles. Our first step follows the same procedures proposed by Campbell et al. (2019). First, the long positions are identified by searching for the terms "long," "hold," or "own stock/shares." Then, we capture the text after these words, stopping when we encounter a period or the word "may" or "short," which indicate the beginning of a new disclosed position (i.e., "I am long XXXX and may …"). We repeat a similar process for possible short positions, looking for the word "short" and then capturing tickers until the word "long" or "may," or a period. We do not allow negating or qualifying words (no, not, none, neither, never, nobody, may, or plan) to appear within the five words before the position indicator for long or short positions. Finally, we look for clues that the author does not hold any position in any stocks. These include the terms “No position,” “None,” or “May.”

Inspection of results shows that the first step procedures are relatively accurate. However, we do come across complex cases. For instance, some articles result in multiple classifications (i.e., long, short, and/or no position) given rise by complex information disclosure. There are also situations where none of these three positions can be identified. In addition, the disclosed positions can be related to stock tickers other than the stock tickers included in “About” or “Include.” As a result, we apply the following procedures to refine our disclosure coding. First, we require a ticker in the disclosed positions to appear in “About” or “Include” to be considered for a long or short position. If the ticker does not show up in “About” or “Include,” we code the disclosed position of the ticker as “no position.” Second, if we identify multiple classifications or fail to recognize a long, short, or no position, we code the disclosed position as “no position.”

SA users are allowed to post comments on published articles. We collected all the comments posted on all the single-ticker articles. Chen et al. (2014) demonstrate that sixty percent of the comments are posted on the day of article publication. Another twenty percent are posted on the next day, and the last twenty percent are posted occasionally in the following weeks. Therefore, we have collected 3,424,953 commentaries written within the first 48 hours after the article publication time. The information that we collect about each comment includes article ID, comment ID, content, user ID, created date-time, discussion ID, and username.

1.3.4 WRDS Data

A firm’s ticker is a standard stock identifier. However, ticker-firm combinations change over time. For instance, the ticker of a company, which ceases to exist, maybe reassigned to another company. Another example is that in the case of a merger. After the merger is completed, the stock ticker of the acquired company is usually changed to the stock ticker of the acquirer.

Therefore, we used PERMNO, which is a permanent identifier. PERMNO does not change throughout a firm's life span regardless of the name change or other circumstances and is never reused. PERMNO is a unique stock (share class) level identifier. Although most companies only own one class of shares, some companies own more than one class of shares traded at different prices. As a result, one company can have more than one PERMNO.

CRSP STOCKNAMES file provides a mapping between CRSP permanent identifiers (PERMNOs) and all historical company names, and exchange tickers, along with their effective date ranges. In this research, we use STOCKNAMES to map stock tickers to PERMNOs. We delete 36,036 articles whose ticker cannot be mapped to PERMNO using the CRSP STOCKNAMES file. After mapping stock tickers to PERMNOs, we have 213,041 articles.

For each firm, we collect data on PERMNO, date, share code, share price, share volume, shares outstanding, holding period return, and closing bid and ask prices from CRSP.

1.3.5 Descriptive Statistics

Table 2 shows the beginning SA dataset and the sources of data loss.

Table 2 Sample Attrition

This table describes the dataset used in this study.

SA Articles Contributed Between 2005 and 2018	771,039
Articles with multiple tickers	(491,129)
Articles with Single Ticker	279,910
Articles which are transcripts	(30,743)
Articles with less than or equal to 100 characters	(90)
Articles Which Are Not Transcripts and Have More Than 100 Characters	249,077
Articles whose ticker cannot be mapped to CRSP PERMNO	(36,036)
Articles with Ticker Mapped to PERMNO	213,041
Total Number of Articles in the dataset used in this study	213,041

Figure 1 shows the number of single-ticker articles submitted to SA from January 2005 to December 2018, which sums up to 213,041 articles. As shown in Figure 1, activity on the SA website has been increasing over the years, reaching about 30,000 single-ticker articles in 2015.

Figure 1 Total Number of Articles by Year

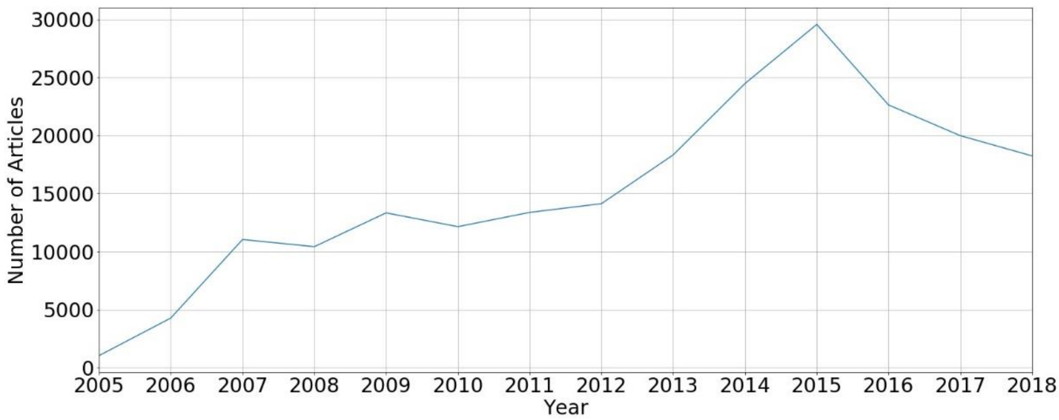


Table 3 reports the growth in coverage of SA using single ticker articles between 2005 and 2018. Column (1) demonstrates that the number of firms covered increases from 320 in 2005 to 3,338 in 2018; column (2) indicates that the number of articles rises from 1,031 in 2005 to 18,245 in 2018; column (3) shows that the number of contributing authors advances from 55 in 2005 to 1,833 in 2018. As shown in column (5) in the last row, the mean number of analysts per firm is 3.5, and column (7) in the last row indicates that the average number of firms per author is 6.7.

Table 3 Descriptive Statistics of the SA Single-Ticker Articles, by Year

This table reports descriptive statistics on single-ticker SA research articles per year. The sample includes 213,041 single-ticker research articles contributed by 11,547 authors. Column (1) shows the number of firms with at least one article for each year. Column (2) indicates the number of articles for each year. Column (3) demonstrates the number of contributing authors for each year. Column (4) reports the average number of articles contributed by each author for each year. The mean and median number of authors who contribute articles for each covered firm are shown in columns (5) and (6), respectively, followed by the average and median numbers of firms each author covers in columns (7) and (8). Average across years is reported as the bottom row. The sample period is from January 2005 to December 2018.

Year	No. of Firms (1)	No. of Articles (2)	No. of Authors (3)	Articles per Author (4)	Authors per Firm		Firms per Author	
					Mean (5)	Median (6)	Mean (7)	Median (8)
2005	320	1031	55	19.074	1.294	1	7.667	2
2006	1284	4270	285	15.018	2.241	1	10.13	3
2007	2131	11040	679	16.26	3.047	1	9.578	2.5
2008	1979	10424	1036	10.062	3.143	1	6.004	2
2009	1936	13337	1166	11.438	3.814	1.5	6.332	2
2010	2151	12149	1040	11.682	3.344	1	6.915	2
2011	2237	13377	1444	9.264	3.779	1	5.855	2
2012	2337	14130	1825	7.742	4.217	2	5.401	2
2013	3320	18337	2348	7.81	3.839	2	5.428	2
2014	3559	24487	2533	9.667	4.408	2	6.193	2
2015	3968	29581	2569	11.515	4.492	2	6.938	2
2016	3230	22640	2378	9.521	3.955	2	5.373	2
2017	3138	19993	2180	9.171	3.964	2	5.706	2
2018	3338	18245	1833	9.954	3.467	2	6.314	2
Average	2494.857	15217.21	1526.5	11.298	3.5	1.536	6.702	2.107

1.4 The Effect of Seeking Alpha Sentiment on Investors' Trading

In this section, we analyze the effect of SA article sentiment on investor trading. We first examine whether SA sentiment influences the direction of investor trading by studying the relation between SA article sentiment and order imbalances.

Order imbalance is defined as the difference between buyer-initiated volume and seller-initiated volume, scaled by total trading volume.

$$OIB_{i,t} = \frac{V_{i,t}^{Buy} - V_{i,t}^{Sell}}{V_{i,t}^{Buy} + V_{i,t}^{Sell}} \tag{1}$$

where,

$OIB_{i,t}$ = the order imbalance for stock *i* on day *t*.

$V_{i,t}^{Buy}$ = the buyer-initiated volume for stock *i* on day *t*.

$V_{i,t}^{Sell}$ = the seller-initiated volume for stock *i* on day *t*.

We only consider single-ticker articles. An article is considered “bullish” if its percentage of negative words is below the median of its overall distribution; an article is considered “bearish” if its percentage of negative words is above the median (Chen et al., 2014a). We construct variables that stand for the number of bullish (bearish) articles and consider a sentiment score,

$$Sentiment_{i,t} = \frac{NumBullishArticles_{i,t} - NumBearishArticles_{i,t}}{NumBullishArticles_{i,t} + NumBearishArticles_{i,t}} \qquad (2)$$

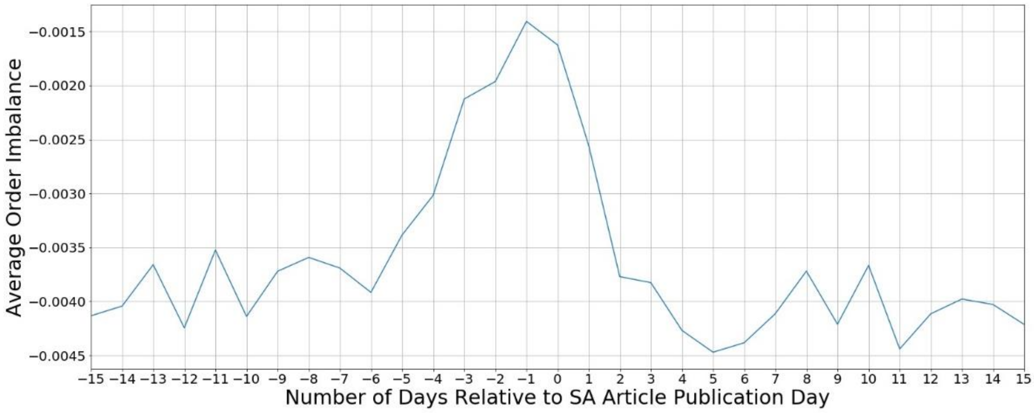
$NumBullishArticles_{i,t}$ = the number of bullish articles for stock *i* on day *t*.

$NumBearishArticles_{i,t}$ = the number of bearish articles for stock *i* on day *t*.

1.4.1 Average Order Imbalances on Different Days

We consider the order imbalance for stock *i* on day $t+n$. day *t* is the SA article publication day. In Figure 2, the x-axis is *n*, which is the number of days that order imbalance is relative to day *t*. *n* is limited to [-15, 15]. The y-axis is the average order imbalance. It is clear that average order imbalance starts to increase continuously when *n* is -6 and peaks when *n* is -1. Average order imbalance starts to decrease continuously when *n* is -1 and stops decreasing when *n* is +5.

Figure 2 Average Order Imbalance on Different Days



1.4.2 Empirical Models

We then estimate the following regression:

$$OIB_{i,t+n} = \beta_0 + \beta_1 PostSA_{i,t+n} + \beta_2 Sentiment_{i,t} + \beta_3 PostSA_{i,t+n} * Sentiment_{i,t} + Controls_{i,t} + \varepsilon_{i,t} \quad (3)$$

where,

$OIB_{i,t+n}$ = the order imbalance for stock *i* on day t+n. day *t* is the SA article publication day. *n* is the number of days that order imbalance is relative to day *t*. *n* is limited to [-6, +6] event window, excluding event day 0. Our choice of *n* is based on Figure 2. Event day 0 is excluded because it's neither before nor after SA article publication.

$Sentiment_{i,t}$ = the sentiment score for stock *i* on day *t* as shown in equation (2).

$PostSA_{i,t+n}$ = the dummy variable equals one if the trading is measured after day *t* and zero if trading is measured before day t. Thus, $PostSA_{i,t+n}$ equals one over the [1, 6] window and zero over the [-6, -1] window.

$\beta_1 PostSA_{i,t+n} * NormalizedPos_{i,t}$ interacts the event time indicators with the sentiment score.

$Controls_{i,t}$ include the following variables:

$Volatility_{i,t}$ = The sum of squared daily returns of stock *i* in the calendar month preceding article's publication day *t*.

$Size_{i,t}$ = The natural log of the market value equity of stock *i* as of the end of the month prior to the article's publication day *t*.

$BTM_{i,t}$ = The book value of equity as of the end of the most recent fiscal year divided by the market value of equity as of the end of the prior year.

$InstOwn_{i,t}$ = The proportion of shares of stock *i* owned by institutional investors per the quarterly Thomson Reuters ownership report closest but prior to the article's publication day *t*.

$MKTRF_{t+n} = R_{MKT,t+n} - R_{f,t+n}$, which is the excess return on the market on day t+n. It is calculated as the value-weight return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate.

SMB_{t+n} = the size factor, historic excess returns of small-cap stocks over large-cap stocks on day t+n.

HML_{t+n} = the value factor, historic excess returns of value stocks (high book-to-market ratio) over growth stocks (low book-to-market ratio) on day t+n.

MOM_{t+n} = the momentum factor, historic excess returns of highest performing stocks over lowest performing stocks on day t+n.

T-statistics are computed using two-way clustered standard errors. The standard errors are clustered by stock and day to account for cross-sectional correlation in residuals[8].

[8] To obtain unbiased estimates in finite samples, the clustered standard errors are adjusted by (N-1)/(N-P)× G/(G-1), where N is the sample size, P is the number of independent variables, and G is the number of clusters (Mark (Shuai), 2020).

So far, we have used sentiment score as shown in equation (2). We also measure sentiment using SA NPAs' disclosed investment positions. We call it the normalized position score and show results in Appendix A.

1.4.3 Empirical Results

Column (1) of Table 4 shows that order imbalances are not significantly related to sentiment scores before the publication of SA articles as the coefficient on $Sentiment_{i,t}$ is not significant (0.047). The difference between order imbalances before and after SA article publication day are not significantly related to the sentiment scores, as the coefficient on the interaction term is not significant (0.095).

We further test if the sum of the coefficient on $Sentiment_{i,t}$ and the coefficient on the interaction term is significantly different from zero and show results in Column (2) of Table 4. Order imbalances are significantly related to sentiment scores after the publication of SA articles, as the sum of the coefficients is significant (0.142). A one-unit increase in the sentiment score is associated with 0.142 percentage points increase in order imbalances after the SA article publication. Another way to measure sentiment using disclosed positions is reported in Appendix A Table A.1. Table A.1 Columns (1) and (2) show similar results as Table 4 Columns (1) and (2).

We have just examined $PostSA_{i,t+n}$ equals one over the [1, 6] window and zero over the [-6, -1] window. We have also tested two other settings: 1) $PostSA_{i,t+n}$ equals one over the [0, 6] window and zero over the [-6, -1] window, 2) $PostSA_{i,t+n}$ equals one over the [1, 6] window and zero over the [-6, 0] window. Our results do not change when using these two settings.

We divide our data into two groups using the median of market cap. A stock is considered "small cap" if its market cap is below the median of its overall distribution. The result for "small cap" stocks is shown in Table 4 Column (3). A stock is considered "big cap" if its market cap is above the median. The result for “big cap” stocks is shown in Table 4 Column (4). The sum of the coefficient on *Sentiment*$_{i,t}$ and the coefficient on the interaction term is significantly (not significantly) different from zero in Table 4 Column 3 (Column 4). For “small cap” stocks, a one-unit increase in the sentiment score is associated with 0.3 percentage points increase in order imbalances after the SA article publication. The evidence shows that SA is more important for small cap stocks.

The evidence shows that investors react to SA article sentiment. In section 3.5, we examine whether investors can gain positive abnormal returns using SA research articles.

Table 4 Seeking Alpha Article Sentiment and Direction of Investor Trading

This table presents results from the estimation of Equation (3). Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level. Coefficients and standard errors are multiplied by 100 for presentation purposes.

Variable	(1)	(2)	(3)	(4)
$PostSA_{i,t+n}$ * $Sentiment_{i,t}$	0.095			
	(0.075)	0.142*	0.3***	0.06
$Sentiment_{i,t}$	0.047	(0.061)	(0.071)	(0.112)
	(0.058)			
$PostSA_{i,t+n}$	-0.242*	-0.242*	-0.202*	-0.414**
	(0.074)	(0.074)	(0.086)	(0.141)
$Volatility_{i,t}$	-0.515	-0.515	-0.497	-0.65
	(0.78)	(0.78)	(0.783)	(5.983)
$Size_{i,t}$	0.017	0.017	0.13***	0.118*
	(0.017)	(0.017)	(0.026)	(0.049)
$BTM_{i,t}$	0.165*	0.165*	-0.133	0.653***
	(0.065)	(0.065)	(0.111)	(0.068)
$InstOwn_{i,t}$	0.783***	0.783***	-0.004	3.149***
	(0.205)	(0.205)	(0.236)	(0.317)
$MKTRF_{t+n}$	114.452***	114.452***	97.568***	133.14***
	(1.979)	(1.979)	(3.204)	(2.266)
SMB_{t+n}	40.818***	40.818***	78.458***	1.746
	(3.328)	(3.328)	(5.313)	(3.914)
HML_{t+n}	0.933	0.933	-3.218	4.701
	(3.602)	(3.602)	(5.82)	(4.117)
MOM_{t+n}	3.744	3.744	9.953**	-2.912
	(2.434)	(2.434)	(3.833)	(2.912)

R2	0.018	0.018	0.012	0.048
Observation	354,899	354,899	177,409	177,490

1.5 Can Retail Investors Capture the Benefits of Retail Financial Advice?

1.5.1 Calendar Time Portfolio Strategies

In order to examine whether investors can profit from crowdsourced research articles on SA from a trader's perspective, we evaluate the profitability of calendar time strategies.

To measure the article sentiment, we calculate the percentage of negative words ($NegSA_{i,j,t-1}$) for each article using word lists from Loughran & McDonald (2011). The average percentage of negative words, $\overline{NegSA}_{i,t-1}$, for firm *i* on day *t-1* is calculated by taking the sum of individual percentage of negative words, $NegSA_{i,j,t-1}$, of the last article published by each author $j = 1$ to $n_{i,t-1}$ on day *t-1*, who have written articles for the firm on day *t-1* and dividing by $n_{i,t-1}$.

$$\overline{NegSA}_{i,t-1} = \frac{1}{n_{i,t-1}} \sum_{j=1}^{n_{i,t-1}} NegSA_{i,j,t-1} \tag{4}$$

Table 5 shows descriptive statistics of the percentage of negative words for the single-ticker articles by year. Columns (2) to (8) of Table 5 show the descriptive statistics of the percentage of negative words in articles. We find that the highest mean and median percentage of negative words are in the year 2009.

Table 5 Descriptive Statistics of Percentage of Negative Words for the Single-Ticker Articles, by Year

Column (1) reports the number of firms with at least one single-ticker article for each year. Columns (2) to (8) show the descriptive statistics of the percentage of negative words in articles. Numbers in columns (2) to (8) are multiplied by 100 for presentation purposes. The sample period is from January 2005 to December 2018.

Year	No. of Firms (1)	Percentage of Negative Words						
		Mean (2)	Median (3)	Std (4)	Min (5)	25th (6)	75th (7)	Max (8)
2005	320	1.313	0.939	1.439	0	0.246	1.882	15.385
2006	1284	1.731	1.418	1.452	0	0.725	2.361	16.000
2007	2131	1.730	1.433	1.418	0	0.717	2.414	20.690
2008	1979	1.959	1.689	1.450	0	0.905	2.744	14.493
2009	1936	2.040	1.786	1.468	0	0.968	2.809	19.231
2010	2151	1.789	1.495	1.404	0	0.800	2.449	13.043
2011	2237	1.768	1.527	1.272	0	0.838	2.439	11.957
2012	2337	1.626	1.419	1.110	0	0.796	2.237	10.652
2013	3320	1.513	1.351	0.962	0	0.826	1.994	8.211
2014	3559	1.456	1.273	0.978	0	0.756	1.942	8.943
2015	3968	1.623	1.449	1.006	0	0.896	2.149	13.333
2016	3230	1.741	1.558	1.019	0	1.011	2.273	12.281
2017	3138	1.528	1.384	0.873	0	0.906	1.983	9.471
2018	3338	1.545	1.397	0.872	0	0.924	1.998	8.347

The U.S. stock market exchanges - the New York Stock Exchange and Nasdaq - are open between 9:30 a.m. and 4 p.m. Eastern Time. We sort firms into quintile portfolios according to the average percentage of negative words ranking of each firm observed before 3:45 pm Eastern Time each trading day. On each day *t-1*, each covered firm is categorized into one of five portfolios. The first portfolio is composed of the stocks with the most positive sentiments (lowest average percentage of negative words), while the fifth portfolio consists of firms with the most negative sentiments (highest average percentage of negative words). We investigate a trading strategy of buying the 1st quintile and shorting the 5th quintile, denoted as the 1-5 portfolio.

When a longer date interval *M* (between day *t-m* (included) and *t-1* (included)) is considered for the quintile portfolios, on each day, we find all stocks that SA articles had covered over the last *m* trading days until day *t-1*. We inspect the trading strategies based on *M* equal to 1 day, 1 week, 1 month or 1 quarter (calendar days). We assume 5 trading days per week and 21 trading days per month for these calculations. As a result, 1-week, 1-month, and 1-quarter date intervals are calculated using 5, 21, and 63 trading days.

The average percentage of negative words, $\overline{NegSA}_{i,M}$, for firm *i* between day *t-m* (included) and *t-1* (included) is calculated by taking the sum of the individual percentage of negative words, $NegSA_{i,j,M}$, of the last article published by each author $j = 1$ to $n_{i,M}$ during date interval *M*, who have written articles for the firm during date interval *M* and dividing by $n_{i,M}$.

$$\overline{NegSA}_{i,M} = \frac{1}{n_{i,M}} \sum_{j=1}^{n_{i,M}} NegSA_{i,j,M} \tag{5}$$

During date interval *M*, the first portfolio is composed of the stocks with the most positive sentiments (lowest average percentage of negative words), while the fifth portfolio consists of firms with the most negative sentiments (highest average percentage of negative words). Again, we investigate the trading strategy of buying the 1st quintile and shorting the 5th quintile, denoted as the 1-5 portfolio.

After determining the composition of each portfolio *p* as the close of trading day *t-1*, we calculate the value-weighted return for day *t*, which is the next trading day. The return of portfolio *p*, $R_{p,t}$, is given by:

$$R_{p,t} = \sum_{i=1}^{n_{p,t-1}} \frac{MktCap_{i,t-1}}{\sum_{j=1}^{n_{p,t-1}} MktCap_{j,t-1}} R_{i,t} \tag{6}$$

where

$MktCap_{i,t-1}$ = the market capitalization for firm *i* at the close of trading on day *t-1*,

$R_{i,t}$ = the return on the common stocks of firm *i* on day *t*, and

$n_{p,t-1}$ = the number of firms in portfolio *p* at the close of trading on day *t-1*.

Following Barber, Lehavy, McNichols, & Trueman (2001), we use value weight instead of equally weight for two reasons: 1) an equal weighting of daily returns will give rise to overstated portfolio returns[9]; 2) a value-weighted daily return will give us the chance to better capture the economic significance of results, as the individual returns of firms with the larger market cap will be more heavily represented in the aggregate return than the returns of firms with smaller market cap.

We examine the trading strategies based on *M* equals to 1 day, 1 week, 1 month or 1 quarter. Each stock that enters a portfolio will be held for date interval *T (holding period)*. In this case, *T* equals to 1 day, 1 month, 1 week or 1 quarter.

Daily return for each portfolio p, $R_{p,t}$, are compounded over the trading days to capture returns with a longer holding period, $R_{p,T}$.

$$R_{p,T} = \prod_{t=1}^{n}(1 + R_{p,t}) - 1 \tag{7}$$

As a result, we end up with 16 different portfolio groups with a different combination of article periods (*M*) and holding periods (*T*), as shown in Table 6.

[9] This problem emerges because of the cycling of the closing price of a company between the bid and ask price (commonly referred to as the bid-ask bounce). For a more detailed discussion, see Blume & Stambaugh (1983), B. M. Barber & Lyon (1997), Canina et al. (1998), and Lyon et al. (1999).

Table 6 Sixteen Portfolio Groups with Different Articles Periods and Holding Periods

Holding Periods Article Periods	1 day	1 week	1 month	1 quarter
1 day	Portfolio Group d-d	Portfolio Group d-w	Portfolio Group d-m	Portfolio Group d-q
1 week	Portfolio Group w-d	Portfolio Group w-w	Portfolio Group w-m	Portfolio Group w-q
1 month	Portfolio Group m-d	Portfolio Group m-w	Portfolio Group m-m	Portfolio Group m-q
1 quarter	Portfolio Group q-d	Portfolio Group q-w	Portfolio Group q-m	Portfolio Group q-q

We use the classic Carhart four-factor model (Carhart, 1997), which is an extension of the Fama-French three-factor model (Fama & French, 1993). Fama & French (1993) indicates three factors, RMRF, HML and SMB which have been proven to explain the cross-section of stock returns. Carhart (1997) introduces another factor, Momentum (Mom) which also helps explain future returns.

For each portfolio composed of differences in returns between stocks favored (based on least negative sentiment) and disfavored (based on most negative sentiment) by SA articles, we compute the alpha using the following regression. Time T is our holding date interval, for which we use 1 month as an example to demonstrate the regression.

$$R_{B,T} - R_{S,T} = \alpha_p + \beta_p MKTRF_T + s_p SMB_T + h_p HML_T + m_p MOM_T + \varepsilon_{p,T} \qquad (8)$$

where

$R_{B,T}$ = the month T return on a portfolio that buys the 1st quintile, which is composed of the most favored stocks.

$R_{S,T}$ = the month T return on a portfolio that buys the 5th quintile, which is composed of the least favored stocks.

$MKTRF_T = R_{MKT,T} - R_{f,T}$, which is the excess return on the market in month *T*. It is calculated as the value-weight return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate.

$SMB_T =$ the size factor, historic excess month T returns of small-cap stocks over large-cap stocks, which is the average of returns on three small portfolios minus the average return on three big portfolios,

$$SMB = \frac{1}{3}(Small\ Value + Small\ Neutral + Small\ Growth) - \frac{1}{3}(Big\ Value + Big\ Neutral + Big\ Growth)$$

$HML_T =$ the value factor, historic excess month *T* returns of value stocks (high book-to-market ratio) over growth stocks (low book-to-market ratio), which is the average return on two value portfolios minus the average return on two growth portfolios,

$$HML = \frac{1}{2}(Small\ Value + Big\ Value) - \frac{1}{2}(Small\ Growth + Big\ Growth)$$

$MOM_T =$ the momentum factor, historic excess month *T* returns of highest performing stocks over lowest performing stocks.

$\alpha_p =$ the difference between the portfolio's return and the expected return.

$\beta_p, s_p, h_p, m_p =$ factor beta (factor loading) is the sensitivity of a portfolio's returns to a particular systematic factor.

Equation (8) uses single-ticker articles and ranking of percentage of negative words to form portfolios. We report another way to form portfolios using all articles (single ticker articles and multiple ticker articles) and disclosed positions as trading signals in Appendix C.

In the following analysis, we use these factor loadings to interpret the nature of stocks in each portfolio. If the value of β_p is greater (less) than 1, then companies in portfolio *p* are on average riskier (less risky) than the market. If the value of s_p is greater (less) than 0, then it indicates a portfolio tilted toward smaller (larger) companies. If the value of h_p is greater (less) than 0, then it shows that a portfolio is tilted toward value (growth) firms. If the value of m_p is greater (less) than 0, then it signifies that a portfolio is composed of stocks which on average performed well (poorly) in the recent past.

1.5.2 Empirical Results

The returns of each portfolio *p* during time *T* are regressed on the factor portfolio returns. Each time period is weighted equally in our portfolios, as each portfolio is built using the last article published by each author during the date interval *M*. In such a way, a time interval with more articles published by an author gets no larger weight in our regression.

Table 6 shows our 16 portfolio groups. For each portfolio group with more than one trading day as the holding period, we can set the holding period to start on different trading days to transform one portfolio into a portfolio group. For instance, regarding 5 trading days, we can set the holding period to begin on Monday, Tuesday, Wednesday, Thursday, or Friday. As a result, one portfolio with 5 trading days as the holding period will be transformed into a portfolio group with 5 portfolios with the holding periods starting on different weekdays. If we apply similar logic to portfolios with 21 or 63 trading dates, we will end up with portfolio groups with 21 or 63 portfolios in the group. For portfolio groups with 1 day as the holding period, we will only have 1 portfolio in each portfolio group. For portfolio groups with 1 week (1 month, 1 quarter) as a holding period, each portfolio group is composed of 5 (21, 63) portfolios.

The above procedures are used to construct the portfolio groups for the following reasons: 1) portfolios will have non-overlapping return periods, 2) the results will not be biased as we have not fixed the holding periods to start on Monday or Tuesday for each portfolio group. For instance, we use Monday, Tuesday, Wednesday, etc., as the start of holding periods for weekly returns and 1st, 2nd, or 3rd, etc., trading day of the month as the start of holding periods for monthly returns. As a result, weekly returns will start on Monday and end on the following Monday, or start on Tuesday and end on the following Tuesday, and so on. And monthly returns will start on the 1st trading day of this month and end on the 1st trading day of the following month, or start on the 2nd trading day of this month and end on the 2nd trading day of the following month, etc. An illustration of how holding periods are constructed can be found in Figure 3 and Figure 4.

Figure 3 Holding Periods of Weekly Return

Weekly returns will start on Monday and end on the following Monday, or start on Tuesday and end on the following Tuesday, and so on.

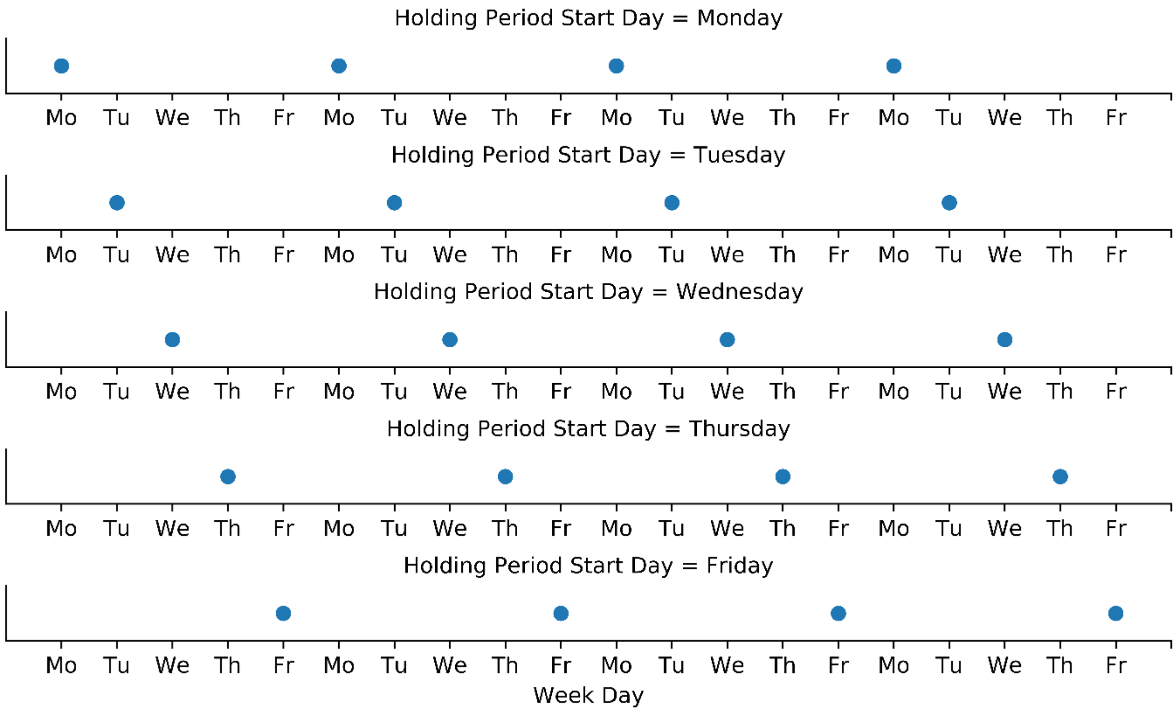


Figure 4 Holding Periods of Monthly Return

Monthly returns will start on the 1st trading day of this month and end on the 1st trading day of the following month, start on the 2nd trading day of this month and end on the 2nd trading day of the following month, etc.

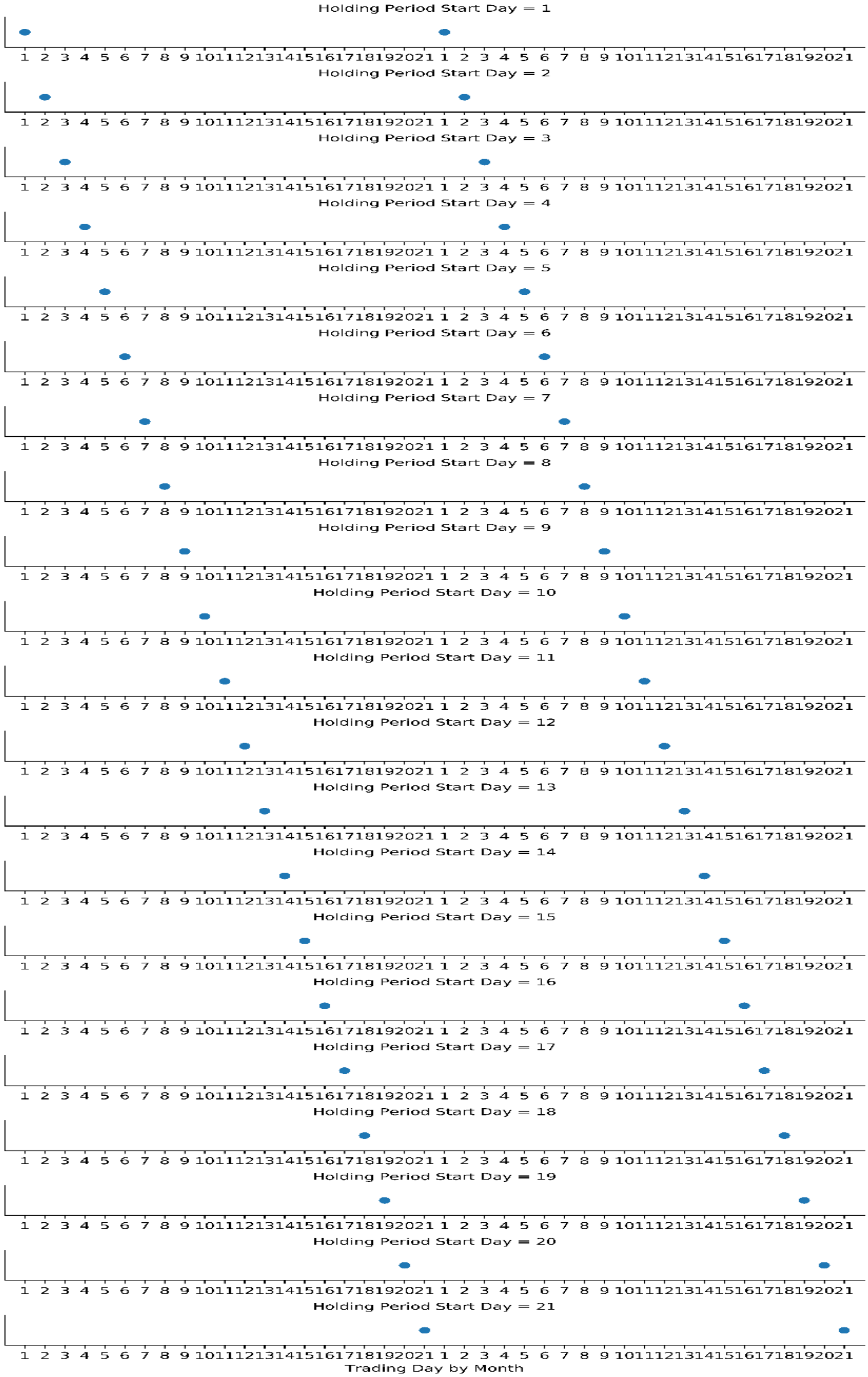

Holding Period Start Day = 1
Holding Period Start Day = 2
Holding Period Start Day = 3
Holding Period Start Day = 4
Holding Period Start Day = 5
Holding Period Start Day = 6
Holding Period Start Day = 7
Holding Period Start Day = 8
Holding Period Start Day = 9
Holding Period Start Day = 10
Holding Period Start Day = 11
Holding Period Start Day = 12
Holding Period Start Day = 13
Holding Period Start Day = 14
Holding Period Start Day = 15
Holding Period Start Day = 16
Holding Period Start Day = 17
Holding Period Start Day = 18
Holding Period Start Day = 19
Holding Period Start Day = 20
Holding Period Start Day = 21
Trading Day by Month

Table 7, columns (1) to (4), demonstrates the estimated alpha values and factor loadings for stocks in portfolio groups (d-d, w-d, m-d, q-d) according to the average percentage of negative words quintile ratings. These portfolio groups have 1 day as the holding period.

The alphas are only significant for portfolio groups d-d and w-d after controlling for the Fama-French factors and momentum. Portfolio group d-d outperforms the model by ten basis points per trading day. The significant coefficients on HML and MOM for portfolio group d-d show that portfolio group d-d behaves like growth stocks that have performed well in the past. The significant coefficients on MKTRF, HML, and MOM for portfolio group w-d show that portfolio group w-d behaves like growth stocks with lower than average market risk, which has performed well in the past. However, the alphas are not significant for portfolio groups m-d or q-d.

Table 7 columns (1) to (4) show that only portfolio groups d-d and w-d with a 1-day holding period have significant alphas. In order to check if the valuable trading information comes from the article period being 1 day or 1 week, we build a portfolio with the article period being 4 trading days (*t-5* to *t-2*) and the trading day after the next trading day as a holding period (*t*). In other words, we are trying to test the portfolio group w-d without the last trading day in the article period. For convenience, we denote our new portfolio group as 4d-d. The result is shown in column (5) of Table 7. We also have another portfolio with an article period being 20 trading days (t-21 to t-2) and the trading day after the next trading day as a holding period (*t*). This is the same as portfolio group m-d without the last trading day in the article period. This new portfolio group is denoted as 20d-d. The result is shown in column (6) of Table 7. It is shown that neither of the alphas from portfolio groups 4d-d nor 20d-d is significant. As a result, it can be shown that the majority of valuable information originates from 1 day article period, which is the last trading day before the holding period.

Table 7 shows results using portfolios formed using single-ticker articles and rankings of the percentage of negative words. Another way to form portfolios using all articles (single-ticker articles and multiple-ticker articles) and disclosed positions as trading signals is reported in Appendix C Table C.1. Table C.1 Columns (1) to (4) do not have any significant alphas. Instead of using portfolios, we examine the association between market return and market-level sentiment. To measure the market level sentiment, we calculate the average percentage of negative words of all single-ticker articles each day and report the results in Appendix B. However, Appendix B Table B.1 does not show any significant correlation between market return and market-level sentiment.

Table 8 shows the number of portfolios and the number of significant alphas in each portfolio group with a holding period longer than 1 day. Each alpha is considered at a 5% confidence level. It shows that the other portfolio groups with 1 week/month/quarter holding periods generally do not have significant alphas.

Based on this result, we decide to use portfolio group d-d as our benchmark and run the following slicing methods in section 1.5.3 with the same article period (1 day) and holding period (1 day). As portfolio group d-d only contains one portfolio, we start to call it portfolio d-d.

Table 7 Fama-French-Carhart Filtering of Portfolio Groups with 1 Day Holding Period

Summary of Fama-French-Carhart four-factor model of portfolio groups with 1 day holding period formed from all single-ticker SA research articles and sorting into quintile ratings based on average percentage of negative words. Each portfolio group only contains 1 portfolio. Quintile 1 is composed of stocks with a relatively low average percentage of negative words, while Quintile 5 consists of firms with a relatively high average percentage of negative words. We investigate the trading strategy of buying the 1st quintile and shorting the 5th quintile, denoted as the 1-5 portfolio. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level. The portfolio is rebalanced daily. The coefficient estimates are those from the Carhart four-factor regression of the portfolio returns (R_B-R_S) on the market excess return (MKTRF), the size factor (SMB), the value factor (HML), and the momentum factor (MOM). Coefficients and standard errors are multiplied by 100 for presentation purposes.

Variable	(1) d-d	(2) w-d	(3) m-d	(4) q-d	(5) 4d-d	（6） 20d-d
Alpha	0.103*	0.052**	0.024.	0.009	0.002	0.02
	(0.041)	(0.019)	(0.013)	(0.01)	(0)	(0)
MKTRF	-0.527	-7.225***	-7.241***	-3.146***	-10.2***	-7.19***
	(3.745)	(1.768)	(1.151)	(0.878)	(1.9)	(1.2)
SMB	-3.32	-5.323	-8.939***	-14.58***	-5.65	-8.71***
	(7.319)	(3.455)	(2.251)	(1.719)	(3.7)	(2.4)
HML	-20.919**	-19.001***	-23.421***	-21.07***	-18.59***	-25.68***
	(7.758)	(3.662)	(2.382)	(1.819)	(3.9)	(2.5)
MOM	47.149***	36.123***	34.595***	33.26***	35.13***	33.48***
	(5.223)	(2.459)	(1.6)	(1.225)	(2.6)	(1.7)
R2	0.05	0.147	0.297	0.381	0.126	0.276
Adjusted R2	0.049	0.146	0.296	0.38	0.125	0.275
Observation	3388	3428	3425	3347	3401	3420

Table 8 Fama-French-Carhart Filtering of Portfolio Groups with Different Holding Periods

Summary of Fama-French-Carhart four-factor model of portfolio groups with 1 day/week/month/quarter holding periods formed from all SA research articles and sorting into quintile ratings based on the average percentage of negative words. Each portfolio group with 1 week/month/quarter holding period contains 5/21/63 portfolios. Quintile 1 is composed of stocks with a relatively low average percentage of negative words, while Quintile 5 consists of firms with a relatively high average percentage of negative words. We investigate the trading strategy of buying the 1st quintile and short selling the 5th quintile, denoted as the 1-5 portfolio. For each portfolio group, the alphas are considered at the 5% confidence level.

Portfolio Group	Number of Portfolios	Number of Significant Alphas
d-d	1	1
w-d	1	1
m-d	1	0
q-d	1	0
4d-d	1	0
20d-d	1	0
d-w	5	0
w-w	5	1
m-w	5	0
q-w	5	0
d-m	21	2
w-m	21	0
m-m	21	0
q-m	21	1
d-q	63	6
w-q	63	5
m-q	63	1
q-q	63	2

1.5.3 Further Tests

This section provides additional checks. All these tests use the Fama-French-Carhart four-factor model on portfolio d-d, relying on subsets of all single-ticker articles using different slicing methods and sorting articles into quintile ratings based on the average percentage of negative words. First, we examine the performance using Pro or Free articles. Second, we also test the performance using articles filtered by the number of comments. Third, we present results using articles contributed by authors who have joined Marketplace. Fourth, author consistency is checked using different criteria, and we test the performance of articles contributed by authors with varying levels of consistency.

1.5.3.1 Pro and Free Articles

We repeat the analysis above on portfolio d-d by separately exploring the performance by using only Pro or Free single-ticker articles. These two portfolios are denoted as "Pro Articles" and "Free Articles." After that, we also test the difference of intercepts (alphas) by deducting the return of portfolio relying on only free articles from the return of portfolio relying on only pro articles on a daily basis and use the same Fama-French-Carhart factors. The test of difference of intercepts portfolio is denoted as "Pro Minus Free Articles."

The results are shown in Table 9 columns (1) to (3). None of the alphas is significant at a 5% confidence level.

1.5.3.2 Comments

Among all the articles in our dataset, there are many which have very few or even no comments. On the other hand, many articles have attracted a wide range of commentaries. As a result, the distribution of the number of comments of each article is quite skewed. It has been shown that commentaries provide feedback on authors' articles. As a result, bad ideas can be corrected, and the informativeness of social media can be boosted in general and mediocre contributors will be discouraged (Chen et al., 2014a). In this section, we test the performance using single-ticker articles filtered by the number of comments.

At Least One Comment on Article Publication Date

We only use single-ticker articles which have at least one comment on the article publication date to make sure the articles considered at least received some feedback from the readers. The portfolio is denoted as "One Comment." The result is shown in Table 9, column (4).

Most Number of Comments in 48 Hours after Article Publication Time

We only rely on single-ticker articles with the most number of comments in 48 hours after article publication time. We sort all single-ticker articles published during our sample period into tertiles based on the number of comments written in the first 48 hours after each article publication. We use the articles in tertile 3, which have a relatively high number of comments published in the first 48 hours, to form our portfolios. This portfolio is denoted as "Comments in 48 Hours." The result is shown in Table 9, column (5).

None of the alphas is significant at a 5% confidence level.

1.5.3.3 Market Place Authors

Investors have to subscribe to each Marketplace to be able to see premium Marketplace research articles. It's natural to assume that these authors do hold some valuable insights which are not fully reflected in the market yet. In this section, we test results using single-ticker articles which are contributed by authors who have joined marketplaces. This portfolio is denoted as "MarketPlace," with the result shown in Table 9 column (6).

The alpha is not significant at a 5% confidence level.

1.5.3.4 Author Consistency

We assess analyst skill by calculating the hypothetical abnormal returns that an investor would receive by following an NPA's stock recommendation for a fixed holding period (Crane & Crotty, 2020). For every three years Y_1, Y_2, and Y_3 (consecutive years), we calculate the subsequent one-day abnormal return following each single-ticker article published by author *i* during these three years. We classify articles into "bullish" or "bearish" recommendations using the sentiment of the article (Chen et al., 2014a). If the percentage of negative words of an article is lower than the median of its overall distribution, the article is regarded as "bullish"; if the percentage of negative words is higher than the median, the article is regarded as "bearish." If a positive abnormal return follows a bullish article, or if a negative abnormal return follows a bearish article, then the article is considered as being consistent. The article-level consistency equals one if the article is consistent, otherwise zero. $Consistency_i$ is the percentage of articles published by author *i*, which are consistent (Chen et al., 2014a).

After we calculate the consistency of each author using all single-ticker articles published by each author between years Y_1 and Y_3, we sort all authors who have contributed articles in all three years into tertiles based on each author's consistency between years Y_1 and Y_3. Authors in tertile 3 will have relatively high consistency, while authors in terile 1 will have relatively low consistency.

Authors in tertile 3/2/1 are considered as best/mid/worst authors between years Y_1 and Y_3. We form portfolio d-d separately using single-ticker articles contributed by the best, mid or worst authors in year Y_4, which is the next year of Y_3. We compare the consistency of authors every three consecutive years and form portfolio d-d using single ticker articles contributed by best, mid, or worst authors in the next year. Portfolios using articles contributed by the best authors in each year will be aggregated into the best_authors portfolio group. A similar rule is applied to the mid authors to form the mid_authors portfolio group and the worst authors to form the worst_authors portfolio group. For each portfolio group, data points from each portfolio will be merged and further tested by the Fama-French-Carhart four-factor model. Results are shown in Table 10. Columns (1) and (2) show results for the best_author and worst_author portfolio groups formed using 3 years of consistency based on the sentiment of the articles. Columns (3) and (4) show results using 1 year of consistency based on the sentiment of the articles. None of the alphas is significant at a 5% confidence level.

Using all SA articles, we also use the authors' disclosed positions (Campbell et al., 2019) as "bullish" or "bearish" recommendations. Columns (5) and (6) show results for the best_author and worst_author portfolio groups formed using 3 years of consistency based on the disclosed positions. Columns (7) and (8) show results using 1 year of consistency based on the disclosed positions. None of the alphas is significant at a 5% confidence level.

Table 9 Fama-French-Carhart Filtering of Portfolio Group d-d Formed Using Subsets of All Articles Generated by Different Slicing Methods

This table shows the summary of the Fama-French-Carhart four-factor model of portfolio group d-d formed using subsets of all the single-ticker articles generated by different slicing methods and sorting them into quintile ratings based on the average percentage of negative words. Quintile 1 is composed of the stocks with a relatively low average percentage of negative words, while Quintile 5 consists of firms with a relatively high average percentage of negative words. We investigate the trading strategy of buying the 1st quintile and shorting the 5th quintile, denoted as the 1-5 portfolio. Column (1) reports the result using Pro SA research articles, denoted as "Pro Articles." Column (2) reports the result using free SA research articles, denoted as "Free Articles." On a daily basis, we calculate the return of portfolio relying on only pro articles minus the return of portfolio relying on only free articles and use the same Fama-French-Carhart filtering to test the difference of intercepts (alphas). We denote this portfolio as "Pro Minus Free Articles," and the result is shown in column (3). Column (4) shows the regression of returns using articles with at least one comment on the article publication date, which is denoted as "One Comment." Column (5) shows the regression of returns using articles with the most number of comments in 48 hours after article publication time. We sort all articles published during our sample period into tertiles based on the number of comments generated in the first 48 hours after each article publication time. We use the articles in tertile 3, which have a relatively high number of comments published in the first 48 hours. This portfolio is denoted as "Comments in 48 Hours." Column (6) shows the results using articles which are contributed by authors who have joined marketplaces, which is denoted as "Marketplace". Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level. The portfolio is rebalanced daily. The coefficient estimates are those from the Carhart four-factor regression of the portfolio returns (R_B-R_S) on the market excess return (MKTRF), the size factor (SMB), the value factor (HML), and the momentum factor (MOM). Coefficients and standard errors are multiplied by 100 for presentation purposes.

Variable	(1) Pro Articles	(2) Free Articles	(3) Pro Minus Free Articles	(4) One Comment	(5) Comments in 48 Hours	(6) Marketplace
Alpha	0.17	0.07.	0.17	0.05	0.08	-0.02
	(0.1)	(0)	(0.1)	(0.1)	(0.1)	(0.1)
MKTRF	-11.84	6.03.	-3.34	-11.42*	-5.82	7.18
	(12)	(3.6)	(12.7)	(5.7)	(8.9)	(9)
SMB	-11.28	-13.65.	7.67	-4	-6.18	-18.1
	(22.2)	(7)	(23.4)	(10.9)	(16.4)	(16.2)
HML	44.99.	-50.51***	60.28*	-34.43**	4.72	35.12*
	(27.1)	(7.4)	(28.7)	(11.3)	(17.1)	(17.8)

MOM	49.94**	36.61***	26.12	20.32**	35.48**	42.94***
	(16.7)	(5)	(17.7)	(7.4)	(11.9)	(12.4)
R2	0.011	0.061	0.005	0.019	0.005	0.009
Adjusted R2	0.008	0.059	0.001	0.018	0.003	0.007
Observation	1051	3388	1051	2781	2154	1718

Table 10 Fama-French-Carhart Filtering of Portfolio Group d-d Formed Using Subsets of All Articles Ranked by Author Historical Consistency

Columns (1) and (2) show results for the best_author and worst_author portfolio groups formed using 3 years of consistency based on the sentiment of the articles. Columns (3) and (4) show results using 1 year of consistency based on the sentiment of the articles. Columns (5) and (6) show results for the best_author and worst_author portfolio groups formed using 3 years of consistency based on the disclosed positions. Columns (7) and (8) show results using 1 year of consistency based on the disclosed positions. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level. Coefficients and standard errors are multiplied by 100 for presentation purposes.

	Percentage of Negative Words				Disclosure Position			
	3-year consistency		1-year consistency		3-year consistency		1-year consistency	
	(1) Best Authors	(2) Worst Authors	(3) Best Authors	(4) Worst Authors	(5) Best Authors	(6) Worst Authors	(7) Best Authors	(8) Worst Authors
Alpha	0.18.	-0.08	-0.04	-0.09	-0.05	0.02	0.15	0.27*
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.2)	(0.1)	(0.1)
MKTRF	34.57**	0.68	5.15	-23.42.	20.59	-3.05	12.29	-19.5
	(11.1)	(12.1)	(10.6)	(12.1)	(16.3)	(28.8)	(11.4)	(12.4)
SMB	-52.69*	-29.45	-6.1	-54.01*	-54.72.	30.05	-13.81	28.93
	(20.6)	(22.3)	(20.1)	(22.4)	(28.5)	(49.5)	(20.8)	(22.2)
HML	-2.54	-38.74	-3.26	35.11	-15.36	26.14	-13.08	-2.2
	(21.9)	(23.8)	(20.4)	(22.9)	(28.2)	(54)	(21.4)	(24.2)
MOM	13.85	42.96**	21.16	69.75***	7.98	52.58	28.19.	64.19***
	(15)	(16.4)	(13.5)	(15)	(20.4)	(37.4)	(15)	(16.4)
R2	0.01	0.013	0.002	0.017	0.006	0.003	0.004	0.017
Adjusted R2	0.007	0.01	0	0.016	0.001	0.002	0.002	0.015
Observation	1390	1348	1603	2288	856	833	1529	1456

1.6 Conclusion

We examine whether social media stock recommendations influence the direction of investor trading. We study the relation between article sentiment on the social media outlet Seeking Alpha and order imbalances. We find that article sentiment contribute directly to short-window order imbalances after the article's publication.

We further examine if investors can capture the benefits of stock recommendations contributed by nonprofessional analysts on Seeking Alpha, if there are any. We build portfolios by buying the stocks with the most positive sentiment and short-selling the ones with the worst sentiment, with various portfolio rebalancing windows. We discover that the portfolio that considers articles on a daily basis together with a one-day holding period yields a positive abnormal gross return of 10 basis points per trading day. This positive abnormal return is statistically significant but is not economically significant after factoring in the bid-ask spread. We apply the calendar time strategy with non-overlapping return periods to get unbiased test statistics, which is a more robust method than panel regression with clustered standard errors. In order to better capture the economic significance of our findings, we use value-weighted abnormal returns instead of equal-weighted abnormal returns. In order to determine the value of the stock recommendations, we perform several additional tests. None of these portfolios, however, produces significant abnormal returns.

We contribute to the literature in two ways: 1) we have proposed a new approach to building the portfolio to assess the value offered by equity review platforms like SA; 2) we question whether the wisdom of crowds exists in a financial context.

The results of order imbalances shows that investors react to the stock recommendations on Seeking Alpha. However, the results of abnormal returns suggest that they are not earning money.

CHAPTER 2 - The Effect of Financial Incentive on the Quality of Stock Recommendations on Social Media

Paper (II) The Effect of Financial Incentive on the Quality of Stock Recommendations on Social Media

2.1 Introduction

Professional financial analysts (hereafter PAs) play an essential role in capital markets by supplying new information and interpreting existing information (Brown et al., 2015, 2016; Bradshaw et al., 2017). In addition, individual investors rely more and more on each other as peer-to-peer information sources, resulting in the proliferation of nonprofessional analysts (hereafter, NPAs) on social media (Campbell et al., 2019; Chen et al., 2014a). PAs and NPAs are different in their motivations, incentives, audiences and expertise. PAs make a living by covering specific firms and industries. They acquire technical and financial skills and are monitored and rewarded by their employers and their customers. PAs are incentivized by bonuses, which are linked to their performance (Groysberg et al., 2011; Stickel, 1992). All these factors give PAs motivation to devote time, effort, and resources to generate information of good quality.

In comparison, NPAs are likely to be paid less for their work. NPAs are not employed by social media platforms and act independently. NPAs, unlike their professional counterparts, are subject to little regulation, which could lead to market manipulation (Campbell et al., 2019). Furthermore, they cater to a very different audience, which is most likely made up of retail investors looking for low-cost and easy-to-access information. They are less likely to have the same degree of technical competence as PAs, and they might devote less time, resources, and commitment to their analyses than PAs. As a result, their analyses and comments are expected to be less accurate.

This study uses data from Seeking Alpha (hereafter SA), a crowd-sourced online equity research platform for investors. Social media platforms use a variety of market mechanisms, from free to subscription-based business models. In April 2015, SA launched Seeking Alpha Investor Marketplace (hereafter, Marketplace). Marketplace is an investment services platform provided by SA's individual authors and author teams. The Marketplace is an author-level subscription plan. The authors offer ideas for investments, research, and guidance on several different styles. Marketplace authors are paid from $25 to $375 per month per subscriber for their premium articles in Marketplace. Meanwhile, SA contributors can continue to publish free articles (regular articles). NPAs are required to disclose their stock positions in stocks discussed in the research articles, possibly to signal SA's commitment to fairness and ethics and avoid any conflict of interest (Campbell, DeAngelis, & Moon, 2019).

We examine two specific research questions. First, we examine what drives NPAs' decision to start offering financial advice on Marketplace? Second, how do financial incentives provided by social media platform owners affect the behavior of NPAs on social media? In particular, we examine the effect of financial incentives on the quality of stock recommendations from NPAs.

The sample period is from January 2005 to December 2018. We examine the factors that affect the NPAs' decision to offer their services on Marketplace. The decision is mainly determined by the number of years the NPA has been active on SA (*NumYearsOnSA*) and the total number of articles contributed by the NPA per year (*NumArticles*). In terms of economic significance, we find that a one-year increase in *NumYearsOnSA* is associated with a 13.5% increase in the probability that an author will join Marketplace. One article increase in *NumArticles* is associated with a 1.6% increase in the probability that an author will join Marketplace. These findings suggest that an NPA is more likely to join Marketplace if he or she has contributed to SA for a longer period of time and has published more articles each year.

We use a research design with NPA fixed effects to control for time-invariant variation in skill levels of NPAs. The financial market context allows us to construct an objective measure of quality, i.e., abnormal returns of disclosed stock holding positions in SA articles. To answer the second research question, we compare the abnormal returns corresponding to NPAs' disclosed stock positions in free articles and Marketplace articles. We find that these disclosures contribute to three-month abnormal returns after the article's publication. Using three-month abnormal returns after the article's publication, we find that financial incentives have negative effects on the quality of free articles. When NPAs have joined Marketplace, the quality of their recommendations in Marketplace articles is better than the quality of their recommendations in free articles. We also conduct additional analyses using different holding periods. Our results still mostly hold with a one-month holding period, but lose some significance with a one-week or one-day holding period. In addition, our results are robust to the use of different asset pricing models for calculating abnormal returns.

We build on work of Chen et al. (2019), who explore the effect of providing financial incentives to NPAs on SA, as well as the implications for online investor communities. They show that financial payments from the platform[10] increase the number of articles but do not affect the quality of the articles. Our paper extends Chen et al. (2019) by showing that financial incentives provided by social media platform owners affect the quality of stock recommendations in SA articles. We show that financial incentives lower the quality of non-exclusive (free) stock recommendations. NPAs react to financial incentives and can tell the difference of quality of their recommendations. When NPAs have joined Marketplace, the quality of their long (short) position recommendations in Marketplace articles is worse (better) than the quality of their

[10] SA launched a premium partnership program in January 2011 that allows contributors to earn $10 per 1,000 page views for their "premium" articles, which are published only on SA and are not freely available elsewhere on the Internet. It is a website-level subscription, which is included in the "pro" subscription. We call these article "pro" articles and we don't include these articles in this study.

long (short) position recommendations in free articles prior to joining Marketplace. Chen et al. (2019) argue that the financial compensation offered in their context back to 2011 is relatively small. In comparison, the monetary compensation offered in our context can be high. One example is that Rida Morwa's service became the first service on SA to reach $1M in annual revenue (*Seeking Alpha's First Millionaire*, 2019). Our findings have implications for social media owners on attracting more valuable stock recommendations to be generated and disseminated by NPAs. Social media outlets show growing popularity among investors. The findings of this study demonstrate the usefulness and value relevance of NPAs' stock recommendation in the investment domain.

We contribute to the literature on the role of social media in financial markets, particularly for equity research platforms such as SA. These studies show that NPAs contribute value-relevant information to the market in that their articles and SA comments predict future stock returns (Chen et al., 2014a), their articles lead to more informed retail trading (Farrell et al., 2020), improve liquidity during earnings announcements (Gomez et al., 2020) and their disclosed positions magnify investor responses to article tone (Campbell et al., 2019). Most of the prior studies have focused on the role of social media in financial markets without considering financial incentives. We contribute to this literature by showing that financial incentives offered by SA have an effect on the quality of stock recommendations in SA articles.

This study also contributes to the literature on the role of sell-side analysts in financial markets. Jame et al. (2016) show that crowdsourced earnings forecasts are incrementally useful to investors beyond those of sell-side analysts. Jame et al. (2017) show that crowdsourced earnings forecasts can discipline sell-side analysts and result in less biased forecasts. Each SA Marketplace NPA makes recommendations to SA's clients (Marketplace subscribers) and is paid based on the number of Marketplace subscribers, making Marketplace NPAs comparable to sell-side analysts. This research contributes to the literature by being the first study to examine the article quality in SA Marketplace and show that the quality of stock recommendations in SA articles is influenced by the financial incentives supplied by SA. This evidence is important because the amount of information available to investors via social media is likely to expand over time, while budgets and headcounts of sell-side equity research departments are likely to decline (Morris, 2017). The number of sell-side analysts has been steadily declining while the number of NPAs has been steadily increasing (M. S. Drake et al., 2020).

The remainder of this study is organized as follows. Section 2 conducts a literature review. Section 3 discusses data sources. Section 4 and 5 discuss two research questions and associated results. Section 6 concludes.

2.2 Related Literature

In this section, we first discuss the literature on "wisdom of the crowds." The present study is closely related to the literature on "the effect of free financial advice on social media on stock markets," "the effect of financial incentive on the quality of user-generated content (UGC)," and "the role of sell-side analysts in capital markets."

2.2.1 Wisdom of Crowds

The concept of the Wisdom of Crowds can be traced back over a century ago and pertains to the phenomenon that the aggregation of information provided by many individuals usually yields better predictions than any single member of the group or even experts. Surowiecki (2004) documents that in the early 20th century, Sir Francis Galton discovered that the crowd in county fairs could accurately predict the weight of an ox when they averaged their individual guesses. The average (or median) prediction of the crowd was closer to the true weight of the ox than any estimate by cattle experts. Berg, Forsythe, Nelson, & Rietz (2008) conducted a study on the ability of the Iowa Electronic Markets in forecasting election results and showed that the markets' predictions are not biased and have a significant ability to forecast elections, surpassing expert opinion polls. Many other important decisions are made in a group setting. Consider trials by juries, interest rates by the Federal Open Market Committee, and the choice of a CEO by a board of directors. Moldoveanu & Martin (2010) conclude: “A collection of heterogeneous problem solvers will always beat out a single expert problem solver.”

2.2.2 The Effect of Free Financial Advice on Social Media on Stock Markets

Social media has changed the way investors discuss stocks in financial markets, and investors are increasingly relying on stock recommendations offered on social media platforms. Most of the previous studies have focused on free financial advice on different social media platforms. Although multiple papers have demonstrated the value of free financial advice, few studies have focused on the quality of fee-based financial recommendations on social media platforms. We first discuss the research related to the value of free financial advice on varied kinds of social media platforms.

Message Board

Access to message boards is generally free, perhaps because message boards have been shown to generate mostly noise. Tumarkin & Whitelaw (2001) find no link between message board activities on Raging Bull and stock returns or abnormal trading volume. Antweiler and Frank (2004) find a statistically significant yet economically insignificant relationship between internet message board activity and stock returns. Das and Chen (2007) claim there is no strong association between sentiment and stock prices.

Twitter

Tweets are in short format and convenient to search using hashtags or cashtags[11]. Twitter is a free medium for sharing views and information in a timely manner, while the longer format of

[11] For instance, $AAPL or $GOOG.

articles can possibly reduce the timeliness. The first stream of literature on Twitter investigates whether Twitter can predict the overall stock market. Bollen, Mao, & Zeng (2011) find that aggregate sentiment inferred from textual analysis of daily Twitter feeds can help predict changes in the Dow Jones index. Mao, Wei, Wang, & Liu (2012) argue that the number of daily tweets mentioning S&P 500 stocks is significantly related to the levels, changes, and absolute changes of the S&P 500 index. The second stream examines whether tweets about specific companies can help investors predict the firms' earnings and sales. Bartov, Faurel, & Mohanram (2018) show that the aggregate view from individual tweets successfully predicts quarterly earnings. Tang (2018) demonstrates that user-generated content related to products and brands on Twitter can be used to forecast company-level sales and points out that the predictive power relies on the "wisdom of the crowd." The third stream studies examine how Twitter activity influences investor response to earnings. Curtis, Richardson, & Schmardebeck (2014) focus on Twitter and StockTwits activity over 30-day rolling windows, and find that high levels of investor attention are linked to higher sensitivity of earnings announcement returns to earnings surprises, with the effect being strongest for companies that outperform analysts' forecasts.

Estimize.com

Estimize.com is an open web-based platform founded in 2011. Anyone can register to become a contributor and have free access to their data and users can make earnings forecasts on this platform. Jame, Johnston, Markov, & Wolfe (2016) demonstrate that the Estimize.com crowd-sourced consensus earnings forecasts provide value-relevant information to the capital market in predicting earnings. Studies by Adebambo & Bliss (2015) and Jame et al. (2016) show that crowd forecasts on Estimize.com are more accurate than traditional earnings consensus. By collaborating with Estimize.com to conduct experiments to limit the information available to randomly selected stocks and users, Da & Huang (2020) document that viewing more public information before making forecasts improves individual forecast accuracy but decrease group consensus forecast accuracy because useful private information is prevented from entering the consensus.

Seeking Alpha

SA is one of the largest social media platforms and has become a popular site for NPAs to share their analyses of stocks. As of January 2021, SA had 10 million registered users and over 17 million unique visitors every month. SA's average visit duration is 4 times more than The Economist, Barron's, or the Wall Street Journal (*Seeking Alpha Media Kit 01/06/2021*, 2021). SA articles are generally reviewed by a panel of editors and are subject to editorial changes. SA pays contributors based on the number of users accessing their articles.

Chen, De, Hu, & Hwang (2014) show that the information in NPA contributed research articles and user-generated comments on SA helps predict earnings and long-term stock abnormal returns. Campbell, DeAngelis, & Moon (2019) document that stock positions of SA contributors are directly related to short-window returns surrounding the article's publication, convey information to investors, and enhance investors' perception of the credibility of SA authors. Drake, Thornock, & Twedt (2017) suggest that coverage by internet intermediaries, including SA, can reduce the level of information asymmetry in stock markets. Wang et al. (2015) investigate the effect of information published on SA and StockTwits and demonstrate that although the correlation between information and stock returns is generally very moderate, the correlation is stronger for authors who have predicted returns in the past.

In addition to having a broad readership, SA is also different from other social media platforms because the articles bring in substantial, edited analysis in long-form that may also contain a formal disclosed position. Platforms such as Estimize.com provide earnings estimates without any analysis. Users of message boards or Twitter are able to post any information without proper quality control. In comparison, SA articles provide in-depth analyses that are reviewed by an editorial board for quality assurance.

Our research builds on previous research on the role of social media in financial markets and examines how financial incentives provided by online community owners may influence NPAs' behavior on SA. In particular, we examine whether financial incentives improve the quality of NPAs' stock recommendations in SA articles.

2.2.3 The Effect of Financial Incentive on the Quality of User-Generated Content (UGC)

Our research contributes to the increasing literature about how financial incentive influences the quality of UGC. Previous research on this topic has primarily focused on online reviews and has produced mixed findings. Multiple studies demonstrate that financial incentives induce a larger volume of reviews (Burtch et al., 2017; Khern-am-nuai et al., 2018). Wang et al. (2012) show no significant differences in quality between paid and unpaid reviews. Khern-am-nuai et al. (2018) demonstrate that financial incentives have a negative effect on the quality of customer reviews. Liu & Feng (2021) find that the influence of financial incentives on contributor participation and total content volume is not monotonic. Qiao et al. (2021) discover that small financial incentives, when used in conjunction with the appropriate intervention strategies, can motivate users to increase contribution volume while maintaining good quality.

Although financial incentives may result in a quantitative increase in engagement volume, they do not guarantee quality improvement and may even undermine contributor performance. Multiple studies reveal that after the initiation of financial incentives, reviewers tend to shorten reviews and express biased opinions in their postings (Burtch et al., 2017; Cabral & Li, 2015; Khern-am-nuai et al., 2018; Qiao et al., 2020; Zhao et al., 2016).

Prior research has shown how financial incentive motivates people in many contexts, and Chen et al. (2019) are among the first to look at the effect of financial incentives on social media content output for financial markets. SA established a premium partnership program in January 2011 that allows NPAs to earn $10 per 1,000 page views for their “premium” articles, which are published only on SA and are not freely available elsewhere on the Internet. Using SA free and “premium” articles, Chen et al. (2019) discover that financial incentives are beneficial in increasing the volume of content output and generating greater community interest, although it does not lead to better or worse stock recommendations.

2.2.4 The Role of Sell-side Analysts in Capital Markets

There are two types of PAs in general: buy-side analysts and sell-side analysts (hereafter, SSA). SA articles are relatively long and similar in format to those written by professional SSAs (Chen et al., 2014a).

SSAs have played an important role in capital markets for decades. Research of SSAs aids in establishing the market’s expectations for earnings and stock price, supports trading recommendations, and gives investors important information on crucial stock investing debates. Their predictions and views are prominently reported in the business press and news media (Rees et al., 2015). Many studies examine the effect of their activities on markets, and show that their reports move markets (Beyer et al., 2010; Frankel et al., 2006; Gleason & Lee, 2003; Li et al., 2015). SSAs face several incentives, including pleasing management, generating trading commissions, and increasing investment banking transactions, which affect the objectivity of their recommendations and forecasts negatively (Jackson, 2005; Lin & McNichols, 1998; Mayew, 2008). Despite these conflicting interests, SSAs have been widely considered as the principal source of stock investment research for investors for years.

Regulatory changes and changes in the market’s supply and demand for information have shifted the landscape of sell-side stock research over the past few years (M. Drake et al., 2020). Regulations limited equity research departments’ ability to support and facilitate investment banking transactions for their brokerages, and as a result, many of the best analysts left the industry or moved to the buy-side (Guan et al., 2019). SSAs are now more concentrated on their efforts to generate revenue from their research via trading commissions (Groysberg & Healy, 2020; Kadan et al., 2008). For institutional clients with high commissions, SSAs devote more time to meeting their needs by providing them with more specialized and personalized services (Brown et al., 2015; Green et al., 2014). Budgets and headcounts of equity research departments have been continuously declining in recent years as a result of these changes (M. S. Drake et al., 2020; Groysberg & Healy, 2020).

Regulatory changes that affect the demand for and supply of sell-side research have contributed to the budgets cut for equity research. Meanwhile, alternative forms of equity research have emerged on social media, such as that contributed by NPAs. In comparison to the decreasing number of SSAs, the number of NPAs contributing to SA has increased in recent years (M. S. Drake et al., 2020). These NPAs provide value-relevant information at least on average, which we have shown in the section "The Effect of Free Financial Advice on Social Media on Stock Markets." A more detailed literature review on PAs can be found in Appendix F.

Each SA Marketplace NPA makes recommendations to SA's clients (Marketplace subscribers) and gets paid based on the number of Marketplace subscribers, which makes Marketplace NPAs comparable to SSAs. This research contributes to the literature on the role of paid analysts in financial markets by being the first study to examine the article quality in SA Marketplace.

2.3 Data

This section shows the data construction and our main variables. We use articles and associated comments from SA, as well as abnormal returns from WRDS (Wharton Research Data Service). The sample period is from January 2005 to December 2018.

2.3.1 The Seeking Alpha Data

Founded in 2004, SA is one of the largest investment-related social media websites in the US (Chen et al., 2014a; Farrell et al., 2018). As of 2020, SA has accumulated more than 15,000 contributors. Content on SA is offered via multiple subscription plans: Basic, Pro, and Marketplace. The Basic subscription gives the users access to all free articles but not any "pro" articles, which can be accessed via Pro subscription. Both Basic and Pro subscription plans are site-level subscription plans. Further, individual authors or a team of authors can offer investing services via Marketplace subscription. The Marketplace subscription plan is a contributor-level subscription plan.

To access articles from several services in Marketplace, the user will have to subscribe to each of these services separately. In this study, we obtain free articles (available at https://seekingalpha.com/articles) from SA. We obtain 747,832 free articles published between January 2005 and December 2018. By subscribing to all available services in Marketplace, we obtain Marketplace articles (available at https://seekingalpha.com/marketplace). A total of 41,001 Marketplace articles are obtained between April 2015 and December 2018.

Every SA article, whether it is free or belongs to Marketplace, has two categories of ticker groups: "About" and "Include." "About" are only identified when a firm is the focus of the article, and "Include" demonstrates the tickers that the article mentions. Table 1 and 2 describe the dataset used in this study in detail.

To prepare the data for analysis, we follow the literature (Campbell et al., 2019). However, we improve on their approach, which is somewhat inadequate when articles' disclosures are unstructured. Next, we explain how we modify their data pre-processing approach.

2.3.2 Data Pre-processing

Long and Short Disclosed Stock Positions

In general, disclosed stock positions follow the same basic format. The author incorporates a disclosure statement at the beginning or end of each article. For instance, an author may include "I am/we are long XXXX," "I/we have no positions in any stocks mentioned, and no plans to initiate any positions within the next 72 hours," or "I am/we are short XXXX." However, the positions are sometimes less clear because the author may disclose complex option holdings or multiple positions in different stocks (i.e., long XXXX and short YYYY). As a result, I use the following procedure to parse these disclosures.

The first step follows the procedure proposed by Campbell et al. (2019). First, the long positions are identified by searching for the terms "long," "hold," or "own stock/shares." Then, we capture the text after these words, stopping when we encounter a period or the word "may" or "short," which indicates the beginning of a new disclosed position (i.e., "I am long XXXX and may …"). We repeat a similar process for possible short positions, looking for the word "short," and then capture tickers until the word "long" or "may," or a period. We do not allow negating or qualifying words (no, not, none, neither, never, nobody, may, or plan) to appear within the five words before the position indicator for long or short positions. Finally, we look for clues that the author does not hold any position in any stocks. These include the terms "No position," "None," or "May."

Although inspection of results shows that the first step procedures are relatively accurate, we do come across articles with multiple classifications (i.e., long, short, and/or no position) because of complex information disclosure, or situations where none of these three positions can be identified. In addition, the position can be related to stock tickers other than the stock tickers mentioned in "About" or "Include." As a result, I apply the following procedure to refine the disclosure coding. First, I require a ticker in the disclosed positions to appear in "About" or "Include" to be considered for a long or short position. If the ticker does not show up in "About" or "Include," I code the position of the ticker as "no position." Second, if I identify multiple classifications or fail to recognize long, short, or no position, I code the position as "no position."

After parsing the disclosed positions, there are 124,265 free articles that include long or short disclosed positions, which correspond to 226,468 disclosed positions. There are 11,195 Marketplace articles that include long or short disclosed positions, corresponding to 22,444 disclosed positions.

Any interested user can comment and share their views, and may agree or disagree with the author's opinions on the stocks being discussed. I download all comments written in response to all the collected articles. Chen et al. (2014) show that sixty percent of the comments are posted on the day of article publication, another twenty percent are posted on the next day, and the last twenty percent are posted occasionally in the following weeks. Thus, I focus on comments which are posted within the first two days (48 hours) after each article publication time. I collect 12,936,355 comments in response to free articles and 78,214 comments in response to Marketplace articles.

Merging Based on Ticker

A stock ticker is a unique abbreviation assigned to publicly traded shares of a particular stock on a specific stock market. The ticker symbol allocation and formatting convention are specific to each stock exchange. For instance, in the United States, stock tickers are usually between one and four letters and represent the company name when possible. After a merger is completed, the ticker of an acquired company is usually changed to the ticker of the acquirer. A company that changes its name may change its ticker. If a firm ceases to exist, its ticker may be reassigned to another firm, and therefore we use PERMNO to uniquely identify each stock.

PERMNO refers to the unique permanent identification number assigned to each stock by CRSP (Center for Research in Security Prices). Unlike ticker or company name, PERMNO will not be changed during the stock's trading history or be re-assigned after the stock ceases trading. Users can use a PERMNO to track a stock's entire trading history in the CRSP files without having to consider changes in name. PERMNO is a five-digit integer for stocks in the CRSP file. Different share classes in the same company have different PERMNOs. Therefore, PERMNO is stock-level, not company-level.

Abnormal Returns

CRSP STOCKNAMES file provides a mapping between PERMNOs and all historical company names and exchange tickers, along with their effective date ranges. In this research, I use STOCKNAMES to map stock tickers to PERMNOs. After uploading PERMNO and event date combinations as our trading events to US Daily Event Study on WRDS, using different pricing models, we obtain abnormal returns. Appendix D shows how abnormal returns are calculated.

Table 1 Sample Attrition

This table describes the SA free articles dataset and sources of data loss.

SA Free and Pro Articles Contributed Between 2004 and 2018	771,127
Pro Articles	(23,295)
Free Articles	747,832
Articles that are transcripts or contain less than or equal to 100 characters	(34,666)
Free Articles Which Are Not Transcripts and Contain More Than 100 Characters	713,166
Articles that have no-position as disclosed positions or no disclosed positions	(588,901)
Articles with "Long Position" or "Short Position" as Disclosed Positions	124,265
Number of Disclosed "Long Position" and "Short Position"	226,468

Table 2 Sample Attrition

This table describes the SA Marketplace articles dataset and sources of data loss.

SA Marketplace Articles Contributed Between 2015 and 2018	41,001
Articles that are transcripts or contain less than or equal to 100 characters	(390)
Articles Which Are Not Transcripts and Contain More Than 100 Characters	40,611
Articles that have no-position as disclosed positions or no disclosed positions	(29,416)
Articles with “Long Position” or “Short Position” as Disclosed Positions	11,195
Number of Disclosed “Long Position” and “Short Position”	22,444

2.4 What Drives the Authors’ Decision to Charge for Their Financial Advice?

In this section, we examine the factors that motivate authors to join Marketplace. Using a probit regression model, Chen et al. (2019) find contributors who have longer tenure on SA, fewer followers, more comments per article, and lower standard deviation of article sentiment are more likely to join the premium program, which provides monetary payments. Based on Chen et al. (2019), we use variables to control the tenure on SA, the number of comments per article, and article sentiment. We also add the variables which stand for the number of articles contributed as well as the measures of the average quality of past articles. The number of followers is not included in the model because we do not have this data before each NPA joined Marketplace.

2.4.1. The Cox Proportional Hazards Model

We use a Cox proportional hazard model, which is proposed by Cox (1972). Cox proportional hazard model is a method of time-to-event analysis while probit regression model does not include time variable.

The hazard function is given by:

$$h(t|Z) = h_0(t)\exp(\beta' Z) \qquad (1)$$

$h_0(t) =$ the baseline hazard function

$Z =$ a vector of covariates

β = a vector of regression coefficients

In the rest of this paper, the model with hazard function given by (1) will be called the Cox model. *t* corresponds to the time to event variable, which is the calendar year (e.g., 2017 and 2018). The censoring indicator variable is *MarketPlaceStarted*, which switches from 0 to 1 after an author joins Marketplace.

The vector Z corresponds to the following covariates:

$AvgNumComments_{i,t}$ = the average number of comments in 48 hours after the article publication time in response to the articles published by author *i* in year *t*. This variable thus quantifies the potential readership a contributor can receive.

$NumYearsOnSA_i$ = the number of years the author *i* has been active on SA. The longer the tenure, the bigger this variable.

$NumArticles_{i,t}$ = the total number of articles contributed by author *i* in year *t*. This variable shows how active a contributor is on SA.

$AvgPerNegWords_{i,t}$ = the average percentage of negative words of articles contributed by author *i* in year *t*. This variable reflects on average how negative a contributor's articles are.

$AvgPerPosWords_{i,t}$ = the average percentage of positive words of articles contributed by author *i* in year *t*. This variable reflects on average how positive a contributor's articles are.

$AvgPerNegWords_{i,t}$ and $AvgPerPosWords_{i,t}$ are related to an article's sentiment that captures both the writing style of a contributor.

$AvgLogWordCount_{i,t}$ = the average natural logarithm of the word count of articles contributed by author *i* in year *t*. In addition to the simple number of articles, this variable is another measure for the effort of a contributor in sharing their views and opinions in stock recommendations.

$AvgPerNumbers_{i,t}$ = the average percentage of numbers in articles contributed by author *i* in year *t*. This variable captures the degree of specificity of the analysis.

CAR stands for cumulative abnormal return. To calculate the average CAR, CARs corresponding to long positions are multiplied with 1 and CARs corresponding to short positions are multiplied with -1.

$AvgCAR1D_{i,t}$ = the average CAR with 1 day holding period corresponding to disclosed stock positions in articles contributed by author *i* in year *t*.

$AvgCAR1W_{i,t}$ = the average CAR with 1 week holding period corresponding to disclosed stock positions in articles contributed by author *i* in year *t*.

$AvgCAR1M_{i,t}$ = the average CAR with 1 month holding period corresponding to disclosed stock positions in articles contributed by author *i* in year *t*.

$AvgCAR1Q_{i,t}$ = the average CAR with 1 quarter holding period corresponding to disclosed stock positions in articles contributed by author *i* in year *t*.

$AvgCAR1D_{i,t}, AvgCAR1W_{i,t}, AvgCAR1M_{i,t}$ *and* $AvgCAR1Q_{i,t}$ are the measures of the average quality of past articles.

2.4.2. Descriptive Statistics

Table 3 shows descriptive statistics of the main variables in equation (1). Table 3 Panel A presents the descriptive statistics of the main variables for 6,589 author-year observations for NPAs that do not go into Marketplace. Table 3 Panel B presents the descriptive statistics of the main variables for 297 author-year observations for NPAs that go into Marketplace. NPAs who go into Marketplace on average contribute more articles each year and have longer tenure on SA than NPAs who do not go into Marketplace.

Table 3 Descriptive Statistics of Variables

Table 3 Panel A reports descriptive statistics for NPAs that do not go into Marketplace. Table 3 Panel B reports descriptive statistics for NPAs that go into Marketplace. Each observation represents a unique author-year combination.

Panel A: Descriptive Statistics of Variables for NPAs That Do Not Go Into Marketplace

	N	Mean	Std. Dev.	25%	50%	75%
AvgNumComments	6,589	25.143	38.591	6.000	13.667	28.182
NumYearOnSA	6,589	3.637	2.553	2.000	3.000	5.000
NumArticles	6,589	7.343	19.199	1.000	2.000	5.000
AvgPerNegWords	6,589	1.609	0.758	1.104	1.491	1.965
AvgPerPosWords	6,589	1.494	0.574	1.113	1.438	1.816
AvgLogWordCount	6,589	7.168	0.516	6.829	7.133	7.470
AvgPerNumbers	6,589	0.045	0.042	0.025	0.036	0.053
AvgCAR1D	6,589	0.001	0.027	-0.005	0.000	0.007
AvgCAR1W	6,589	0.002	0.069	-0.013	0.001	0.016
AvgCAR1M	6,589	0.004	0.123	-0.031	0.001	0.036
AvgCAR1Q	6,589	0.006	0.245	-0.063	0.002	0.077

Panel B: Descriptive Statistics of Variables for NPAs That Go Into Marketplace

	N	Mean	Std. Dev.	25%	50%	75%
AvgNumComments	297	29.293	32.859	0.500	10.333	19.550
NumYearOnSA	297	4.729	2.507	0.000	3.000	4.000
NumArticles	297	34.165	48.913	1.000	5.000	16.000
AvgPerNegWords	297	1.510	0.536	0.168	1.139	1.479
AvgPerPosWords	297	1.473	0.495	0.151	1.182	1.440
AvgLogWordCount	297	7.131	0.427	4.980	6.860	7.151
AvgPerNumbers	297	0.044	0.048	0.003	0.027	0.036
AvgCAR1D	297	0.000	0.014	-0.077	-0.002	0.000
AvgCAR1W	297	-0.001	0.025	-0.165	-0.007	0.000
AvgCAR1M	297	0.001	0.053	-0.232	-0.017	0.002
AvgCAR1Q	297	0.000	0.113	-0.633	-0.045	-0.001

2.4.3. Empirical Results

We report the results of the Cox model in Table 4. For each variable in the model, the table gives the estimate of the coefficient, the standard error, the p-value, and the hazard ratio. In the Cox model, a positive coefficient means that an increase in the variable is associated with an increase in the probability that the author will join Marketplace. We find the coefficient of *NumYearsOnSA* is significantly positive (0.127), and the coefficient of *NumArticles* is significantly positive (0.015). The larger *NumYearsOnSA*, the longer the NPA has been active on SA. Each 1-year increase in *NumYearsOnSA* is associated with a 13.5% increase in the probability that the author will join Marketplace. Each 1-article increase in *NumArticles* is associated with a 1.6% increase in the probability that the author will join Marketplace. These results indicate that if an NPA has contributed to SA for a longer time and has written more articles each year, he or she is more likely to join Marketplace.

The remaining explanatory variables are not significant at the 5 percent level. The number of comments, article sentiment, the effort of a contributor, degree of specificity of analysis, and quality of past articles do not affect the probability that the contributor will join Marketplace.

Table 4 The Cox Proportional Hazards Model

Table 4 shows 6,886 author-year observations between January 2005 and December 2018. The dependent variable is an NPA's decision whether or not to join the Marketplace.

Parameter	Parameter Estimate	Standard Error	P-Value	Hazard Ratio
AvgNumComments	0.002	0.002	0.432	1.002
NumYearOnSA	0.127	0.030	<.0001	1.135
NumArticles	0.015	0.001	<.0001	1.016
AvgPerNegWords	-0.092	0.149	0.539	0.912
AvgPerPosWords	-0.203	0.194	0.294	0.816
AvgLogWordCount	0.091	0.197	0.645	1.095
AvgPerNumbers	0.541	2.073	0.794	1.717
AvgCAR1D	1.975	4.631	0.670	7.209
AvgCAR1W	0.079	2.719	0.977	1.082

AvgCAR1M	-0.457	1.425	0.749	0.633
AvgCAR1Q	0.004	0.539	0.993	1.004
Number of Observations	6,886			
Pseudo R2	0.016			

The evidence shows that NPAs who go into Marketplace contribute more articles per year and have a longer tenure on SA than those who do not. However, these two types of NPAs are not different in terms of abnormal returns. The Marketplace NPAs are more active, but they are not particularly more skilful than other NPAs. This finding helps to remove the endogeneity concerns that NPAs that tend to publish on Marketplace might be better NPAs. In this section, we have a deeper understanding of which NPAs are more likely to join Marketplace and receive the financial payments. In the next section, we examine if financial incentives affect the quality of stock recommendations of NPAs on SA.

2.5 Do financial incentives affect the quality of stock recommendations?

In this section, we test whether financial incentives provided by social media platform owners affect the quality of stock recommendations. We use abnormal returns as the proxy for quality of stock recommendations.

2.5.1 Research Design

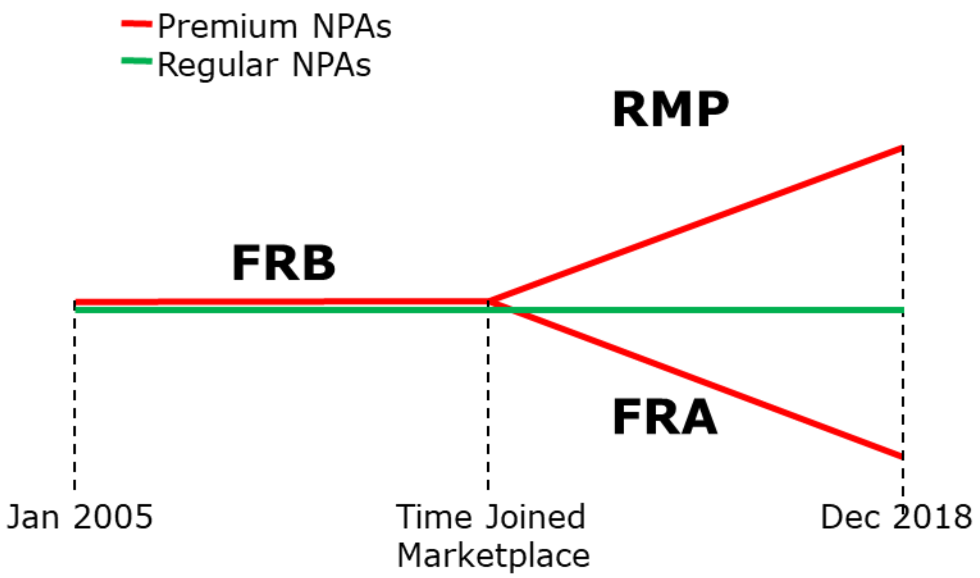
Before the start of Marketplace in 2015, SA had free and pro articles, among which we only consider the free articles[12]. In 2015, SA started Marketplace[13] to allow NPAs to charge for their premium advice in the Marketplace. Note that NPAs could decide to join the Marketplace whenever they want; for example, some may have joined Marketplace in 2015, while others may have joined in 2016, or not at all. Irrespective of their decision to join the Marketplace, NPAs can continue offering free stock recommendations. Indeed, we observe that some NPAs continued offering free recommendations after they joined the Marketplace. We consider all free and Marketplace articles after 2015. Using a unique research design, we are able to capture the differences in the quality of NPAs' stock recommendations belonging to different groups, as shown below in Figure 1. This research design allows us to compare the quality of NPAs' stock recommendations before and after they join the Marketplace. We call NPAs who only contribute free recommendations since they joined SA as "regular NPAs." We call NPAs who joined Marketplace "premium NPAs." Premium NPAs contribute free recommendations before they joined Marketplace and Marketplace recommendations after they joined Marketplace. They can also continue contributing free recommendations after joining Marketplace.

[12] Pro articles are fee based and need website-level subscription. We wanted to compare free articles (not fee based) and Marketplace articles (fee based and need author-level subscription).

[13] https://seekingalpha.com/marketplace

Figure 1 Research Design

The sample period is from January 2005 to December 2018. Regular NPAs only contribute free stock recommendations since joining SA. Premium NPAs are the NPAs who contribute Marketplace articles. "Time Joined Marketplace" is the time each NPA joined Marketplace and is different for each NPA. FRB is free recommendations before a premium NPA joined the Marketplace. FRA is free recommendations after a premium NPA joined the Marketplace. RMP is recommendations in Marketplace.



We examine how financial incentives provided by social media platform owners affect the behavior of NPAs on SA. We test if premium NPAs can tell the difference in the quality of their advice by comparing FRA and RMP while controlling for the average skill level (i.e., the green line). We also test whether financial incentives affect the quality of free stock recommendations from premium NPAs by comparing FRA and FRB.

Author-specific properties might affect the quality of stock recommendations besides the financial rewards. For instance, the inherent skills and expertise of NPAs differ from one to another. While controlling for the average skill level of analysts, we use a research design controlling for NPA fixed effects, which absorbs the time-invariant variation in skill levels of NPAs. We also control for year-month fixed effects.

2.5.2 Empirical Models

We organize our analysis around the following regression specification:

$$AbRet_{i,t+1,t+m} = \beta_0 + \beta_1 MarketPlaceStatus_{i,t} + \beta_2 MarketPlaceArticle_i + \beta_3 NumOfComments_{i,t} + \beta_4 Volatility_{i,t} + \beta_5 Size_{i,t} + \beta_6 BTM_{i,t} + Author_i + \varepsilon_{i,t} \qquad (2)$$

The unit of observation in this analysis is a stock-day combination, which indicates an NPA is long or short a stock on a day that advice is given on SA.

$AbRet_{i,t+1,t+m}$ = the cumulative abnormal return[14] (hereafter, CAR) of stock *i* during the holding period from day *t+1* to day *t+m*. *t* is the article publication day.

$MarketPlaceStatus_{i,t}$ = a dummy variable which indicates whether the author, who has contributed an article containing a position of stock *i*, has joined Marketplace on day *t* or not.

$MarketPlaceArticle_i$ = a dummy variable that indicates whether the article containing a disclosed position of stock *i* is a Marketplace article or not.

$Author_i$ controls for author-specific fixed effects that account for unobserved heterogeneity across authors.

$NumOfComments_{i,t}$ =the number of SA comments posted over day *t* to *t+1* in response to the article, which contains a disclosed position of stock *i*.

$Volatility_{i,t}$ = The sum of squared daily returns of stock *i* in the calendar month preceding day *t*.

[14] The definition of cumulative abnormal return can be found in Appendix D.

$Size_{i,t}$ = The natural log of the market value equity of stock *i* as of the end of the month prior day t.

$BTM_{i,t}$ = The book value of equity of stock *i* as of the end of the most recent fiscal year divided by the market value of equity of stock *i* as of the end of the prior year.

2.5.3 Descriptive Statistics

Table 5 shows descriptive statistics of the main variables in equation (2). The observations include positions of premium NPAs and free NPAs. Table 5 Panel A presents the descriptive statistics of the main variables for 194,508 firm-day observations with long positions. We show the CARs with 1 quarter holding period. The mean and the median of our abnormal returns are near zero, suggesting fairly symmetric return distributions. Statistics for *NumOfComments* suggest that articles have on average 31.563 comments in 2 days after the article publication time. The median of *Volatility*, *Size,* and *BTM* are similar to Campbell et al. (2019).

Table 5 Panel B presents the descriptive statistics of the main variables for 2,776 firm-day observations with short positions. The mean and median of abnormal returns in Table 5 Panel B are smaller than the mean and median of abnormal returns in Table 5 Panel A, suggesting that NPAs try to long (short) stocks with relatively better (worse) future abnormal returns. Statistics for *Volatility* and *Size* are similar in Table 5 Panel A and Panel B. The mean and median of *BTM* in Table 5 Panel B are smaller than the mean and median of *BTM* in Table 5 Panel A, suggesting that stocks with short positions tend to have lower book-to-market ratios.

Table 5 Descriptive Statistics of the Main Variables

Table 5 Panel A reports descriptive statistics for long positions. Table 5 Panel B reports descriptive statistics for short positions. Each observation represents a unique firm-day combination. We show the abnormal returns with 1 quarter holding period.

Panel A: Descriptive Statistics for Long Positions						
Variable	N	Mean	Std. Dev.	25%	50%	75%
$AbRet_{i,t+1,t+m}$	194,508	-0.005	0.184	-0.078	-0.002	0.073
$MarketPlaceStatus_{i,t}$	194,508	0.144	0.351	0	0	0
$MarketPlaceArticle_{i}$	194,508	0.077	0.266	0	0	0
$NumOfComments_{i,t}$	194,508	31.563	52.200	4	13	35
$Volatility_{i,t}$	194,508	0.012	0.076	0.002	0.004	0.010
$Size_{i,t}$	194,508	17.469	2.362	16.130	18.301	19.178
$BTM_{i,t}$	194,508	0.461	0.592	0.192	0.293	0.577

Panel B: Descriptive Statistics for Short Positions						
Variable	N	Mean	Std. Dev.	25%	50%	75%
$AbRet_{i,t+1,t+m}$	8,607	-0.015	0.267	-0.141	-0.016	0.122
$MarketPlaceStatus_{i,t}$	8,607	0.183	0.387	0	0	0
$MarketPlaceArticle_{i}$	8,607	0.071	0.257	0	0	0
$NumOfComments_{i,t}$	8,607	80.224	123.814	6	24	103
$Volatility_{i,t}$	8,607	0.029	0.122	0.006	0.011	0.025
$Size_{i,t}$	8,607	16.488	2.194	14.814	17.258	17.875
$BTM_{i,t}$	8,607	0.329	0.540	0.081	0.138	0.361

2.5.4 Empirical Results

Table 6 shows the results of equation (2) for abnormal returns calculated based on three asset pricing models[15] with one quarter holding period. Table 6 presents results using CAR as the dependent variable. Table 6 Columns (1) and (2) report results using abnormal returns defined with respect to Carhart 4-factor model. Columns (3) and (4) report results using abnormal returns defined with respect to the Fama-French 3-factor model. Columns (5) and (6) report results using abnormal returns defined according to CAPM. "Long Position" demonstrates the

[15] Model definitions can be found in Appendix D.

results of buying stocks with disclosed long positions. “Short Position” demonstrates the results of buying stocks with disclosed short positions.

For example, we can see the Carhart 4 factor model using stocks with long positions, as shown in Table 6 Column (1). These findings compare the three-month abnormal returns if investors buy stocks that have disclosed long positions in SA articles. The coefficient of *MarketPlaceStatus* is significantly negative (-2.072%), suggesting after NPAs have joined Marketplace, the three-month abnormal returns corresponding to their long positions in free articles are on average 2.072% lower than before Marketplace. The coefficient of *MarketPlaceArticle* is significantly positive (0.538%), suggesting that the three-month abnormal returns corresponding to NPAs’ long positions in Marketplace articles are on average 0.538% higher than their free articles after they joined Marketplace. The three-month abnormal returns corresponding to NPAs’ long positions in Marketplace articles are -1.534% (0.538%-2.072%) relative to their free articles before they were on Marketplace.

Also, we can see the Carhart 4 factor model using stocks with short positions, as shown in Table 6 Column (2). These findings compare the abnormal returns if investors buy stocks that have disclosed short positions in SA articles. The coefficient of *MarketPlaceStatus* is positive (1.764%). The coefficient of *MarketPlaceArticle* is significantly negative (-4.298%). As stocks with short positions are bought instead of stocks with long positions, both signs are opposite with Table 6 Column (1). After NPAs have joined Marketplace, the three-month abnormal returns corresponding to their short positions in free articles are on average 1.764% higher than before Marketplace. The three-month abnormal returns corresponding to NPAs’ short positions in Marketplace articles are on average 4.298% lower than their free articles after they joined Marketplace. The three-month abnormal returns corresponding to NPAs’ short positions in Marketplace articles are -2.534% (1.764%-4.298%) relative to their free articles before they were on Marketplace.

Our findings add to the growing literature about how financial incentives affect the quality of UGC. Previous research on this subject has mostly focused on online reviews and provided mixed results. Our study is among the first studies to examine the effect of financial incentives on UGC in a financial market context. We build on the work of Chen et al. (2019), who examine the effect of providing financial incentives to NPAs on SA. They show that financial payments from SA boost the number of articles but have no effect on the quality of those articles. We demonstrate that financial incentives offered by SA have an effect on the quality of stock recommendations in SA articles.

We find that: 1) When NPAs have joined Marketplace, their free recommendation quality is lower than before they joined Marketplace; 2) When NPAs have joined Marketplace, the quality of their recommendations in Marketplace articles is better than the quality of their recommendations in free articles; 3) When NPAs have joined Marketplace, the quality of their recommendations of long (short) positions in Marketplace articles is worse (better) than the quality of their recommendations of long (short) positions in free articles before they were on Marketplace.

Our findings have implications for social media owners in terms of encouraging NPAs to develop and share more valuable stock recommendations. Investors are becoming more interested in social media. The findings of this study show how useful and valuable NPA stock recommendations are in the investment world.

Table 6 Columns (3) to (6) provide similar results as Columns (1) and (2). These results suggest that financial incentives affect the quality of NPAs' stock recommendations on SA. Table 6 shows results with CAR as the dependent variable. The results for buy and hold abnormal returns (hereafter, BHAR) can be found in Appendix E Table E.1, which provide similar results as Table 6.

Table 6 Effect of Financial Incentive on Stock Recommendation Quality (CAR with One Quarter Holding Period)

Coefficients and standard errors are multiplied by 100 for presentation purposes. This table presents the results using CAR with one quarter holding period as the dependent variable. The "CAPM" uses abnormal returns defined according to CAPM. "Fama-French 3" uses abnormal returns defined with respect to the Fama-French 3-factor model. "Carhart 4" uses abnormal returns defined with respect to the Carhart 4-factor model. "Long Position" demonstrates the results of buying stocks with disclosed long positions. "Short Position" demonstrates the results of buying stocks with disclosed short positions. The sample period is from 2004 to 2018. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level.

Parameter	(1) Carhart 4 Long Position	(2) Carhart 4 Short Position	(3) Fama-French 3 Long Position	(4) Fama-French 3 Short Position	(5) CAPM Long Position	(6) CAPM Short Position
MarketPlaceStatus	-2.072***	1.764	-1.898***	2.31	-1.404***	1.184
	(0.288)	(1.484)	(0.288)	(1.479)	(0.294)	(1.498)
MarketPlaceArticle	0.538*	-4.294**	0.52*	-3.987**	0.181	-3.389*
	(0.253)	(1.359)	(0.253)	(1.354)	(0.258)	(1.372)
NumOfComments	-0.002*	-0.008*	-0.001	-0.007*	-0.002.	-0.005
	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)	(0.003)
Volatility	-9.393***	-11.794***	-8.762***	-11.108***	-8.674***	-10.919***
	(0.55)	(2.462)	(0.549)	(2.453)	(0.56)	(2.485)
Size	-0.336***	0.022	-0.291***	0.082	-0.295***	-0.172
	(0.026)	(0.223)	(0.026)	(0.222)	(0.027)	(0.225)
BTM	1.31***	2.366***	1.158***	2.186**	2.458***	2.209**
	(0.086)	(0.682)	(0.086)	(0.679)	(0.088)	(0.688)
R2	0.143	0.448	0.142	0.450	0.145	0.440
Observation	194,508	8,607	194,508	8,607	194,508	8,607
Author fixed effect	Y	Y	Y	Y	Y	Y
Year-Month fixed effect	Y	Y	Y	Y	Y	Y

2.5.5 Further Tests

We examine the model represented in equation (2) based on abnormal returns using different holding periods: one day, one week, or one month.

Table 7, 8 and 9 shows the results using CAR as dependent variable with one month, one week and one day holding periods. In Table 7, 8 and 9, Columns (1) and (2) report results using abnormal returns defined with respect to the Carhart 4-factor model, Columns (3) and (4) report results using abnormal returns defined with respect to the Fama-French 3-factor model, and Columns (5) and (6) report results using abnormal returns defined according to CAPM. “Long Position” demonstrates the results of buying stocks with disclosed long positions. “Short Position” demonstrates the results of buying stocks with disclosed short positions.

Table 7 has similar results (sign and significance level) as Table 6 for long positions. However, Table 7 has opposite sign on *MarketPlaceStatus* in comparison with Table 6 for short positions. Our results still mostly hold when we use one month holding period. In Table 8 and 9, more variables are not significant. We show that our results work better for a longer holding period like one month or one quarter, not a short holding period like one day or one week. The results using BHAR as the dependent variable with less than 1 quarter holding period can be found in Appendix E Table E.2 to E.4, which provide similar results (same sign, similar magnitude, and similar significance level for each corresponding column) as Table 7 to 9.

Table 7 Effect of Financial Incentive on Stock Recommendation Quality (CAR with One Month Holding Period)

Coefficients and standard errors are multiplied by 100 for presentation purposes. This table presents the results using CAR with one month holding period as the dependent variable. The "CAPM" uses abnormal returns defined according to CAPM. "Fama-French 3" uses abnormal returns defined with respect to the Fama-French 3-factor model. "Carhart 4" uses abnormal returns defined with respect to the Carhart 4-factor model. "Long Position" demonstrates the results of buying stocks with disclosed long positions. "Short Position" demonstrates the results of buying stocks with disclosed short positions. The sample period is from 2004 to 2018. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level.

Parameter	(1) Carhart 4 Long Position	(2) Carhart 4 Short Position	(3) Fama-French 3 Long Position	(4) Fama-French 3 Short Position	(5) CAPM Long Position	(6) CAPM Short Position
MarketPlaceStatus	-0.604***	-0.669	-0.554***	-0.548	-0.399*	-1.266
	(0.157)	(0.857)	(0.158)	(0.857)	(0.161)	(0.877)
MarketPlaceArticle	0.355*	-1.675**	0.344*	-1.613*	0.196	-1.458.
	(0.138)	(0.785)	(0.139)	(0.785)	(0.141)	(0.803)
Control variables	Y	Y	Y	Y	Y	Y
Author fixed effect	Y	Y	Y	Y	Y	Y
Year-Month fixed effect	Y	Y	Y	Y	Y	Y

Table 8 Effect of Financial Incentive on Stock Recommendation Quality (CAR with One Week Holding Period)

Coefficients and standard errors are multiplied by 100 for presentation purposes. This table presents the results using CAR with one week holding period as the dependent variable. The "CAPM" uses abnormal returns defined according to CAPM. "Fama-French 3" uses abnormal returns defined with respect to the Fama-French 3-factor model. "Carhart 4" uses abnormal returns defined with respect to the Carhart 4-factor model. "Long Position" demonstrates the results of buying stocks with disclosed long positions. "Short Position" demonstrates the results of buying stocks with disclosed short positions. The sample period is from 2004 to 2018. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level.

Parameter	(1) Carhart 4 Long Position	(2) Carhart 4 Short Position	(3) Fama-French 3 Long Position	(4) Fama-French 3 Short Position	(5) CAPM Long Position	(6) CAPM Short Position
MarketPlaceStatus	-0.097	-0.549	-0.085	-0.53	-0.067	-0.763
	(0.079)	(0.474)	(0.08)	(0.474)	(0.081)	(0.487)
MarketPlaceArticle	0.156*	-0.602	0.145*	-0.605	0.111	-0.568
	(0.07)	(0.434)	(0.07)	(0.434)	(0.071)	(0.446)
Control variables	Y	Y	Y	Y	Y	Y
Author fixed effect	Y	Y	Y	Y	Y	Y
Year-Month fixed effect	Y	Y	Y	Y	Y	Y

Table 9 Effect of Financial Incentive on Stock Recommendation Quality (CAR with One Day Holding Period)

Coefficients and standard errors are multiplied by 100 for presentation purposes. This table presents the results using CAR with one day holding period as the dependent variable. The "CAPM" uses abnormal returns defined according to CAPM. "Fama-French 3" uses abnormal returns defined with respect to the Fama-French 3-factor model. "Carhart 4" uses abnormal returns defined with respect to the Carhart 4-factor model. "Long Position" demonstrates the results of buying stocks with disclosed long positions. "Short Position" demonstrates the results of buying stocks with disclosed short positions. The sample period is from 2004 to 2018. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level.

Parameter	(1) Carhart 4 Long Position	(2) Carhart 4 Short Position	(3) Fama-French 3 Long Position	(4) Fama-French 3 Short Position	(5) CAPM Long Position	(6) CAPM Short Position
MarketPlaceStatus	-0.03	0.208	-0.02	0.239	-0.016	0.104
	(0.035)	(0.244)	(0.035)	(0.245)	(0.036)	(0.247)
MarketPlaceArticle	-0.035	0.013	-0.045	0.037	-0.07*	0.043
	(0.031)	(0.223)	(0.031)	(0.224)	(0.032)	(0.227)
Control variables	Y	Y	Y	Y	Y	Y
Author fixed effect	Y	Y	Y	Y	Y	Y
Year-Month fixed effect	Y	Y	Y	Y	Y	Y

2.6 Conclusion

We investigate how financial incentives offered by social media owners affect the quality of stock recommendations contributed by nonprofessional analysts (NPAs) on the social media platform Seeking Alpha (SA). After adjusting for information in the articles and returns following these articles, we find that an NPA is more likely to start a paid financial service (Marketplace) on SA if they have been a member of SA for a longer time and contributed more articles. We show that financial incentives reduce the quality of non-exclusive (free) stock recommendations. NPAs respond to financial incentives and the quality of their recommendations in Marketplace articles is higher than the quality of their recommendations in free articles after they have joined Marketplace. NPAs put their finest work out where it makes the most income. When NPAs have joined Marketplace, the quality of their recommendations of long (short) positions in Marketplace articles is worse (better) than the quality of their recommendations of long (short) positions in free articles before they were on Marketplace.

This research adds to the literature on the role of social media in financial markets and role of sell-side analysts in financial markets. The sell-side equity research landscape is changing due to a variety of supply and demand factors, including budget constraints and new regulations. Individuals are able to share their thoughts and analyses with a large audience powered by investment-focused social media sites like SA. In the future, these changes could have a significant effect on how investors obtain company-specific research.

CHAPTER 3 - Social Media Sentiment Analysis, Short Text Classification

A huge number of short texts are generated on social media. Microblogging websites have become a rich source for amateur investors, which makes them an ideal source for sentiment analysis. The rapid development of textual analysis techniques has paved the way for automatic sentiment analysis.

HotCopper (hereafter, HC) and StockTwits (hereafter, ST) have been widely used by researchers for sentiment analysis and stock market return predictions. Many studies have focused on using sentiments disclosed on social media to predict the financial market performance. However, the underlying value of the messages collected from social media is still not fully understood due to the lack of a comprehensive apprehension of the performance of different textual analysis methods. The quality of the sentiment analysis is highly dependent on the textual analysis techniques used as well as the pre-processing of data. This chapter consists of two studies. In the first study, we shed light on these problems by comparing the classification accuracy of a dictionary and machine-learning techniques used in analyzing textual data in financial social media. In the second study, we propose a way to conduct pre-processing and compare performance of multiple machine learning classifiers in classifying messages from ST.

Paper (III) The Performance Evaluation of Textual Analysis Tools in Financial Markets

3.1.1 Introduction

Social media platforms and online discussion forums have become popular places to share and learn about products, services, and even financial markets. Therefore, one would expect that the sentiments expressed in social media and investor discussion forums may contain value-relevant information, which will be incorporated by the financial market. Indeed, recent studies (e.g., Tetlock, 2007; Loughran and McDonald, 2011; Hu and Tripathi, 2015) show that the percentage of negative words used in an article captures the tone of varied financial reports and even influence the stock market. Antweiler and Frank (2004) and Li (2008) have also shown the effect of qualitative information on equity valuation. The words selected by news article authors, social media content contributors and company reports have been proved to have explanatory or even predictive power for stock returns, earnings and even managing deceptive activities. There is much research towards how to extract the sentiments from the financial articles and user-generated content on social media using machine-learning algorithms, however, we are yet to have a good understanding of performance accuracy of textual analysis tools.

Textual analysis, especially sentiment analysis, is a domain-dependent problem. An expression that has a clear sentiment in one domain may be ambiguous in other domains. This issue is particularly strong in the financial context analysis, as there are specialized concepts and limited use of effective words. It has been shown that dictionaries developed for other disciplines misclassify common words in the financial context (Loughran & McDonald, 2011). For instance, 'liability' is a neutral word in a financial context, and therefore the L&M dictionary developed by Loughran & McDonald (2011) has been widely used (Chen et al., 2014) in financial context analysis since its first appearance in academia.

Besides the L&M dictionary, researchers have used other machine learning algorithms for classification. For example, Leung and Ton (2015) use Naïve Bayes (NB) algorithm to classify message sentiments and find that sentiments positively relate to the returns of small-cap stocks. Malo et al. (2013) compare the performance of the support vector machine (SVM) with varied underlying pattern analysis algorithms and find that performance is significantly improved when different algorithms are combined. Tirunillai and Tellis (2012) implement both NB and SVM to show that SVM outperform NB in five markets and underperform in one market.

Although previous studies have tested the sentiments from different social media posts or articles, we are still unaware of the strengths and limitations of various approaches for sentiment analysis of user-generated content (UGC) in financial markets. In this research, we contribute to the literature by comparing the performance efficiency of a widely used dictionary in the financial context (Loughran & McDonald, 2011) and two classifiers, Naïve Bayes Classifier and Support Vector Machine Classifier.

The rest of the paper is organized as follows: The second section presents the previous studies; the third section describes the dataset, the fourth section explains the method and the last two sections present the results and discussion.

3.1.2 Literature Review

Textual analysis in financial markets is an emerging area, therefore, the corresponding taxonomies are still not clear. Researchers have used various dictionaries and machine learning algorithms to extract user sentiments from posts and articles in different contexts with varied degrees of success. To address these issues, recently, Loughran and McDonald (2015) compare the four most widely-used dictionaries, which are Henry (2008), Harvard's General Inquirer (GI), DICTION, and the L&M (Loughran & McDonald, 2011). They showed that each word list has its expertise in different contexts, but the L&M dictionary is better than the three other dictionaries in financial contexts for the following two reasons: First, it is more comprehensive than the rest, without missing commonly-appearing negative or positive words. Second, the L&M dictionary was created for the financial context analysis, while GI and DICTION are not specifically designed to analyze financial communications.

Loughran & McDonald (2011) create six different word lists, out of which we will only use the "negative" and "positive" word lists, which have been most widely used in literature (Gurun and Butler 2012; Garcia 2013; and Chen et al. 2014). Our main focus is the fraction of negative words rather than the positive words, as positive words are often negated to convey negative feelings (e.g., not good). A recent work (Chen et al. 2014) using the L&M dictionary shows that the fraction of negative words in articles and comments from SA (SeekingAlpha.com) is negatively related to the subsequent abnormal return with three-month holding period. Besides social media, researchers have also tested other traditional media like newspapers. Based on the earlier work of Tetlock (2007), Dougal et al. (2012) study the Wall Street Journal's (WSJ) columns and find that more pessimistic column tones are linked with more negative market returns on the following day. The research has also been extended to study documents that are related to corporate fundamentals. Huang et al. (2014) find that sentiments of an earnings press release misinform market participants. Although these papers have used the L&M dictionary in many varied financial contexts, we have little understanding of whether this dictionary performs better or worse compared to machine learning techniques, such as Naïve Bayes and support vector machines (SVM).

Besides the wide use of predefined dictionaries, machine-learning algorithms are also becoming popular for analyzing financial articles and comments. In the pool of algorithms, two that are most extensively used to analyze the financial context are Naïve Bayes, and Support Vector Machine. Machine learning algorithms are trained on the training set. We could then apply the "knowledged" algorithms to the remaining sample dataset or out-of-sample dataset. When all sentences are classified, by comparing the classified sentiments and self-disclosed sentiments, one could compute the accuracy of the algorithms. A common challenge in a couple of studies is the unavailability of self-disclosed sentiments, which makes it difficult to assess the accuracy of the classification scheme. For example, using a dataset from Yahoo! Finance message board, Kim and Kim (2014) find only 25.9% of the total messages have self-disclosed sentiments, which makes it is hard to test the accuracy of the out-of-sample dataset. In comparison, our research could compute the exact accuracy, as all of the posts in our dataset have self-disclosed sentiments by the investors.

The Naïve Bayes algorithm, one of the oldest, is called “naive” because it assumes the words are independent of each other, even though that is quite unlikely. It has been used to classify messages (Leung & Ton, 2015) as bullish, neutral, or bearish. However, Leung & Ton (2015) don’t compare the performance of Naïve Bayes with other classification approaches. Antweiler and Frank (2004) is one of the first studies to implement an NB classifier and demonstrate that positive message board posts are followed by negative returns on the next day. Although the authors have reported significant misclassification using the NB algorithm, they don’t use other classifiers for comparison or propose other alternative high-accuracy algorithms.

SVM is another extensively-used classifier in the financial context. Malo et al. (2013) show that substantial performance could be improved by combining different underlying pattern analysis algorithms. Instead of comparing different classifiers, this research focuses on comparing the underlying algorithms and the combined effect of algorithms.

A few studies have implemented both NB and SVM classifiers. Tirunillai and Tellis (2012) compare the accuracy and sensitivity of the two algorithms. The overall results show that SVM performs better than NB in the products review context. Zhang et al. (2012) conduct comprehensive research on eight widely applied text classifiers to stock message board data. They find that NB performs better than SVM in the out-of-sample test.

There are two streams of study in the literature. One stream is about comparing the performance of predefined dictionaries, which are used to analyze sentiments in a financial context. The other stream has focused on different machine learning algorithms to classify the sentiments into several groups. Tirunillai and Tellis (2012) report the classification accuracy using NB and SVM, but not in a financial context. Leung and Ton (2015) fail to report the accuracy for the out-of-sample data and were only using NB. To sum up, none of the studies has compared the performance of dictionaries and the machine learning algorithms in analyzing the financial context in one research. Our research contributes to the literature by comparing the performance of one dictionary and two classifiers that are widely used in the financial context.

3.1.3 Data

The data for this study is obtained from Hot Copper (HC) message board (Hotcopper.com.au). HC allows users to make recommendations (posts) and comment on existing posts about stocks listed in ASX. For this study, we focus on all the fifty stocks, which are listed in the S&P/ASX 50 index. The sample contains all the messages for these stocks from January 2014 to March 2015.

We download and save the messages in a database with date and time stamp, user name, length of the post, content, and sentiment. HC requires users to disclose a sentiment and position along with their posts. Users can choose sentiment from “Sell,” “Hold,” “Buy,” or “None,” and position from “Not Held” and “Held.” Users have to make a choice before the posts are eligible to be seen by other HC users. There is no such requirement on many other message boards, such as Yahoo! Finance. Kim and Kim (2014) show that only 26% of messages from Yahoo Finance have sentiments disclosed by the users, compared to 100% for our dataset.

3.1.4 Methodology

This study compares the performance of a dictionary and two classifiers in a financial context. Specifically, we compare the performance of the L&M dictionary, Naïve Bayes classifier, and Support Vector Machine based classification.

When using the L&M dictionary, we calculate the percentage of negative words for each post and get the median of its overall distribution. A post is bullish if its fraction of negative words is below this median; a post is bearish if its fraction of negative words is above the median (Chen et al., 2014b).

For the NB classifier, we define a message $\{W_k\}(k=1,...K)$ as a sequence of words in which each word $\{W_k\}$ is indexed by k. The words from W_k could be categorized in a message of class C (e.g., bullish) or in counter class $\tilde{C}$ (e.g., bearish). The number of occurrence of words from W_k in class C or counter class $\tilde{C}$ are denoted as m_k and $\tilde{m}_k$ .The total number of words that exist in class C or counter class $\tilde{C}$ are described as n and $\tilde{n}$. We observe the conditional probabilities of words contained in messages from the training set as $P(W_k \mid C) = m_k / n$ and $P(W_k \mid \tilde{C}) = \tilde{m}_k / \tilde{n}$. Then we could calculate the posterior probability $P(C \mid W_k)$ from the prior probability $P(C)$ with:

$$P(C \mid W_k) = \frac{P(W_k \mid C)P(C)}{P(W_k)} = \frac{P(W_k \mid C)P(C)}{P(W_k \mid C)P(C) + P(W_k \mid \tilde{C})P(\tilde{C})} \tag{1}$$

where $P(C)$ is the initial possibility that a message belongs to class C and $P(\tilde{C}) = 1 - P(C)$, while $P(C \mid W_k)$ is the probability that a message belongs to class C given that word W_k is observed. *PS* stands for a post. Using the naïve independence assumption, the possibility for a sentence in class *C* is:

$$P(C \mid PS) = P(C \mid W_1,...,W_n) = \frac{P(C)\prod_{i=1}^{n} P(W_i \mid C)}{P(W_1,...,W_n)} \tag{2}$$

Since $P(W_1,...,W_n)$ is constant given the input, we use the following classification rule:

$$P(C \mid W_1,...,W_n) \propto P(C)\prod_{i=1}^{n} P(W_i \mid C) \tag{3}$$

Finally, the classification with the highest posterior probability is chosen.

$$\hat{C} = \arg\max_{C} P(\mathrm{C}) \prod_{i=1}^{n} P(W_i \mid C) \tag{4}$$

The main difference between NB classifiers lies in the assumptions they make towards the distribution of $P(W_i \mid C)$. For example, Multinomial NB utilizes the NB algorithm on multinomially distributed data. $P(W_i \mid C)$ is the probability that feature *i* appearing in a message that belongs to class *C*. The parameters $P(W_i \mid C)$ may be zero. In this case, the NB algorithm uses Lidstone smoothing or Laplace smoothing.

Bernoulli Naïve Bayes focus on the data that is multivariate Bernoulli distributed. Thus, samples to be processed by this classifier need to be represented as binary-valued feature vectors. If the samples are other kinds of data, the Bernoulli NB will binarize the input. And $P(W_i \mid C)$ is calculated based on:

$$P(W_i \mid C) = P(i \mid C)W_i + (1 - P(i \mid C))(1 - W_i) \tag{5}$$

This method gives a penalty for the non-occurrence of a feature *i* and while the multinomial NB only ignores a non-occurrence feature.

Support Vector Machines (SVMs) are a set of supervised learning methods widely used for classification. A separation boundary will be produced by an SVM in a feature set. We consider *n* training observations, x_i, each of which is a p-dimensional vector of features. Each training set has an associated class label, which could be self-disclosed or manually classified. In our case, the labels y_i are “Buy” or “Sell.” Thus, (x_i, y_i) represents pairs of features and labels. Then a hyperplane is constructed that could separate the training dataset “perfectly” according to the labels. To alleviate the problem of over-fitting, new parameters are introduced into the model: slack values ε_i, and budget *C*. To get the maximal margin hyperplane (MMH), we need to find the largest margin, which is the aggregate smallest perpendicular distance to a training observation from the hyperplane (as shown by the first expression in expression set (6)).

$$y_i(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ij}) \geq M, \forall i = 1, \ldots, n \quad \sum_{j=1}^{p} \beta_j^2 = 1 \quad \varepsilon_i \geq 0, \sum_{i=1}^{n} \varepsilon_i \leq C \tag{6}$$

M is the margin, ε_i states the location for the ith observation relative to the margin and hyperplane, C controls how much the individual ε_i can be modified to violate the margin. In essence, C governs the bias-variance trade-off for the SVM. Besides, there are multiple kernels that SVM could implement, including linear, polynomial, rbf, sigmoid, or precomputed.

3.1.5 Results and Contribution

The dataset is pre-processed before classification. Our raw data contains a total of 36,557 posts. We deleted all the posts with self-disclosed sentiments “None,” removed all the posts with zero “Length_of_Post” and got 19,985 posts, which will be our main dataset (dataset M). Then we removed all the posts with “Hold” sentiments, and the remaining dataset contains 12,920 messages (dataset S).

As all the posts have self-disclosed sentiments (“Buy,” “Hold,” or “Sell”) in our dataset M, therefore, we don’t need to classify the training set for NB and SVM manually. Following the literature (Leung & Ton, 2015), we randomly draw 1000 posts and make an N-fold classification run. In N-fold cross-validation, the original 1000 posts are partitioned in N equal size subsamples. In the N subsamples, a single subsample is retained as the testing dataset, and the rest N-1 subsamples are considered as the training dataset. Then we repeat the cross-validation process until each of the N subsamples has been used as the testing data for one time. We report the average of all these runs. However, if we only focus on the posts with “Buy” and “Sell” sentiments (dataset S), the accuracy is substantially increased (see Figure 1).

Figure 1 10-Fold Validation on Dataset with 1,000 Posts

Figure 1a 10-Fold Validation on Dataset with “Hold” Sentiment

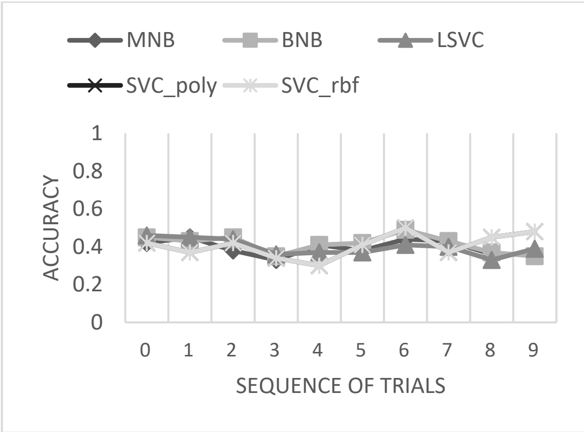
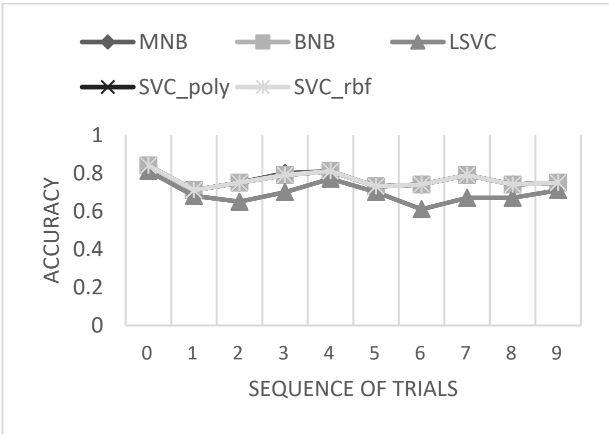


Figure 1b 10-Fold Validation on Dataset without “Hold” Sentiment



In Figure 1, we report results for Multinomial Naïve Bayes (MNB), Bernoulli Naïve Bayes (BNB), Linear Support Vector Classifier (LSVC), SVM with radial basis function kernel (SVC_rbf) (Chang et al., 2010), and SVM with the polynomial kernel (SVC_poly) (Shashua, 2009). The x-axis is the sequence number of trials, and the y-axis is the accuracy. Figure 1a shows the classification accuracy of 1,000 posts randomly chosen from the dataset M (with “Hold” sentiments). Figure 1b shows the classification accuracy of 1,000 posts randomly chosen from the dataset S (without “Hold” sentiments).

It is clear that after removing the posts with “Hold” sentiments from dataset M, we get much better accuracy for all the classifiers. The reason could be that, although authors of these posts with “Hold” sentiment didn’t indicate clear sentiment, the sentences that they use contain words that will show the tone tendency, which will make it hard to classify these kinds of posts. To explore a bit further, we run a ten-fold validation on all the posts with only “Buy” and “Sell” sentiments. Dataset S is used for Figure 2. The results for SVM with rbf and poly kernel are quite similar, so the lines are almost identical. It is clear that the overall accuracy for the SVM with rbf or poly kernel is better than NB or other SVM classifiers.

Figure 2 10-Fold Validation for Dataset S

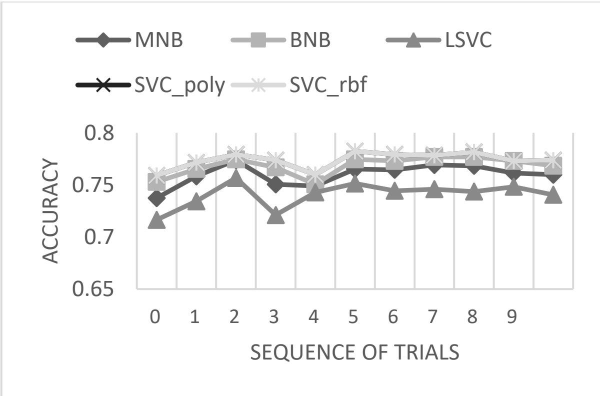


Figure 3 10-Fold Validation for Different Size of Dataset

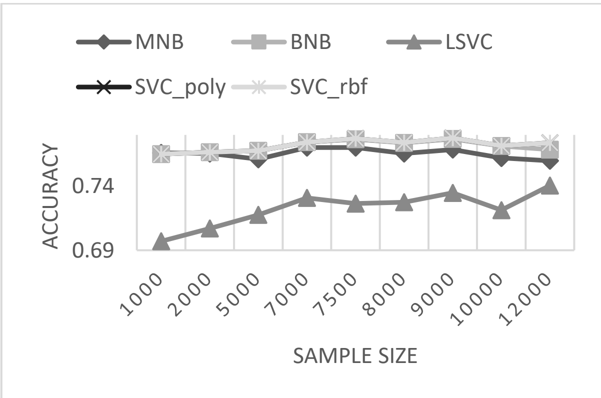


Figure 3 shows the result for 10-fold cross-validation on dataset S with different sample sizes. An in-sample test means testing the accuracy on the training dataset. An out-of-sample test means testing the accuracy on a dataset other than the training dataset. The x-axis tells the number of messages taken into consideration for the 10-fold cross-validation. The y-axis shows the performance accuracy. It is clear that the accuracy does not always increase with the accumulation of the number of messages. The accuracy for SVMs reaches peaks at 7500 and 9000 messages. Antweiler and Frank (2004) and Leung and Ton (2015) hand-code 1000 messages as the training set, but they didn’t report the accuracy for the out-of-sample classification as their dataset didn’t have self-disclosed sentiments (e.g., buy or sell). Tirunillai and Tellis (2012) collected 347,628 reviews and reported a 78% average precision. Their result is comparable to ours (77.4%), but they didn’t disclose their choice of training set, and they were not focusing on the financial context. Kim and Kim (2014) use 4000 messages as the training set and report only 62.7% out-of-sample accuracy. In this case, the researchers need to choose the size of the training set wisely.

Figure 4 shows the accuracy for machine learning classifiers with different numbers of characters in each message taken into consideration. We can see that for most of the algorithms, the accuracy reaches the top with 1000 characters in each message considered. This means that we don’t need to consider all the characters in all of the messages to get the best accuracy. This is consistent with the result that 93.1% of the messages consume less than 1000 characters.

Figure 4 Accuracy for Machine Learning Classifiers with Different Number of Characters Taken into Consideration

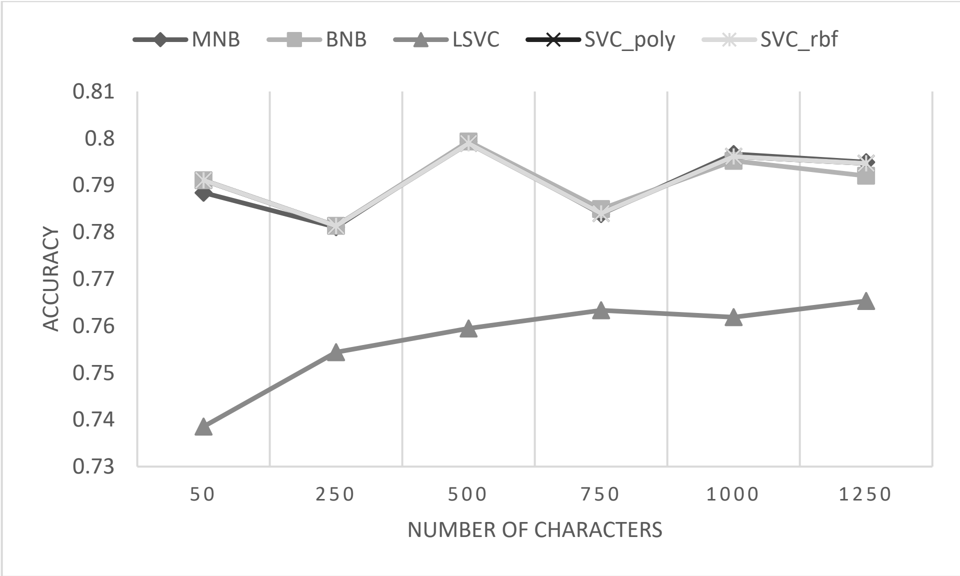
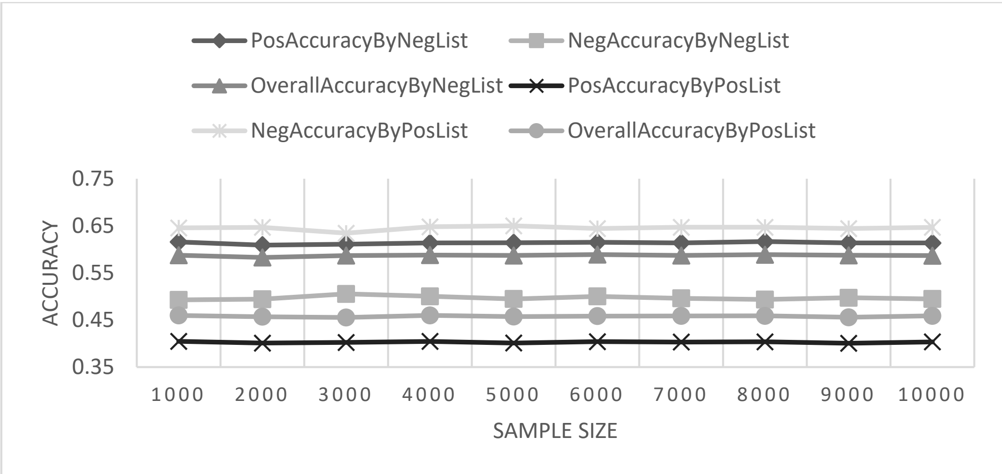


Figure 5 Accuracy for L&M Dictionary with Different Sample Size



After demonstrating the results for the machine learning classifiers, we test the accuracy of the classification by using the L&M dictionary. As we have discussed before, we classify a post as bullish if the percentage of negative words is below the median of the overall distribution; a post is bearish if the percentage of negative words is above the median (Chen et al., 2014b). Also, we try to use the median of the percentage of positive words as another discriminative value. If the percentage of positive words is above the median of the overall distribution, the post is bullish. If not, the post is considered bearish. Figure 4 shows the result for the L&M dictionary using dataset S.

Table 1 Comparison of Accuracy

Classifier	L&M	MNB	BNB	LSVC	SVC_poly	SVC_rbf
Accuracy	0.587	0.759799	0.768474	0.740511	0.773819	0.773819

In Figure 5, PosAccuracyByNegList means the classification accuracy for positive messages using the negative word list, and the rest coordinates on the x-axis are named accordingly. The accuracy is 59% by using the negative word list. The overall accuracy using negative word list is better than the positive word list as the positive words could be negated to express the negative sentiment. It is somewhat surprising that the classification accuracy for the negative messages using the positive word list is much higher than the rest classification using the L&M word list. The reason could be that the messages without any positive words have a higher tendency to convey negative sentiment.

As shown in Table 1, we find that the overall accuracy of machine learning classifiers is much better than the L&M dictionary, even though the L&M is designed for the financial context. Among the tested NB and SVM classifiers, SVM with rbf or poly kernel performs best in the financial context.

3.1.6 Conclusion

This study examines the different textual analysis powers of one dictionary and two machine-learning classifiers in the financial context. Most of the dataset used by the previous research does not have self-disclosed sentiment for all the messages in their testing set. Thus, it is hard for these researchers to report the classification accuracy. Even when some researchers reported their classification accuracy; their out-of-sample accuracy is very poor, or they fail to describe the number of messages in their training set. Because of these shortcomings, it is challenging for the previous studies to compare the classification accuracy of the L&M dictionary and these machine-learning classifiers.

In this research, we fill the gap by showing that SVM with rbf or poly kernel performs best in the financial context analysis of short messages. In addition, researchers should try to train the machine learning algorithms on a reasonable size dataset to get better classification accuracy. This finding could help information systems and finance researchers to get better results when trying to analyze the short messages in financial context.

Paper (IV) The Performance Evaluation of Machine Learning Classifiers on Financial Microblogging Platforms

3.2.1 Introduction

Social media, especially microblogging services, are becoming popular sources of information in almost all domains. For example, millions of Tweets are generated on Twitter every day. Users create, share and discuss information on various topics, from personal life and healthcare problems to societal issues and politics. Financial analysis and investment strategies, which used to be limited to domain experts, are now provided by retail investors on social media (Chen et al., 2014b). The quality of information available on social media platforms is comparable to expert opinions. In fact, many studies have established connections between sentiments on social media platforms and market returns (Oh & Sheng, 2011); Chen et al., 2014; Leung & Ton, 2015). Many studies have analyzed tweets from Twitter, but since Twitter covers a very broad range of topics, it's difficult to filter and choose the right Tweets concentrating on the desired topic. We argue that financial microblogging platforms, such as StockTwits for the stock market, provide a better data source to study discussions and analyze market sentiments.

In recent years, researchers have shown the effect of sentiments derived from microblogging platforms on stock markets (Bollen, Mao, & Zeng, 2011; Leung & Ton, 2015). Social media users use microblogging services to share their opinion about stock markets. This huge amount of data on microblogging platforms like StockTwits, is a treasure trove for market analysts, becomes a new market sentiment indicator and competes with traditional sources (newspaper, online news media, and blogs written by experts). Furthermore, the short length of each message (maximum 140 characters per message) and the use of cashtags (an identifier like hashtag but starts with '$') make it less noisy and easier to analyze. Furthermore, high frequency of content creation by users also allows analysts to track user behavior at a different level, in real-time, during trading.

Given the untrusted content, it’s very challenging for an average person to process the huge amount of data and estimate market sentiments. These shortcomings can be addressed by using machine learning techniques. There has been increasing interest in stock market predictions using various machine learning techniques. Different machine learning algorithms have been used to classify messages into different sentiment groups. However, we are yet to understand the classification efficiency of these algorithms for analyzing messages from a microblogging platform. In this research, we compare the classification performance of different classifiers used for classifying posts on a microblogging platform StockTwits.

Section 2 reviews the related literature on feature selection and sentiment analysis methods, Section 3 describes the data used in this research, Section 4 explains the machine learning classifiers used in this research and Section 5 presents the results. We conclude with a discussion in Section 6.

3.2.2 Literature Review

In literature, many approaches have been used to conduct sentiment analysis in social media. Researchers have used various pre-defined dictionaries and machine learning classifiers to extract user sentiments from social media messages and articles in different contexts. To deal with this issue, Loughran and McDonald (2015) compare the four most widely used dictionaries, which are Henry (2008), Harvard’s General Inquirer (GI), DICTION, and L&M (Loughran & McDonald, 2011). Each dictionary has its expertise, but the L&M is better than the other three in a financial context for the following two reasons. First, the L&M dictionary does not miss common positive and negative words, which makes it more comprehensive than the rest. Second, the L&M dictionary is created for financial context analysis. It has been shown that L&M does really poorly in short message classification in comparison with machine learning classifiers (Hu & Tripathi, 2015a). Thus, we will only compare machine-learning classifiers in this study.

Regarding state of the art for machine learning classification in financial markets, Antweiler and Frank (2004) came up with a novel idea to compute a bullishness index using the computational linguistics method and showed that stock messages can predict market volatility. Bollen et al. (2011) measure collective mood state in term of two states (positive vs negative) and 6 dimension (Cal, Alert, Sure, Vital, Kind and Happy) from Twitter data using OpinionFinder and Google Profile of Mood States, and find an accuracy of 86.7% in predicting the directional changes in the closing price of Dow Jones Industrial Average. Sprenger, Tumasjan, Sandner, and Welpe (2014) collect Twitter messages containing cashtags of S&P 100 companies and classify each message using Naïve Bayes (NB) trained with a set of 2,500 tweets. The results demonstrate that the bullishness index is correlated with the abnormal return, and message volume is associated with trading volume. Oh and Sheng (2011) collect data from StockTwits for three months. The messages are classified by a 'bag of words' approach which apply a machine learning algorithm J48 classifiers. They argued that the sentiments appear to have strong forecasting power over the future market directions. Tirunillai and Tellis (2012) collect data from consumer reviews and classify the reviews using NB and Support Vector Machine (SVM). The results show that negative user-generated content (UGC) has a significant negative effect on abnormal returns with a short "wear-in" and long "wear-out" effects; positive UGC has no significant effect on these metrics. Oliveira, Cortez, and Areal (2013) collect data from StockTwits for 605 trading days. Messages are counted as "bullish" if they contain the words "bullish," and the same logic is applied to messages containing "bearish" words. In contrast with previous studies, they find no evidence of return predictability using sentiment indicators, and of the information content of posting volume for forecasting volatility. Leung and Ton (2015) collect 2.5 million messages from Hotcopper (the biggest Australian stock discussion forum). The messages are classified using NB with a manually classified training set of 10,000 messages. They find that the number of board messages and message sentiment significantly are positively related to the contemporaneous returns of underperforming (low ROE, EBIT margin, EPS) small-capitalization stocks with high market growth potential.

The goal of this paper is to overcome the limitation of previous studies. Prior studies have used varied machine learning classifiers, but no comprehensive comparison has been made between different classifiers. Also, the nature of microblogging (short in length, use of slang, and typo errors) calls for sophisticated pre-processing before the messages can be fed to machine learning algorithms. Finally, many metadata from messages can be used to increase the performance of these algorithms.

3.2.3 Data

We focus on the top ten US stocks based on market capitalization: Apple (AAPL), Alphabet (GOOG, GOOGL), Microsoft (MSFT), Amazon (AMZN), Berkshire Hathaway (BRK.A, BRK.B), Exxon Mobil (XOM), Facebook (FB), Johnson & Johnson (JNJ), General Electric (GE), Wells Fargo (WFC). For each stock, we collect messages posted on StockTwits from January 01, 2016, to June 31, 2016. We randomly select 20,000 tweets for this research.

StockTwits (http://stocktwits.com/) is selected as our data source for this study. StockTwits is a social media platform designed for sharing ideas between various stakeholders, such as investors, traders, and entrepreneurs, etc. It is a popular platform, which had 230,000 active users in June 2013. Messages are limited to 140 characters but may contain links, charts, or even video, which is similar to Twitter. However, in contrast to Twitter, StockTwits only focuses on the stock market and stock investment, which makes it a less noisy data source than other general microblogging services, such as Twitter. Each message contains at least one $cashtag (i.e., $AAPL, $AMZN, $GOOG). Since September 2012, users have been able to disclose their sentiment for each message (post) as “Bullish” or “Bearish.” Since this data contains self-disclosed sentiments, it can be used to test machine-learning algorithms without manual classification.

Pre-processing of data

To remove noise from messages, we have pre-processed all the messages (Agarwal, et. al., 2011) as follows: 1) replace all URLs with a tag ||U||, 2) replace all targets (e.g. “@Sam”) and all cashtags (e.g. “$AAPL”) with tag ||T|| 3) replace all negations (e.g., not, no, never, n’t, cannot) with the notation “NOT,” and 4) replace a sequence of repeated characters by three characters, for instance, convert goooood to goood.

Afterward, we process the tweets using natural language processing tools: 1) we use Stanford tokenizer (Klein & Manning, 2003) to tokenize the tweets, 2) we use a part-of-speech tagger to process tokenized messages and attach a part-of-speech tag to each word, 3) we use the stopword list in Python NLTK to identify and remove stopwords from each message, 4) punctuations are also removed from messages, 5) we use WordNet (Miller & Fellbaum, 1998) to find English words, 6) we get the stem of each word using Porter stemmer.

Prior polarity scoring

We base some of our features on the prior polarity of words (Agarwal et al., 2011). In this case, Dictionary of Affect in Language (DAL) is used and extended by WordNet. DAL contains about 8,000 English words with a pleasantness score between 1 to -3 (negative to positive) for each word. We normalize the scores by dividing all the scores by 3. Words with polarity less than 0.5 are treated as negative, while words with polarity higher than 0.8 are treated as positive, and the rest is treated as neutral. When a word is not found in the DAL dictionary, all synonyms are retrieved from WordNet. We then search for each of the synonyms in DAL. If any synonym is from DAL, the same pleasantness score of the original word in DAL is assigned to its synonym. If none of the synonyms appears in DAL, then the word is not linked with any prior polarity.

Features

Following Agarwal et al. (2011), the features that we use could be divided into four classes: first, a list of words from the training set, and the occurrence of these words for each tweet as Boolean values. Second, counts of primary features, which result in a natural number ($\in$ N). Third, features whose value is a real number ($\in$ R). Fourth, features whose values are Boolean ($\in$ B). Each of these general classes is further divided into two subclasses: Polar features vs Non-polar features. We classify a feature as polar if we find its prior polarity by searching DAL (extended by WordNet). All the other features, which do not have any prior polarity, fall in the Non-polar category. Finally, each of the Polar and Non-Polar features are divided into two subclasses: POS and Other. POS is features that are parts-of-speech (POS) of words, with types of JJ (Adjective), RB (Adverb), VB (Verb), NN (Noun).

Like Agarwal et al. (2011), row f_1 belongs to class Polar POS and is the count of the number of positive and negative POS in messages. f_2, f_3, f_4 all belong to class Polar Other. f_2 is the number of negation words and positive and negative prior polarity. f_3 is the number of (+/-) hashtags, capitalized words, and words with exclamation marks. f_4 belongs to Non-Polar POS and is the number of the different part-of-speech tags. f_5, f_6 belong to Non-Polar Other. f_5 is other words without polarity; f6 is the number of hashtags, URLS, targets, and cashtags. f_7 belongs to Polar POS and is the sum of prior polarity scores of words with POS of JJ, RB, VB, and NN. f_8 belongs to Polar Other and is the sum of prior polarity scores of all words. f_9 refers to class Non-Polar Other and is the percentage of tweets that is capitalized. Finally, f_{10} belongs to class Non-Polar Other and is the presence of exclamation and presence of capitalized words. The descriptions are shown in Table 1.

Table 1 Summary Statistics

N	Polar	POS	# of (+/-) POS (JJ, RB, VB, NN)	f_1
		Other	# of negation words, positive words, negative words	f_2
			# of (+/-) hashtags, capitalized words, exclamation words	f_3
	Non-Polar	POS	# of POS (JJ, RB, VB, NN)	f_4
		Other	# of words without prior polarity	f_5
			# of hashtags, URLs, targets, cashtags	f_6
P	Polar	POS	For POS, ∑ prior polarity score of words that POS	f_7
		Other	∑ prior polarity scores of all words	f_8
	Non-Polar	Other	Percentage of capitalized text	f_9
B	Non-Polar	Other	Exclamation, capitalized text	f_{10}

3.2.4 Machine Learning Models

In this research, we use three different classifiers: Naïve Bayes (NB), Logistic Regression (LR), and Support Vector Machine (SVM). We choose these three classifiers, as NB and SVM are the two most widely used classifiers in the social media sentiment analytics in a financial context, and LR is a good approach for 2-way classification (classify dataset into two groups), although it has not been explored in comparison with the other two classifiers in the social media sentiment analytics in a financial context. Each classifier is tested using 10-fold cross-validation, which is common practice with machine-learning classifiers. For Naïve Bayes, we use Multinomial NB and Bernoulli NB. For SVM, we use three different kernels, which are linear, poly, and rbf kernels.

Naïve Bayes

NB is based on Bayes' theorem with the naïve assumption of independence between every pair of features. Given C stands for a class and W_1 to W_n are the feature vector, Bayes' theorem states the following:

$$P(C|W_1, \dots, W_n) = \frac{P(C)P(W_1, \dots, W_n|C)}{P(W_1, \dots, W_n)} \tag{1}$$

The naïve assumption gives that:

$$P(W_1, \dots, W_n|C) = \prod_{i=1}^{n} P(W_i|C) \tag{2}$$

The relationship of Equation (1) is then simplified to:

$$P(C|W_1, \dots, W_n) = \frac{P(C)\prod_{i=1}^{n} P(W_i|C)}{P(W_1, \dots, W_n)} \tag{3}$$

As $P(W_1, \dots, W_n)$ is always a constant value given the input (W_1 to W_n), we can apply the following classification rule:

$$P(C|W_1, \dots, W_n) \propto P(C)\prod_{i=1}^{n} P(W_i|C) \tag{4}$$

Finally, the classification with the highest posterior probability is chosen.

$$\hat{C} = argmaxP(C)\prod_{i=1}^{n} P(W_i|C) \tag{5}$$

The main difference between NB classifiers is the assumptions that they make regarding the distribution of $P(W_i|C)$.

Multinomial NB uses the NB algorithm for multinomial distributed data. $P(W_i|C)$ is estimated by a smoothed version of maximum likelihood:

$$P(W_i|C) = \frac{N_{Ci} + \alpha}{N_C + \alpha n} \tag{6}$$

where N_{Ci} is the total number of times feature W_i falls in a sample of class C in the training set, and N_C is the total number of all features for class C. α is the smoothing parameter and prevents zero probabilities.

Bernoulli NB uses the NB classifier for multivariate Bernoulli distributed data. The decision rule for Bernoulli NB is based on:

$$P(W_i|C) = P(i|C)W_i + (1 - P(i|C))(1 - W_i) \tag{7}$$

which penalized the non-occurrence of a feature i that is an indicator for class C.

Logistic Regression

The logistic function $\sigma(t)$ is defined as follows:

$$\sigma(t) = \frac{1}{1 + e^{-t}} \tag{8}$$

We presume that t is a function of the independent variables $(W_1, \ldots, W_n)$, where:

$$t = f(W_1, \ldots, W_n) \tag{9}$$

And the logistic function could be written as:

$$F(W_1, \ldots, W_n) = \frac{1}{1 + e^{-f(W_1, \ldots, W_n)}} \tag{10}$$

F(x) is described as the probability of the dependent variable *(C)* is a "Bullish" or "Bearish."

Support Vector Machine

SVMs are a group of supervised learning algorithms widely used for classification. To have an overview of SVMs, SVMs provide a separation boundary (linear or non-linear) in the dataset. We consider a training set with n observations (x_i). Each of the observations is a p-dimensional vector of features. Each training set has a self-disclosed label (y_i) in this research. Then a hyperplane or a hypersurface is constructed that could separate the training dataset with respect to the labels. To balance the problem of over-fitting and under-fitting, a parameter is introduced into the model: penalty parameter *C* of the error term. The lower your *C* value, the smoother and more generalized your decision boundary is going to be. But if you have a large *C* value, the classifier will attempt to do whatever is in its power to perfectly separate each sample to correctly classify it.

Kernel methods enable SVMs to be functional in a higher dimensional, implicit feature space, without calculating the coordinates of data in that space, but rather by calculating the inner products between all pairs of data.

Measures

We measure the accuracy, precision, recall, and F1 measures for all the classifiers.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{11}$$

$$Precision = \frac{tp}{tp + fp} \tag{12}$$

$$Recall = \frac{tp}{tp + fn} \tag{13}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{14}$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \; (15) \quad Precision = \frac{tp}{tp + fp} \; (16)$$

$$Recall = \frac{tp}{tp + fn} \; (17) \quad F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \; (18)$$

where *tp* is true positive, *tn* is true negative, *fp* is false positive, and *fn* is false negative.

3.2.5 Results

We use 20,000 messages for this research. Each message has a self-disclosed sentiment, which is “Bullish” or “Bearish.” This provides good training sets as well as testing sets for the classifiers. We do 10-cross validations for this study. The original dataset is partitioned into 10 equal size subsamples. In the ten subsamples, a single subsample is used as the testing dataset, while the rest nine subsamples are used as the training set. Then we report the average accuracy, precision, recall, and F1 measure for all the experiments with different sizes of data. Figure 1 shows the learning curve for the 2-way classification. “MNB” is Multinomial NB, “BNB” is Bernoulli NB, “LR” is Logistic Regression, “LSVC” is Linear SVM, “SVC_poly” is SVM using the poly kernel, and “SVC_rbf” is SVM using radial basis function kernel.

It is clear that Logistic Regression Classifier outperforms all the other classifiers in this 2-way classification. However, Logistic Regression is not widely used in the classification of messages from social media in the financial context. In this case, we encourage researchers to use more classifiers and compare the accuracy of the classifiers instead of only focusing on one or two classifiers with one kind of kernel. Figure 1 also shows that there is a quite sharp increase in accuracy when the size of the dataset moves over 7,500. Thus, we encourage researchers to use a training set of more than 7500 to have good accuracy in classification.

Overall accuracy, precision, recall, and F-Measure are summarised in Figure 2. There is a trade-off between recall and precision. Thus researchers have used F-Measure to determine which method is superior to others. The Logistic Regression classifier has the highest value for accuracy (0.844) and F-Measure (0.901). This means that LR outperforms other classifiers in social media sentiment analytics in a financial context. We also notice that SVMs with poly and rbf kernel have a recall of value one and the lowest precision among all the classifiers. This means that these two classifiers have no false-negative classification and have a great amount of false-positive classification, which makes these two classifiers have very poor performance.

Figure 1 Learning Curve

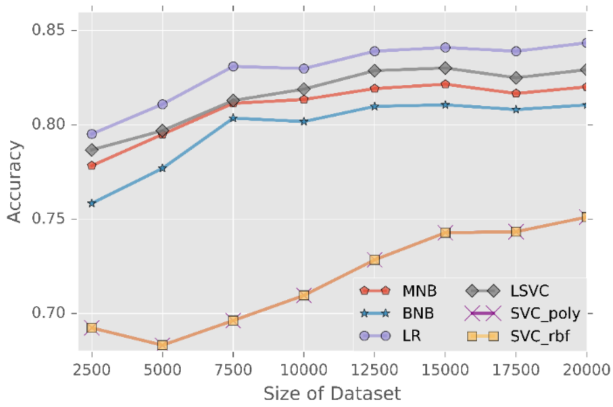
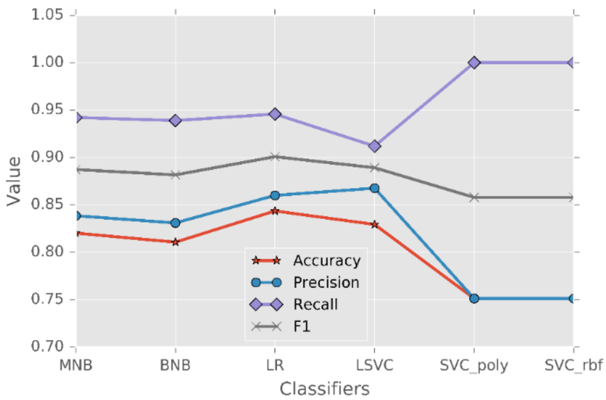


Figure 2 Accuracy, Precision, Recall and F1 Measure



Previous research on Twitter has used SVM to classify tweets from Twitter into two sentiment groups and obtained an accuracy of 75.39%. They streamed the data in real-time. No language, location, or any other kind of restriction is made during the streaming process. Tweets in foreign languages are converted into English using Google translate before the annotation process. They manually annotated 11,875 tweets. In comparison, our research comes up with an accuracy of 81.9% using Linear SVM, with a dataset of 10,000 tweets. Using almost the same method (unigram and metadata features), the accuracy for StockTwits outperforms Twitter. The reasons could be: 1) StockTwits is focusing on the financial market, which has less noise. 2) There is a greater portion of users in StockTwits who are investors or traders.

These people use more formal and accurate words than average users of Twitter. In this case, StockTwits is considered as a better data source to conduct sentiment analysis, especially in a financial context.

3.2.6 Conclusion

In this study, we achieve the following: First, we find that among the three classifiers, Logistic Regression performs the best in classifying messages on StockTwits. Though prior research studies analyzing financial microblogging services have been using NB or SVM, we report a superior performance of Logistic Regression in this environment. Second, we get a better accuracy using messages from StockTwits than from Twitter as a data source. When we want to find the correlation between social media sentiment and stock market variables, we want to include as many messages from social media platforms as we can. This gives rise to the need to classify all messages (with or without a self-disclosed sentiment) from a social media platform. Thus, we posit that StockTwits could be a better data source than Twitter to analyze sentiments in financial markets.

CONCLUSION

This thesis contains three studies that explore different aspects of the value of financial advice on social media. The thesis starts by examining the effect of the long or short positions of nonprofessional analysts (hereinafter, NPAs) contributing to the social media outlet Seeking Alpha (hereafter SA) on the direction of investor trading and subsequent stock returns. The second study explores the effect of financial incentives provided by social media platform owners on the quality of stock recommendations from NPAs. Finally, the third study compares the performance of textual analysis methods on textual data collected from financial social media websites.

In the first study, we discover that NPA positions lead to short-window order imbalances following SA article publication. We form portfolios by buying stocks with the most favorable sentiment and short selling stocks with the least favorable sentiment, with various rebalancing windows. However, there is no indication that the information on Seeking Alpha can be used to generate economically significant abnormal returns.

The second study adds to the literature about the role of social media in financial markets, the role of sell-side analysts in financial markets, and the effect of financial incentives in affecting the quality of user-generated content produced by NPAs on SA. We examine the effect of offering financial incentives to NPAs on the quality of stock recommendations. We find that NPAs who have been with SA for a longer period and that have published more articles are more likely to join the premium partnership program and get monetary payments. Financial incentives lower the quality of free stock recommendations. NPAs respond to financial incentives by putting their best work where it generates the most money. After entering the premium partnership program, the quality of NPAs' long (short) stock position recommendations in fee-based articles is worse (better) than the quality of their long (short) stock position recommendations in free articles before joining the premium partnership program.

In the final study, we examine how textual analysis approaches perform on data from the financial microblogging sites HotCopper (hereinafter, HC) and StockTwits (hereafter, ST), which are extensively used for sentiment analysis and stock market return predictions. When classifying short text from HC, we show that machine learning classifiers outperform the Loughran & McDonald (2011) dictionary. Researchers should try to use a financial social media like ST rather than more informal social media like Twitter when examining the effect of social media sentiment on stock market abnormal returns.

APPENDICES

Appendix A: The Effect of Seeking Alpha Disclosed Positions on Investors' Trading

We construct variables that stand for the number of long (short) disclosed positions and consider a normalized position score,

$$NormalizedPos_{i,t} = \frac{NumLongPos_{i,t} - NumShortPos_{i,t}}{NumLongPos_{i,t} + NumShortPos_{i,t}} \qquad (A.1)$$

$NumLongPos_{i,t}$ = the number of disclosed long positions for stock *i* on day *t*.

$NumShortPos_{i,t}$ = the number of disclosed short positions for stock *i* on day *t*.

We then estimate the following regression:

$$OIB_{i,t+n} = \beta_0 + \beta_1 PostSA_{i,t+n} + \beta_2 NormalizedPos_{i,t} + \beta_3 PostSA_{i,t+n} * NormalizedPos_{i,t} + Controls_{i,t} + \varepsilon_{i,t} \qquad (A.2)$$

where,

$OIB_{i,t+n}$ = the order imbalance for stock *i* on day *t+n*. Day *t* is the SA article publication day. *n* is the number of days that order imbalance is relative to day *t*. *n* is limited to [-6, +6] event window, excluding event day 0. The choice of *n* is based on Chapter 1, Figure 2.

$NormalizedPos_{i,t}$ = the normalized position score for stock *i* on day *t* as shown in equation (A.1).

$PostSA_{i,t+n}$ = the dummy variable equals one if the trading is measured after day *t* and zero if trading is measured before day *t*. Thus, $PostSA_{i,t+n}$ equals one over the [1,6] window and zero over the [-6, -1] window.

$\beta_1 PostSA_{i,t+n} * NormalizedPos_{i,t}$ interacts the event time indicators with the normalized position score.

$Controls_{i,t}$ are defined the same as in Chapter 1 equation (3).

T-statistics are computed using two-way clustered standard errors. The standard errors are clustered by stock and day to account for cross-sectional correlation in residuals[16].

[16] To obtain unbiased estimates in finite samples, the clustered standard errors are adjusted by (N-1)/(N-P)× G/(G-1), where N is the sample size, P is the number of independent variables, and G is the number of clusters (Mark (Shuai), 2020).

Table A.1 Seeking Alpha Stock Recommendations and Direction of Investor Trading

This table presents results from the estimation of Equation (A.2). Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level. Coefficients and standard errors are multiplied by 100 for presentation purposes.

Variable	(1)	(2)
$PostSA_{i,t+n}$ * $NormalizedPos_{i,t}$	0.084	
	(0.059)	
$NormalizedPos_{i,t}$	0.062	0.146**
	(0.046)	(0.047)
$PostSA_{i,t+n}$	-0.132*	-0.132*
	(0.059)	(0.059)
$Volatility_{i,t}$	-0.178	-0.178
	(0.143)	(0.143)
$Size_{i,t}$	0	0
	(0.011)	(0.011)
$BTM_{i,t}$	0.194***	0.194***
	(0.038)	(0.038)
$InstOwn_{i,t}$	0.634***	0.634***
	(0.131)	(0.131)
$MKTRF_{t+n}$	105.937***	105.937***
	(1.205)	(1.205)
SMB_{t+n}	37.182***	37.182***
	(2.005)	(2.005)
HML_{t+n}	12.305***	12.305***
	(2.174)	(2.174)
MOM_{t+n}	-2.817.	-2.817.
	(1.491)	(1.491)
R2	0.016	0.016
Observation	835,916	835,916

Column (1) of Table A.1 shows that order imbalances are not significantly related to normalized position scores before the publication of SA articles as the coefficient on $NormalizedPos_{i,t}$ is not significant (0.062). The difference between order imbalances before and after SA article publication day are not significantly related to the normalized scores, as the coefficient on the interaction term is not significant (0.084).

We further test if the sum of the coefficient on $NormalizedPos_{i,t}$ and the coefficient on the interaction term is significantly different from zero and shows results in Column (2) of Table A.1. Order imbalances are significantly related to normalized position scores after the publication of SA articles as the sum of the coefficients is significant (0.146). A one-unit increase in the normalized position score is associated with 0.146 percentage points increase in order imbalances after the SA article publication.

Appendix B: Market Prediction

In this section, we try to predict the market return using the average percentage of negative words across all single-ticker research articles. Time *T* can be 1 day, 1 week or 1 month. We use *T* as 1 month as an example to demonstrate the model.

To measure the market level sentiment, we calculate the average percentage of negative words of all single-ticker articles published during month *T*, $\overline{NegSA}_T$, which is calculated by taking the sum of the individual percentage of negative words, $NegSA_{j,T}$, of each article published during month *T* and divided by n_T, which is the total number of single-ticker articles published during month *T*.

$$\overline{NegSA}_T = \frac{1}{n_T}\sum_{j=1}^{n_T} NegSA_{j,T} \qquad (B.1)$$

To predict the market level return, we use the following regression model:

$$R_{MKT,T} = \alpha + \beta\overline{NegSA}_T + \beta R_{MKT,T-1} \qquad (B.2)$$

where

$R_{MKT,T}$ = the market return of month *T*.

$R_{MKT,T-1}$ = the market return of month *T-1*.

$\overline{NegSA}_T$ = average percentage of negative words of all single-ticker articles published during month *T*.

Table B.1 shows the result of market-level prediction when *T* is 1 day. We remove all days with less than *n* articles published, with *n* can be 1, 5, 10, 20, 30, 40, or 50. The coefficients of $\overline{NegSA}_T$ are never significant at a 5% confidence level. In untabulated results, we find the coefficients of $\overline{NegSA}_T$ are not significant at 5% confidence level when *T* is 1 week or 1 month. We show that the percentage of negative words does not work well when aggregated on a market level.

Table B.1 Day Level Market Prediction

Summary of Market Prediction model with T = 1, 1 day article period and 1-day holding period, formed using average percentage of negative words across all SA single-ticker research articles. With the same model, we remove days that have less than *n* articles published. Choices of *n* are listed on the x-axis of the table. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level. $\overline{NegSA}_T$ is the average percentage of negative words of all single-ticker articles published on daty T. $R_{MKT,T-1}$ is the market return of day *T-1*. Coefficients and standard errors are multiplied by 100 for presentation purposes.

Number of Articles Limit of Each Day	1	5	10	20	30	40	50
$\overline{NegSA}_T$	0.007	-0.008	-0.003	-0.02	-0.04	-0.04	-0.05
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
$R_{MKT,T-1}$	-7.88***	-7.89***	-7.94***	-7.89***	-7.93***	-8.22***	-7.91***
	(1.7)	(1.7)	(1.7)	(1.7)	(1.8)	(1.8)	(1.8)
R2	0.006	0.006	0.006	0.006	0.006	0.006	0.006
Adjusted R2	0.006	0.006	0.006	0.006	0.006	0.006	0.006
Observation	3562	3523	3475	3163	3246	3203	3163

Appendix C: Portfolio Construction Using Articles with Disclosed Positions

Data

In this chapter, we consider all articles no matter they are single-ticker articles or multiple-ticker articles. As before, we remove SA transcripts and articles whose length is less than 100 characters. Following the procedure in 1.3.3, each article is classified with a disclosed position: "long," "short," "complex" or "no position." We remove all articles with "no position" or "complex" disclosed positions and use only articles with "long" or "short" disclosed positions. One article can have disclosure of multiple stock tickers. Afterward, we map each ticker with a disclosed position to PERMNO using the CRSP STOCKNAMES file.

Research Design

We count the number of "long" disclosed positions and the number of "short" disclosed positions of firms covered in SA articles published on day *t-1* and before the close of the trading day. The "LONG" portfolio is composed of stocks with "long" disclosed positions. The "SHORT" portfolio is composed of stocks with "short" disclosed positions. If a stock has *a1* times "long" and *b1* times "short" disclosed positions on the same day, it is considered *a1* times in the "LONG" portfolio and *b1* times in the "SHORT" portfolio accordingly. We investigate the trading strategy of buying the "LONG" portfolio and short selling the "SHORT" portfolio, denoted as the "LONG-SHORT" portfolio.

When a longer date interval *M* (between day *t-m* (included) and *t-1* (included)) is considered for the portfolios, we consider all stocks which have been covered by SA articles over the last *m* trading days, until day *t-1*. During date interval *M*, we count the number of "long" disclosed positions and the number of "short" disclosed positions of each covered firm. Afterward, we construct the "LONG" portfolio and the "SHORT" portfolio and investigate the trading strategy of buying the "LONG" portfolio and short selling the "SHORT" portfolio, similar to what we have done for the daily portfolio using disclosed positions.

After determining the composition of each portfolio *p* as the close of trading day *t-1*, we calculate the value-weighted return for day *t* as proposed in section 1.5.1. Daily returns for each portfolio p, $R_{p,t}$, are compounded over the trading days to capture returns with a longer holding period, $R_{p,T}$.

We inspect the performance of trading strategies based on the disclosed positions of each firm. We inspect the trading strategies based on *M (article period)* equals to 1 day, 1 week, 1 month or 1 quarter (calendar day). Each stock that enters a portfolio will be held for date interval *T (holding period)*. In this case, *T* equals to 1 day, 1 month, 1 week or 1 quarter. As a result, we end up with 16 different portfolio groups with a different combination of article periods (*M*) and holding periods (*T*), as shown in Chapter 1 Table 6. We assume 5 trading days per week and 21 trading days per month for these calculations. As a result, 1-week, 1-month, and 1-quarter date intervals are calculated using 5, 21, and 63 trading days.

For each portfolio composed of differences in returns between the "LONG" portfolio and the "SHORT" portfolio, we regress it on alpha using the following regression. Time *T* is our holding date interval, for which we use 1 month as an example to demonstrate the regression.

$$R_{LONG,T} - R_{SHORT,T} = \alpha_p + \beta_p MKTRF_T + s_p SMB_T + h_p HML_T + m_p MOM_T + e_{p,T} \quad (C.1)$$

where

$R_{LONG,T} =$ the month *T* return on the "LONG" portfolio, which buys all the stocks with disclosed long positions during month *T*.

$R_{SHORT,T} =$ the month *T* return on the "SHORT" portfolio, which buys all the stocks with disclosed short positions during month *T*.

while $MKTRF_T$, SMB_T, HML_T, MOM_T, α_p, β_p, s_p, h_p, m_p are the same as shown in section 1.5.1.

Empirical Results

Chapter 1 Table 6 shows our 16 portfolio groups. We shuffle the holding periods the same way as shown in Chapter 1empirical results section. As a result, for portfolio groups with 1 day as the holding period, we will only have 1 portfolio in each portfolio group. For portfolio groups with 1 week (1 month, 1 quarter) as holding periods, each portfolio group is composed of 5 (21, 63) portfolios.

Table C.1, columns (1) to (4), demonstrate the estimated alpha values and factor loadings for stocks in portfolio groups (d-d, w-d, m-d, q-d) according to the disclosed positions. These portfolio groups have 1 day as the holding period. However, the alphas are not significant for all these portfolio groups.

Table C.1 Fama-French-Carhart Filtering of Portfolio Groups with 1 Day Holding Period

Summary of Fama-French-Carhart four-factor model of portfolio groups with 1 day holding period formed using all SA research articles and formed into portfolios using disclosed positions. Each portfolio group only contains 1 portfolio. The "LONG" portfolio is composed of stocks with disclosed long positions. The "SHORT" portfolio is composed of stocks with disclosed short positions. We investigate the trading strategy of buying the "LONG" portfolio and short selling the "SHORT" portfolio, denoted as the "LONG-SHORT" portfolio. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level. The portfolio is rebalanced daily. The coefficient estimates are those from the Carhart four-factor regression of the portfolio returns (R_{LONG}-R_{SHORT}) on the market excess return (MKTRF), the size factor (SMB), the value factor (HML), and the momentum factor (MOM). Coefficients and standard errors are multiplied by 100 for presentation purposes.

Variable	(1) d-d	(2) w-d	(3) m-d	(4) q-d
Alpha	-0.02	0.03	0.02	0.02
	(0)	(0)	(0)	(0)
MKTRF	84.15***	83.9***	81.95***	83.96***
	(2.3)	(1.4)	(1.4)	(1.3)
SMB	-37.61***	-32.23***	-28.28***	-25.89***
	(4.4)	(2.8)	(2.7)	(2.5)
HML	0.83	-6.68*	2.11	0.24
	(4.6)	(3)	(2.9)	(2.6)
MOM	-4.57	-5.16*	-4.83*	-4*
	(3.2)	(2)	(1.9)	(1.7)
R2	0.36	0.544	0.545	0.606
Adjusted R2	0.359	0.543	0.544	0.606
Observation	3052	3360	3516	3531

Appendix D: Asset Pricing Models

The dependent variable is $AbRet_{i,t+1,t+m}$, which is the abnormal return of stock *i* during the holding period from day *t+1* to day *t+m*. *t* is the article publication day. For instance, an NPA may say, "I am/we are long AAPL." If an investor trade on this disclosed position and buy stock AAPL, an abnormal return $AbRet_{AAPL,t+1,t+m}$ will be generated if he/she holds it for *m* days. $AbRet_{AAPL,t+1,t+m}$ is the measurement of the quality of the article.

I apply three different asset pricing models to obtain the abnormal returns following the event days. First, I use a simple capital asset price model (CAPM), which only includes a market risk factor (MKTRF), as shown in equation (D.2). Second, I use the Fama-French three-factor model (Fama & French, 1993), which adds a size factor (SMB) and a value factor (HML) to the CAPM, as shown in equation (D.4). Third, I use the Carhart four-factor model (Carhart, 1997), which adds a momentum factor (MOM) to the three-factor model, as shown in equation (D.6). For each model, we report the results for two types of abnormal return: cumulative abnormal return (CAR) and buy-hold abnormal return (BHAR).

More specifically, we obtain abnormal returns using the following three models. We use the holding period as month *T* to demonstrate the regression.

$$AbRet_{i,T} = R_{i,T} - E\left(R_{i,T}\right) \qquad (D.1)$$

where,

$AbRet_{i,T}$ = abnormal return of stock *i* in month *T*.

$R_{i,T}$ = actual return of stock *i* in month *T*,

$E\left(R_{i,T}\right)$ = expected return of stock *i* in month *T*.

The first model is the CAPM:

$$E\left(R_{i,T}\right) - R_{f,T} = \alpha + \beta_{1,i} * MKTRF_T \qquad (D.2)$$

$$AbRet_{i,T} = R_{i,T} - E\left(R_{i,T}\right) = R_{i,T} - \left(R_{f,T} + \alpha + \beta_{1,i} * MKTRF_T\right) \qquad (D.3)$$

where $R_{f,T}$ is the one-month Treasury bill rate in month *T* and $MKTRF_T = R_{MKT,T} - R_{f,T}$, which is the excess return on the market in month *T*. It is calculated as the value-weighted return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate.

The second model is the Fama-French 3-factor model:

$$E\left(R_{i,T}\right) - R_{f,T} = \alpha + \beta_{1,i} * MKTRF_T + \beta_{2,i} * SMB_T + \beta_{3,i} * HML_T \qquad (D.4)$$

$$AbRet_{i,T}$$
$$= R_{i,T} - E(R_{i,T})$$
$$= R_{i,T} - (R_{f,T} + \alpha + \beta_{1,i} * MKTRF_T + \beta_{2,i} * SMB_T + \beta_{3,i} * HML_T) \quad (D.5)$$

where SMB_T is the size factor, which is the historic excess returns of small-cap stocks over large-cap stocks in month *T* and HML_T is the value factor, which is the historic excess returns of value stocks (high book-to-market ratio) over growth stocks (low book-to-market ratio) in month *T*.

The final model is the Carhart 4-factor model, which is an extension of the Fama-French 3-factor model:

$$E(R_{i,T}) - R_{f,T} = \alpha + \beta_{1,i} * MKTRF_T + \beta_{2,i} * SMB_T + \beta_{3,i} * HML_T + \beta_{4,i} * MOM_T \quad (D.6)$$

$$AbRet_{i,T}$$
$$= R_{i,T} - E(R_{i,T})$$
$$= R_{i,T} - (R_{f,T} + \alpha + \beta_{1,i} * MKTRF_T + \beta_{2,i} * SMB_T + \beta_{3,i} * HML_T + \beta_{4,i} * MOM_T) \quad (D.7)$$

where MOM_T is the momentum factor, which is the historic excess returns of highest performing stocks over lowest performing stocks in month *T*.

We calculate the ensuing cumulative three-month abnormal return for each disclosed position in our sample.

Appendix E: Effect of Financial Incentive on Stock Recommendation Quality using BHAR

Table E.1 to E.4 show results using BHAR as the dependent variable with different holding periods.

Table E.1 Effect of Financial Incentive on Stock Recommendation Quality (BHAR with One Quarter Holding Period)

Coefficients and standard errors are multiplied by 100 for presentation purposes. This table presents the results using BHAR with one quarter holding period as the dependent variable. The "CAPM" uses abnormal returns defined according to CAPM. "Fama-French 3" uses abnormal returns defined with respect to the Fama-French 3-factor model. "Carhart 4" uses abnormal returns defined with respect to the Carhart 4-factor model. "Long Position" demonstrates the results of buying stocks with disclosed long positions. "Short Position" demonstrates the results of buying stocks with disclosed short positions. The sample period is from 2004 to 2018. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level.

Panel B: BHAR as the Dependent Variable

Parameter	(1) Carhart 4 Long Position	(2) Carhart 4 Short Position	(3) Fama-French 3 Long Position	(4) Fama-French 3 Short Position	(5) CAPM Long Position	(6) CAPM Short Position
MarketPlaceStatus	-2.238***	2.986.	-2.085***	3.496*	-1.541***	2.317
	(0.309)	(1.567)	(0.311)	(1.558)	(0.313)	(1.579)
MarketPlaceArticle	0.566*	-4.717**	0.563*	-4.378**	0.187	-3.668*
	(0.271)	(1.435)	(0.273)	(1.426)	(0.275)	(1.446)
NumOfComments	-0.005***	-0.009**	-0.003**	-0.008*	-0.004***	-0.005
	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)	(0.003)
Volatility	-13.799***	-13.31***	-13.457***	-12.545***	-14.375***	-12.412***
	(0.59)	(2.6)	(0.593)	(2.584)	(0.598)	(2.619)
Size	-0.158***	0.323	-0.126***	0.36	-0.141***	0.114
	(0.028)	(0.235)	(0.028)	(0.233)	(0.029)	(0.237)
BTM	1.456***	2.633***	1.089***	2.327**	2.54***	2.319**
	(0.092)	(0.72)	(0.093)	(0.715)	(0.094)	(0.725)

R2	0.145	0.422	0.143	0.424	0.148	0.416
Observation	194,508	8,607	194,508	8,607	194,508	8,607
Author fixed effect	Y	Y	Y	Y	Y	Y
Year-Month fixed effect	Y	Y	Y	Y	Y	Y

Table E.2 Effect of Financial Incentive on Stock Recommendation Quality (BHAR with One Month Holding Period)

Coefficients and standard errors are multiplied by 100 for presentation purposes. This table presents the results using BHAR with one month holding period as the dependent variable. The "CAPM" uses abnormal returns defined according to CAPM. "Fama-French 3" uses abnormal returns defined with respect to the Fama-French 3-factor model. "Carhart 4" uses abnormal returns defined with respect to the Carhart 4-factor model. "Long Position" demonstrates the results of buying stocks with disclosed long positions. "Short Position" demonstrates the results of buying stocks with disclosed short positions. The sample period is from 2004 to 2018. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level.

Panel B: BHAR as the Dependent Variable

Parameter	(1) Carhart 4 Long Position	(2) Carhart 4 Short Position	(3) Fama-French 3 Long Position	(4) Fama-French 3 Short Position	(5) CAPM Long Position	(6) CAPM Short Position
MarketPlaceStatus	-0.599***	-0.835	-0.566***	-0.706	-0.398*	-1.455
	(0.16)	(0.867)	(0.162)	(0.869)	(0.164)	(0.887)
MarketPlaceArticle	0.31*	-1.73*	0.307*	-1.674*	0.135	-1.5070.
	(0.141)	(0.794)	(0.142)	(0.796)	(0.144)	(0.812)
Control variables	Y	Y	Y	Y	Y	Y
Author fixed effect	Y	Y	Y	Y	Y	Y
Year-Month fixed effect	Y	Y	Y	Y	Y	Y

Table E.3 Effect of Financial Incentive on Stock Recommendation Quality (BHAR with One Week Holding Period)

Coefficients and standard errors are multiplied by 100 for presentation purposes. This table presents the results using BHAR with one week holding period as the dependent variable. The "CAPM" uses abnormal returns defined according to CAPM. "Fama-French 3" uses abnormal returns defined with respect to the Fama-French 3-factor model. "Carhart 4" uses abnormal returns defined with respect to the Carhart 4-factor model. "Long Position" demonstrates the results of buying stocks with disclosed long positions. "Short Position" demonstrates the results of buying stocks with disclosed short positions. The sample period is from 2004 to 2018. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level.

Panel B: BHAR as the Dependent Variable						
Parameter	(1) Carhart 4 Long Position	(2) Carhart 4 Short Position	(3) Fama-French 3 Long Position	(4) Fama-French 3 Short Position	(5) CAPM Long Position	(6) CAPM Short Position
MarketPlaceStatus	-0.092	-0.564	-0.08	-0.545	-0.059	-0.763
	(0.08)	(0.476)	(0.081)	(0.476)	(0.083)	(0.488)
MarketPlaceArticle	0.158*	-0.658	0.147*	-0.661	0.109	-0.626
	(0.071)	(0.436)	(0.071)	(0.436)	(0.073)	(0.447)
Control variables	Y	Y	Y	Y	Y	Y
Author fixed effect	Y	Y	Y	Y	Y	Y
Year-Month fixed effect	Y	Y	Y	Y	Y	Y

Table E.4 Effect of Financial Incentive on Stock Recommendation Quality (BHAR with One Day Holding Period)

Coefficients and standard errors are multiplied by 100 for presentation purposes. This table presents the results using BHAR with one day holding period as the dependent variable. The "CAPM" uses abnormal returns defined according to CAPM. "Fama-French 3" uses abnormal returns defined with respect to the Fama-French 3-factor model. "Carhart 4" uses abnormal returns defined with respect to the Carhart 4-factor model. "Long Position" demonstrates the results of buying stocks with disclosed long positions. "Short Position" demonstrates the results of buying stocks with disclosed short positions. The sample period is from 2004 to 2018. Standard errors are in parentheses. . indicates significance at the 10% confidence level. * indicates significance at the 5% confidence level, ** indicates significance at the 1% confidence level, and *** indicates significance at the 0.1% confidence level.

Panel B: BHAR as the Dependent Variable						
Parameter	(1) Carhart 4 Long Position	(2) Carhart 4 Short Position	(3) Fama-French 3 Long Position	(4) Fama-French 3 Short Position	(5) CAPM Long Position	(6) CAPM Short Position
MarketPlaceStatus	-0.03	0.208	-0.02	0.239	-0.016	0.104
	(0.035)	(0.244)	(0.035)	(0.245)	(0.036)	(0.247)
MarketPlaceArticle	-0.035	0.013	-0.045	0.037	-0.07*	0.043
	(0.031)	(0.223)	(0.031)	(0.224)	(0.032)	(0.227)
Control variables	Y	Y	Y	Y	Y	Y
Author fixed effect	Y	Y	Y	Y	Y	Y
Year-Month fixed effect	Y	Y	Y	Y	Y	Y

Appendix F: Professional Analysts Recommendations

The Nature of Analyst Expertise and the Distributions of Earnings Forecasts

Clement (1999) shows that forecasting accuracy increases with employer size (proxying for research resources) but declines with the number of industries and firms followed (proxying for specialization). Jacob et al. (1999) show the number of forecasts made in a forecasting interval (proxying for effort) and analyst aptitude (analyst-target alignments) are both positively related to forecast accuracy. Brown (2001) shows that a simple model based on analysts' past performance as a predictor of future accuracy performs as well as advanced models shown by Clement (1999) and Jacob et al. (1999).

Accurate long-term forecasts are important for company valuation as most terminal value estimations are based on assumptions about long-term growth. Dechow et al. (2000) show that analysts are often evaluated on the accuracy of their buy and sell recommendations and annual earnings forecasts, but not their long-term growth forecasts. Given the studies demonstrating mispricing due to optimistic long-term growth forecasts, identifying analysts who consistently issue more accurate long-term growth forecasts should also be attractive to investors.

Another area of study is the differences in decision-making between buy-side and sell-side analysts and the differences between experienced and less experienced analysts. Maines et al. (1997) show that experienced analysts are more effective than MBA students in processing the segmental disclosures in footnotes to companies' financial statements. Bouwman et al. (1995) find that buy-side analysts value the research reports of sell-side analysts by showing that buy-side analysts combine their independent analyses with sell-side analyses to make portfolio decisions. Cheng et al. (2006) find that fund managers place a higher weight on buy-side analyses when sell-side analyses are biased or when uncertainty about sell-side report bias increases.

Several publications examine characteristics that make forecasts more useful. Forecast timing, in addition to accuracy, plays an important role in the usefulness of forecasts. Cooper et al. (2001) find that performance rankings based on forecast timeliness are more informative than those based on abnormal trading volume and forecast accuracy. Cooper et al. (2001) also show that lead analysts, defined as analysts who provide timely forecasts, have a bigger effect on stock prices than follower analysts. Gleason & Lee (2003) show that lead analysts can make price adjustments faster than follower analysts. Mozes (2003) shows that forecast immediacy (the speed with which analysts react to significant public information changes) improves forecast accuracy as compared to outstanding forecasts, implying that forecast timeliness is important in price discovery. When evaluating the usefulness of analysts' forecasts and accuracy relative to the existing consensus, research should consider both accuracy and timeliness.

Bold forecasts have a greater effect on pricing and provide more useful private information than herding forecasts (Clement & Tse, 2005). Analysts who are confident are more likely to make bold forecasts, whereas less confident analysts are more likely to herd. Hong, Kubik, et al. (2000) show that career concerns may discourage boldness, which makes less experienced analysts are more likely to herd. Analysts with relatively good or poor prior performance are more likely to make bold forecasts (Clarke & Subramanian, 2006). Graham (1999) shows that analysts herd if analysts' private information is inconsistent with public information, suggesting that analysts are conservative in forecasting.

Studies examine the characteristics of analysts and investors that are linked to forecasting dispersion, which is calculated as the standard deviation of analysts' forecasts. If analyst disagreement reflects general investor disagreement, forecast dispersion is a proxy for investor uncertainty. Barron (1995) claims that trading can occur even if the level of dispersion does not vary because analysts' relative positions change from one forecast period to the next.

Market and Analyst Efficiency

Many studies find that analysts underreact to a wide range of accounting and other economic information. Analyst forecasting errors are on average in the same direction as previous revisions, implying that the revisions are incomplete. However, not all research suggests that analysts underreact to information. Easterwood & Nutt (1999) show that analysts are systematically optimistic in response to new information, underreacting to negative information and overreacting to positive information.

Researchers find that investors tend to underreact to analysts' forecast revisions (Gleason & Lee, 2003) and stock recommendations (Womack, 1996). Thus, there seems to be a delay in the response of investors to both company information and direct signals from analysts. Barber et al. (2001) argue that, while markets may be inefficient in terms of analysts' stock recommendations, exploiting such inefficiencies is unprofitable once transaction costs are taken into account.

The stock market is often slower than financial analysts in integrating new information. For instance, 40% of the market's underestimation of current accruals' transitory component can be explained by analysts' forecasts (Elgers et al., 2003). As a result, analysts are better at noticing the difference between the persistence of accrual and cash flow components of earnings than investors.

Analysts' Incentives and Behavioral Biases

The research establishes that forecast accuracy is directly associated with the possibility of being promoted, especially for less experienced analysts (Hong, Kubik, et al., 2000). Analysts are motivated to work hard to improve forecast accuracy. However, less experienced analysts are more likely to be fired for being bold (deviating from the consensus). Less experienced analysts have incentives to sacrifice some accuracy and timeliness in exchange for the safety of being close to the consensus.

Multiple studies suggest the existence of selection bias (Hayes, 1998; McNichols & O'Brien, 1997). Hayes (1998) shows that the incentives for analysts to follow companies they have favorable opinions about increases with the amount of stock that investors currently own, which should increase with the size of the company followed and the effect of analysts' recent buy recommendations. Hayes (1998) also shows that short-selling restrictions on the stock and the dispersion of ownership among investors will also increase asymmetry. Market inefficiency described in the behavioral finance literature could be explained by selection bias. Hayes (1998), Hong, Lim, et al. (2000), and McNichols & O'Brien (1997) show that the effect of low analyst coverage is most pronounced in stocks that have historically underperformed.

Several studies examine how employers' incentives to gain/keep underwriting business or generate trading commissions affect analysts' forecasts and recommendations. Dugar & Nathan (1995) and Lin & McNichols (1998) show that affiliated analysts, who work for companies that have existing underwriting connections, provide relatively optimistic recommendations. Cowen et al. (2006) show that analysts at retail brokerage companies are more optimistic than those who only work for institutional clients.

There are mixed results on whether the market adjusts analysts' forecasts for potential bias. Lin & McNichols (1998) find the market unravels analysts' incentives to give favorable recommendations due to underwriting relationships. Hayes & Levine (2000) show that the market does not unravel the effects of analysts' incentives to withdraw coverage of companies for which they have pessimistic opinions.

The evidence is mixed on whether psychological biases or economic incentives affect analysts' forecasts. Analyst incentives may cause analysts to underreact to the publicly available information. Trueman (1990) models underreaction as a function of analysts' incentives to hide the fact that they are unable to develop private information about a firm's prospects. Raedy et al. (2006) model an underreaction resulting from asymmetric loss functions, which creates incentives for analysts to alter their forecasts in a direction consistent with the analysts' present research reports' interpretation of firms' prospects. Mozes (2003) shows that forecasts released more quickly after a news event are associated with greater uncertainty and underreaction. Loeffler (1998) shows that underreaction and overconfidence, two types of bias discussed in the psychological literature, are likely to cause forecast rationality violations; however, these cognitive errors do not appear to be of major economic significance. Loeffler (1998) finds that analysts distort estimates when they believe their clients misunderstand the forecasts' true precision in order to adjust for investor perceptions of the forecasts.

REFERENCES

About Seeking Alpha. (2020). https://seekingalpha.com/page/about_us

Adebambo, B. N., & Bliss, B. (2015). The value of crowdsourcing: Evidence from earnings forecasts. *Working Paper*.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of twitter data. *Proceedings of the Workshop on Languages in Social Media*, 30–38.

Ammann, M., & Schaub, N. (2020). Do Individual Investors Trade on Investment-related Internet Postings? *Management Science*.

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

Asquith, P., Pathak, P. A., & Ritter, J. R. (2005). Short interest, institutional ownership, and stock returns. *Journal of Financial Economics*, *78*(2), 243–276. https://doi.org/10.1016/j.jfineco.2005.01.001

Author Testimonials. (2020). https://seekingalpha.com/page/testimonials

Avery, C. N., Chevalier, J. A., & Zeckhauser, R. J. (2015). The "CAPS" prediction system and stock market returns. *Review of Finance*, *20*(4), 1363-1381at least one comment on article publicati.

Barber, B., Lehavy, R., McNichols, M., & Trueman, B. (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns. *The Journal of Finance*, *56*(2), 531–563.

Barber, B. M., & Loeffler, D. (1993). The "Dartboard" Column: Second-Hand Information and

Price Pressure. *The Journal of Financial and Quantitative Analysis*, *28*(2), 273. https://doi.org/10.2307/2331290

Barber, B. M., & Lyon, J. D. (1997). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics*, *43*(3), 341–372.

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, *55*(2), 773–806.

Barber, B. M., & Odean, T. (2013). The behavior of individual investors. In *Handbook of the Economics of Finance* (Vol. 2, pp. 1533–1570). Elsevier.

Barron, O. E. (1995). Trading volume and belief revisions that differ among individual analysts. *Accounting Review*, 581–597.

Bartov, E., Faurel, L., & Mohanram, P. S. (2018). Can Twitter help predict firm-level earnings and stock returns? *The Accounting Review*, *93*(3), 25–57.

Berg, J., Forsythe, R., Nelson, F., & Rietz, T. (2008). Results from a dozen years of election futures markets research. *Handbook of Experimental Economics Results*, *1*, 742–751.

Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, *50*(2–3), 296–343.

Blume, M. E., & Stambaugh, R. F. (1983). Biases in computed returns: An application to the size effect. *Journal of Financial Economics*, *12*(3), 387–404.

Boehmer, E., Jones, C. M., Zhang, X., & Zhang, X. (2019). Tracking retail investor activity. *Available at SSRN 2822105*.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Bouwman, M. J., Frishkoff, P., & Frishkoff, P. A. (1995). The relevance of GAAP-based information: a case study exploring some uses and limitations. *Accounting Horizons*, *9*(4), 22.

Bradshaw, M. T., Wang, X., & Zhou, D. (2017). *Soft information in the financial press and analysts' recommendation revisions*. Working Paper, Boston College.

Brown, L. D. (2001). How important is past analyst forecast accuracy? *Financial Analysts Journal*, *57*(6), 44–49.

Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2015). Inside the "black box" of sell-side financial analysts. *Journal of Accounting Research*, *53*(1), 1–47.

Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2016). The activities of buy-side analysts and the determinants of their stock recommendations. *Journal of Accounting and Economics*, *62*(1), 139–156.

Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (2017). Stimulating Online Reviews by Combining Financial Incentives and Social Norms. *Management Science*. https://doi.org/10.1287/mnsc.2016.2715

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

Cabral, L., & Li, L. (2015). A dollar for your thoughts: Feedback-conditional rebates on eBay. *Management Science*, *61*(9), 2052–2063.

Campbell, J. L., DeAngelis, M. D., & Moon, J. R. (2019). Skin in the game: Personal stock holdings and investors' response to stock analysis on social media. *Review of Accounting*

Studies, *24*(3), 731–779.

Canina, L., Michaely, R., Thaler, R., & Womack, K. (1998). Caveat compounder: A warning about using the daily CRSP equal-weighted index to compute long-run excess returns. *The Journal of Finance*, *53*(1), 403–416.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, *52*(1), 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Chang, Y.-W., Hsieh, C.-J., Chang, K.-W., Ringgaard, M., & Lin, C.-J. (2010). Training and Testing Low-degree Polynomial Data Mappings via Linear SVM. *The Journal of Machine Learning Research*, *11*, 1471–1490. http://dl.acm.org/citation.cfm?id=1756006.1859899

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

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

Chen, H., Hu, Y. J., & Huang, S. (2019). Monetary Incentive and Stock Opinions on Social Media. *Journal of Management Information Systems*, *36*(2), 391–417.

Cheng, Y., Liu, M. H., & Qian, J. (2006). Buy-side analysts, sell-side analysts, and investment decisions of money managers. *Journal of Financial and Quantitative Analysis*, *41*(1), 51–83.

Clarke, J., & Subramanian, A. (2006). Dynamic forecasting behavior by analysts: Theory and evidence. *Journal of Financial Economics*, *80*(1), 81–113.

Clement, M. B. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics*, *27*(3), 285–303.

Clement, M. B., & Tse, S. Y. (2005). Financial analyst characteristics and herding behavior in forecasting. *The Journal of Finance*, *60*(1), 307–341.

Cookson, J. A., Engelberg, J., & Mullins, W. (2020). Echo Chambers. *Available at SSRN*.

Cooper, R. A., Day, T. E., & Lewis, C. M. (2001). Following the leader:: a study of individual analysts' earnings forecasts. *Journal of Financial Economics*, *61*(3), 383–416.

Cowen, A., Groysberg, B., & Healy, P. (2006). Which types of analyst firms are more optimistic? *Journal of Accounting and Economics*, *41*(1–2), 119–146.

Cox, D. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, *34*(2), 187–220.

Crane, A., & Crotty, K. (2020). How skilled are security analysts? *The Journal of Finance, 75*(3), 1629–1675.

Curtis, A., Richardson, V. J., & Schmardebeck, R. (2014). Investor attention and the pricing of earnings news. *Handbook of Sentiment Analysis in Finance, Forthcoming*.

Da, Z., & Huang, X. (2020). Harnessing the wisdom of crowds. *Management Science, 66*(5), 1847–1867.

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

Dechow, P. M., Hutton, A. P., & Sloan, R. G. (2000). The relation between analysts' forecasts of long-term earnings growth and stock price performance following equity offerings. *Contemporary Accounting Research*, *17*(1), 1–32.

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

Drake, M., Joos, P., Pacelli, J., & Twedt, B. (2020). Analyst forecast bundling. *Management Science*, *66*(9), 4024–4046.

Drake, M. S., Moon, J., Twedt, B. J., & Warren, J. (2020). Social Media Analysts and Sell-Side Analyst Research. *Available at SSRN 3456801*.

Drake, M. S., Thornock, J. R., & Twedt, B. J. (2017). The internet as an information intermediary. *Review of Accounting Studies*, *22*(2), 543–576.

Dugar, A., & Nathan, S. (1995). The effect of investment banking relationships on financial analysts' earnings forecasts and investment recommendations. *Contemporary Accounting Research*, *12*(1), 131–160.

Easterwood, J. C., & Nutt, S. R. (1999). Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism? *The Journal of Finance*, *54*(5), 1777–1797.

Elgers, P. T., Lo, M. H., & Pfeiffer Jr, R. J. (2003). Analysts' vs. investors' weightings of accruals in forecasting annual earnings. *Journal of Accounting and Public Policy*, *22*(3), 255–280.

Engelberg, J. (2008). Costly Information Processing: Evidence from Earnings Announcements. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1107998

Engelberg, J. E., & Parsons, C. A. (2011). The Causal Impact of Media in Financial Markets. *The Journal of Finance*, *66*(1), 67–97. https://doi.org/10.1111/j.1540-6261.2010.01626.x

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, *49*(3), 283–306.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

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

Farrell, M., Green, T. C., Jame, R., & Markov, S. (2018). The Democratization of Investment Research: Implications for Retail Investor Profitability and Firm Liquidity. *Available at SSRN 3222841*.

Farrell, M., Green, T. C., Jame, R., & Markov, S. (2020). The Democratization of Investment Research and the Informativeness of Retail Investor Trading. *Available at SSRN 3222841*.

Frankel, R., Kothari, S. P., & Weber, J. (2006). Determinants of the informativeness of analyst research. *Journal of Accounting and Economics*, *41*(1–2), 29–54.

Garcia, D. (2013). Sentiment during Recessions. *The Journal of Finance*, *68*(3), 1267–1300. https://doi.org/10.1111/jofi.12027

Gleason, C. A., & Lee, C. M. C. (2003). Analyst forecast revisions and market price discovery. *The Accounting Review*, *78*(1), 193–225.

Gomez, E., Heflin, F., Moon, J., & Warren, J. (2018). Crowdsourced financial analysis and information asymmetry at earnings announcements. *Georgia Tech Scheller College of Business Research Paper*, *18–45*.

Gomez, E., Heflin, F., Moon, J., & Warren, J. (2020). Can financial analysis on social media help level the playing field among investors? Evidence from Seeking Alpha. *SSRN Electronic Journal*, *18–45*.

Graham, J. R. (1999). Herding among investment newsletters: Theory and evidence. *The*

Journal of Finance, *54*(1), 237–268.

Green, T. C., Jame, R., Markov, S., & Subasi, M. (2014). Broker-hosted investor conferences. *Journal of Accounting and Economics*, *58*(1), 142–166.

Groysberg, B., & Healy, P. M. (2020). *Wall Street Research*. Stanford University Press.

Groysberg, B., Healy, P. M., & Maber, D. A. (2011). What drives sell-side analyst compensation at high-status investment banks? *Journal of Accounting Research*, *49*(4), 969–1000.

Guan, Y., Li, C., Lu, H., & Wong, M. H. F. (2019). Regulations and brain drain: Evidence from Wall Street star analysts' career choices. *Management Science*, *65*(12), 5766–5784.

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

Hayes, R. M. (1998). The impact of trading commission incentives on analysts' stock coverage decisions and earnings forecasts. *Journal of Accounting Research*, *36*(2), 299–320.

Hayes, R. M., & Levine, C. B. (2000). An approach to adjusting analysts' consensus forecasts for selection bias. *Contemporary Accounting Research*, *17*(1), 61–83.

Heimer, R. Z. (2016). Peer pressure: Social interaction and the disposition effect. *The Review of Financial Studies*, *29*(11), 3177–3209.

Henry, E. (2008). Are Investors Influenced By How Earnings Press Releases Are Written? *Journal of Business Communication*, *45*(4), 363–407. https://doi.org/10.1177/0021943608319388

Hong, H., Kubik, J. D., & Solomon, A. (2000). Security analysts' career concerns and herding of earnings forecasts. *The Rand Journal of Economics*, 121–144.

Hong, H., Lim, T., & Stein, J. C. (2000). Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies. *The Journal of Finance, 55*(1), 265–295. https://doi.org/10.1111/0022-1082.00206

Hu, T., & Tripathi, A. K. (2015). *The Performance Evaluation of Textual Analysis Tools in Financial Markets*. http://papers.ssrn.com/abstract=2661064

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

Hvidkjaer, S. (2008). Small trades and the cross-section of stock returns. *The Review of Financial Studies*, *21*(3), 1123–1151.

Jackson, A. R. (2005). Trade generation, reputation, and sell-side analysts. *The Journal of Finance*, *60*(2), 673–717.

Jacob, J., Lys, T. Z., & Neale, M. A. (1999). Expertise in forecasting performance of security analysts. *Journal of Accounting and Economics*, *28*(1), 51–82.

Jame, R., Johnston, R., Markov, S., & Wolfe, M. C. (2016). The value of crowdsourced earnings forecasts. *Journal of Accounting Research*, *54*(4), 1077–1110.

Jame, R., Markov, S., & Wolfe, M. (2017). Does crowdsourced research discipline sell-side analysts? *SMU Cox School of Business Research Paper*, *18–6*.

Kadan, O., Madureira, L., Wang, R., & Zach, T. (2008). Conflicts of interest and stock recommendations: The effects of the global settlement and related regulations. *The Review of Financial Studies*, *22*(10), 4189–4217.

Kaniel, R., Liu, S., Saar, G., & Titman, S. (2012). Individual investor trading and return patterns around earnings announcements. *The Journal of Finance*, *67*(2), 639–680.

Kelley, E. K., & Tetlock, P. C. (2017). Retail short selling and stock prices. *The Review of Financial Studies*, *30*(3), 801–834.

Khern-am-nuai, W., Kannan, K., & Ghasemkhani, H. (2018). Extrinsic versus Intrinsic Rewards for Contributing Reviews in an Online Platform. *Information Systems Research*. https://doi.org/10.1287/isre.2017.0750

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

Klein, D., & Manning, C. D. (2003). Accurate unlexicalized parsing. *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1,* 423–430.

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

Li, E. X., Ramesh, K., Shen, M., & Wu, J. S. (2015). Do analyst stock recommendations piggyback on recent corporate news? An analysis of regular-hour and after-hours revisions. *Journal of Accounting Research*, *53*(4), 821–861.

Lin, H., & McNichols, M. F. (1998). Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics*, *25*(1), 101–127.

Liu, Y., & Feng, J. (2021). Does money talk the impact of monetary incentives on user generated content contributions. *Information Systems Research*.

https://doi.org/10.1287/isre.2020.0971

Loeffler, G. (1998). Biases in analyst forecasts: cognitive, strategic or second-best? *International Journal of Forecasting*, *14*(2), 261–275.

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2015). Textual Analysis in Finance and Accounting: A Survey. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2504147

Lyon, J. D., Barber, B. M., & Tsai, C. (1999). Improved methods for tests of long-run abnormal stock returns. *The Journal of Finance*, *54*(1), 165–201.

Maines, L. A., McDaniel, L. S., & Harris, M. S. (1997). Implications of proposed segment reporting standards for financial analysts' investment judgements. *Journal of Accounting Research*, *35*, 1–24.

Malo, P., Sinha, A., Takala, P., Ahlgren, O., & Lappalainen, I. (2013). Learning the roles of directional expressions and domain concepts in financial news analysis. *Proceedings - IEEE 13th International Conference on Data Mining Workshops, ICDMW 2013*, 945–954. https://doi.org/10.1109/ICDMW.2013.36

Mao, Y., Wei, W., Wang, B., & Liu, B. (2012). Correlating S&P 500 stocks with Twitter data. *Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research - HotSocial '12*, 69–72. https://doi.org/10.1145/2392622.2392634

Mark (Shuai), M. (2020). *Two-Way Cluster-Robust Standard Errors and SAS code*. https://sites.google.com/site/markshuaima/home/two-way-clustered-standard-errors-and-sas-code?authuser=0

Mayew, W. J. (2008). Evidence of management discrimination among analysts during earnings conference calls. *Journal of Accounting Research*, *46*(3), 627–659.

McNichols, M., & O'Brien, P. C. (1997). Self-selection and analyst coverage. *Journal of Accounting Research*, *35*, 167–199.

Miller, G., & Fellbaum, C. (1998). *Wordnet: An electronic lexical database*. MIT Press Cambridge.

Mitchell, M. L., & Stafford, E. (2000). Managerial decisions and long-term stock price performance. *The Journal of Business*, *73*(3), 287–329.

Moldoveanu, M. C., & Martin, R. L. (2010). *Diaminds: Decoding the mental habits of successful thinkers*. University of Toronto Press.

Morris, S. (2017). Banks to Cut $1.2 Billion in Research Spending, McKinsey Says. *Bloomberg, June*, *21*, 2017.

Mozes, H. A. (2003). Accuracy, usefulness and the evaluation of analysts' forecasts. *International Journal of Forecasting*, *19*(3), 417–434.

Oh, C., & Sheng, O. (2011). *Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement*.

Oliveira, N., Cortez, P., & Areal, N. (2013). On the Predictability of Stock Market Behavior using StockTwits Sentiment and Posting Volume. *Portuguese Conference on Artificial Intelligence*, 355–365.

Qiao, D., Lee, S.-Y., Whinston, A. B., & Wei, Q. (2020). Financial Incentives Dampen Altruism in Online Prosocial Contributions: A Study of Online Reviews. *Information*

Systems Research, *31*(4), 1361–1375.

Qiao, D., Lee, S.-Y., Whinston, A. B., & Wei, Q. (2021). Mitigating the Adverse Effect of Monetary Incentives on Voluntary Contributions Online. *Journal of Management Information Systems*, *38*(1), 82–107.

Raedy, J. S., Shane, P., & Yang, Y. (2006). Horizon-Dependent Underreaction in Financial Analysts' Earnings Forecasts. *Contemporary Accounting Research*, *23*(1), 291–322.

Rees, L., Sharp, N., & Twedt, B. (2015). Who's heard on the Street? Determinants and consequences of financial analyst coverage in the business press. *Review of Accounting Studies*, *20*(1), 173–209.

SA Marketplace. (2020). https://seekingalpha.com/author/sa-marketplace#regular_articles

Seeking Alpha's First Millionaire. (2019). https://seekingalpha.com/article/4267212-seeking-alphas-first-millionaire

Seeking Alpha's Marketplace of guided investment communities. (2020). https://seekingalpha.com/marketplace?sort=price-high-to-low&source=marketplace_nav_right

Seeking Alpha Media Kit 01/06/2021. (2021). https://static.seekingalpha.com/uploads/pdf_income/sa_media_kit_01.06.21.pdf

Shashua, A. (2009). *Introduction to Machine Learning: Class Notes 67577* [Learning]. 109. http://arxiv.org/abs/0904.3664

Solomon, D. H. (2012). Selective Publicity and Stock Prices. *The Journal of Finance*, *67*(2), 599–638. https://doi.org/10.1111/j.1540-6261.2012.01726.x

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, *20*(5), 926–957.

Stickel, S. E. (1992). Reputation and performance among security analysts. *The Journal of Finance*, *47*(5), 1811–1836.

Surowiecki, J. (2004). The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. *Economies, Societies and Nations*, *296*, 5.

Tang, V. W. (2018). Wisdom of crowds: Cross-sectional variation in the informativeness of third-party-generated product information on Twitter. *Journal of Accounting Research*, *56*(3), 989–1034.

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

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

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

Top Financial and Stock Market News Sites. (2020). https://www.investopedia.com/articles/investing/112514/top-sites-latest-stock-market-news.asp

Trueman, B. (1990). On the incentives for security analysts to revise their earnings forecasts. *Contemporary Accounting Research*, *7*(1), 203–222.

Tumarkin, R., & Whitelaw, R. F. (2001). News or Noise? Internet Postings and Stock Prices. *Financial Analysts Journal*, *57*(3), 41–51. https://doi.org/10.2469/faj.v57.n3.2449

Wang, G., Wang, T., Wang, B., Sambasivan, D., Zhang, Z., Zheng, H., & Zhao, B. Y. (2015). Crowds on wall street: Extracting value from collaborative investing platforms. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 17–30.

Wang, J., Ghose, A., & Ipeirotis, P. G. (2012). Bonus, Disclosure, and Choice: What Motivates the Creation of High-Quality Paid Reviews? *No Journal Available*.

West, J. (2012). Financial advisor participation rates and low net worth investors. *Journal of Financial Services Marketing*, *17*(1), 50–66.

Who reads Seeking Alpha? (2020). https://seekingalpha.com/page/who_reads_sa

Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, *18*(2), 107–126.

Womack, K. L. (1996). Do Brokerage Analysts' Recommendations Have Investment Value? *The Journal of Finance*, *51*(1), 137–167. https://doi.org/10.1111/j.1540-6261.1996.tb05205.x

Zhang, Y., Swanson, P. E., & Prombutr, W. (2012). MEASURING EFFECTS ON STOCK RETURNS OF SENTIMENT INDEXES CREATED FROM STOCK MESSAGE BOARDS. *Journal of Financial Research*, *35*(1), 79–114. https://doi.org/10.1111/j.1475-6803.2011.01310.x

Zhao, L., Detlor, B., & Connelly, C. E. (2016). Sharing knowledge in social Q&A sites: The unintended consequences of extrinsic motivation. *Journal of Management Information Systems*, *33*(1), 70–100.