Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media

Hailiang Chen
Department of Information Systems, City University of Hong Kong

Prabuddha De
Krannert School of Management, Purdue University

Yu (Jeffrey) Hu
Scheller College of Business, Georgia Institute of Technology

Byoung-Hyoun Hwang
Krannert School of Management, Purdue University and Korea University Business School, Korea University

Social media has become a popular venue for individuals to share the results of their own analysis on financial securities. This paper investigates the extent to which investor opinions transmitted through social media predict future stock returns and earnings surprises. We conduct textual analysis of articles published on one of the most popular social media platforms for investors in the United States. We also consider the readers' perspective as inferred via commentaries written in response to these articles. We find that the views expressed in both articles and commentaries predict future stock returns and earnings surprises. (*JEL* G11, G12, G14)

"The issue for the pros is that the institution of [financial] analysis risks becoming de-professionalized. In the same way many jobs … became commoditized by the use of new tools or access to information, the era of DIY [do-it-yourself] financial analysis is dawning."[1]

This paper was formerly entitled: "Customers as Advisors: The Role of Social Media in Financial Markets." This paper has benefited greatly from comments by the referees, David Hirshleifer (the editor), Mara Faccio, Seoyoung Kim, Jose Liberti, Dong Lou, Tim Loughran, Ilya Polak, Jin Xu, and seminar participants at Jefferies (Quant Group), Korea University, University of Notre Dame, SAC Capital (Quant Group), University of Sydney, University of Technology-Sydney, the 3rd Annual Behavioral Finance Conference at Queen's University, the 2013 Boulder Summer Conference on Consumer Financial Decision Making, the 2013 CityU IS Research Summer Workshop, the 2013 Conference on Information Systems and Technology, and the 2011 IEEE International Workshop on Statistical Signal Processing. The work described in this paper was partially supported by a grant from the City University of Hong Kong [Project No. 7200306]. Send correspondence to Byoung-Hyoun Hwang, Krannert School of Management, Purdue University, West Lafayette, IN 47907, phone: 404-580-3978, email: bhwang@purdue.edu.

[1] Quote by Horace Dediu, former analyst, now blogger at Asymco, January 19, 2011.

doi:10.1093/rfs/hhu001 Advance Access publication February 4, 2014

Instead of relying on expert advice, consumers increasingly turn to fellow customers when choosing among products, a trend facilitated by the emergence of social media and the associated creation and consumption of user-generated content (e.g., Chen and Xie 2008; Gartner 2010). Deloitte (2007), for instance, finds that 82% of Internet consumers in the United States report to be directly influenced by peer reviews in their purchasing decisions. Empirical evidence suggests that the influence of peer-based advice, such as user-generated ratings on Yelp.com or Amazon.com, is increasing, while the influence of traditional advice sources, such as the *Michelin Guide* or *Consumer Reports*, is decreasing (Datamonitor 2010).

Peer opinions have also begun to play a greater role in financial markets. Traditionally the domain of professional forecasters, financial analysis is increasingly being performed and broadcast by investors themselves. As of 2008, nearly one in four adults in the U.S. reports to directly rely on investment advice transmitted via social media outlets (Cogent Research 2008) and regulators conclude that "social media is landscape-shifting," with its relevance to financial markets only growing (SEC 2012, p. 1). But do peer opinions actually impart value-relevant news? Or do they merely constitute "random chatter" in a task best left to professional analysts? Or worse, are some users taking advantage of the lack of regulation inherent in social media outlets and attempting to intentionally spread false "information" and mislead fellow market participants? The goal of this study is to assess the performance of investors-turned-advisors and to test whether investors can turn to their peers for genuine, useful investment advice.

To examine the role of peer-based advice, we extract user-generated opinions from Seeking Alpha (hereafter, SA; http://seekingalpha.com). Our choice of SA as the focus of this study was motivated by its popularity. As of August 2013, SA had 500,000 to 1 million unique visitors per day (comScore - ScorecardResearch) and, as such, was one of the biggest investment-related social media websites in the U.S. The website's goal is to provide "opinion and analysis rather than news, and [it] is primarily written by investors who describe their personal approach to stock picking and portfolio management, rather than by journalists" (Seeking Alpha 2012). The channels through which investors can voice their opinions and exchange investment ideas are twofold: (a) Users can submit opinion articles to SA, which are generally reviewed by a panel and subject to editorial changes. If deemed of adequate quality, these articles are then published on the SA website. (b) In response to these articles, any interested user can write a commentary, sharing his or her own view, which may agree or disagree with the author's view on the company in question. Over our 2005-2012 sample period, SA articles and SA commentaries were written by around 6,500 and 180,000 different users, respectively, and cover more than 7,000 firms.

To quantify and study the views disseminated through SA, we employ textual analysis. Specifically, we build on prior literature, suggesting that the frequency of negative words used in an article captures the tone of the report (e.g., Das and Chen 2007; Tetlock 2007; Tetlock, Saar-Tsechansky, and Macskassy 2008; Li 2008; Loughran and McDonald 2011; Davis, Piger, and Sedor 2012).[2]

To preview our findings, we observe that the fraction of negative words in SA articles and the fraction of negative words in SA comments both negatively predict stock returns over the ensuing three months.[3] In our analysis, we only consider comments posted in the first two days of article publication, which comprise roughly 80% of all comments posted. We move the beginning of the three-month holding period forward to ensure that the days over which abnormal returns and SA views are computed do not overlap. The predictability arising from SA comments is particularly evident when the number of comments over which the fraction of negative words is computed is relatively high. Our results are robust to the inclusion of control variables, reflecting analyst recommendation upgrades/downgrades, positive/negative earnings surprises, and the average fraction of negative words in Dow Jones News Services (DJNS) articles.

One interpretation of our findings is that views expressed in SA articles and SA commentaries contain pieces of value-relevant information, which, as of the article publication date, are not fully factored into the price. As investors subsequently adopt the SA view, either through the SA platform itself or through news that arrives following the article publication, prices gradually adjust. As a result, SA views predict future stock market performance. Such an interpretation would point to the usefulness of social media outlets as a source of genuine, value-relevant advice.

An alternative perspective is that SA views incite naïve investor reaction. That is, SA views reflect false or spurious information yet still cause investors to trade in the direction of the underlying articles and comments and move prices accordingly. Our methodology (skipping the first two days after article publication and focusing on a three-month horizon) and our observed lack of a return reversal are somewhat at odds with this interpretation. Moreover, whether followers of SA have enough capital by themselves to cause market prices to move in the manner that we document in this study is unclear.

To further differentiate between the "value relevance-" and the "naïve investor reaction-" interpretations of the data, we examine whether SA views predict subsequent earnings surprises. Specifically, we regress a measure of price-scaled earnings surprise on the fraction of negative words in SA articles from 30 days to three days prior to the earnings announcement, as well as the fraction of negative words in the corresponding SA comments.

[2] We use the negative word list compiled by Loughran and McDonald (2011) to characterize the views expressed in SA articles and SA commentaries.

[3] The views expressed in SA articles and SA commentaries also predict returns over a one-month, six-month, one-year, and three-year horizons.

Earnings surprise is the difference between the reported earnings-per-share (EPS) and the average of financial analysts' EPS forecasts issued/updated within 30 days prior to the earnings announcement.

If opinions expressed through SA were unrelated to firms' fundamentals, or if the information was spurious and already fully incorporated by financial analysts into their reported EPS forecasts, then no association should be observed between our earnings-surprise variable and our measure of peer-based advice. In contrast to this view, we find that the fraction of negative words in SA articles and comments strongly predict subsequent scaled earnings surprises. Given that earnings are unlikely to be caused by SA users' opinions, the earnings-surprise predictability suggests that the opinions expressed in SA articles and comments indeed provide value-relevant information (beyond that provided by financial analysts).

What are the mechanisms behind this predictability? Ex ante, it is unclear that social media platforms should work in the domain of financial markets. In particular, the openness and lack of regulation inherent in social media outlets implies that uninformed actors can easily spread erroneous "information" among market participants.[4] Moreover, it is not obvious what incentives truly informed actors would have to share their insights with others.

Several factors might play a role in making social media a valuable source of investment advice. First, users could derive significant utility from the attention and recognition they receive from posting opinions that subsequently are confirmed by the stock market. On occasions, SA contributors and their articles are referred to and discussed in prominent outlets such as *Forbes*, *WSJ-Marketwatch*, and *Morningstar* (press clips are available upon request), and based on anecdotal accounts, many users strive to become online celebrities. Second, each SA contributor earns $10 per 1,000-page views that his/her article receives.[5] An article deemed to be of particularly high quality by the SA editors earns the contributor at least $500 and potentially more depending on the number of page views the article receives. Concurrently, articles are reviewed by an editorial board. If SA editors are educated and if the crowd allocates more attention to authors that, historically, have produced good articles, this creates an incentive to share good advice. It would also discourage authors from posting uninformative content (for benign or not-so-benign reasons) as by doing so, authors risk being rejected by the editorial board; even if allowed

4 See Boehme, Danielsen, and Sorescu (2006), Frieder and Zittrain (2008), and Hanke and Hauser (2008) for related evidence.

5 In order to receive compensation, SA contributors have to give SA exclusive rights to their articles. More specifically, SA contributors are not allowed to re-post their SA articles on other for-free websites. Should SA contributors have their own subscriber-based blogs, they are only allowed to post headlines and an excerpt of no more than one-third of the total article on their blog; they have to include a link to the full SA article.

to publish, they may receive less attention and suffer both monetary and non-monetary consequences.[6] Third, social media platforms are unique in allowing users to directly interact with each other and to provide immediate and publicly visible feedback on the author's view on a company. This feature enables users to intervene and correct bad articles, which, on the assumption that the crowd is educated, increases the informativeness of social media outlets and further discourages the involvement of malignant contributors.

Finally, to the degree that SA users' reading an article and trading on it can have *some* price impact and expedite the convergence of market prices to what authors perceive to be the fair fundamental value, informed actors have an incentive to contribute to SA to publicize their investment ideas and to convince other investors to follow their investment approach.

The relative importance of some of the aforementioned mechanisms is difficult to assess empirically. In this study, we attempt to examine the relevance of the second and third mechanisms by utilizing data provided by SA. For each article published in the second half of 2012, our data contain the number of page views each article receives and the number of times each article is read-to-end. We construct a measure of article-level consistency, which equals one if a more positive (more negative) article, subsequently, is followed by positive (negative) abnormal returns, and zero otherwise. We then compute the average article-level consistency for each author across all the articles that the author publishes in the three-year period prior to 2012 and relate this variable to measures of attention that the author receives in the second half of 2012.

Our first analysis shows that both the number of page views, which directly impacts the level of monetary compensation the author receives from SA, and the number of times an article is read-to-end increase with the author's historical level of consistency. This pattern is in line with the argument that followers can differentiate between authors that offer historically good versus bad advice and the "popularity" of these authors' changes accordingly.

Our second analysis computes to what degree SA commentaries are of a different tone than the underlying SA article and we relate this measure of author/follower disagreement to the author's historical track record. Our evidence implies that followers disagree with authors more when the authors' articles have been inconsistent. For these historically inconsistent authors, our evidence also suggests that in instances where the tone of comments is in disagreement with the tone of the underlying article, it is the tone of the comments that, statistically speaking, more reliably predicts subsequent stock market performance.

In the end, all of the findings in this study point to the usefulness and value relevance of peer-based advice in the investment domain, and they hint at

[6] Relatedly, many authors maintain their own subscriber-based financial blogs. These authors would appear to have a genuine interest in producing consistent, high-quality research reports, which would increase their network of clients and paying subscribers.

the possibility that social media outlets specializing in financial markets may eventually mirror the development of other "bottom-up knowledge generators" such as Wikipedia and the way they have changed how information is produced, evaluated, and disseminated (Tyckoson et al. 2011). The popular press has broached this issue when reporting that social media outlets, through their growing influence among the investor population, are already creating a rivalry with traditional advice sources, such as professional sell-side analysts,[7] with far-reaching implications for financial market participants (SEC 2012).

Our study speaks to several lines of research. First, our paper relates to the literature on the usefulness of peer-based advice (e.g., Chevalier and Mayzlin 2006; Liu 2006; Chen and Xie 2008; Zhu and Zhang 2010). In particular, we provide initial evidence that social media outlets also play a valuable role in the domain of financial markets.

By showing that our measure of views expressed in SA articles and SA commentaries predicts earnings above and beyond those forecasted by sell-side analysts, we also add to the literature on professional forecasters (e.g., Womack 1996; Barber et al. 2006). Arguably, financial securities are complicated products and perhaps best analyzed by investment professionals. At the same time, it is plausible that large crowds (in our case, close to 200,000 users) sometimes possess insight that did not fully factor into the earnings forecasts of sell-side analysts (who number fewer than 1,000).[8] Moreover, financial analysts' jobs are fraught with built-in conflicts of interest and competing pressures (e.g., Daniel, Hirshleifer, and Teoh 2002).

Our study also contributes to the literature analyzing the media's effect on the stock market (e.g., Barber and Loeffler 1993; Huberman and Regev 2001; Busse and Green 2002; Tetlock 2007; Engelberg 2008; Tetlock, Saar-Tsechansky, and Macskassy 2008; Fang and Peress 2009; Engelberg and Parsons 2011; Dougal et al. 2012; Gurun and Butler 2012; Solomon 2012). Our study distinguishes itself from the aforementioned studies through its focus on a social media platform. Social media outlets are unique in the sense that they enable direct and immediate interaction among users. As alluded to above, these interactions, combined with the seeming intelligence of the "crowd," may be one of the primary reasons social media platforms are able to produce value-relevant content that is incremental to that revealed through traditional news channels.

Finally, we propose a new laboratory for investigating questions about social interactions and investing. Social interactions among investors and the

---

[7] Examples include: "Apple's 'Underdog' Analysts Outperform Wall Street From Helsinki, Caracas," Bloomberg, January 19, 2011; "Apple and Wall Street: Six quarters of lousy estimates," CNN, September 26, 2011.

[8] The notion is akin to the one proposed in a study on a popular TV game show called "Who Wants to Be a Millionaire?" (Surowiecki 2005). In this show, the participant is asked a question and given four answers to choose from. As a lifeline, the participant can ask the audience which of the four answers they think is the correct answer. The study finds that, when asked, the audience has the right answer 91% of times, presumably, not because the individuals in the audience are more knowledgeable than the participant, but because, as a collective entity, they know more than a single individual.

information so transmitted are generally unobservable to researchers. Recent studies rely on proxies such as geographic distance to capture word-of-mouth effects (e.g., Feng and Seasholes 2004; Hong, Kubik, and Stein 2004; Ivkovic and Weisbenner 2007).[9] Social media websites make information shared among investors accessible and, as such, pose an interesting setting to conduct further research on social interactions, information exchange and diffusion, and their implications for financial markets.[10]

By utilizing a social media outlet to examine the value relevance of peer opinions, our study is perhaps most closely related to those of Tumarkin and Whitelaw (2001), Antweiler and Frank (2004), and Das and Chen (2007), who examine how conversations on Internet message boards associate with stock returns. Tumarkin and Whitelaw (2001) detect no association; Das and Chen (2007, p. 1385) find "no strong relationship from sentiment to stock prices on average across the individual stocks," while Antweiler and Frank (2004) find a statistically significant, yet economically meaningless, association. Together, the results presented in these studies suggest that stock opinions transmitted through social media outlets do not predict stock returns. In sharp contrast, the stock return predictability documented here is both statistically significant and economically meaningful.

This difference in results may, in part, be explained by our broader sample: Tumarkin and Whitelaw (2001) study 73 Internet service companies from April 1999 to February 2000; Antweiler and Frank (2004) consider 45 large-cap companies in the year 2000; and Das and Chen (2007) study 24 tech-sector stocks from July 2001 to August 2001. In comparison, our analysis encompasses more than 7,000 companies and our sample period ranges from 2005 to 2012. More crucially, social media outlets have evolved dramatically since the late 1990s, providing a substantially greater and more meaningful channel through which users share information and ideas (e.g., Boyd and Ellison 2007; Chapman 2009). In particular, all of the aforementioned studies examine the usefulness of message boards, which, by design, are open, unstructured, and characterized by very short messages ("chatter"). SA and other related social media outlets represent the more recent venue of allowing investors to share the results of their own financial analysis with peers. The reports summarizing the results of their analyses are relatively long and similar in format to those of professional sell-side analysts. Fellow users can respond via commentaries, which themselves are substantially longer than the average message board post. Combined with the editorial board and some of the other aforementioned features, this setup is perhaps more conducive to a focused and structured debate.

---

[9] For an overview of the literature on social interactions and investing, see Hirshleifer and Teoh (2009) and Seasholes (2010), among others.

[10] For related evidence, see Giannini (2011), who studies how information about the stock market flows among users of Twitter.

1. Data

This section describes the sample construction and introduces our main variables of interest. Our study uses data collected from SA articles, SA commentaries, and DJNS articles, as well as financial analyst data from the Institutional Brokers' Estimate System (IBES), and financial-statement and financial market data from Compustat and the Center for Research in Security Prices (CRSP), respectively. The sample period is from 2005 to 2012 and is determined by the availability of SA data.[11]

1.1 Seeking alpha

As of August 2013, SA is one of the largest investment-related social media websites in the U.S. (comScore – ScorecardResearch). Articles submitted to SA are generally reviewed by a panel and are subject to editorial changes. The review process is intended to improve the quality of published articles without interfering with the author's original opinion. Authors are required to disclose their identity and, as of 2010, have to report their holdings on the stocks they discuss.

We downloaded all opinion articles that were published between 2005 and 2012 on the SA website. Specifically, we wrote a computer program to automate the process of downloading articles from SA and extracting relevant information from the downloaded HTML files. The program can directly access a MySQL database and store the extracted information in database tables.[12] The SA website has a separate section containing news announcements; these news announcements are not part of our analysis.

SA assigns a unique ID to each article. In addition, SA editors tag each article with one or more stock tickers prior to publication. Single-ticker articles focus solely on one stock, making it relatively easy to extract the author's opinion on that company. Multiple-ticker articles discuss more than one stock in the same article, rendering extraction of the author's various opinions for each of the tagged stocks difficult, if not impossible. We therefore focus our analysis on the 97,070 single-ticker articles, which comprise roughly one-third of all articles published on SA. The information we collect about each article includes the following items: article ID, title, main text, date of publication, author name, and stock ticker.

SA allows any interested investor to not only write and read articles, but also to post commentaries in response to an article. We download all commentaries written in response to the 97,070 single-ticker articles in our sample. Sixty

[11] SA was founded in 2003. However, there were only a total of 241 articles published from 2003 to 2004. There were 4,796 articles published in 2005, and the number has continued to increase substantially.

[12] One potential situation that we cannot rule out entirely is that SA editors remove articles that, ex post, turn out to be wrong. When re-downloading SA articles at different points in time, we do not notice that any of the articles from our previous download are missing, suggesting that ex post removals are not common. At the same time, it is difficult to reject this scenario with absolute certainty and we acknowledge that it might possibly play a role.

A "Negative" Article about Google (12 negative words, 494 total words, NegSA = 2.43%):

Does Google Uphold 'Do No Evil' with shareholders?

January 12, 2010 | about: GOOG

Author: Ravi Nagarajan (http://seekingalpha.com/author/ravi-nagarajan)
Article URL: http://seekingalpha.com/article/182037-does-google-uphold-do-no-evil-with-shareholders

As we discussed recently in an article on Ken Auletta's new book, *Googled: The End of the World as We Know It,* the story of Google's (GOOG) founding and astounding growth is one that has a secure place in the history books. A major part of Google's success has been attributed to its unique way of doing business. The motto "Do No Evil" has been enshrined into Google's core philosophy. Google has been positioned by its founders as more than just a business but as an institution that seeks to promote a better world for society.
This type of pronouncement from a corporation was always certain to bring about a great deal of skepticism. After all, Google is now a large corporation presumably seeking to maximize shareholder wealth. Or is it?

Wonderful Timing, Just Not For Shareholders
As the Wall Street Journal reminds us Monday, in early 2009 Google re-priced a large number of options at much lower strike prices. 7.6 million options with an average strike price of $522 were exchanged for an equivalent number exercisable at $308.57. This narrowly **missed** the low for the year of $282.75. Google now trades at just under $600.
Google's founders were supposedly influenced by Warren Buffett when they published an "owner's manual" shortly before Google's IPO. It is, therefore, even more surprising that management reacted to what proved to be a temporary share price **decline** by massively re-pricing options at the expense of Google's shareholders.

No Justification
Did Google's management believe that the share price **decline** was temporary and did not reflect a **decline** in intrinsic value? If so, how could a re-price of the options be justified? Eventually, the share price would recover to reflect intrinsic value and option holders would benefit even at the original strike price.
Or did Google's management believe that intrinsic value had **declined** and the share price accurately reflected the **decline**? If so, how could management possibly justify resetting the option strike price and providing employees and managers who presided over the **decline** in intrinsic value any benefit from a subsequent recovery?
The likely response to this **criticism** is that management "had no choice" because they had to "retain key employees". There are ample reasons for skepticism regarding such a claim. But even if the **concern** had merit, why use stock options to promote retention? Does the average recipient of Google options have any direct control over Google's share price or intrinsic value?
This sorry episode is only another reason to be highly skeptical of companies that use stock options as currency for paying employees. Other than for top management (who presumably are accountable for progress in overall corporate results and intrinsic value progress), options are a very **poor** way of aligning employee incentives with shareholder interests. Of course, this is even more true when management creates a "heads I win, tails you **lose**" situation by re-pricing options when the share price **declines**.

Disclosure: Author has no position.

Figure 1
Two sample articles from Seeking Alpha
This figure presents two articles published on Seeking Alpha. The first article is "negative" (the fraction of negative words is 2.43%: 12 out of 494); the second article is "positive" (the fraction of negative words is 0%: 0 out of 447).

percent of the commentaries are posted on the day of article publication, an additional 20% are posted on the next day, and the remaining 20% are posted sporadically over the ensuing weeks.[13] In our analysis, we focus on the 459,679 commentaries written within the first two days of article publication. We assume that most commentaries written in response to a single-ticker article pertain to that article and the company discussed in that article. The information we collect about each commentary includes the following items: article ID, comment ID,

[13] Based on our reading of a random sample of 200 commentaries, comments posted sporadically over the ensuing weeks generally do not bear much relevance to the article and the company in question.

A "Positive" Article about Google (0 negative word, 447 total words,*NegSA* = 0%):

Android: Potentially the Greatest Gaming Platform

June 1, 2009 | about: GOOG

Author: Bruce Everiss (http://seekingalpha.com/author/bruce-everiss)
Article URL: http://seekingalpha.com/article/140631-android-potentially-the-greatest-gaming-platform

With all this talk of Android here and elsewhere on the web, it is perhaps worth looking at what it is. Especially as it has the potential to very rapidly become one of the biggest gaming platforms.
Android is a Linux based operating system for smart phones championed by Google (GOOG). It is open source and is developed by the Open Handset Alliance, whose 47 members include nearly all the major organisations in the smartphone industry. Sony Ericsson (SNE), Toshiba (TOSBF.PK), LG, Samsung, Motorola (MOT), HTC, Garmin (GRMN), Intel (INTC), Nvidia (NVDA), ARM, Google, [[eBay]], Vodafone (VOD), Sprint Nextel (S), etc etc. So there are more major players behind it than there are behind all the other smartphone standards put together. So it has the makings of becoming a standard.
Android has also been implemented by users on a wide range of devices that it was not installed on by the manufacturer. This includes devices from Nokia (NOK), Dell (DELL), Asus and Motorola. This is possible because Android is open source. Expect users to implement it on just about every device that they can!
The Android Software Development Kit (SDK) is available for free download and works on a wide variety of platforms including Windows XP, Vista, Mac OS and Linux.
Android can use touch screens, still & video cameras, accelerometers, GPS and accelerated 3D graphics. It works with most media standards.
The application store is called Android Market. Initially everything was free, but since February 2009 it can handle paid for applications with developers getting 70% and carriers getting 30%.
Android is the new kid on the block when it comes to smartphones. However it already works amazingly well. Just look at an HTC Magic or Samsung i7500 to see just how amazingly well. Android has a very strong potential to end up beating competing smartphone systems from Nokia, Microsoft (MSFT), Blackberry (RIMM), Palm (PALM) and Apple (AAPL), and here's why:

- Because it is open source and the SDK is freely available, there will be a massive number of people developing for it. So there will very soon be more applications available for it than for the competitors.
- Handsets will be available from nearly every handset manufacturer. There will be a huge choice of such devices with different specifications and price points. Android will also be used on netbook devices.
- With the backing of Google there is already the huge array of Google applications that run on it. These make Android phones immensely useful even before you start downloading applications from other people.

This is exciting and important stuff, everybody involved in the game industry should be watching it very closely indeed.

Disclosure: No positions

Figure 1
Continued

main text, date the comment is made, and author name. Figures 1 and 2 provide sample SA articles and sample SA commentaries.

To extract authors' opinions, we build on prior literature, which suggests that the frequency of negative words used in an article captures the tone of the report (e.g., Das and Chen 2007; Tetlock 2007; Li 2008; Tetlock, Saar-Tsechansky, and Macskassy 2008; Loughran and McDonald 2011; Davis, Piger, and Sedor 2012).[14] We use the negative words list compiled by Loughran and McDonald (2011), which they designed for use in studies on financial markets (http://www.nd.edu/~mcdonald/Word_Lists.html). $NegSA_{i,t}$ is the average fraction of negative words across all single-ticker articles published on SA about company i on day t. $NegSA\text{-}Comment_{i,t}$ is the average fraction of negative words across all SA comments posted over days t to t+1 in response

[14] Following the literature, we focus on the fraction of negative words rather than the fraction of positive words as positive words are often negated to convey negative feelings (e.g., not perfect).

Two "Negative" Comments about Google:

1. Commenter: SC Investor (http://seekingalpha.com/user/1115868/comments) (4 negative words, 167 total words, *NegSA-Comment* = 2.40%):

Google's long term **problem** is the **declining** quality of its crown jewel - its browser. The bulk of Google's revenue still derives from advertising revenues around its browser which still maintains a market share of over 70%. This will **decline** long term partially due to mobil which is the author's point. The main issue for me is the Google search engine is becoming less and less usable due to the pre-paid ranking of search results. A prime example is my search for a restaurant's web site. Using Google search the restaurant's own web site can usually be found on page three of the search results. What are the first two pages? Yelp and twenty other rating sites that paid for their place in line. The point is I don't use Google's search nearly as often now. Where do I go? Less commercialized search engines or **worse** for Google - Facebook. Maybe I am overreacting to my personal experience but I have friends coming to the same conclusion.

2. Commenter: joshpritchard (http://seekingalpha.com/user/758651/comments) (15 negative words, 287 total words, *NegSA-Comment* = 5.23%):

@rubicon59

The trial has progressed quite a bit from then. It was **broken** into three phases: Copyright, Patents, then **Damages**. The copyright component, which would have been where Oracle had its chance to win an **injunction**, is over. The ruling is out. The jury found Google to have **infringed** by copying 9 lines of code, out of millions. That's it. They said Google did **infringe** the SSO of the API, but couldn't decide whether it was covered by Fair Use. Then Oracle brought a motion to the judge to decide on Fair Use as a matter of law. He **rejected** the motion yesterday.

The judge in the Oracle v Google case stated yesterday that the max **damages** for the 9 lines of Rangecheck code would be $150K (the max for statutory **damages**). Now the court is on the patent phase of the trial, with only 2 of the remaining 7 patents in suit still standing. All three court experts (Oracle, Google, and the Court have all provided one) say that the combined value of the 2 patents is ~5M at most. Even if Oracle were to win 3x treble for willful **infringement**, you're still looking at **damages** <$20M. An **injunction** *is* off the table. Oracle has almost definitely spent more on the case than they can hope to recoup in **damages**.

See groklaw.net for court filing and transcripts... it was just yesterday (Thurs) that Judge Alsup said explicitly that Oracle was going to get a max of $150K in **damages** for the copyright **infringement**, and suggested they settle instead of putting before the jury (which will almost certainly be a more cost effective move for Oracle at this point, though they'd **lose** a lot of face for their baseless case).

Figure 2
Sample comments from Seeking Alpha
This figure presents four selected comments made on the same article published on Seeking Alpha (Article URL: http://seekingalpha.com/article/578061-new-reasons-google-could-plummet-by-2013). The title of the article is "New Reasons Google Could Plummet By 2013."

to single-ticker articles about company i on day t, if there were any such comments, and zero otherwise. In our regression analysis, we include $NegSA\text{-}Comment_{i,t}$, as well as an indicator variable, $I(NegSA\text{-}Comment_{i,t})$, denoting whether there were any comments posted in response to SA articles discussing company i on day t.

In separate tests, we examine whether other word categories also predict future stock market performance. In particular, we count the fraction of words in SA articles and SA commentaries that fall into the positive, uncertain, litigious, strong modal, or weak modal word categories as defined by Loughran and McDonald (2011) and we correlate these variables with future abnormal returns (to be defined below). In short, we observe no reliable predictability for any of these word categories (results are available upon request).

1.2 Dow Jones News Service

To explore whether views expressed in SA articles and SA commentaries have an effect above and beyond news released in more traditional media outlets, we construct a measure of information revelation through DJNS articles. We

Two "Positive" Comments about Google:

3. Commenter: gmmpa (http://seekingalpha.com/user/2323881/comments) (3 negative words, 239 total words, *NegSA-Comment* = 1.26%):

As an investor I am not ready to give up on Google just yet. The Internet is still a very large expanding and evolving entity. It is the biggest change in human history since the printing press. It represents among other things the sum total of human knowledge and repesents a new way to store, access and present the information to everyone on the planet. It is not even close to being done. It is natural to believe there will be **problems**, legal issues, an governmental **interference** in its growth. This will not stop its growth. It is too important to mankind.

Like IBM, Microsoft, Intel, Apple, AT&T, MCI and dozens companies before it Google will figure it out and remain a player. It has a huge amount of free cash flow and needs a mature management team to figure out how best to deploy it wisely. It shows signs that it may be able to do that. Facebook is just another new player in the scheme of things and it has and will have the same growing **problems** as did the companies listed above. It needs to figure out how to add value to the sector to remain a long term player.

As an investor I have made money investing in everyone of these companies starting with IBM in the early 60s. It has been a fun ride. Google is a player and the ride is not yet over.

4. Commenter: XRTrader (http://seekingalpha.com/author/xrtrader/comments) (1 negative word, 171 total words, *NegSA-Comment* = 0.58%):

What?

1. GOOG can do $50 in EPS this next year, and has $140 in cash. At 600, it is trading at <10x PE ex cash.

2. Yes, it has eccentric founders who are dabbling in too many industries. But, no one can **doubt** the search business.

3. Android business continues to ramp, and will become one of the dominant business. This could become a cash cow. The mobility acq will aid with this.

4. Chrome, mobile, and cloud businesses are all huge possible avenues for growth.

In the article, you use a lot of broad language and generalities, but you dont mention the earnings power, cash position, and low valuation of GOOG. In fact, i think the issues you bring up (Gov attention, etc) are already baked into the stock price. When these lift, there will be multiple expansion. And, if the multiple expands to even 12 to 14 ex cash, you would be talking about 750 to 850 stock price. Thats not counting continued growth and the new markets.

Figure 2
Continued

access DJNS articles for the stocks covered by single-ticker SA articles via the Factiva database. Since DJNS articles are not tagged by company name or stock ticker, we formulate a search query to find matched news articles for each stock from 2005 to 2012. We start with each company's name as it appears in the CRSP database and require the CRSP company name to show up at least once in the first 50 words of the DJNS news article.[15] To improve the query performance, we adjust the CRSP company names to match Factiva's coding of company names. For example, we change "Svcs" to "Services," "Tech" to "Technology," and "Intl" to "International" in our queries. If a company changes its name during our sample period, we query all possible names and combine the search results.

Despite our best effort, the matching of SA articles with the corresponding DJNS articles is not perfect. In particular, if a DJNS article discusses Cadbury (a subsidiary of Kraft Foods), but does not mention Kraft Foods itself, our search query will treat the Kraft Foods SA article as not having a matching DJNS article. Similarly, if a DJNS article only uses the product/brand (e.g., Camry)

15 We observe similar results using 25-word and 100-word cutoff points.

Table 1
Descriptive statistics of Seeking Alpha and Dow Jones News Service articles

	Seeking Alpha (SA) Articles	Seeking Alpha (SA) Comments	Dow Jones News Service (DJNS) Articles
Panel A: Single-Ticker SA Articles, SA Comments and DJNS Articles			
Total # Stock tickers	7,422	5,031	4,507
Total # Articles (or Comments)	97,070	459,679	322,046
Avg. # Words per article	675	82	380
StDev. # Words per article	466	104	934
Avg. % Negative words	1.25%	1.75%	1.48%
StDev % Negative words	0.96%	2.74%	1.49%
Panel B: SA and DJNS Articles with Word Stem "*Earn*" and Corresponding SA Comments			
Total # Stock tickers	5,054	3,406	3,889
Total # Articles	45,239	200,546	100,403
Avg. # Words per article	741	79	455
StDev. # Words per article	520	101	836
Avg. % Negative words	1.20%	1.63%	1.49%
StDev % Negative words	0.88%	2.62%	1.20%

This table reports summary statistics for single-ticker *Seeking Alpha* (SA) articles, SA comments written in response to single-ticker SA articles and *Dow Jones News Service* (DJNS) articles. In Panel B, we report summary statistics for the subset of single-ticker SA articles with the word stem "*earn*" and the SA comments written in response to these SA articles, as well as DJNS articles with the word stem "*earn*." The sample period is 2005–2012.

without mentioning the underlying company (e.g., Toyota), we treat the Toyota SA article as not having a matching DJNS article. We acknowledge the noise in our matching procedure. At the same time, we suspect that most DJNS articles of relevance would mention the underlying company, which renders our DJNS variable still informative.

In addition to the matched company name, we collect the following information about each DJNS article: article title, main text, and date of publication. The DJNS variable, $NegDJNS_{i,t}$, is the average fraction of negative words across all articles published in the DJNS about company i on day t, if there were any such articles, and zero otherwise. In our regression analysis, we include $NegDJNS_{i,t}$, as well as an indicator variable, $I(DJNS_{i,t})$, denoting whether there were articles published in the DJNS about company i on day t.

Table 1 illustrates a few features of our data. The average length of an SA article is 675 words, which is longer than the average length of a DJNS article (380 words). The average length of comments posted in response to SA articles is 82 words. This length is meaningful and significantly longer than that of messages posted on Internet message boards which, according to Antweiler and Frank (2004, p. 1263), "is most frequently between 20 and 50 [words]." The average fraction of negative words used in SA articles is 1.25%; and the average fraction of negative words used in SA comments is 1.75%. In comparison, the average fraction of negative words used in DJNS articles is 1.48%. The correlation between *NegSA* and *NegSA-Comment* within the subset of observations with comments posted to an SA article is 0.170. In Section 3, we examine what factors determine the magnitude of the correlation between *NegSA* and *NegSA-Comment* and the degree to which readers appear to challenge the author's viewpoint on the company in question.

1.3 Abnormal returns and other variables

We obtain financial statement and financial market data from Compustat and CRSP, respectively. Following the literature, we compute abnormal returns as the difference between raw returns minus returns on a value-weighted portfolio of firms with similar size, book-to-market ratio, and past returns (Daniel et al. 1997). Because our main variable of interest, $NegSA\text{-}Comment_{i,t}$, is the average fraction of negative words across SA comments posted over days t to t+1, we compute three-month holding period returns from trading day t+3 to t+60, $ARet_{i,t+3,t+60}$. If the SA article is published on a non-trading day, we move the beginning of the three-month holding period forward to ensure that the days over which abnormal returns and SA views are computed do not overlap.

Our choice of a three-month holding period is motivated by the literature on financial analysts and whether their recommendations have investment value (Womack 1996). Womack considers one-month post-event returns, three-month post-event returns, and six-month post-event returns; we choose three months as the "middle ground" of these timeframes. In later analyses, we also report results for alternate holding periods.

Other variables include: $Volatility_{i,t}$, which is the sum of squared daily returns in the calendar month prior to day t; and $ARet_{i,t}$, $ARet_{i,t-1}$, $ARet_{i,t-2}$, and $ARet_{i,t-60,t-3}$, which are abnormal returns on day t, day t-1, day t-2, and cumulative abnormal returns over the three calendar months prior to day t, respectively.

We obtain data on sell-side analyst recommendations and earnings forecasts from the IBES detail recommendation file and the IBES unadjusted U.S. detail history file, respectively. The IBES recommendation file tracks each recommendation made by each analyst, where recommendations are standardized and converted to numerical scores ranging from 1 (strong buy) to 5 (strong sell). We use the recommendation file to compute the number of recommendation upgrades/downgrades for company i on day t ($Upgrade_{i,t}$, $Downgrade_{i,t}$). The IBES unadjusted detail history file tracks each EPS forecast made by each analyst. We use this dataset to compute our earnings surprise measure, which is the difference between the reported EPS and the average quarterly EPS forecast. In our regression analysis, we include two binary variables indicating whether a positive earnings surprise was announced ($PosES_{i,t}$) and whether a negative earnings surprise was announced ($NegES_{i,t}$).

Table 2 presents the descriptive statistics of the main variables. The mean and the median of our abnormal return measure are slightly negative, which is the result of larger firms outperforming smaller firms during our sample period and the use of a *value-weighted* portfolio return in our abnormal return calculation.

Table 3 reports the characteristics of the firms in our sample. The average market capitalization is $10.3 billion, the average book-to-market ratio is 0.640, the average one-year holding period return is 14%, the average analyst coverage is 10.9, and the average retail holdings are 26%. In comparison, the average

Table 2
Summary statistics: Firm/trading day level

	N	Mean	Std. Dev.	25th Pctl	50th Pctl	75th Pctl
$ARet_{i,t+3,t+60}$	40,946	−0.004	0.162	−0.079	−0.006	0.067
$ARet_{i,t}$	40,946	0.000	0.040	−0.010	0.000	0.010
$ARet_{i,t-1}$	40,946	0.000	0.041	−0.010	0.000	0.010
$ARet_{i,t-2}$	40,946	0.000	0.031	−0.009	0.000	0.009
$ARet_{i,t-60,t-3}$	40,946	0.004	0.241	−0.086	−0.008	0.073
$Upgrade_{i,t}$	40,946	0.030	0.167	0.000	0.000	0.000
$Downgrade_{i,t}$	40,946	0.034	0.180	0.000	0.000	0.000
$PosES_{i,t}$	40,946	0.060	0.238	0.000	0.000	0.000
$NegES_{i,t}$	40,946	0.016	0.126	0.000	0.000	0.000
$Volatility_{i,t}$	40,946	0.018	0.059	0.004	0.008	0.017

This table reports the summary statistics of the main variables. The observations are on a firm-day level. Abnormal returns ($ARet_i$) are company i's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market/past-return-characteristics; t is the day an article about company i is published on the *Seeking Alpha* website, or the ensuing trading day if the article is published on a non-trading day. $Upgrade_{i,t}$ and $Downgrade_{i,t}$ are the number of financial analysts upgrading and downgrading company i on day t. $PosES_{i,t}$ and $NegES_{i,t}$ are indicator variables denoting whether company i experienced a positive (negative) earnings surprise on day t. $Volatility_{i,t}$ is the sum of squared daily returns in the calendar month prior to day t.

Table 3
Summary statistics: Firm/calendar year level

	N	Mean	Std. Dev.	25th Pctl	50th Pctl	75th Pctl
Size	7,773	10,291	29,424	529	1,930	7,204
BM	7,773	0.640	1.080	0.274	0.470	0.760
Past Return	7,773	0.140	1.300	−0.200	0.070	0.310
Coverage	7,773	10.870	7.840	5.000	10.000	16.000
Retail Holdings	7,773	0.260	0.230	0.083	0.210	0.390

This table reports the summary statistics of various firm characteristics. The observations are on a firm-year level. Every year t (from 2005 to 2012), we compile a list of firms in our sample. We then compute the respective firms' characteristics as of December. *Size* is the firm's market capitalization in millions. *BM* is the firm's book-to-market ratio. *Past Return* is the firm's cumulative one-year return. *Coverage* is the firm's analyst coverage, which is set equal to zero if the firm is not covered by any analysts in that year. *Retail Holdings* is one minus the fraction of shares held by institutional investors.

firm in the full CRSP/Compustat sample from 2005 to 2012 has a market capitalization of \$3.3 billion, a book-to-market ratio of 0.820, one-year holding period returns of 7.7%, an analyst coverage of 5.5, and retail holdings of 45%. Compared to the average CRSP/Compustat firm, our average sample firm is, therefore, larger, has a higher market-to-book ratio, and has higher past stock returns.

2. Main Results

We organize our main analysis around the following regression specification:

$$ARet_{i,t+3,t+60} = \alpha + \beta_1 NegSA_{i,t} + \beta_2 NegSA\text{-}Comment_{i,t} + X\delta + \varepsilon_{i,t}. \qquad (1)$$

The dependent variable is our measure of abnormal returns, $ARet_{i,t+3,t+60}$, where i indexes firms and t denotes the day on which the article appears on the SA website or the ensuing trading day if the article is published on a non-trading day.

Our two key independent variables are: $NegSA_{i,t}$, which is the average fraction of negative words across all single-ticker articles published on SA about company i on day t, and $NegSA\text{-}Comment_{i,t}$, which is the average fraction of negative words across SA comments posted over days t to t+1 in response to single-ticker SA articles, if there were any such comments (and zero otherwise). In other words, the observations in our regression specification are on a firm-day level. That is, if, in a given week, there were 18 SA articles on Apple and the articles were all published on Monday, we would have one observation for Apple in that week; if the 18 articles were spread across Monday, Tuesday, and Wednesday, we would have three observations. In total, our regression analysis encompasses 40,946 firm-days (and as such 40,946 observations) with SA articles and with the data necessary to construct our dependent and independent variables.

X includes the following variables, all of which are described in Section 1: $I(SA\text{-}Comment_{i,t})$, $NegDJNS_{i,t}$, $I(DJNS_{i,t})$, $Upgrade_{i,t}$, $Downgrade_{i,t}$, $PosES_{i,t}$, $NegES_{i,t}$, $Volatility_{i,t}$, $ARet_{i,t}$, $ARet_{i,t-1}$, $ARet_{i,t-2}$, and $ARet_{i,t-60,t-3}$. X also contains year-month fixed effects. T-statistics are computed using standard errors clustered by firm and year-month to account for serial- and cross-correlation, as well as heteroscedasticity.

The regression results in Table 4 indicate that the views expressed on SA are generally confirmed by subsequent stock market performance. The coefficient estimate on $NegSA_{i,t}$, by itself, equals –0.379 (t-statistic = –2.24), suggesting that future abnormal returns are 0.379% lower when the fraction of negative words in SA articles is 1% higher. When including $NegSA\text{-}Comment_{i,t}$ as an additional independent variable, the coefficient estimate on $NegSA_{i,t}$ is –0.332 (t-statistic = –2.03). The coefficient estimate on $NegSA\text{-}Comment_{i,t}$ equals –0.194 (t-statistic = –3.44), which implies that future abnormal returns are 0.194% lower when the fraction of negative words in SA comments is 1% higher.

Our results in Table 4 are robust to the inclusion of variables reflecting earnings surprises and analyst upgrades/downgrades. Our results also hold whether we control for views in DJNS articles (column 3) or not (column 2). Including leads and lags of the average fraction of negative words in DJNS articles does not alter this observation (results are available upon request). In untabulated analyses, we experiment with controlling for the tone in WSJ articles and our results continue to hold. We obtain similar results when including day-of-the-week fixed effects. We also obtain similar results when including firm-fixed effects, albeit in this specification we require firms to be above the median in terms of number of days discussed on SA; this restriction ensures that we have meaningful within-firm variation (all results are available upon request).

The coefficient estimates on the control variables are generally consistent with expectations. The estimates on $Upgrade_{i,t}$ and $PosES_{i,t}$ are positive and the estimates on $Downgrade_{i,t}$ and $NegES_{i,t}$ are negative, albeit not statistically

Table 4
Seeking Alpha and abnormal returns

	(1)	(2)	(3)
$NegSA_{i,t}$	−0.379	−0.332	−0.320
	(−2.24)	(−2.03)	(−1.98)
$NegSA\text{-}Comment_{i,t}$		−0.194	−0.196
		(−3.44)	(−3.55)
$I(SA\text{-}Comment_{i,t})$		0.001	0.001
		(0.25)	(0.17)
$NegDJNS_{i,t}$			−0.254
			(−1.44)
$I(DJNS_{i,t})$			0.009
			(1.33)
$Upgrade_{i,t}$	0.003	0.003	0.003
	(0.59)	(0.60)	(0.50)
$Downgrade_{i,t}$	−0.005	−0.005	−0.005
	(−1.08)	(−1.06)	(−1.10)
$PosES_{i,t}$	0.0014	0.001	−0.002
	(0.38)	(0.35)	(−0.41)
$NegES_{i,t}$	−0.004	−0.004	−0.006
	(−0.44)	(−0.49)	(−0.66)
$Volatility_{i,t}$	−0.044	−0.043	−0.042
	(−0.52)	(−0.50)	(−0.49)
$ARet_{i,t}$	−0.068	−0.070	−0.071
	(−1.64)	(−1.68)	(−1.71)
$ARet_{i,t-1}$	−0.077	−0.077	−0.077
	(−2.00)	(−2.00)	(−2.01)
$ARet_{i,t-2}$	−0.021	0.022	−0.022
	(0.35)	(0.37)	(−0.38)
$ARet_{i,t-60,t-3}$	−0.021	−0.022	−0.022
	(−1.41)	(−1.42)	(−1.43)
# Obs.	40,946	40,946	40,946
Adj. R^2	1.20%	1.23%	1.24%

This table reports coefficient estimates from regressions of abnormal returns on measures of the views reflected in *Seeking Alpha* (SA) articles and comments. The sample period is 2005–2012. Abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market/past return-characteristics from t+3 to t+60, where t is the day of article appearance or the ensuing trading day if the article is published on a non-trading day. $NegSA_{i,t}$ is the average fraction of negative words across all articles published on SA about company i on day t. $NegSA\text{-}Comment_{i,t}$ is the average fraction of negative words across SA comments posted over days t to t+1 in response to the SA articles, if there were any such comments, and zero otherwise. $NegDJNS_{i,t}$ is the average fraction of negative words across all articles published in the DJNS about company i on day t, if there were any such articles, and zero otherwise. $I(SA\text{-}Comment_{i,t})$ and $I(DJNS_{i,t})$ are indicator variables denoting whether there were comments posted on SA articles and whether there were articles published in the DJNS. We include year-month fixed effects. Other independent variables are as described in Table 2. T-statistics are computed using standard errors clustered by firm and year-month and are reported in parentheses.

significantly so. The estimates on $ARet_{i,t}$ and $ARet_{i,t-1}$ are negative, consistent with the presence of a short-term reversal. The estimate on $NegDJNS_{i,t}$ is negative, but with a t-statistic of -1.44 it is not reliably different from zero. The difference in results compared to the literature (Tetlock 2007; Tetlock, Saar-Tsechansky, and Macskassy 2008) is due to the difference in the return window over which the value relevance of DJNS and SA is assessed. DJNS articles are news articles and, as such, can be expected to have more of an immediate impact on prices. SA articles, on the other hand, resemble analyst reports (both in terms of format and character) and reflect more of a medium- or long-term view. Given the focus of this study, we choose our return window to match the window typically employed in the literature on the investment value of analyst recommendations. We note that when shortening the return window to a week as in Tetlock (2007) or to a day as in Tetlock, Saar-Tsechansky, and Macskassy (2008) and when not skipping the first two days after article publication, the coefficient estimate on $NegDJNS_{i,t}$ becomes statistically significant.[16]

In general, SA articles and comments predict returns over various horizons. Figure 3 is a plot of the coefficient estimates on $NegSA_{i,t}$ and $NegSA\text{-}Comment_{i,t}$ and the associated 95% confidence intervals for the following holding periods: one month, three months, six months, one year, and three years. The coefficient estimates, generally, are reliably different from zero; they also tend to increase with the length of the holding period. This finding is somewhat distinct from the observation made in the financial analyst literature that most of the abnormal performance after a recommendation upgrade/downgrade accrues around the date of the recommendation change. One potential explanation for this difference in results is that despite SA's growing popularity, financial analysts receive substantially more attention among investors; their opinions are therefore incorporated into the market price at a faster pace. If the views reflected in social media outlets are indeed value relevant and if social media continues to grow in popularity, one may speculate that, in the future, more of the abnormal performance predicted through social media platforms will accrue around the initial article's publication date.

Much of the literature on financial analysts examines whether stocks followed by analysts that receive a recommendation upgrade subsequently outperform stocks that receive a recommendation downgrade. In a similar fashion, we focus on the subset of firms receiving a SA recommendation in the form of SA articles and test whether stocks discussed more favorably on SA subsequently outperform those that receive less favorable coverage.

While not the focus of this study, in separate analyses, we test whether the selection of stocks also is associated with abnormal stock market performance. Put differently, our previous tests show that stocks with SA articles that have a high fraction of negative words subsequently underperform those that have SA articles with a low fraction of negative words. What remains unanswered is whether the more negatively viewed securities actually earn negative abnormal returns, or if they, instead, simply earn less positive abnormal returns as SA contributors tend to cover stocks that, on average, earn abnormally high returns. An analogous statement would apply if SA contributors tended to cover stocks that, on average, earn abnormally low returns.

To examine this question, we re-estimate our regression equations on the full CRSP/Compustat sample and add an indicator variable that equals one if the stock is covered by SA on a particular trading day, and zero otherwise. The combination of the slope on the indicator variable and the slope on the fraction

16 The coefficient estimate on $NegDJNS_{i,t}$ is similar to the one reported in this study, but the standard error decreases noticeably.

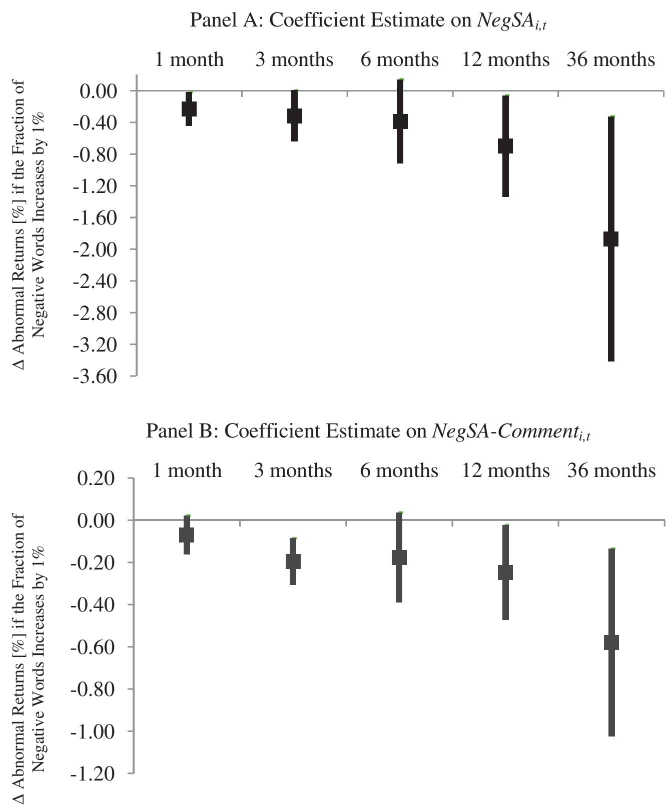

Figure 3
Seeking Alpha and abnormal returns over different holding periods
This figure reports coefficient estimates from regressions of abnormal returns on measures of the views reflected in *Seeking Alpha* (SA) articles and comments. The sample period is 2005–2012. Abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market/past return-characteristics. The horizons over which cumulative abnormal returns are computed are 1 month, 3 months, 6 months, 12 months, and 36 months. The regression equation is identical to the one in column (3) of Table 4. Here, we plot the coefficient estimates on $NegSA_{i,t}$ and $NegSA\text{-}Comment_{i,t}$ along with their corresponding 95% confidence intervals. Standard errors are clustered by firm and year-month.

of negative words in SA articles and commentaries provides an estimate of the *net* SA effect within a regression framework.

In untabulated results, we observe that the coefficient estimate on $NegSA_{i,t}$ ranges from –0.245 (t-statistic = –2.21) to –0.278 (t-statistic = –2.43), depending on the regression specification, indicating that future abnormal returns are 0.25% to 0.28% lower when the fraction of negative words in SA articles is 1% higher; the estimate on $NegSA\text{-}Comment_{i,t}$ ranges from –0.161 (t-statistic = –2.07) to -0.162 (t-statistic = –2.11), which indicates that future abnormal returns are 0.16% lower when the fraction of negative words in SA comments is 1% higher. In comparison, the estimate on the indicator variable ranges from 0.008 (t-statistic = 1.46) to 0.009 (t-statistic = 1.92) or less than 0.01% over a three-month holding period. The net SA effect, therefore, is dominated by the estimates on $NegSA_{i,t}$ and $NegSA\text{-}Comment_{i,t}$.

Our result holds within a calendar-time framework. In particular, we assign, at the end of each trading day t, stocks into quintile (quartile) portfolios based on $NegSA_{i,t}$. We skip two days and hold each stock in its respective portfolio for three months. We compute the spread between the daily average abnormal return for the bottom quintile (quartile) portfolio and the top quintile (quartile) portfolio, $aret_{long-short,NegSA,quintile}$ ($aret_{long-short,NegSA,quartile}$). The average $aret_{long-short,NegSA,quintile}$ in our sample period is 2.6 basis points (bps) (t-statistic = 2.87); the average $aret_{long-short,NegSA,quartile}$ is 2.4 bps (t-statistic = 2.89).

When repeating our analysis for $NegSA\text{-}Comment_{i,t}$, we observe that the average $aret_{long-short,NegSA-Comment,quintile}$ is 2.2 bps (t-statistic = 1.87); the average $aret_{long-short,NegSA-Comment,quartile}$ is 1.7 bps (t-statistic = 2.92).

These numbers compare well with those presented in Tetlock, Saar-Tsechansky, and Macskassy (2008). When forming quartile portfolios based on the news content of each firm's DJNS stories during the prior trading day and holding the portfolio for one trading day, Tetlock, Saar-Tsechansky, and Macskassy find that the corresponding long-short portfolio earns abnormal returns of 10.1 bps a day (the effect for WSJ stories is much weaker).[17] Compared to Tetlock, Saar-Tsechansky, and Macskassy, we evaluate the strength of an "opinion signal" over a relatively long time horizon; it appears reasonable to entertain the notion that the average *daily* effect of a SA opinion article over a period of three months should be noticeably smaller than the ensuing one-day effect of a DJNS news article.

Figure 4 depicts how much $1 invested in the aforementioned long-short strategy would have evolved over our sample period. Figure 4 illustrates that while the profits from going long (short) stocks with relatively positive (negative) SA views are stronger in some years than in others, they are not exclusive to a brief time period either, which suggests that our results hold more generally across time.

In the end, our finding that a measure of tone in SA articles and commentaries predicts future stock returns suggests that the opinions transmitted via this particular social media outlet impart value-relevant information. Generalizing this interpretation, our results suggest that investment-related social media websites provide a meaningful platform for people to help each other make more informed investment decisions; they also hint at the possibility that, going forward, these outlets will eventually mirror the development of other bottom-up knowledge generators such as Wikipedia and the way these knowledge generators have changed how information is produced and shared.

---

[17] See Tables II and III in Tetlock, Saar-Tsechansky, and Macskassy (2008).

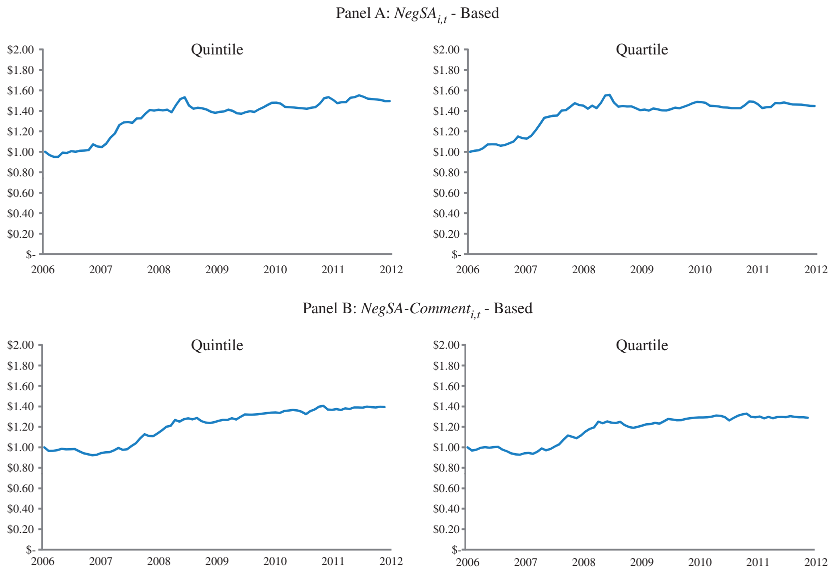

Figure 4
Calendar-time trading strategy based on Seeking Alpha
This figure depicts how \$1 invested in a simple calendar-time trading strategy would have evolved. The trading strategy is as follows: At the end of each trading day t, we assign stocks into quintile (quartile) portfolios based on the average fraction of negative words across all articles published on SA about company i on day t ($NegSA_{i,t}$); we also form quintile (quartile) portfolios based on the average fraction of negative words across SA comments posted over days t to t+1 in response to the SA articles ($NegSA\text{-}Comment_{i,t}$). We skip two days and hold each stock in its respective portfolio for three months. Based on the daily returns of a long-short portfolio, where we go long stocks in the bottom quintile (quartile) and short stocks in the top quintile (quartile), we plot how much \$1 would have grown/shrunk through calendar time.

In this regard, our study speaks to the growing literature in household finance and discussions about the degree to which retail investors make informed investment decisions. Much of the early literature suggests that retail investors are uninformed and taken advantage of by institutional investors; retail investors also have been found to suffer from various behavioral biases (e.g., Odean 1998; Barber and Odean 2000; Benartzi 2001). However, a growing body of more recent works detects patterns in the data which, taken together, imply that retail traders are skilled and able to identify and trade on novel, value-relevant information (e.g., Coval and Shumway 2005; Kaniel, Saar, and Titman 2008; Griffin et al. 2011; Kaniel et al. 2012; Kelley and Tetlock 2013). Social media may represent one channel through which retail investors, as a group, have become more informed.

2.1 Number of comments

In an attempt to better understand our results, we explore whether the return predictability depends on the number of comments over which aggregate views on a stock are computed. We focus on the subset of observations with SA comments and assign each observation its tercile rank based on the number of comments over which $NegSA\text{-}Comment_{i,t}$ is computed. We then re-estimate our main regression with the addition of this new tercile-rank variable and its interaction term with $NegSA\text{-}Comment_{i,t}$; the tercile-rank variable either equals zero, one, or two. The average number of comments across the bottom-tercile observations is 1.39; the average number of comments across the medium-tercile observations is 4.18; and the average number of comments across the top-tercile observations is 20.63.

As reported in Table 5, the regression produces a negative slope on the interaction term, suggesting that the predictive power of $NegSA\text{-}Comment_{i,t}$ for future abnormal stock returns is stronger when $NegSA\text{-}Comment_{i,t}$ is computed over many comments. The coefficient estimates on $NegSA\text{-}Comment_{i,t}$ and its interaction term are –0.120 (t-statistic = –1.74) and –0.196 (t-statistic = –2.18), respectively. These numbers imply that when the average fraction of negative words in SA comments is 1% higher, future abnormal returns for firms in the top tercile are 0.512% lower,[18] whereas future abnormal returns for firms in the bottom tercile are only 0.120% lower.[19]

2.2 Noise or value-relevant information?

Despite the lack of a reversal in future stock market performance (Figure 3) and the methodological steps we have taken (skipping the first two days after article publication and assessing abnormal returns over a medium-/long-run horizon), we cannot conclude with confidence whether stock opinions revealed through social media contain value-relevant news (predictability channel), or whether followers react to false or spurious publicity, which then moves market prices over the ensuing three months (clout channel). Both channels point to the importance of social media outlets, but with very different implications.

Table 6 provides additional evidence on this matter. In particular, we regress a firm's price-scaled quarterly earnings surprise on the fraction of negative words in SA articles published from 30 days to three days prior to the earnings announcement, the fraction of negative words in the corresponding SA comments, and various control variables. Earnings surprise is the difference between the reported quarterly EPS and the average EPS forecast across all analysts issuing estimates. We do not consider "stale" forecasts issued more than 30 days prior to the earnings announcement. SA views and earnings consensus forecasts are, thus, computed over the same horizon. We winsorize the absolute value of the price-scaled earnings surprise at the 99th percentile to mitigate the influence of outliers on our results. In total, our regression analysis encompasses 3,621 quarterly earnings announcements (and as such 3,621 observations) with the data necessary to construct our dependent and independent variables.

[18] Calculation for firms in the top tercile, i.e., $Rank(\#SA\text{-}Comment_{i,t}) = 2$: $(0.120+0.196\times 2)*1\%) = 0.512\%$.

[19] Calculation for firms in the bottom tercile, i.e., $Rank(\#SA\text{-}Comment_{i,t}) = 0$: $(0.120+0.196\times 0)*1\%) = 0.120\%$.

Table 5
Seeking Alpha, abnormal returns, and number of Seeking Alpha comments

	(1)	(2)
$NegSA_{i,t}$	−0.393	−0.381
	(−1.95)	(−1.92)
$NegSA\text{-}Comment_{i,t}$	−0.120	−0.122
	(−1.74)	(−1.77)
$NegSA\text{-}Comment_{i,t}$ * $Rank(\#SA\text{-}Comment_{i,t})$	−0.196	−0.196
	(−2.18)	(−2.21)
$Rank(\#SA\text{-}Comment_{i,t})$	0.009	0.009
	(1.93)	(2.05)
$NegDJNS_{i,t}$		−0.226
		(−1.05)
$I(DJNS_{i,t})$		0.007
		(0.98)
$Upgrade_{i,t}$	0.007	0.006
	(0.92)	(0.84)
$Downgrade_{i,t}$	−0.006	−0.006
	(−0.84)	(−0.84)
$PosES_{i,t}$	0.003	0.000
	(0.59)	(0.01)
$NegES_{i,t}$	−0.013	−0.015
	(−0.79)	(−0.82)
$Volatility_{i,t}$	−0.037	−0.036
	(−0.40)	(−0.40)
$ARet_{i,t}$	−0.118	−0.120
	(−2.13)	(−2.15)
$ARet_{i,t-1}$	−0.071	−0.071
	(−1.39)	(−1.40)
$ARet_{i,t-2}$	−0.103	−0.104
	(−1.38)	(−1.38)
$ARet_{i,t-60,t-3}$	−0.031	−0.031
	(−1.70)	(−1.71)
# Obs.	21,124	21,124
Adj. R^2	2.26%	2.27%

This table reports coefficient estimates from regressions of abnormal returns on measures of the views reflected in *Seeking Alpha* (SA) articles and comments. The sample period is 2005–2012. The sample consists of observations with SA comments. Abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market/past return-characteristics from t+3 to t+60, where t is the day of article appearance or the ensuing trading day if the article is published on a non-trading day. $NegSA_{i,t}$ is the average fraction of negative words across all articles published on SA about company i on day t. $NegSA\text{-}Comment_{i,t}$ is the average fraction of negative words across SA comments posted over days t to t+1 in response to the SA articles, if there were any such comments, and zero otherwise. $Rank(\#SA\text{-}Comment_{i,t})$ is the tercile rank of number of SA-comments posted; realizations of $Rank(\#SA\text{-}Comment_{i,t})$ either equal zero, one, or two. $NegDJNS_{i,t}$ is the average fraction of negative words across all articles published in the DJNS about company i on day t, if there were any such articles, and zero otherwise. $I(SA\text{-}Comment_{i,t})$ and $I(DJNS_{i,t})$ are indicator variables denoting whether there were comments posted on SA articles and whether there were articles published in the DJNS. We include year-month fixed effects. Other independent variables are as described in Table 2. T-statistics are computed using standard errors clustered by firm and year-month and are reported in parentheses.

The appealing feature of this setting is that a company's reported quarterly earnings are unlikely to be affected by opinions posted on the SA website. Sell-side analysts are also unlikely to revise their earnings forecasts *upward* in direct response to negative SA views (thereby inducing negative SA views to predict negative earnings surprises). SA views predicting future earnings surprises would, thus, point more towards the predictability channel.

Following Tetlock, Saar-Tsechansky, and Macskassy (2008), our independent variables include: (a) the average fraction of negative words across all

Table 6
Seeking Alpha and earnings surprises

	(1)	(2)	(3)	(4)	(5)
$NegSA_{i,t-30,t-3}$	−0.266	−0.232			
	(−2.45)	(−2.27)			
$NegSA\_EA_{i,t-30,t-3}$			−0.306	−0.267	−0.229
			(−2.54)	(−2.34)	(−1.72)
$I(NegSA\_EA_{i,t-30,t-3})$			0.001	0.001	0.005
			(0.72)	(0.64)	(1.62)
$NegSA\_NoEA_{i,t-30,t-3}$			−0.209	−0.193	0.020
			(−1.48)	(−1.43)	(0.18)
$I(NegSA\_NoEA_{i,t-30,t-3})$			0.000	0.001	0.002
			(0.15)	(0.36)	(0.61)
$NegSA\text{-}Comment_{i,t-30,t-3}$	−0.095	−0.094			
	(−1.72)	(−1.72)			
$I(SA\text{-}Comment_{i,t-30,t-3})$	−0.002	−0.002			
	(−1.43)	(−1.31)			
$NegSA\text{-}Comment\_EA_{i,t-30,t-3}$			−0.146	−0.144	−0.100
			(−2.25)	(−2.28)	(−1.59)
$I(SA\text{-}Comment\_EA_{i,t-30,t-3})$			0.000	0.000	0.000
			(0.45)	(0.49)	(0.05)
$NegSA\text{-}Comment\_NoEA_{i,t-30,t-3}$			−0.023	−0.019	−0.008
			(−0.30)	(−0.25)	(−0.08)
$I(SA\text{-}Comment\_NoEA_{i,t-30,t-3})$			−0.003	−0.003	0.001
			(−1.75)	(−1.66)	(0.25)
$NegDJNS_{i,t-30,t-3}$		−0.113			
		(−1.52)			
$I(DJNS_{i,t-30,t-3})$		−0.001			
		(−0.57)			
$NegDJNS\_EA_{i,t-30,t-3}$				−0.127	−0.073
				(−1.68)	(−1.50)
$I(DJNS\_EA_{i,t-30,t-3})$				0.001	0.001
				(0.78)	(0.99)
$NegDJNS\_NoEA_{i,t-30,t-3}$				−0.039	−0.110
				(−0.63)	(−0.78)
$I(DJNS\_NoEA_{i,t-30,t-3})$				−0.002	−0.001
				(−1.35)	(−0.31)
$Lagged(DependentVar_{i,t})$	0.158	0.158	0.157	0.156	0.380
	(1.24)	(1.25)	(1.26)	(1.27)	(3.47)
$ForecastDispersion_{i,t}$	0.224	0.227	0.238	0.241	−0.374
	(1.14)	(1.15)	(1.20)	(1.22)	(−11.94)
$ln(MarketCapital_{i,lagged})$	0.001	0.001	0.001	0.002	0.001
	(3.45)	(3.77)	(3.65)	(3.87)	(2.03)
$ln(Market/Book_{i,lagged})$	0.001	0.001	0.002	0.001	0.000
	(1.78)	(1.57)	(1.79)	(1.52)	(0.15)
$PastReturn_{i,t-30,t-3}$	0.005	0.004	0.004	0.004	0.09
	(0.37)	(0.33)	(0.35)	(0.34)	(0.95)
# Obs.	3,621	3,621	3,621	3,621	1,077
Adj. R^2	6.57%	6.82%	6.64%	7.35%	25.41%

We estimate a regression of price-scaled earnings surprise on measures of views reflected in *Seeking Alpha* (SA). The sample period is 2005–2012. Earnings surprise is the difference between reported quarterly EPS and the consensus EPS forecast across all analysts issuing earnings estimates from 30 to three calendar days prior to the earnings announcement. $NegSA_{i,t-30,t-3}$ is the fraction of negative words in SA articles about company i from 30 to three days prior to the earnings announcement on day t. $NegSA\text{-}Comment_{i,t-30,t-3}$ and $NegDJNS_{i,t-30,t-3}$ are the fraction of negative words in SA comments in response to the SA articles and the fraction of negative words in *Dow Jones News Service* articles about company i from 30 to three days prior to the earnings announcement; if there were no comments or no DJNS articles, these variables are set equal to zero. $I(SA\text{-}Comment_{i,t-30,t-3})$ and $I(DJNS_{i,t-30,t-3})$ are indicator variables denoting whether there were any SA comments posted and whether there were any DJNS articles written about company i. $NegSA\_EA_{i,t-30,t-3}$, $NegSA\text{-}Comment\_EA_{i,t-30,t-3}$, $I(SA\text{-}Comment\_EA_{i,t-30,t-3})$, $NegDJNS\_EA_{i,t-30,t-3}$, $I(DJNS\_EA_{i,t-30,t-3})$ are the analogues for articles that contain the word stem "*earn*". $NegSA\_NoEA_{i,t-30,t-3}$, $NegSA\text{-}Comment\_NoEA_{i,t-30,t-3}$, $I(SA\text{-}Comment\_NoEA_{i,t-30,t-3})$, $NegDJNS\_NoEA_{i,t-30,t-3}$, $I(DJNS\_NoEA_{i,t-30,t-3})$ are the analogues for articles that do not contain the word stem "*earn*". $I(NegSA\_EA_{i,t-30,t-3})[I(NegSA\_NoEA_{i,t-30,t-3})]$ is an indicator variable denoting whether there were any SA articles that contain [do not contain] the word stem "*earn*." $Lagged(DependentVar_{i,t})$ is the price-scaled earnings surprise (our dependent variable) from the previous quarter. $ForecastDispersion_{i,t}$ is the price-scaled standard deviation of analysts' EPS forecasts. $ln(MarketCapital_{i,lagged})$ is the logarithm of the market capitalization as of the quarterly earning's corresponding fiscal quarter end. $ln(Market/Book_{i,lagged})$ is the logarithm of the market-to-book ratio as of the most recent fiscal year end. $PastReturn_{i,t-30,t-3}$ is the cumulative stock market performance from 30 to three calendar days prior to the earnings announcement. We include year-month fixed effects. In Column (5), we only consider SA content published prior to analysts' most recent earnings forecast to compute the earnings surprise variable. T-statistics are computed using standard errors clustered by firm and year-month and are reported in parentheses.

single-ticker SA articles about company i from 30 days to three days prior to the earnings announcement ($NegSA_{i,t-30,t-3}$); (b) the average fraction of negative words across comments posted in response to these single-ticker SA articles, if there are any such comments, and zero otherwise ($NegSA\text{-}Comment_{i,t-30,t-3}$); (c) the average fraction of negative words across all DJNS articles about company i from 30 days to three days prior to the earnings announcement, if there are any such articles, and zero otherwise ($NegDJNS_{i,t-30,t-3}$); and (d) indicator variables denoting whether any comments were posted in response to SA articles and whether any DJNS article appeared about company i from 30 days to three days prior to the earnings announcement ($I(SA\text{-}Comment_{i,t-30,t-3})$ and $I(DJNS_{i,t-30,t-3})$).

Tetlock, Saar-Tsechansky, and Macskassy (2008) provide evidence that much of the predictability from WSJ and DJNS articles to subsequent earnings surprises is generated by articles that contain the word stem "*earn*." In additional tests, we examine whether this finding carries over to our setting. We separate our textual analysis-based variables by whether the underlying SA and DJNS articles contain the word stem "*earn*" or not ($NegSA\_EA_{i,t-30,t-3}$ versus $NegSA\_NoEA_{i,t-30,t-3}$; $NegSA\text{-}Comment\_EA_{i,t-30,t-3}$ versus $NegSA\text{-}Comment\_NoEA_{i,t-30,t-3}$; $NegDJNS\_EA_{i,t-30,t-3}$ versus $NegDJNS\_NoEA_{i,t-30,t-3}$). We do not require SA comments to contain the word stem "*earn*" as such requirement would dramatically lower the number of SA comments available for analysis. However, we do separate SA comments by whether they are made in response to SA articles that contain the word stem "*earn*" versus those that do not. If, in the period from 30 days to three days prior to the earnings announcement, there are no articles that contain [do not contain] the word stem "*earn*," the respective variables are set equal to zero and we include indicator variables denoting these cases.

Our control variables represent various firm characteristics: (a) lagged scaled earnings surprises; (b) price-scaled standard deviations of analysts' earnings-per-share forecasts; (c) the logarithm of market capitalization; (d) the logarithm of market-to-book ratios as of December of the calendar year prior to the earnings announcement; and (e) cumulative abnormal returns from 30 to three calendar days prior to the earnings announcement. As in all of our previous regression equations, we include year-month fixed effects and compute t-statistics using standard errors clustered by firm and year-month to account for serial- and cross-correlation, as well as heteroscedasticity.

As reported in columns (1)–(2) of Table 6, the coefficient estimate on $NegSA_{i,t-30,t-3}$ ranges from −0.232 (t-statistic = –2.27) to –0.266 (t-statistic = –2.45), depending on the set of control variables chosen, which suggests that when the fraction of negative words in SA articles is 1% higher, subsequent scaled earnings are between 0.232% and 0.266% further below market expectations as measured via financial analysts' forecasts. For reference, the mean scaled earnings surprise is –0.069% and the median is 0.052%.[20] The coefficient estimate on $NegSA\text{-}Comment_{i,t-30,t-3}$ ranges from −0.094 (t-statistic = –1.72) to –0.095 (t-statistic = –1.72).

When we separate our textual analysis-based variables by whether the underlying articles contain the word stem "*earn*" or not, we observe in columns (3)–(4) of Table 6 that more of the predictability from articles (comments) to subsequent earnings surprises comes from articles that contain the word stem "*earn*." The coefficient estimate on $NegSA\_EA_{i,t-30,t-3}$ ranges from −0.267 (t-statistic = –2.34) to −0.306 (t-statistic = –2.54) versus –0.193 (t-statistic = –1.43) to –0.209 (t-statistic = –1.48) for $NegSA\_NoEA_{i,t-30,t-3}$. Similarly, the coefficient estimate on $NegSA\text{-}Comment\_EA_{i,t-30,t-3}$ ranges from −0.144 (t-statistic = –2.28) to –0.146 (t-statistic = –2.25) versus –0.019 (t-statistic = –0.25) to –0.023 (t-statistic = –0.30) for $NegSA\text{-}Comment\ \_NoEA_{i,t-30,t-3}$.

In general, our results appear robust. In untabulated analyses, we re-compute the earnings surprise variable based on the *median* EPS forecast across all analysts issuing estimates and the results are similar. We also obtain qualitatively similar results when re-computing the earnings estimate component in the earning surprise variable based on a seasonal random walk; in this case, the coefficient estimates on $NegSA_{i,t-30,t-3}$ and $NegSA\text{-}Comment_{i,t-30,t-3}$ are –0.645 (t-statistic = –2.43) and −0.354 (t-statistic = –1.73), respectively.

The coefficient estimates on the control variables are generally consistent with expectations. In particular, earnings surprises tend to be more positive for larger, growth-oriented firms. This pattern is in line with the literature and consistent with the notion that these firms have stronger incentives to manage earnings upward and guide analyst earnings forecasts downward in an attempt to avoid negative earnings surprises (e.g., Richardson, Tuna, and Wu 2002; Skinner and Sloan 2002). Consistent with the literature, we also observe that the fraction of negative words in DJNS articles indicates subsequent negative earnings surprises.

While SA views and earnings consensus forecasts are computed over the same horizon, there remains the concern that SA articles are published closer to the earnings announcement date than analyst earnings forecasts and that, as a result, SA articles have a "timing advantage" based on their closer proximity to the earnings release. To get a sense of the impact of SA's timing advantage, we conduct the following test for each quarterly earnings announcement. Consider a quarterly earnings announcement i covered by x analysts and associated with y SA articles in the month prior to the announcement. For each of the x analysts, we compute the number of calendar days that pass from the analyst's most recent earnings forecast to the quarterly earnings announcement date. We then compute the average across the x analysts, $analyst\text{-}horizon_i$.

[20] The 10th percentile is –0.282%, the 25th percentile is –0.022%, the 75th percentile is 0.189%, and the 90th percentile is 0.462%.

Correspondingly, we compute the number of calendar days that pass from a SA article's publication date to the quarterly earnings announcement date for each of the *y* SA articles. We then compute the average across the y articles, $SA\text{-}horizon_i$. The average $analyst\text{-}horizon_i$ is 14.99 (median = 14.62), while the average $SA\text{-}horizon_i$ is 16.19 (median = 16.33), suggesting that SA articles actually have a slight timing disadvantage.

In column (5) of Table 6, we report results excluding SA articles that are published after the first estimate used to compute the consensus forecast. In particular, consider a firm with a quarterly earnings announcement on May 31, 2013 that is covered by two analysts: analyst A makes his/her last earnings forecast on May 20; analyst B makes his/her last earnings forecast on May 8. In this last specification, we only consider the SA articles that are published *prior to* May 8 (if there are any) to ensure that SA has no timing advantage relative to any of the analysts whose forecast is used to construct the consensus forecast. When doing so, the average $analyst\text{-}horizon_i$ becomes 12.37 (median = 12.00) and the average $SA\text{-}horizon_i$ becomes 24.53 (median = 25.75); that is, SA articles now have a distinct timing disadvantage relative to financial analysts. The estimates on $NegSA\_EA_{i,t-30,t-3}$ and $NegSA\text{-}Comment\_EA_{i,t-30,t-3}$ are now –0.229 (t-statistic = –1.72) and –0.100 (t-statistic = –1.59), respectively.

3. The Mechanisms

In the last part of this study, we examine the potential mechanisms behind social media's seeming value relevance. In Section 3.1, we provide evidence that followers tend to ignore authors whose articles have proven to be inaccurate. On the assumption that skill is persistent, this particular mechanism leads readers to assign less weight to articles of lower quality. Moreover, since authors presumably do not like to be ignored, this mechanism potentially motivates them to produce good and honest articles for both monetary and non-monetary reasons. In Section 3.2, we point to a second channel. Our evidence suggests that when articles have the potential to be of lower quality, readers disagree more through their commentaries, potentially dampening the impact of articles that are likely to be less accurate.

3.1 Author track record and following by readers

We obtain proprietary data from SA. The dataset contains the number of page views an article receives and the number of times it is read-to-end for each article published in the second half of 2012.[21] Our observations are now on an author level and we estimate a single cross-sectional regression equation. There are 308 SA authors who compose single-ticker articles in the second half

[21] The mechanism that allows SA to gauge the number of read-to-ends is as follows: Whenever the reader scrolls down and reaches the end of the article, the webpage automatically searches for and loads all comments made in response to the article. Whenever a search for comments is triggered, it is counted as one read-to-end.

of 2012 and for whom we have data to construct our measure of author track record.

The dependent variable in the first regression equation is author i's number of page views for articles published in the second half of 2012 ($PageView_i$). The dependent variable in the second regression equation is the number of times an article is read-to-end ($Read\text{-}to\text{-}End_i$). Both $PageView_i$ and $Read\text{-}to\text{-}End_i$ are expressed in thousands.

Our independent variable of primary interest is $Consistency_i$. For each single-ticker article published by author i over the three-year period prior to 2012, we compute the ensuing cumulative three-month abnormal return; as in our main analysis, we skip the first two days after article publication. An article is considered "bullish" if its fraction of negative words is below the median of its overall distribution; an article is considered "bearish" if its fraction of negative words is above the median. An article is *consistent* if a bullish article is followed by positive abnormal returns OR if a bearish article is followed by negative abnormal returns. $Consistency_i$ is the fraction of articles published by author i that are consistent.

In robustness checks, we experiment with an alternate measure, $Consistency\text{-}Portfolio_i$, which is computed as the average cumulative three-month abnormal returns following bullish articles (set equal to zero if the author has no bullish articles) minus the average cumulative three-month abnormal returns following bearish articles (set equal to zero if the author has no bearish articles). In short, we observe similar results to those obtained when using $Consistency_i$.[22]

We note that realizations of $Consistency_i$ display a mild degree of persistence. When sorting authors into terciles based on their $Consistency_i$ over the three-year period prior to 2012 and re-computing $Consistency_i$ for 2012, we observe that the average $Consistency_i$ of top-tercile authors is 5.7% higher than that of their bottom-tercile counterparts.

Other independent variables include the natural logarithm of the average number of words used by author i across articles published in the second half of 2012 ($Article\ Length_i$) and the corresponding average fraction of negative words ($NegSA_i$). We expect the coefficient estimate on $Article\ Length_i$ to be negative as long articles could deter readers, thereby lowering the page view. Long articles may also be more difficult to finish, thereby lowering the number of times an article is read-to-end. Whether more negative articles attract more reader attention than more positive articles is an empirical question.

We also include an indicator variable denoting whether author i maintains an investment-related blog ($I(Blog_i)$) and an indicator variable denoting

[22] In particular, the coefficient estimate on $Consistency\text{-}Portfolio_i$ equals 172.400 (t-statistic = 2.40) when the dependent variable is the number of page views and 97.686 (t-statistic = 2.54) when the dependent variable is the number of times an article is read-to-end. The reason the estimates on $Consistency\text{-}Portfolio_i$ are of different magnitude than those on $Consistency_i$ is due to the difference in the unit of the underlying variables. When gauging the effect of a one-standard deviation increase in $Consistency\text{-}Portfolio_i$ and $Consistency_i$ on the dependent variables, the estimates imply effects of similar economic significance.

Table 7
The mechanisms: Author-track record and following

	Page View (1)	Read-to-End (2)
Consistency$_i$	49.151	27.754
	(2.34)	(2.33)
Article Length$_i$	−27.663	−20.523
	(−1.94)	(−2.31)
NegSA$_i$	−570.508	−300.345
	(−0.73)	(−0.66)
I(Blog$_i$)	54.901	31.270
	(3.50)	(3.28)
I(Company$_i$)	−24.978	−12.598
	(−1.33)	(−1.19)
# Obs.	308	308
Adj. R^2	3.75%	4.18%

We estimate cross-sectional regressions of measures of an author's following on a measure of an author's past track record. In Column (1), the dependent variable is *PageView$_i$*, which is the number of page views [in thousands] across articles published by author i in the second half of 2012. In Column (2), the dependent variable is *Read-to-End$_i$*, which is the number of times [in thousands] an article published by author i in the second half of 2012 was read-to-end. Our main independent variable, *Consistency$_i$*, is computed as follows: for each article published on the SA website by author i over the past three years, we compute the ensuing cumulative three-month abnormal returns (skipping two days after article publication). An article is bullish if its fraction of negative words is below the median of its overall distribution; an article is bearish if its fraction of negative words is above the median. An article is defined to be consistent if a bullish article is followed by positive abnormal returns or if a bearish article is followed by negative abnormal returns. *Consistency$_i$* is the fraction of all single-ticker articles published by author i that are consistent. Other independent variables are: *Article Length$_i$*, which is the natural logarithm of the average number of words in articles published by author i in the second half of 2012, *NegSA$_i$*, which is the average fraction of negative words in single-ticker articles published by author i in the second half of 2012, *I(Blog$_i$)*, which is an indicator variable equal to one if author i posts on his/her SA profile the URL of an investment-related blog that he/she maintains., and *I(Company$_i$)*, which is an indicator variable equal to one if author i posts on his/her SA profile the company he/she works for. T-statistics are computed using White (1980) standard errors and are reported in parentheses.

whether author i reveals the name of the company name he/she works for ($I(Company_i)$). Having an investment blog could increase the author's incentive to produce consistent, high-quality research reports. Together, with disclosing the name of the company the author works for, it may also serve to enhance the author's credibility and, hence, his or her page views and the number of times his or her articles are read-to-end. *T*-statistics are computed using White (1980) standard errors.

The results are presented in Table 7. When the dependent variable is the number of page views, the coefficient estimate on *Consistency$_i$* is 49.15 (t-statistic = 2.34), which suggests that when the fraction of consistent articles increases by 10%, subsequent page views, on average, increase by 4,915. To put this number in perspective, such an increase would move the median article (in terms of number of page views) to the 56$^{\text{th}}$ percentile. Given SA's current compensation scheme of $10 per 1,000 page views, such an increase translates to an increase in monetary compensation of $49.15. When the dependent variable is the number of times an article is read-to-end, the coefficient estimate on *Consistency$_i$* equals 27.75 (t-statistic = 2.33), which suggests that when the fraction of consistent articles increases by 10%, the number of read-to-ends, on average, increases by 2,775. Such an increase would move the median article (in terms of read-to-ends) to the 55$^{\text{h}}$ percentile.

The analysis up to this point exploits across-author variation. We also examine whether we arrive at similar conclusions when exploiting within-author variation. Specifically, we look at how measures of incremental author popularity evolve with the performance of recent stock picks. Given the short sample period of our proprietary SA data, which renders the estimation of the time series effect of past article performance on author popularity difficult, our time series results should be interpreted with caution.

We construct the following variables. For each author i, we take his/her first article published during the second half of 2012 and save the corresponding page view (in thousands), $\textit{Baseline-PageView}_i$, and the number of times (in thousands) the article was read-to-end, $\textit{Baseline-Read-to-End}_i$. For the first single-ticker article and for each subsequent single-ticker article, we compute the cumulative three-month abnormal returns (skipping two days after article publication) and check whether the article is consistent. An article $j-1$ composed by author i is defined to be consistent ($\textit{Consistency}_{i,j-1}$=1) if a bullish article is followed by positive abnormal returns or if a bearish article is followed by negative abnormal returns, and is not consistent otherwise ($\textit{Consistency}_{i,j-1}$=0). An article is bullish if its fraction of negative words is below the median of its overall distribution; an article is bearish if its fraction of negative words is above the median.

We test how the rolling performance measure of recent stock picks $j-1$, $\textit{Consistency Recent Articles}_{i,j-1}$, relates to subsequent $\Delta \textit{PageView}_{i,j}$ (=$\textit{PageView}_{i,j}$-$\textit{Baseline-PageView}_i$) and to subsequent $\Delta\textit{Read-to-End}_{i,j}$ (=$\textit{Read-to-End}_{i,j}$-$\textit{Baseline-Read-to-End}_i$). We require article(s) $j-1$ to have been published at least three months prior to article j.[23] The standard errors are clustered by author and by year-month. Our regression equation is as follows:

$$\Delta Y_{i,j} = \alpha + \beta_1 \textit{Consistency Recent Articles}_{i,j-1} + \varepsilon_{i,j}. \tag{2}$$

In short, we find that our measures of incremental author popularity increase with the performance of recent stock picks. The coefficient estimate on $\textit{Consistency Recent Articles}_{i,j-1}$ is 14.911 (t-statistic = 3.37) when the dependent variable is based on the number of page views and 9.042 (t-statistic = 4.27) when the dependent variable is based on the number of times an article is read-to-end. The results are in line with our regression results exploiting across-author variation.[24]

One explanation for our findings is that intelligent followers can differentiate between authors that offer historically good versus bad advice and the popularity of these authors changes accordingly. This interpretation assumes that $\textit{Consistency}_i$ captures the author's ability to predict future abnormal

[23] The results are similar when shortening the window needed to evaluate the quality of recent stock picks from three months to two months or one month (results are available upon request).

[24] In robustness checks, we also re-estimate our regression equations on the panel of author/article-level observations. The results are similar to those presented in this paper and are available upon request.

stock market performance (predictability channel). However, if we assume that SA followers can significantly alter market prices, one may also argue that $Consistency_i$ captures an author's completely undeserved clout and the degree to which he/she can persuade uninformed, naïve readers to follow completely non-informative advice; the author's undeserved clout on SA, in turn, relates to his/her SA page views (completely undeserved clout channel).

Because we focus on a three-month-return window and because we choose to skip the first two days of article publication (during which most of the undeserved, clout-induced price impact may be expected to occur), our evidence, perhaps, more points towards the predictability channel than the completely undeserved clout channel. We also note that we observe weaker, but, overall, still similar results when re-computing the $Consistency_i$ measure skipping the entire first month (as opposed to skipping the first two days).[25]

In separate tests, we also re-compute $Consistency_i$ based on the degree to which positive (negative) articles prior to an earnings announcement are followed by positive (negative) earnings surprises. A company's reported quarterly earnings are unlikely to be affected by naïve investor following. One downside of this alternate measure of author track record is that not all authors compose articles shortly before an earnings announcement and our sample size drops noticeably. We observe qualitatively similar results under this alternate measure of $Consistency_i$.[26]

3.2 Author track record and author/reader interaction

To the degree that $Consistency_i$ captures an author's predictive ability, another potential mechanism behind social media's seeming value relevance becomes testable. The incremental predictive power of SA articles and SA commentaries documented in Section 2 arises from authors' and commentators' views not always being perfectly aligned. The correlation between *NegSA* and *NegSA-Comment* within the subset of observations with comments is 0.170. Here, we examine factors that determine the magnitude of the correlation between *NegSA* and *NegSA-Comment* and the degree to which readers adopt or, conversely, challenge the author's viewpoint. Although this exercise does not allow us to make causal statements, it enables us to better understand which scenarios are more likely to be associated with author/follower "disagreement."

To be consistent with our previous test, we estimate a single cross-sectional regression equation for the second half of 2012; our observations are again

[25] The coefficient estimate on $Consistency_i$ is 40.026 (t-statistic = 2.01) when the dependent variable is the number of page views and 22.655 (t-statistic = 1.97) when the dependent variable is the number of times an article is read-to-end.

[26] The coefficient estimate on $Consistency_i$ equals 37.066 (t-statistic = 1.87) when the dependent variable is the number of page views and 21.025 (t-statistic = 1.91) when the dependent variable is the number of times an article is read-to-end.

Table 8
The mechanisms: Author-track record and follower disagreement

	Coefficient Estimate (t-statistic)
$Consistency_i$	−0.004
	(−2.22)
$Article\ Length_i$	−0.002
	(−2.47)
$NegSA_i$	0.240
	(3.08)
$I(Blog_i)$	−0.000
	(−0.12)
$I(Company_i)$	0.001
	(0.94)
# Obs.	265
Adj. R^2	7.51%

We estimate a cross-sectional regression of author/follower disagreement on a measure of an author's past track record. The dependent variable is *$Disagreement_i$*, which is the average author/follower disagreement across all single-ticker articles j published by author i in the second half of 2012. Author/follower disagreement on article j is measured as the absolute difference between the fraction of negative words in article j and the average fraction of negative words across all comments posted in response to article j. Our main independent variables are as in Table 7. T-statistics are computed using White (1980) standard errors and are reported in parentheses.

on an author level. For each single-ticker article published by author i in the second half of 2012 that receives at least one comment, we measure author/follower disagreement as the absolute difference between the fraction of negative words in the article and the average fraction of negative words across all comments written in response to the article; as in our main analysis, we only consider comments written in the first two days of article publication. Our dependent variable is the average author/follower disagreement across articles published by author i. We include the same set of independent variables as in our previous test: *$Consistency_i$*, *$Article\ Length_i$*, *$NegSA_i$*, $I(Blog_i)$, and $I(Company_i)$. Again, t-statistics are computed using White (1980) standard errors.

The results are presented in Table 8. Several features are noteworthy. In line with expectations, the coefficient estimate on *$Consistency_i$* is negative (−0.004, t-statistic = –2.22), suggesting that historically more accurate (inaccurate) authors, subsequently, are faced with comments that disagree with the authors' articles to a smaller (greater) degree. The regression produces a positive slope on *$NegSA_i$*, implying that followers are more prone to challenge the author's viewpoint when the author writes a more negative article. The coefficient estimate on *$Article\ Length_i$* is negative. One possible interpretation of this finding is that longer articles are more detailed and more convincing; thereby they elicit less disagreement from the followers' side.

All qualifications made to the results on author following presented in the previous subsection also apply to the results on author/reader interaction. In particular, our results are weaker, but, overall, similar when using alternate measures of *$Consistency_i$*, including *$Consistency\text{-}Portfolio_i$*, *$Consistency_i$* when skipping the entire first month (as opposed to skipping the first two days) and earnings-based $Consistency_i$.[27] In addition, when exploiting within-author variation (as opposed to across-author variation), we find that our measure of incremental author-follower disagreement decreases with the performance of recent stock picks; our results also hold when estimating our regression equations on the panel of author/article-level observations (all results are available upon request).

In a tangential, yet related vein, we observe that the absolute difference between the average fraction of negative words in SA articles about stock i published on date t and the average fraction of negative words across all comments written in response to these articles over days t and t+1 weakly positively correlates with share turnover in stock i over days t and t+1 (ρ = 0.02; p-value = 0.01). Share turnover also increases with the standard deviation in the fraction of negative words across the comments (ρ = 0.35; p-value $<$ 0.01). To the degree that trading intensity increases with disagreement, the positive correlations suggest that author/follower disagreement and follower disagreement are accompanied with disagreement among investors in general.

Our final test provides indirect evidence on the possibility that readers, through their interaction, can help improve articles composed by historically inaccurate authors. We re-estimate our main regression equation of subsequent abnormal stock returns on $NegSA_{i,t}$ and $NegSA\text{-}Comment_{i,t}$, but we now focus on the subset of firm-day observations for which (1) there is disagreement about the single-covered firm between author and commentators AND (2) the SA author has a "poor track record." Author and commentators are defined to disagree if $NegSA_{i,t}$ is below the median and $NegSA\text{-}Comment_{i,t}$ is above the median of its distribution in year t OR if $NegSA_{i,t}$ is above the median and $NegSA\text{-}Comment_{i,t}$ is below the median of its distribution in year t. To gauge an author's track record, we compute each author i's $Consistency_{i,t-3,t-1}$ over the previous three years. An author is defined to have a "poor track record" as of year t if his/her $Consistency_{i,t-3,t-1}$ is below the median of its distribution.

The results presented in Table 9 show that the regressions produce strong negative slopes on $NegSA\text{-}Comment_{i,t}$, all of which are statistically significant at the 1% level. In two out of three regressions, the slope on $NegSA_{i,t}$ is large by economic standards; however, in all cases, the estimates lack statistical significance. At least from a statistical perspective then, Table 9 suggests that when authors have a poor track record and when authors and commentators disagree, it is the tone of the comments that more reliably predicts subsequent stock market performance.

[27] The coefficient estimates equal –0.007 (t-statistic = –1.35), –0.006 (t-statistic = –1.61), and –0.005 (t-statistic = –1.55).

Table 9
The mechanisms: Predictability when author-track record is poor and followers challenge the author

	(1)	(2)	(3)
$NegSA_{i,t}$	−0.013	−0.687	−0.664
	(−0.03)	(−1.21)	(−1.16)
$NegSA\text{-}Comment_{i,t}$		−0.560	−0.557
		(−3.30)	(−3.27)
$NegDJNS_{i,t}$			−0.611
			(−1.15)
$I(DJNS_{i,t})$			0.019
			(1.20)
$Upgrade_{i,t}$	0.017	0.015	0.014
	(0.78)	(0.66)	(0.60)
$Downgrade_{i,t}$	0.013	0.012	0.012
	(0.71)	(0.70)	(0.63)
$PosES_{i,t}$	−0.013	−0.010	−0.017
	(−1.01)	(−0.85)	(−1.21)
$NegES_{i,t}$	−0.050	−0.046	−0.052
	(−1.37)	(−1.32)	(−1.39)
$Volatility_{i,t}$	−0.026	−0.026	−0.026
	(−0.86)	(−0.88)	(−0.84)
$ARet_{i,t}$	−0.099	−0.099	−0.109
	(−0.76)	(−0.76)	(−0.83)
$ARet_{i,t-1}$	−0.123	−0.124	−0.123
	(−0.64)	(−0.65)	(−0.65)
$ARet_{i,t-2}$	−0.128	−0.136	−0.132
	(−0.90)	(−0.97)	(−0.94)
$ARet_{i,t-60,t-3}$	−0.017	−0.016	−0.017
	(−0.55)	(−0.54)	(−0.56)
# Obs.	2,439	2,439	2,439
Adj. R^2	1.12%	1.48%	1.49%

This table reports coefficient estimates from regressions of abnormal returns on measures of the views reflected in *Seeking Alpha* (SA) articles and comments. The sample period and the regression equations mirror Table 4, but we now focus on the subset of firm-day observations for which there is disagreement about the single-covered firm between SA author and commentators AND the SA author has a "poor track record." Author and commentators are defined to disagree if $NegSA_{i,t}$ is below the median and $NegSA\text{-}Comment_{i,t}$ is above the median of its distribution in year t OR if $NegSA_{i,t}$ is above the median and $NegSA\text{-}Comment_{i,t}$ is below the median of its distribution in year t. To gauge an author's track record, we compute each author i's $Consistency_{i,t-3,t-1}$ over the previous three years. An author is defined to have a "poor track record" as of year t if his/her $Consistency_{i,t-3,t-1}$ is below the median of its distribution. Otherwise, all variables are as described in Tables 2 and 4. We include year-month fixed effects. T-statistics are computed using standard errors clustered by firm and year-month and are reported in parentheses.

4. Conclusion

The Internet has become increasingly popular both as a venue to place trades and as a source of information. Da, Engelberg, and Gao (2011), for instance, provide evidence of a strong link between aggregate search frequency of stock tickers in Google and trading by retail investors. This study examines how views expressed on a popular social media website for investors pertain to security prices. We find that the opinions revealed on this website strongly predict future stock returns and earnings surprises. The predictability holds even after controlling for the effect of traditional advice sources, such as financial analysts and news media. Together, our findings point to the usefulness of peer-based advice in financial markets.

References

Antweiler, W., and M. Z. Frank. 2004. Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance* 59:1259–94.

Barber, B. M., R. Lehavy, M. McNichols, and B. Trueman. 2006. Buys, holds, and sells: The distribution of investment banks' stock ratings and the implications for the profitability of analysts' recommendations. *Journal of Accounting and Economics* 41:87–117.

Barber, B. M., and D. Loeffler. 1993. The 'dartboard' column: Second-hand information and price pressure. *Journal of Financial and Quantitative Analysis* 28:273–84.

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

Benartzi, S. 2001. Excessive extrapolation and the allocation of 401(k) accounts to company stock. *Journal of Finance* 56:1747–64.

Boehme, R. D., B. R. Danielsen, and S. M. Sorescu. 2006. Short-sale constraints, differences of opinion, and overvaluation. *Journal of Financial and Quantitative Analysis* 41:455–87.

Boyd, D. M., and N. B. Ellison. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication* 13:210–30.

Busse, J. A., and T. C. Green. 2002. Market efficiency in real time. *Journal of Financial Economics* 65:415–37.

Chapman, C. 2009. The history and evolution of social media. Available from: http://www.webdesignerdepot.com/2009/10/the-history-and-evolution-of-social-media/.

Chen, Y., and J. Xie. 2008. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science* 54:477–91.

Chevalier J. A., and D. Mayzlin. 2006. The effect of Word of Mouth on sales: Online book reviews, *Journal of Marketing Research* 43:345–54.

Cogent Research. 2008. Social media's impact on personal finance & investing. Available from: http://www.cogentresearch.com.

Coval, J. D., and T. Shumway. 2005. Do behavioral biases affect prices? *Journal of Finance* 60:1–34.

Da, Z., J. Engelberg, and P. Gao. 2011. In search of attention. *Journal of Finance* 66:1461–99.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52:1035–58.

Daniel, K., D. Hirshleifer, and S. H. Teoh. 2002. Investor psychology in capital markets: Evidence and policy implications. *Journal of Monetary Economics* 49:139–209.

Das, S. R., and M. Y. Chen. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science* 53:1375–88.

Datamonitor. 2010. Social media in financial services: The customer as the advisor.

Davis, A. K., J. M. Piger, and L. M. Sedor. 2012. Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research* 29:845–68.

Deloitte. 2007. Most consumers read and rely on online reviews; companies must adjust.

Dougal, C., J. Engelberg, D. Garcia, and C. Parsons. 2012. Journalists and the stock market. *Review of Financial Studies* 25:639–79.

Engelberg, J. 2008. Costly information processing: Evidence from earnings announcements. SSRN eLibrary.

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

Fang, L., and J. Peress. 2009. Media coverage and the cross-section of stock returns. *Journal of Finance* 64:2023–52.

Feng, L., and M. Seasholes. 2004. Correlated trading and location. *Journal of Finance* 59:2117–44.

Frieder, L., and J. Zittrain. 2008. Spam works: Evidence from stock touts and corresponding market activity. *Hastings Communications & Entertainment Law Journal* 30:479–520.

Gartner. 2010. User survey analysis: Consumer marketing using social network analysis.

Giannini, R. C. 2011. Twitter: An investigation of the impact of network communication. SSRN eLibrary.

Griffin, J. M., J. Harris, T. Shu, and S. Topaloglu. 2011. Who drove and burst the tech bubble? *Journal of Finance* 66:1251–90.

Gurun, U. G., and A. W. Butler. 2012. Don't believe the hype: Local media slant, local advertising, and firm value. *Journal of Finance* 67:561–98.

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

Hirshleifer, D., and S. H. Teoh. 2009. Thought and behavior contagion in capital markets. In *Handbook of financial markets: Dynamics and evolution*. Eds. T. Hens and K. R. Schenk-Hoppé. Elsevier/North-Holland.

Hong, H., J. D. Kubik, and J. C. Stein. 2004. Social interaction and stock-market participation. *Journal of Finance* 59:137–63.

Huberman, G., and T. Regev. 2001. Contagious speculation and a cure for cancer: A non-event that made stock prices soar. *Journal of Finance* 56:387–96.

Ivkovic, Z., and S. Weisbenner. 2007. Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices. *Review of Financial Studies* 20:1327–57.

Kaniel, R., S. Liu, G. Saar, and S. Titman. 2012. Individual investor trading and return patterns around earnings and announcements. *Journal of Finance* 67:639–80.

Kaniel, R., G. Saar, and S. Titman. 2008. Individual investor trading and stock returns *Journal of Finance* 63:273–310.

Kelley, E. K., and P. C. Tetlock. 2013. How wise are crowds? Insights from retail orders and stock returns. *Journal of Finance* 68:1229–65.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45:221–47.

Liu, Y. 2006. Word of Mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing* 70:74–89.

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

Odean, T. 1998. Are investors reluctant to realize their losses? *Journal of Finance* 53:1775–98.

Richardson, S., I. Tuna, and M. Wu. 2002. Predicting earnings management: The case of earnings restatements. Working Paper, University of Pennsylvania.

Seasholes, M. 2010. *Social interactions and investing.* Hoboken, NJ: John Wiley & Sons, Inc.

SEC, 2012. Investment adviser use of social media. *National Examination Risk Alert* 2:1–7.

Seeking Alpha. 2012. *About Seeking Alpha*. Available from: http://seekingalpha.com/page/about_us.

Skinner, D. J., and R. G. Sloan. 2002. Earnings surprises, growth expectations, and stock returns. *Review of Accounting Studies* 7:289–312.

Solomon, D. H. 2012. Selective publicity and stock prices. *Journal of Finance* 67:599–638.

Surowiecki, J. 2005. *The wisdom of crowds: Why the many are smarter than the few*. New York: Anchor Books.

Tetlock, P. C. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62:1139–68.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 63:1437–67.

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

Tyckoson, D., D. Hoffman, P. Kobasa, and P. Ayers. 2011. The Wikipedia effect: How Wikipedia has changed the way the world finds and evaluates information. ALA Annual Conference and Exhibition.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48:817–38.

Womack, K. L. 1996. Do brokerage analysts' recommendations have investment value? *Journal of Finance* 51:137–67.

Zhu, F., and X. M. Zhang. 2010. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing* 74:133–48.