

# CUSTOMERS AS ADVISORS: THE ROLE OF SOCIAL MEDIA IN FINANCIAL MARKETS

Hailiang Chen, Prabuddha De, Yu (Jeffrey) Hu, and Byoung-Hyoun Hwang\*

October 30, 2011

## Abstract

This paper investigates the extent to which peer-based advice transmitted through social media affects the stock market. We conduct textual analysis of articles published on Seeking Alpha, a popular social-media platform among investors. We find that the views expressed in these articles associate strongly with contemporaneous and subsequent stock returns, and help predict earnings surprises. The social media effect is stronger for articles that receive more attention and for companies held mostly by retail investors, the primary generators and consumers of social-media content. Together, these findings point to the importance of social media as both a source of peer-based advice and a channel through which views become reflected in stock prices.

**JEL Classification:** G11, G12, G14

**Keywords:** Social media, Peer-based advice, Financial markets.

---

\*All authors are from the Krannert School of Management, Purdue University, 403 W State Street, West Lafayette, IN 47907. Email: chenhl@purdue.edu, pde@purdue.edu, yuhu@purdue.edu, and bh-wang@purdue.edu. The authors would like to thank Mara Faccio, David Hirshleifer, Seoyoung Kim, Tim Loughran, Ilya Polak, Jin Xu, and seminar participants at Jefferies & Company, Korea University and the 2011 IEEE International Workshop on Statistical Signal Processing for helpful comments.

# 1 Introduction

*“The issue for the pros is that the institution of [financial] analysis risks becoming de-professionalized. In the same way many jobs that took specialized skills became commoditized by the use of new tools or access to information, the era of DIY [do-it-yourself] financial analysis is dawning.”*<sup>1</sup>

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., Deloitte (2007); Chen and Xie (2008); Datamonitor (2010); Gartner (2010)). Gartner (2010), for instance, shows that opinions of fellow consumers disseminated through social media play a crucial role in purchasing decisions across a wide range of products. Deloitte (2007) finds that 82% of US Internet consumers report to be directly influenced by peer reviews in their purchasing decisions. Furthermore, empirical evidence suggests that the influence of peer-based advice is increasing, while that of traditional advice sources is decreasing (Datamonitor (2010)).

This study examines whether peer-based advice also plays a role in financial markets. On the one hand, financial securities are complicated products and perhaps best analyzed by investment professionals with an adequate level of training and experience. On the other hand, traditional advice sources (such as financial analysts) are fraught with built-in conflicts of interest and competing pressures (e.g., Daniel et al. (2002)).<sup>2</sup> Moreover, traditional advice sources have little economic incentive to cover smaller, less popular companies; they may also lack the timeliness and accessibility that customers often demand (e.g., Datamonitor (2010)).

To examine the role of peer-based advice, we extract user-generated stock opinions from the most frequently visited personal finance social-media website,<sup>3</sup> *Seeking Alpha* (hereafter, SA; <http://seekingalpha.com>), and examine how the views expressed on this website pertain

---

<sup>1</sup>Quote by Horace Dediu, former analyst, now blogger at *Asymco*, January 19th 2011.

<sup>2</sup>For instance, the Securities and Exchange Commission (SEC) advises investors on its website “to do their homework before investing. If you purchase a security solely because an analyst said the company was one of his or her ‘top picks’, you may be doing yourself a disservice. Especially if the company is one you’ve never heard of, take time to investigate” (<http://www.sec.gov/investor/pubs/analysts.htm>).

<sup>3</sup>This is as of October 2011.

to investor trading and security prices. Our setting allows us to uncover fresh insights into information dissemination and price formation related to stocks. Specifically, we detect a strong link between the views expressed on SA and contemporaneous and subsequent stock returns. This link is stronger for articles that receive more attention and for companies held mostly by retail investors. The views expressed on SA also predict earnings surprises, suggesting that this form of peer-based advice not only affects investor behavior, but also provides meaningful information with regard to company fundamentals. Together, our findings point to the importance of peer-based advice and social media in financial markets, and hint at the possibility that social-media outlets specializing in financial markets may eventually mirror the development of other peer-based information sources (e.g., Wikipedia) and the way they have changed how information is produced, found, and evaluated (Tyckoson et al. (2011)). The popular press has broached this issue when reporting that bloggers, and their growing clout among the investor population, are already creating a rivalry with traditional advice sources, such as professional sell-side analysts (Bloomberg (2011); CNN (2011)).<sup>4</sup>

Our choice of SA as the focus of this study was motivated by its popularity among investors in voicing their opinions and exchanging investment ideas. As of October 2011, SA had over 5.0 million monthly visitors and, as such, ranked #1 among personal finance social-media websites (<http://www.quantcast.com/seekingalpha.com>). 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" (SeekingAlpha, 2011).<sup>5</sup> Articles submitted to SA are reviewed by a panel and subject to editorial changes.<sup>6</sup> The review process is intended to improve the quality of published articles, without interfering with the authors' original opinions. Authors are required to disclose their identity and, as of 2010, have to disclose their holdings on the stocks they discuss. Many authors also maintain their own subscriber-based financial-blog websites and thus have a genuine incentive to produce high-quality research reports, which would increase their network

---

<sup>4</sup>"Apple's 'Underdog' Analysts Outperform Wall Street From Helsinki, Caracas," Bloomberg, Jan 19th 2011; "Apple and Wall Street: Six quarters of lousy estimates," CNN, Sep 26th 2011.

<sup>5</sup>[http://seekingalpha.com/page/about\\_us](http://seekingalpha.com/page/about_us)

<sup>6</sup>According to SA, approximately 25% of article submissions are accepted.

of clients and paying subscribers.

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 et al. (2008); Li (2008); Davis et al. (2011); Loughran and McDonald (2011)), and we use the negative word list compiled by Loughran and McDonald (2011) to characterize the views expressed in SA articles.<sup>7</sup>

We find that the fraction of negative words (i.e., the number of negative words divided by the total number of words) contained in SA articles negatively correlates with contemporaneous and subsequent stock returns. The association is statistically significant and economically meaningful. For instance, the portfolio of stocks in the top tercile based on the fraction of negative words (“bearish stocks”), on average, experience abnormal returns of -42 basis points (bp) on the day the article is published and an additional -8bp abnormal return on the ensuing trading day; in comparison, the portfolio of stocks in the bottom tercile based on the fraction of negative words (“bullish” stocks), on average, experience abnormal returns of +42bp on the day the article is published and an additional +9bp abnormal return on the ensuing trading day. The differential market performance we observe does not revert. We observe very similar results when we augment our data with articles appearing in the *Wall Street Journal* (hereafter, WSJ) and control for the average fraction of negative words in WSJ articles, suggesting that the association between abnormal returns and the views expressed in SA articles (on a given day) does not simply reflect company-specific news announced on that day. Including leads and lags of the average fraction of negative words in WSJ articles does not alter our findings.

The concern that the association observed here is caused by company-specific news, which simultaneously determines both abnormal returns and the types of SA articles written, is further mitigated by the fact that SA articles are first reviewed by a panel, resulting in a multiple-day delay from article composition to eventual publication on the SA website. This

---

<sup>7</sup>Prior literature focuses on the fraction of negative words as positive words are often negated to convey negative feelings (e.g., “not perfect”).

delay also implies that SA articles are unlikely to be written *in response* to abnormal returns on that day. In additional tests, we find that the association between abnormal returns and the views expressed on SA is stronger for companies held mostly by retail investors (SA's primary audience) and for articles that receive more attention. Together, the results point to the importance of social media as a channel through which views disseminate into the stock market; they also point to the possibility that the associations presented here are not spurious, but causally linked (from SA articles to stock prices).

To explore whether peer-based advice contains value-relevant news (or merely incites naïve investor reaction), we examine whether the views expressed through social media predict subsequent earnings surprises. Specifically, we estimate regressions of price-scaled earnings surprise on the fraction of negative words from thirty days to three days prior to the earnings announcement. Earnings surprise is the difference between reported earnings-per-share (EPS) and the average (or median) of financial analysts' most recent EPS forecasts. If opinions expressed through SA were unrelated to firms' fundamentals, or if information were spurious and already fully incorporated by financial analysts into their reported EPS forecasts, then no association should be observed between subsequent scaled earnings surprise and the measure of social-media views. In contrast, we find that the fraction of negative words in articles published prior to the earnings announcement strongly predicts subsequent scaled earnings surprises: On average, a one standard deviation increase in the fraction of negative words is associated with a scaled earnings surprise that is 0.204% lower; for reference, the mean scaled earnings surprise is -0.190% and the standard deviation is 2.610%. These findings suggest that peer-based advice not only affects investor behavior, but also provides value-relevant information beyond what is provided by traditional advice sources, such as financial analysts.

This study contributes to the growing body of literature that examines the media's effect on stock markets (e.g., Barber and Loeffler (1993); Busse and Green (2002); Antweiler and Frank (2006); Tetlock (2007); Engelberg (2008); Tetlock et al. (2008); Fang and Peress (2009); Engelberg and Parsons (2011)). Most studies in this literature examine the role of traditional media outlets, i.e., newspaper and television, on financial markets, where the investor is

(merely) the recipient of information. In contrast, our analysis focuses on the role of an emerging and relatively new media outlet, social media, which enables investors to not only consume, but also produce information and generate peer-based advice.<sup>8</sup>

As such, our study adds to those of Tumarkin and Whitelaw (2001), Antweiler and Frank (2004) and Das and Chen (2007), who examine how the views expressed in messages posted on *Yahoo Finance* and *Raging Bull* relate to stock returns. Tumarkin and Whitelaw (2001) detect no association; Das and Chen (2007) find “no strong relationship from sentiment to stock prices on average across the individual stocks” (page 1385); Antweiler and Frank (2004) find a statistically significant, yet economically meaningless, association.<sup>9</sup> Together, the results presented in these studies suggest that social-media outlets have little impact. In sharp contrast, the effect 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 3,030 companies from 2006 to 2010. Perhaps more importantly, social-media sites 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)). The stronger effect we document thus may not come as a complete surprise.

Our study also contributes to the debate on how retail traders affect market efficiency. Much of the early literature suggests that retail investors are uninformed and suffer from various behavioral biases (e.g., Odean (1998); Barber and Odean (2000); Benartzi and Thaler (2001)). More recently, however, a growing body of work detects patterns in the data which,

---

<sup>8</sup>Social media also makes the communication two-dimensional as readers can provide feedback and directly interact with the content producer, which, as discussed in Section 3.5, has potentially important implications on the type and quality of information shared.

<sup>9</sup>The examination of how the views expressed in social media outlets affect stock returns is not the only analysis performed by the aforementioned studies. For instance, Antweiler and Frank (2004) and Das and Chen (2007) also test how message board activity relates to stock return volatility and they both detect a reliable association.

taken together, imply that retail traders are skilled and able to identify and trade on novel, value-relevant information (e.g., Coval et al. (2005); Kaniel et al. (2008); Griffin et al.; Kaniel et al. (2011); Kelley and Tetlock (2011)). As Kelley and Tetlock (2011) suggest, one explanation for these two seemingly contradictory sets of findings is that “through learning or attrition, the average skill of retail traders may have changed over time.” Social media may represent one channel through which retail investors have become more informed.

Last but not least, this study proposes a new laboratory for investigating questions about social interactions and investing. Social interactions among investors and the 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., Hong et al. (2004); Ivkovic and Weisbenner (2007)), or analyst coverage to capture the speed of information diffusion (Hong et al. (2000)). Feng and Seasholes (2004) take advantage of a feature of the Chinese stock market, namely, that investors can place trades only at the branch office where they opened their brokerage account, and present evidence of herding among investors in the same locale/branch office.<sup>10</sup> Social-media websites make information shared among investors accessible to researchers and, as such, represent an interesting laboratory for conducting further research on social interactions, information exchange/diffusion, and their implications for financial markets.

The remainder of this paper is organized as follows. In Section 2, we discuss our data. Section 3 presents our main results, followed by a discussion of why investors may be willing to share private value-relevant information. We discuss potential trading strategies in Section 4. Section 5 concludes the paper.

## 2 Data

This section describes the sample construction and introduces our main variables of interest. Our study uses data collected from SA and WSJ articles, as well as financial-analyst data

---

<sup>10</sup>For an overview of the literature on social interactions and investing, please see Hirshleifer and Teoh (2009) and Seasholes (2010), among others.

from the Institutional Brokers' Estimate System (IBES) file, financial-statement and financial-market data from COMPUSTAT and the Center for Research in Security Prices (CRSP), respectively, and institutional-holdings data from Thomson Financial. The sample period spans from 2006 to 2010 and is determined by the availability of SA data.<sup>11</sup>

## 2.1 *Seeking Alpha*

We downloaded all articles that were published between 2006 and 2010 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 the database tables.

SA assigns a unique id to each article. To categorize articles, SA editors tag each article with one or more stock tickers before 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 very difficult, if not impossible. We, therefore, focus our analysis on single-ticker articles only, which comprise roughly one third of all articles published on SA. Excluding multiple-ticker articles biases our sample, but we believe that this bias may work against us and weaken our reported findings. SA users can share their thoughts with others (including the article's author) by commenting on an article. The number of comments likely correlates with the level of attention the article receives. From 2006 to 2010, single-ticker articles on average received 3.77 comments; in comparison, multiple-ticker articles received 5.99 comments, suggesting that multiple-ticker articles attract more attention from SA users than single-ticker articles, and, as such, have greater potential to affect security prices.

The information we collect about each article includes the following items: article id, title,

---

<sup>11</sup>SA was founded in 2003. However, there were less than 4,700 articles in total published from 2003 to 2005. More than 15,000 articles were published in 2006 alone, and the number continuously increased over the years thereafter.

main text, date of publication, author name, and stock ticker. (Please see Figure 1 for two sample SA articles.) After merging with other data (to be discussed below), our final sample comprises 3,030 distinct U.S. common stocks listed on NYSE, NASDAQ, or AMEX. The sample covers not only the most-popular and largest companies, such as Citigroup, Google, Starbucks, and Wal-Mart, but also a large number of lesser-known and smaller companies.<sup>12</sup>

To extract the 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); Tetlock et al. (2008); Li (2008); Davis et al. (2011); Loughran and McDonald (2011)). We use the negative words list compiled by Loughran and McDonald (2011), which they designed specifically for use in studies on financial markets.<sup>13</sup> The SA views for company  $i$  on day  $t$ ,  $NegSA_{i,t}$ , is the average fraction of negative words across all articles published on SA about company  $i$  on day  $t$ . Following prior 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). As pointed out by Tetlock et al. (2008), focusing on the fraction of negative words also has the potential to increase the power of the analysis as the psychology literature argues that "negative information has more impact and is more thoroughly processed than positive information across a wide range of contexts" (pages 1442-1443).

Figure 2 reports the number of SA articles used in our study by the day of the week. In general, substantially more articles are published on weekdays than on weekends. In our analysis, we aggregate articles published over the weekend and other non-trading days with articles published on the ensuing trading day. Our results do not change if we drop all articles published on non-trading days, which represent 8.86% of the articles in our sample.

## 2.2 *Wall Street Journal*

To explore whether SA views have an effect above and beyond the news released in traditional-media outlets, we also construct a measure of information revealed by articles published in

---

<sup>12</sup>A complete list of the companies in our sample is available upon request.

<sup>13</sup>[http://www.nd.edu/~mcdonald/Word\\_Lists.html](http://www.nd.edu/~mcdonald/Word_Lists.html)

the WSJ. We access WSJ articles for the 3,030 stocks covered by SA via the *Factiva* database. Since WSJ articles are not tagged by company name or stock ticker, we formulate a search query to find matched news articles for each stock from 2006 to 2010. 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 WSJ news article.<sup>14</sup> To improve the query performance, we adjust the CRSP company names to match Factiva’s coding of company names, for example, by changing “Company” to “Co” and “Intl” to “International”. If a company changes its name during our sample period, we query all possible names and combine the search results for that stock. A program is written to send our search queries to Factiva and retrieve the WSJ articles appearing in the search result for each query. The information collected about each WSJ article includes: article title, main text, and date of publication. To ensure that the WSJ article discusses one company only, we again exclude articles that can be tied to two or more stock tickers. The WSJ-variable,  $NegWSJ_{i,t}$ , is the average fraction of negative words across all WSJ articles about company  $i$  on day  $t$  – if there are any such articles – and zero otherwise. In our regression analysis, we include  $NegWSJ_{i,t}$ , as well as an indicator variable,  $DummyWSJ_{i,t}$ , denoting whether the WSJ publishes an article about company  $i$  on day  $t$ .

Table 1 presents descriptive statistics of the SA and WSJ articles used in this study. The average length of an SA article is 489 words, which is comparable to the average length of a WSJ article (482 words) and significantly longer than that of an average message posted on Internet message boards or forums (which, according to Antweiler and Frank (2004), is between 20 and 50 words). The average fraction of negative words used in SA articles is 1.54%. In comparison, the average fraction of negative words used in WSJ articles is 2.08%. The fraction of negative words used fluctuates over time, with the fraction being highest in 2008 and 2009 when the stock market performed poorly.

---

<sup>14</sup>We observe similar results using 25-word and 100-word cutoff points.

### 2.3 Abnormal Returns and Other Variables

We obtain financial-statement and financial-market data from COMPUSTAT and CRSP, respectively. We use these data to construct measures of abnormal returns, share turnover, volatility, and past returns. Following prior literature, we compute abnormal returns as the difference between raw returns minus returns on a value-weighted portfolio of firms with similar size and book-to-market ratio. Specifically, we follow the categorization outlined on Kenneth French’s website<sup>15</sup> and form six portfolios based on the 30th and 70th NYSE book-to-market ratio percentiles and on the median NYSE market equity.<sup>16</sup> We compute one-day holding-period returns for day  $t$  and day  $t+1$  ( $ARet_{i,t}$  and  $ARet_{i,t+1}$ , respectively), as well as two-day holding-period returns from day  $t$  to  $t+1$  and day  $t+1$  to  $t+2$  ( $ARet_{i,t,t+1}$  and  $ARet_{i,t+1,t+2}$ , respectively). We also consider cumulative abnormal returns over one calendar week, one calendar month, and three calendar months.

Other variables based on COMPUSTAT and CRSP include:  $Turnover_{i,t}$ , which is the average number of shares traded from thirty to three calendar days prior to day  $t$  divided by the number of shares outstanding;  $Volatility_{i,t}$ , which is the sum of squared daily returns from thirty to three calendar days prior to day  $t$ ; and  $PastReturns_{i,t}$ , which is the cumulative abnormal return from thirty to three calendar days prior to day  $t$ .

We obtain data on quarterly institutional-holdings from Thomson Financial to compute measures of retail holdings, where retail holdings equal 100% minus the institutional holdings reported by Thomson Financial. Finally, we gather information regarding sell-side analyst recommendations and earnings forecasts from the Institutional Brokers Estimate System (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  $t$ . The IBES unadjusted detail-history file tracks

---

<sup>15</sup>[http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data\\_Library/six\\_portfolios.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/six_portfolios.html)

<sup>16</sup>Here, size equals the market value of equity as of the most recent June, and book-to-market ratio equals the book equity for the most recent fiscal year ending in year  $t-1$ , divided by the market value of equity as of December in year  $t-1$ .

each EPS forecast made by each analyst (among others). We use this dataset to compute our earnings-surprise measure, which is the difference between reported EPS and the average annual EPS forecast (computed across all forecasts issued by financial analysts in the three months prior to the earnings announcement).

Table 2 presents the descriptive statistics of the main variables used in this study. Both the mean and the median of our abnormal return measures are zero, indicating that most of the expected component in raw returns is removed, providing a relatively clean estimate of the stock returns' unexpected or abnormal component.

### 3 Main Results

Table 3 reports our main results. Every trading day, we assign stocks with articles published on the SA website into terciles based on the average fraction of negative words used in the corresponding articles. Stocks are held for one day, two days, or one week. We report the average cumulative raw- and abnormal returns.  $T$ -statistics are computed using Newey and West (1987) standard errors with one lag (for the one-day holding period), two lags (for the two-day holding period) and five lags (for the one-week holding period).

Stocks with a relatively low fraction of negative words (i.e., “bullish” securities) experience raw returns of +0.461% ( $t$ -statistic=9.31) on the day of publication on the SA website, while stocks with a relatively high fraction of negative words (i.e., “bearish” securities) experience raw returns of -0.385% ( $t$ -statistic=-5.47). The difference is +0.846% with a  $t$ -statistic of 9.76. The differential market performance grows to 1.019% ( $t$ -statistic=8.97) the day after publication on the SA website and equals 1.133% ( $t$ -statistic=8.69) two days after, and 1.167% ( $t$ -statistic=6.11) one week after publication. In untabulated analysis, we find that the differential market performance equals 0.794% ( $t$ -statistic=2.06) one month after and 1.177% ( $t$ -statistic=1.68) three months after publication. That is, the differential market performance we observe does not revert.

We observe similar results when examining abnormal returns: Stocks with a relatively low fraction of negative words (i.e., “bullish” securities) experience abnormal returns of +0.422%

( $t$ -statistic=9.17) on the day of publication on the SA website, while stocks with a relatively high fraction of negative words (i.e., “bearish” securities) experience abnormal returns of -0.418% ( $t$ -statistic=-6.31). The difference is +0.839% with a  $t$ -statistic of 10.33. The differential market performance grows to 1.006% ( $t$ -statistic=9.35) the day after publication on the SA website and equals 1.113% ( $t$ -statistic=8.82) two days after and 1.130% ( $t$ -statistic=6.72) one week after publication. In untabulated analysis, we find that the differential market performance equals 0.835% ( $t$ -statistic=2.75) one month after and 1.313% ( $t$ -statistic=2.53) three months after publication. The substantial market reaction associated with the publication of SA articles combined with the lack of return reversal suggests that SA articles largely contain value-relevant non-redundant information and that investors react appropriately.

Much of the differential market reaction we detect occurs on the day the SA article(s) is (are) published.<sup>17</sup> The immediate market response agrees with the observation that close to 60% of the comments made in response to an article are posted on the day of publication with an additional 20% being posted the ensuing calendar day. Coupled with the fact that during our sample period, 109 new articles are posted each day on the SA website, these numbers indicate that, in general, the level of attention an article receives is immediate and ebbs quickly. Relatedly, we examine the number of SA articles published (in one-hour increments) over the course of a day for the September 21st 2011 to October 16th period.<sup>18</sup> The results, presented in Figure 3, show that 57% of the articles are published by 1pm EST, and 81% are published before the market closes (4pm EST). Taken together, it appears plausible that at least part of the market reaction on day  $t$  can be attributed to articles published on day  $t$ .

Moreover, note that SA articles are first reviewed by a panel and subject to editorial changes, resulting in a multiple-day delay from article composition to eventual publication on the SA website. The association between SA articles published on day  $t$  and returns on day  $t$  is, thus, unlikely to have been generated by articles written *in response* to abnormal returns on day  $t$ . Because the editorial board is unlikely to monitor the market movement of the

---

<sup>17</sup>Nevertheless, our later analysis hints at significant profits attainable even from trading on days  $t+1$  up to  $t+5$ .

<sup>18</sup>We have no time-stamp data for the SA articles published between 2006 and 2010. September 21st 2011 is the first day we started recording the time of publication for each SA article.

thousands of stocks covered by SA contributors and time the release of articles accordingly, it also seems unlikely that the observed association results from the *editorial board* publishing positive/negative SA articles on day  $t$  in response to high/low abnormal returns on day  $t$ .

We have several additional analyses to help bolster our interpretation of the results that at least part of the observed market reaction can be attributed to SA articles. For one, we augment our data with articles appearing in the WSJ. To the extent that important company news is covered by the WSJ, we should observe weaker results once controlling for the views expressed in the WSJ if the effect documented here were fully generated by company-specific news. In contrast, we detect little difference in results when controlling for WSJ articles. Perhaps more importantly, we find that the association between abnormal returns and the views expressed on SA is stronger for companies held mostly by retail investors (SA’s primary audience) and for articles that receive more attention.

### 3.1 Regression Analysis

Table 4 extends our analysis within a regression framework. Our observations are on a firm/trading-day level. Depending on the regression specification, we have between 30,212 and 30,255 observations. Our dependent variable is our measure of abnormal returns. We consider one-day holding-period returns for day  $t$  and day  $t+1$  ( $ARet_{i,t}$  and  $ARet_{i,t+1}$ , respectively), as well as two-day holding-period returns from day  $t$  to  $t+1$  and day  $t+1$  to  $t+2$  ( $ARet_{i,t,t+1}$  and  $ARet_{i,t+1,t+2}$ , respectively), where  $t$  is the day on which the article appears on the SA website. We focus on one- and two-day horizons because of the high level of noise inherent in long-horizon returns and because the effect documented here is mostly generated in the first two days of publication on the SA website.

Our independent variables in the regression equations are:  $NegSA_{i,t}$ , which is the average fraction of negative words across all articles published on SA about company  $i$  on day  $t$ ;  $NegWSJ_{i,t}$ , which equals the average fraction of negative words across all WSJ articles about company  $i$  on day  $t$  – if there are any such articles – and zero otherwise;  $DummyWSJ_{i,t}$ , which is an indicator variable denoting whether no WSJ article is written about company  $i$  on day

$t$ ;  $Upgrade_{i,t}$  and  $Downgrade_{i,t}$ , which are the number of analyst upgrades (downgrades) on company  $i$  on day  $t$ ;  $PosES_{i,t}$  and  $NegES_{i,t}$ , which are indicator variables denoting whether there is a positive (negative) earnings surprise for company  $i$  on day  $t$ ;  $\ln(Turnover_{i,t})$ , which is the natural logarithm of the average share turnover from thirty days to three days prior to day  $t$ ;  $Volatility_{i,t}$ , which is the sum of the squared raw daily returns from thirty days to three days prior to day  $t$ ; and  $PastReturn_{i,t}$ , which is the cumulative abnormal return from thirty days to three days prior to day  $t$ .

To account for cross-correlation and the fact that some of our variables may have a time trend, we include year-week fixed effects. Cross-correlation is further accounted for by the use of abnormal returns, which removes market-wide effects and any correlation across firms that may be linked to firm size and book-to-market ratio. For the two-day holding period return results,  $t$ -statistics are computed using Newey and West (1987) standard errors with one lag to account for the serial correlation due to the overlap in returns.

The regression results in Table 4 show that the views expressed on SA strongly correlate with stock-market returns. For instance, the coefficient estimate on  $NegSA_{i,t}$  in column (1) is -0.292 ( $t$ -statistic=-8.71), which indicates that the same-day abnormal return decreases by 0.292% when the fraction of negative words in SA articles increases by 1%. These results hold regardless of whether the holding period includes the day for which  $NegSA_{i,t}$  are computed (columns 1 to 4) or not (columns 5 to 8).

The results also hold whether we control for company news reported in WSJ articles (columns 2, 4, 6, and 8) or not (columns 1, 3, 5, and 7), suggesting that the association between abnormal returns and  $NegSA_{i,t}$  (on a given day) does not simply reflect company-specific news announced on that day. Including leads and lags of the average fraction of negative words in WSJ articles does not alter our findings. While the views expressed in WSJ and SA articles both relate to abnormal returns, the implied effect is stronger for SA. This observation does not change when we standardize the fractions of negative words in both media outlets by subtracting the current year's mean and dividing the difference by the

current year's standard deviation of the fraction of negative words.<sup>19</sup>

The coefficient estimates on other independent variables are consistent with findings from prior studies. For instance, the coefficient estimates on  $Upgrade_{i,t}$  and  $Downgrade_{i,t}$  in column (1) are 0.033 and -0.042, respectively, suggesting that stock prices rise after an analyst upgrades his/her recommendation and fall after a recommendation downgrade. Similarly, the coefficient estimates on  $PosES_{i,t}$  and  $NegES_{i,t}$  in column (1) are 0.004 and -0.020, respectively, suggesting that stock prices rise on positive earnings surprises and fall on negative earnings surprises.

### 3.2 Stock Ownership

In an attempt to better understand the underlying mechanisms at hand, we explore whether the effect documented here is stronger among stocks predominantly held by retail investors. We conjecture that retail investors are the primary audience of social media and that opinions revealed through SA have a greater potential to alter market prices when securities are primarily held by retail investors. Stocks predominantly held by retail investors also may represent stocks that are neglected by traditional media companies and professional investment research firms, which lack the economic incentive to cover smaller, less popular companies. SA authors are thus more likely to uncover novel information that is not already reflected in the market price for these types of securities. To test our conjecture, we collect quarterly institutional holdings from Thomson Financial and construct the variable  $RetailHoldings_{i,t}$  ( $= 1 - InstitutionalHoldings_{i,t}$ ) to denote the fraction of a company's shares held by retail investors. We then re-estimate our main regression with the addition of this new variable and its interaction term with  $NegSA_{i,t}$ .

The results reported in Table 5 show that the interaction terms in the first three regression specifications are negative and statistically significant, implying that the effect of  $NegSA_{i,t}$  on stock-market returns is substantially stronger for firms whose outstanding shares are held mostly by retail investors. For instance, when the holding period includes the day for which

---

<sup>19</sup>For instance, the coefficient estimates on standardized  $NegSA_{i,t}$  and  $NegWSJ_{i,t}$  in column (2) are -0.003 ( $t$ -statistic=-8.17) and -0.003 ( $t$ -statistic=-2.88), respectively.

$NegSA_{i,t}$  are computed (column (1)), the coefficient estimate on  $RetailHoldings_{i,t} \times NegSA_{i,t}$  is -0.399 ( $t$ -statistic=-2.61). The 25th percentile of retail holdings is 12.8%, and the 75th percentile is 37.3%. These numbers imply that when the average fraction of negative words in SA articles increases by 1%, the implied decrease in abnormal returns for a firm at the 75th percentile with respect to retail holdings is 0.313% ( $=0.164\%+0.399 \times 37.3\%$ ), whereas the implied decrease in abnormal returns for a firm at the 25th percentile with respect to retail holdings “merely” equals 0.215% ( $=0.164\%+0.399 \times 12.8\%$ ). These findings suggest that stock ownership has a moderating effect on the effect of SA on stock returns.

### 3.3 Article Attention

In additional analyses, we examine whether the effect is stronger for articles that attract more attention. Our first proxy for article attention uses the feature that SA users can share their thoughts with others (including the article’s author) by commenting on an article. 60% of the comments are made on the day of publication, and an additional 20% are made on the ensuing calendar day. The number of comments likely correlates with the level of attention an article garners. Our second proxy is based on the number of prior SA articles written by the author of the article in question. If an author is new to the SA community, she may attract less attention; even if users read her articles, they may discount her views on a given stock.

If there is a causal link from SA articles to stock returns, *on the day of publication*, we should observe a stronger association for articles that garner more attention. The prediction on how  $NegSA_{i,t}$  differentially relates to returns on the day(s) *after* publication, however, is ambiguous. On the one hand, SA articles drawing a high level of attention may continue to predict returns more strongly than SA articles that are initially neglected by the SA community. On the other hand, if prices adjust quickly to the information contained in articles receiving high attention, then the views expressed in these articles may already be mostly incorporated in the prices by the end of the publication day, and, thus, only weakly relate to returns beyond the publication day. At the same time, the views expressed in SA articles that receive limited attention may only be partially incorporated in the prices by the

end of the publication day and, thus, relate more strongly to subsequent returns. Whether  $NegSA_{i,t}$  should associate more strongly with subsequent returns for high attention articles than for low attention articles is therefore unclear.

Table 6 reports the results from repeating our regression analysis with the modification that we now include the logarithm of (1 + the average number of comments received by articles on company  $i$  from the day of publication until the end of the ensuing day), the logarithm of (1 + the average number of prior articles written by the authors of articles on company  $i$  on day  $t$ ), as well as two interaction terms of each of these two variables with the  $NegSA_{i,t}$  variable. We take the logarithm because both the number of comments an article receives and the number of prior articles written are highly right-skewed. We add one to each variable in order to treat zero observations.

For the one-day holding-period regression specification (column (1)), the coefficient estimates on the two interaction terms are negative (-0.091 and -0.052, respectively) and statistically significant (with  $t$ -statistics of -2.23 and -3.03, respectively), implying that articles discussed more intensely on SA and written by more established authors tend to have a larger impact on stock returns. In terms of economic interpretations, when the average fraction of negative words in SA articles increases by 1%, the decrease in the same-day abnormal return associated with SA articles that receive 10 comments is 0.218% larger than the decrease in the same-day abnormal return associated with SA articles that receive no comments. Similarly, the decrease in the same-day abnormal return associated with SA articles whose authors have written at least 100 articles before is 0.240% larger than the decrease in the same-day abnormal return associated with SA articles whose authors have not written any articles before.<sup>20</sup> Together, these findings hint at the possibility that the observed relation between the views expressed on SA and stock returns may be causal.

When extending the analysis to holding periods excluding the article publication date, the coefficient estimates on the two interaction terms lose their statistical significance and flip signs. One interpretation of these findings is that the views expressed in SA articles receiving

---

<sup>20</sup>The 25th percentile of prior articles written is 36; the 75th percentile is 604.

a high level of attention are mostly incorporated in the price by the end of the publication day and, as such, relate less strongly to subsequent returns than the views expressed in SA articles that receive less attention and only gradually diffuse through the investor population.

### 3.4 Noise or Value-Relevant Information?

Our current research design does not allow us to infer with confidence whether stock opinions revealed through social media contain value-relevant news, or whether investors react to false or spurious publicity. The lack of reversal in stock returns after a bullish/bearish SA article points to the former. Table 7 provides additional evidence on this matter. Motivated by Tetlock et al. (2008), who examine whether the fraction of negative words in news stories (in the DJNS and WSJ) predicts firm earnings, we estimate a regression of price-scaled earnings surprise on the fraction of negative words in SA articles. Earnings surprise is the difference between the reported quarterly EPS and the average EPS forecast across all analysts issuing estimates for company  $i$ 's EPS reported at time  $t$ . We do not consider “stale” forecasts issued more than 90 days prior to the earnings announcement. We winsorize the absolute value of scaled earnings surprise at the 99th percentile to mitigate the influence of outliers on our results.

Following Tetlock et al. (2008), we compute the average fraction of negative words across all articles published on SA about company  $i$  from thirty days to three days prior to the earnings announcement,  $NegSA_{i,t-30,t-3}$ . Other independent variables include: lagged scaled earnings surprise; the average fraction of negative words across all WSJ articles about company  $i$  from thirty days to three days prior to the earnings announcement – if there are any such articles – and zero otherwise; an indicator variable denoting whether no WSJ article appears about company  $i$  from thirty days to three days prior to the earnings announcement; the logarithm of market capitalization and book-to-market ratio as of December of the calendar year prior to the earnings announcement; and cumulative abnormal return from thirty to three calendar days prior to the earnings announcement. We also include calendar year-/quarter fixed effects.

If opinions expressed through SA were unrelated to firms' fundamentals, or if information were spurious and fully incorporated by financial analysts into their reported EPS forecasts, no association would be observed between scaled earnings surprise and the social-media-sentiment measure. As reported in columns (1) - (3) of Table 7, the coefficient estimates on  $NegSA_{i,t-30,t-3}$  range from -0.137 ( $t$ -statistic=-3.37) to -0.185 ( $t$ -statistic=-4.57) depending on the set of control variables, suggesting that when the fraction of negative words in articles published prior to the earnings announcement increases by one standard deviation (=1.110%), subsequent scaled earnings are between 0.152% and 0.205% below "market expectations," as measured by financial analysts' forecasts; for reference, the mean scaled earnings surprise is -0.190% and the standard deviation is 2.610%. We observe very similar results when restricting our comparison to financial analysts' forecasts that are issued/updated within 10 days of the earnings announcement, suggesting that the results are not an artifact of SA's timing advantage (recall that  $NegSA_{i,t}$  are computed from 30 days to 3 days prior to the earnings announcement). Together, these findings point to the usefulness of peer-based advice over traditional advice sources.

We also explore where SA authors have more of a competitive advantage relative to financial analysts. In particular, we suspect that SA authors do especially well analyzing companies producing goods that they can easily relate to, and that the association between scaled earnings surprise and  $NegSA_{i,t-30,t-3}$  is stronger for companies with high advertising expenditure. Frieder and Subrahmanyam (2005) present evidence that retail investors prefer to invest in stocks with easily-recognized products. Grullon et al. (2004) find that firms with higher advertising expenditure have a wider shareholder base. To test our prediction, we repeat our regression analysis with the modification that we now include the logarithm of advertising expenditure divided by total assets of company  $i$ , as well as an interaction term of this variable with the  $NegSA_{i,t-30,t-3}$  variable. We take the logarithm because advertising expenditure is highly right-skewed. As reported in Table 7, the coefficient estimate on  $\ln(Advertising) \times NegSA_{i,t}$  is -2.376 ( $t$ -statistic=-2.30). This result is consistent with the notion that SA authors do particularly well analyzing companies that spend significant

resources advertising goods that SA authors can relate to.

### 3.5 Discussion

Our interpretation of the results presented in this study is that conversations on SA collectively provide value-relevant information to investors and help improve market efficiency. A natural question arises as to why SA authors would be willing to share such value-relevant information with others. Traditional theories suggest that a rational arbitrageur with valuable private information should keep that information to herself until the market price moves to the perceived true fundamental value (e.g., Friedman (1953)). On the contrary, more recent private information sharing theories (e.g., Dow and Gorton (1994); Stein (2008); Gray and Kern (2011)) suggest that investors want to share investment ideas with each other for the purpose of receiving constructive feedback, gaining access to new ideas, and attracting additional capital. All of these arguments can be applied to explain why SA authors want to share their insights.

In particular, Stein (2008) argues that market players have an incentive to share their ideas with competitors in order to receive valuable feedback to improve their ideas. Applied to our setting, SA authors can benefit from the comments made by others on their articles, which help SA authors confirm or refine their investment ideas.<sup>21</sup>

Another reason for authors to share their insights on SA is that these authors are likely to be resource-constrained. That is, SA authors lack the capital to push market prices to their perceived true fundamental values. Having established a full position in the asset, SA authors thus write articles to publicize their investment ideas and convince other investors to follow their investment approach and expedite the convergence of market prices to fundamental values (e.g., Dow and Gorton (1994); Gray and Kern (2011)).

---

<sup>21</sup>Stein also predicts that less valuable ideas would be shared among a large group of players, while more valuable ideas would be localized within a small group of players. Given the openness of social media outlets, Stein's argument suggests that the best ideas are not shared on SA; nonetheless, our results indicate that even the possibly less valuable ideas shared on SA are useful to investors.

## 4 Trading Strategy

In our last set of analyses, we evaluate whether investors can trade based on SA views. To enhance the implementability of our trading strategy and minimize the effect of trading costs, we exclude stocks with a price below \$5 and market capitalization below \$2 billion. The trading strategy is as follows: At the end of each trading day  $t$ , we assign stocks into tercile 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 report results from stocks being held in the portfolio for one day, two days, and five days. We compute *value-weighted* portfolio returns for the top- and bottom-tercile portfolios.<sup>22</sup> For the top-tercile portfolio (which contains the most bearish stocks), returns are computed from the closing bid price on day  $t$  to the closing ask price on day  $t+1$  (day  $t+2$ , or day  $t+5$ ); for the bottom-tercile portfolio (which contains the most bullish stocks), returns are computed from the closing ask price on day  $t$  to the closing bid price on day  $t+1$  (day  $t+2$ , or day  $t+5$ ). In short, we account for the bid-ask spread.

The results presented in Table 8 hint at significant trading profits if one trades based on the views expressed on SA. For instance, the bottom-tercile portfolio, which contains stocks with a relatively low  $NegSA_{i,t}$ , on average, experienced returns of +0.069% a day. The top-tercile portfolio, which contains stocks with a relatively high  $NegSA_{i,t}$ , on average, experienced returns of -0.066% a day. If one had bought stocks in the bottom-tercile portfolio and shorted stocks in the top-tercile portfolio – ignoring trading commission, but including the bid-ask spread – the ensuing long-short portfolio based on  $NegSA_{i,t}$  would have generated returns of 0.135% a day.

Figure 4 depicts how much \$1 invested in the “long leg” (black line) – i.e., the bottom-tercile portfolio, which contains stocks with a relatively low  $NegSA_{i,t}$  – would have evolved throughout our sample period. We observe that \$1 invested in the long leg at the beginning of 2006 would have grown to more than \$1.80 by the end of 2010. To put things in perspective, \$1 invested in the Dow Jones Industrial Index at the beginning of 2006 would have grown

---

<sup>22</sup>The results for equal-weighted portfolio returns are stronger than the ones presented in this study.

to \$1.06 by the end of 2010 (not shown in the figure). Similarly, \$1 invested in the “short leg” (grey line) – i.e., the top-tercile portfolio, which contains stocks with a relatively high  $NegSA_{i,t}$  – would have shrunk to less than \$0.28 by the end of 2010.

Table 8 also reports results from an isolated trading strategy, where we repeat our exercise but require that the top-tercile portfolio based on  $NegSA_{i,t}$  does not contain stocks that are simultaneously in the top-tercile based on  $NegWSJ_{i,t}$  on day  $t$ . The same applies to the bottom-tercile portfolio. Again, the findings suggest significant trading profits if one trades based on the views expressed on SA. For instance, the bottom-tercile portfolio, which contains stocks with a relatively low  $NegSA_{i,t}$ , on average, experienced returns of +0.060% a day, and the top-tercile portfolio, which contains stocks with a relatively high  $NegSA_{i,t}$ , on average, experienced returns of -0.063% a day. The long-short portfolio based on  $NegSA_{i,t}$  would have generated returns of 0.123% a day.

## 5 Conclusion

The Internet has become increasingly popular both as a venue to place trades and as a source of information. Da et al. (2011), for instance, provide evidence of a strong link between aggregate search frequency of stock tickers in Google and trading by retail investors. Investors now have the option to seek investment advice directly from peer investors through social-media sites. This study examines how the views expressed on a popular social-media site for investors, *Seeking Alpha*, pertain to investor trading and security prices. We find that the opinions revealed on SA strongly associate with the corresponding companies’ stock returns, even after controlling for the effect of traditional advice sources, such as financial analysts and newspaper articles, and other financial-market variables. In addition, our evidence suggests that the effect of peer-based advice on stock returns is stronger for articles that receive more attention and for companies held mostly by retail investors. Together, our findings point to the growing role of social media and peer-based advice in financial markets. Exploring implications of peer-based advice and, more generally, using social media outlets as a laboratory to investigate the effects of social interactions on investment behavior should prove to be interesting avenues for

future research.

## References

- Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *Journal of Finance*, 59:1259–1294.
- Antweiler, W. and Frank, M. Z. (2006). Do us stock markets typically overreact to corporate news stories? *SSRN eLibrary*.
- Barber, B. M. and Loeffler, D. (1993). The 'dartboard' column: Second-hand information and price pressure. *Journal of Financial and Quantitative Analysis*, 28(2):273–284.
- Barber, B. M. and Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance*, 55:773–806.
- Benartzi, S. and Thaler, R. H. (2001). Naïve diversification strategies in defined contribution saving plans. *American Economic Review*, 91:79–98.
- Boyd, D. M. and Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13:210–230.
- Busse, J. A. and Green, T. C. (2002). Market efficiency in real time. *Journal of Financial Economics*, 65(3):415 – 437.
- Chapman, C. (2009). The history and evolution of social media.
- Chen, Y. and Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54:477–491.
- Coval, J. D., Hirshleifer, D. A., and Shumway, T. (2005). Can individual investors beat the market? *SSRN eLibrary*.
- Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *Journal of Finance*, forthcoming.
- Daniel, K., Hirshleifer, D., and Teoh, S. H. (2002). Investor psychology in capital markets: Evidence and policy implications. *Journal of Monetary Economics*, 49:139–209.
- Das, S. R. and Chen, M. Y. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9):1375–1388.
- Datamonitor (2010). Social media in financial services: The customer as the advisor.
- Davis, A. K., Piger, J. M., and Sedor, L. M. (2011). Beyond the numbers: Measuring the information content of earnings press release language. *SSRN eLibrary*.
- Deloitte (2007). Most consumers read and rely on online reviews; companies must adjust.
- Dow, J. and Gorton, G. (1994). Arbitrage chains. *The Journal of Finance*, 49(3):819–849.
- Engelberg, J. (2008). Costly information processing: Evidence from earnings announcements. *SSRN eLibrary*.

- Engelberg, J. and Parsons, C. A. (2011). The causal impact of media in financial markets. *Journal of Finance*, 66:67–97.
- Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. *Journal of Finance*, 64(5):2023–2052.
- Feng, L. and Seasholes, M. (2004). Correlated trading and location. *Journal of Finance*, 59:2117–2144.
- Frieder, L. and Subrahmanyam, A. (2005). Brand perceptions and the market for common stock. *Journal of Financial and Quantitative Analysis*, 40(01):57–85.
- Friedman, M. (1953). *The case for flexible exchange rates*. University of Chicago Press.
- Gartner (2010). User survey analysis: Consumer marketing using social network analysis, worldwide.
- Gray, W. R. and Kern, A. E. (2011). Talking your book: Social networks and price discovery. *SSRN eLibrary*.
- Griffin, J. M., Shu, T., and Topalaglu, S. Examining the dark side of financial markets: Who trades ahead of major announcements?
- Grullon, G., Kanatas, G., and Weston, J. P. (2004). Advertising, breadth of ownership, and liquidity. *Review of Financial Studies*, 17(2):439–461.
- Hirshleifer, D. and Teoh, S. H. (2009). *Thought and Behavior Contagion in Capital Markets*, chapter 1, pages 1–56. Elsevier/North-Holland.
- Hong, H., Kubik, J. D., and Stein, J. C. (2004). Social interaction and stock-market participation. *Journal of Finance*, 59:137–163.
- Hong, H., Lim, T., and Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance*, 55:265–295.
- Ivkovic, Z. and Weisbenner, S. (2007). Information diffusion effects in individual investors common stock purchases: Covet thy neighbors investment choicess. *Review of Financial Studies*, 20:1327–1357.
- Kaniel, R., Liu, S., Saar, G., and Titman, S. (2011). Individual investor trading and return patterns around earnings announcements. *Journal of Finance*, *forthcoming*.
- Kaniel, R., Saar, G., and Titman, S. (2008). Individual investor trading and stock returns. *Journal of Finance*, 63:273–310.
- Kelley, E. K. and Tetlock, P. C. (2011). How wise are crowds? insights from retail orders and stock returns.
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45:221–247.

- Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance*, 66(1):35–65.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.
- Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance*, 53:1775–1798.
- Seasholes, M. (2010). *Social Interactions and Investing*. Hoboken, NJ: John Wiley & Sons, Inc.
- SeekingAlpha (2011). About seeking alpha.
- Stein, J. C. (2008). Conversations among competitors. *American Economic Review*, 98(5):2150–2162.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62:1139–1168.
- Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms’ fundamentals. *Journal of Finance*, 63:1437–1467.
- Tumarkin, R. and Whitelaw, R. F. (2001). News or noise? internet postings and stock prices. *Financial Analysts Journal*, 57:41–51.
- Tyckoson, D., Hoffman, D., Kobasa, P., and Ayers, P. (2011). The wikipedia effects: How wikipedia has changed the way the world finds and evaluates information.

## Figure 1. Two Sample Articles from Seeking Alpha

This figure presents two articles published on Seeking Alpha. The first article is “positive” (the fraction of negative words is 0%: 0 out of 447); the second article is “negative” (the fraction of negative words is 2.43%: 12 out of 494). The format is slightly adjusted; the content, however, remains unaltered.

### A Positive Article about Google

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***

## A Negative Article about Google

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 2. Number of Articles and Percentage Distribution by the Day of the Week

This figure depicts the number of SA articles used in our study by the day of the week, both in raw numbers and as a percentage.

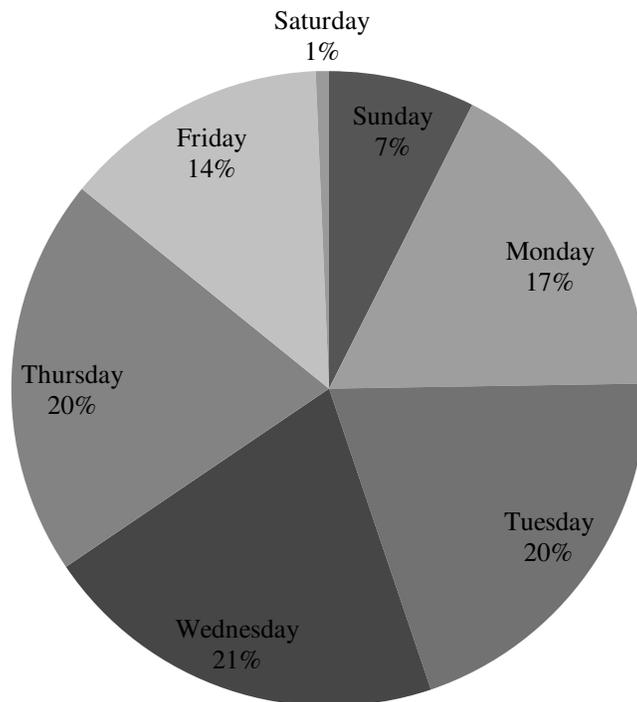
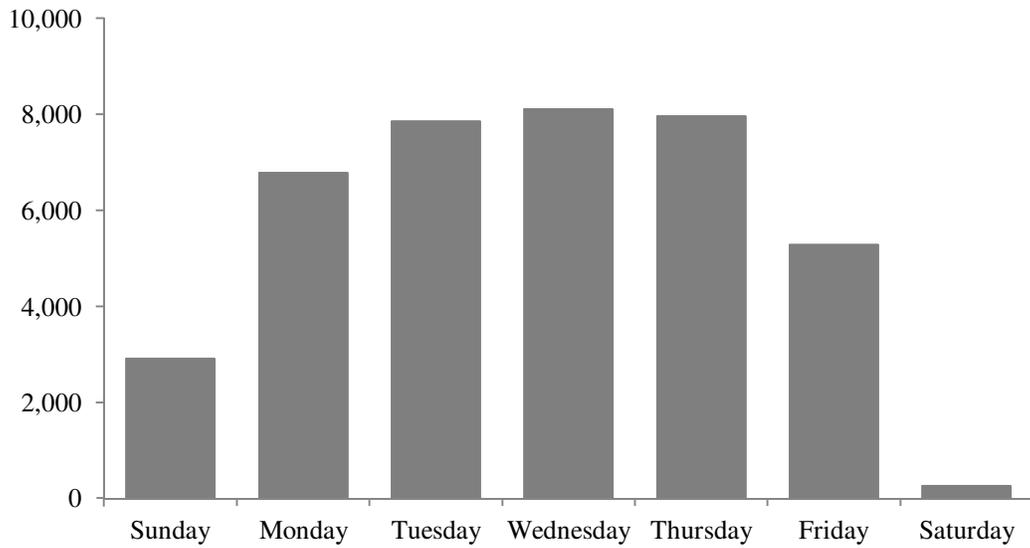


Figure 3. Distribution of New SA Articles across a Day

This figure depicts the average number of SA articles published within each hour in a day and the cumulative percentage for each hourly interval from September 21 to October 16, 2011 (26 days in total).

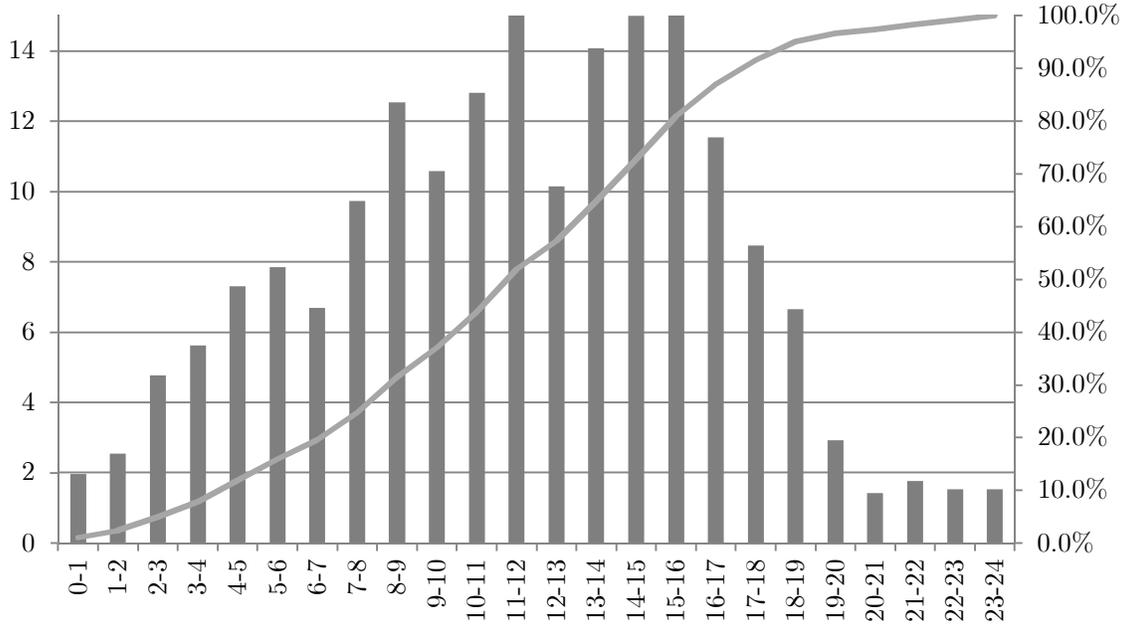


Figure 4. Trading Strategy

This figure depicts how much \$1 invested in the basic trading strategy would have grown over our sample period from 2006 to 2010. We exclude stocks with a price below \$5 and market capitalization below \$2 billion. The SA-trading strategy is as follows: At the end of each trading day  $t$ , we assign stocks into tercile portfolios based on the average fraction of negative words across all articles published on SA about company  $i$  on day  $t$  ( $NegSA_{i,t}$ ). The portfolio of securities in the bottom  $NegSA_{i,t}$ -tercile is referred to as “Bullish”; the portfolio of securities in the top  $NegSA_{i,t}$ -tercile is referred to as “Bearish.” Stocks are held in the portfolio for one day. We compute value-weighted portfolio returns. For the “bullish” (“bearish”) portfolio, returns are computed from the closing ask (bid) price on day  $t$  to the closing bid (ask) price on day  $t+1$ . Based on the daily returns of the “bullish” portfolio (black line) and the daily returns of the “bearish” portfolio (grey line), we plot how much \$1 invested would have grown.



Table 1. Descriptive Statistics of Seeking Alpha and Wall Street Journal Articles

This table reports summary statistics for the *Seeking Alpha* (SA) and *Wall Street Journal* (WSJ) articles covered in this study. The sample period is 2006-2010.

|                                   | 2006  | 2007  | 2008  | 2009  | 2010  | All Years |
|-----------------------------------|-------|-------|-------|-------|-------|-----------|
| Panel A: Seeking Alpha            |       |       |       |       |       |           |
| Total # Stock tickers             | 1,072 | 1,670 | 1,493 | 1,379 | 1,523 | 3,030     |
| Total # Articles                  | 3,767 | 9,341 | 8,371 | 8,871 | 8,131 | 38,481    |
| Avg. # Words per article          | 460   | 452   | 500   | 483   | 542   | 489       |
| Avg. # Negative words per article | 6     | 6     | 9     | 8     | 8     | 8         |
| Avg. % Negative words             | 1.37% | 1.45% | 1.69% | 1.69% | 1.42% | 1.54%     |
| StDev % Negative words            | 1.14% | 1.17% | 1.24% | 1.22% | 1.15% | 1.20%     |
| Panel B: Wall Street Journal      |       |       |       |       |       |           |
| Total # Stock tickers             | 852   | 718   | 745   | 577   | 521   | 1,373     |
| Total # Articles                  | 6047  | 4,895 | 5,220 | 3,706 | 3,374 | 23,242    |
| Avg. # Words per article          | 426   | 494   | 468   | 513   | 553   | 482       |
| Avg. # Negative words per article | 8     | 9     | 10    | 12    | 12    | 10        |
| Avg. % Negative words             | 1.83% | 1.95% | 2.25% | 2.33% | 2.17% | 2.08%     |
| StDev % Negative words            | 1.75% | 1.70% | 1.65% | 1.50% | 1.62% | 1.67%     |

Table 2. Summary Statistics

This table reports the summary statistics of the main variables. The observations are on a company/trading-day level. Abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market-characteristics.  $Upgrade_{i,t}$  is the number of financial analyst upgrades and  $Downgrade_{i,t}$  is the number of financial analyst downgrades on 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.  $\ln(Turnover_{i,t})$  is the natural logarithm of the average share turnover from thirty to three calendar days prior to day  $t$ .  $Volatility_{i,t}$  is the sum of squared raw daily returns in the period of thirty to three calendar days prior to day  $t$ .  $PastReturn_{i,t}$  is the accumulative abnormal return of company  $i$  in the period of thirty to three calendar days prior to day  $t$ .

|                       | N      | Mean   | Std. Dev | 25 <sup>th</sup> Pctl | 50 <sup>th</sup> Pctl | 75 <sup>th</sup> Pctl |
|-----------------------|--------|--------|----------|-----------------------|-----------------------|-----------------------|
| $ARet_{i,t}$          | 30,255 | 0.000  | 0.066    | -0.013                | -0.000                | 0.014                 |
| $ARet_{i,t,t+1}$      | 30,236 | 0.001  | 0.082    | -0.019                | -0.001                | 0.019                 |
| $ARet_{i,t+1}$        | 30,236 | 0.000  | 0.045    | -0.012                | -0.001                | 0.012                 |
| $ARet_{i,t+1,t+2}$    | 30,212 | -0.001 | 0.060    | -0.018                | -0.001                | 0.016                 |
| $Upgrade_{i,t}$       | 30,255 | 0.039  | 0.223    | 0                     | 0                     | 0                     |
| $Downgrade_{i,t}$     | 30,255 | 0.047  | 0.278    | 0                     | 0                     | 0                     |
| $PosES_{i,t}$         | 30,255 | 0.048  | 0.214    | 0                     | 0                     | 0                     |
| $NegES_{i,t}$         | 30,255 | 0.015  | 0.120    | 0                     | 0                     | 0                     |
| $\ln(Turnover_{i,t})$ | 30,255 | 2.512  | 0.926    | 1.970                 | 2.512                 | 3.070                 |
| $Volatility_{i,t}$    | 30,255 | 0.035  | 0.327    | 0.005                 | 0.010                 | 0.024                 |
| $PastReturn_{i,t}$    | 30,255 | -0.000 | 0.179    | -0.061                | -0.006                | 0.050                 |

Table 3. Seeking Alpha and Abnormal Returns: Portfolio Results

This table reports raw returns and abnormal returns for portfolios formed based on the views reflected in *Seeking Alpha* (SA) articles. The sample period is 2006-2010. Abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market-characteristics.  $NegSA_{i,t}$  is the average fraction of negative words across all articles published on SA about company  $i$  on day  $t$ . At the end of each trading day  $t$ , we assign stocks into terciles based on  $NegSA_{i,t}$ . The portfolio of securities in the bottom  $NegSA_{i,t}$ -tercile is referred to as "Bullish"; the portfolio of securities in the top  $NegSA_{i,t}$ -tercile is referred to as "Bearish." We report the average cumulative raw return (Panel A) and the average cumulative abnormal return (Panel B) for the tercile portfolios as well as differences in portfolios for various holding periods. T-statistics are computed using Newey-West (1987) standard errors and are reported in parentheses.

|                                     | $NegSA_i$ | $t=0$   | +1 day  | +2 days | +1 week |
|-------------------------------------|-----------|---------|---------|---------|---------|
| Panel A: Cumulative Raw Return      |           |         |         |         |         |
| "Bullish"                           | 0.474%    | 0.461%  | 0.606%  | 0.662%  | 0.769%  |
|                                     |           | (9.31)  | (9.39)  | (8.85)  | (7.82)  |
|                                     | 1.303%    | 0.177%  | 0.299%  | 0.314%  | 0.552%  |
|                                     |           | (2.14)  | (3.01)  | (2.75)  | (3.91)  |
| "Bearish"                           | 2.717%    | -0.385% | -0.413% | -0.471% | -0.397% |
|                                     |           | (-5.47) | (-4.28) | (-4.01) | (-2.34) |
| $\Delta$ (Bullish, Bearish)         |           | 0.846%  | 1.019%  | 1.133%  | 1.167%  |
|                                     |           | (9.76)  | (8.97)  | (8.69)  | (6.11)  |
| Panel B: Cumulative Abnormal Return |           |         |         |         |         |
| "Bullish"                           | 0.474%    | 0.422%  | 0.511%  | 0.537%  | 0.492%  |
|                                     |           | (9.17)  | (8.53)  | (7.82)  | (5.62)  |
|                                     | 1.303%    | 0.146%  | 0.207%  | 0.186%  | 0.294%  |
|                                     |           | (1.82)  | (2.19)  | (1.71)  | (2.20)  |
| "Bearish"                           | 2.717%    | -0.418% | -0.495% | -0.575% | -0.638% |
|                                     |           | (-6.31) | (-5.51) | (-5.32) | (-4.27) |
| $\Delta$ (Bullish, Bearish)         |           | 0.839%  | 1.006%  | 1.113%  | 1.130%  |
|                                     |           | (10.33) | (9.35)  | (8.82)  | (6.72)  |

Table 4. Seeking Alpha and Abnormal Returns: Regression Results

This table reports coefficient estimates from regressions of abnormal returns on measures of the views reflected in *Seeking Alpha* (SA) articles. The sample period is 2006-2010. Abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market-characteristics.  $NegSA_{i,t}$  is the average fraction of negative words across all articles published on SA about company  $i$  on day  $t$ .  $NegWSJ_{i,t}$  is the average fraction of negative words across all articles published in the WSJ about company  $i$  on day  $t$ , if there were any such articles, and zero otherwise.  $DummyWSJ_{i,t}$  is an indicator variable denoting whether no article was published in the WSJ about company  $i$  on day  $t$ . Other independent variables are as described in Table 2. We include year-week fixed effects. T-statistics are computed using Newey-West (1987) standard errors and are reported in parentheses.

|                       | $ARet_{i,t}$       |                    | $ARet_{i,t,t+1}$   |                    | $ARet_{i,t+1}$    |                   | $ARet_{i,t+1,t+2}$ |                   |
|-----------------------|--------------------|--------------------|--------------------|--------------------|-------------------|-------------------|--------------------|-------------------|
|                       | (1)                | (2)                | (3)                | (4)                | (5)               | (6)               | (7)                | (8)               |
| $NegSA_{i,t}$         | -0.292<br>(-8.71)  | -0.278<br>(-8.20)  | -0.363<br>(-9.39)  | -0.354<br>(-9.43)  | -0.065<br>(-2.79) | -0.070<br>(-2.96) | -0.096<br>(-2.71)  | -0.100<br>(-2.87) |
| $NegWSJ_{i,t}$        |                    | -0.214<br>(-2.96)  |                    | -0.148<br>(-1.71)  |                   | 0.055<br>(1.09)   |                    | 0.034<br>(0.49)   |
| $DummyWSJ_{i,t}$      |                    | -0.004<br>(-2.05)  |                    | -0.003<br>(-1.77)  |                   | 0.000<br>(0.30)   |                    | -0.000<br>(-0.13) |
| $Upgrade_{i,t}$       | 0.033<br>(19.78)   | 0.033<br>(19.66)   | 0.035<br>(6.95)    | 0.034<br>(6.84)    | 0.001<br>(1.24)   | 0.001<br>(1.20)   | 0.002<br>(1.64)    | 0.002<br>(1.53)   |
| $Downgrade_{i,t}$     | -0.042<br>(-30.92) | -0.042<br>(-30.83) | -0.045<br>(-14.21) | -0.045<br>(-14.20) | -0.004<br>(-4.12) | -0.004<br>(-4.17) | -0.005<br>(-2.48)  | -0.005<br>(-2.52) |
| $PosES_{i,t}$         | 0.004<br>(2.41)    | 0.004<br>(2.40)    | 0.007<br>(3.33)    | 0.007<br>(3.35)    | 0.003<br>(2.40)   | 0.003<br>(2.43)   | 0.003<br>(1.54)    | 0.003<br>(1.57)   |
| $NegES_{i,t}$         | -0.020<br>(-6.24)  | -0.020<br>(-6.27)  | -0.037<br>(-6.99)  | -0.037<br>(-6.99)  | -0.018<br>(-8.14) | -0.018<br>(-8.11) | -0.022<br>(-5.50)  | -0.022<br>(-5.49) |
| $\ln(Turnover_{i,t})$ | -0.001<br>(-1.65)  | -0.001<br>(-1.60)  | -0.001<br>(-0.99)  | -0.001<br>(-1.00)  | -0.000<br>(-0.66) | -0.000<br>(-0.71) | -0.001<br>(-1.18)  | -0.001<br>(-1.22) |
| $Volatility_{i,t}$    | -0.002<br>(-1.34)  | -0.002<br>(-1.30)  | -0.001<br>(-0.30)  | -0.001<br>(-0.29)  | -0.000<br>(-0.54) | -0.000<br>(-0.57) | 0.003<br>(1.75)    | 0.003<br>(1.75)   |
| $PastReturn_{i,t}$    | -0.000<br>(-0.05)  | -0.000<br>(-0.11)  | -0.002<br>(-0.30)  | -0.002<br>(-0.31)  | -0.000<br>(-0.13) | -0.000<br>(-0.09) | -0.010<br>(-1.83)  | -0.010<br>(-1.83) |
| # Obs.                | 30,255             | 30,255             | 30,236             | 30,236             | 30,236            | 30,236            | 30,212             | 30,212            |
| Adj. $R^2$            | 4.86%              | 4.88%              | 4.24%              | 4.24%              | 0.61%             | 0.61%             | 0.91%              | 0.91%             |

Table 5. Seeking Alpha, Abnormal Returns and Retail Holdings

This table reports coefficient estimates from regressions of abnormal returns on measures of the views reflected in *Seeking Alpha* (SA) articles. The sample period is 2006-2010. Abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market-characteristics.  $NegSA_{i,t}$  is the average fraction of negative words across all articles published on SA about company  $i$  on day  $t$ .  $RetailHoldings_{i,t}$  is the fraction of company  $i$ 's shares held by retail investors. Other independent variables are as described in Table 2 and 4. We include year-week fixed effects. T-statistics are computed using Newey-West (1987) standard errors and are reported in parentheses.

|                                      | $ARet_{i,t}$<br>(1) | $ARet_{i,t,t+1}$<br>(2) | $ARet_{i,t+1}$<br>(3) | $ARet_{i,t+1,t+2}$<br>(4) |
|--------------------------------------|---------------------|-------------------------|-----------------------|---------------------------|
| $RetailHoldings_{i,t}$               | 0.006<br>(1.96)     | 0.011<br>(2.32)         | 0.004<br>(1.95)       | 0.001<br>(0.40)           |
| $RetailHoldings_{i,t} * NegSA_{i,t}$ | -0.399<br>(-2.61)   | -0.604<br>(-2.48)       | -0.278<br>(-2.62)     | -0.417<br>(-1.91)         |
| $NegSA_{i,t}$                        | -0.164<br>(-2.95)   | -0.181<br>(-2.37)       | 0.012<br>(0.33)       | 0.024<br>(0.33)           |
| $NegWSJ_{i,t}$                       | -0.220<br>(-3.02)   | -0.157<br>(-1.80)       | 0.052<br>(1.03)       | 0.031<br>(0.45)           |
| $DummyWSJ_{i,t}$                     | -0.004<br>(-2.13)   | -0.004<br>(-1.83)       | 0.000<br>(0.25)       | -0.000<br>(-0.27)         |
| $Upgrade_{i,t}$                      | 0.033<br>(19.10)    | 0.034<br>(6.59)         | 0.001<br>(0.99)       | 0.001<br>(1.14)           |
| $Downgrade_{i,t}$                    | -0.042<br>(-30.70)  | -0.046<br>(-14.07)      | -0.004<br>(-4.12)     | -0.005<br>(-2.52)         |
| $PosES_{i,t}$                        | 0.004<br>(2.47)     | 0.007<br>(3.42)         | 0.003<br>(2.29)       | 0.002<br>(1.32)           |
| $NegES_{i,t}$                        | -0.020<br>(-6.46)   | -0.037<br>(-7.05)       | -0.017<br>(-7.96)     | -0.022<br>(-5.46)         |
| $\ln(Turnover)_{i,t}$                | -0.001<br>(-1.36)   | -0.000<br>(-0.55)       | -0.000<br>(-0.54)     | -0.001<br>(-1.48)         |
| $Volatility_{i,t}$                   | -0.002<br>(-1.30)   | -0.001<br>(-0.31)       | -0.000<br>(-0.56)     | 0.004<br>(1.83)           |
| $PastReturn_{i,t}$                   | -0.000<br>(-0.10)   | -0.002<br>(-0.33)       | -0.000<br>(-0.11)     | -0.010<br>(-1.78)         |
| # Obs.                               | 29,956              | 29,937                  | 29,937                | 29,913                    |
| Adj. $R^2$                           | 4.87%               | 4.21%                   | 0.59%                 | 0.93%                     |

Table 6. Seeking Alpha, Abnormal Returns and Proxies for Article Attention

This table reports coefficient estimates from regressions of abnormal returns on measures of the views reflected in *Seeking Alpha* (SA) articles. The sample period is 2006-2010. Abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market-characteristics.  $NegSA_{i,t}$  is the average fraction of negative words across all articles published on SA about company  $i$  on day  $t$ .  $Comment_{i,t}$  is the average number of comments received from day  $t$  until the end of day  $t+1$  for articles published in SA on company  $i$ .  $PriorArticle_{i,t}$  is the average number of prior articles by the authors of articles published on SA about company  $i$ . Other independent variables are as described in Table 2 and 4. We include year-week fixed effects. T-statistics are computed using Newey-West (1987) standard errors and are reported in parentheses.

|   | $ARet_{i,t}$<br>(1) | $ARet_{i,t,t+1}$<br>(2) | $ARet_{i,t+1}$<br>(3) | $ARet_{i,t+1,t+2}$<br>(4) |
|---|---------------------|-------------------------|-----------------------|---------------------------|
| $\ln(1+Comment_{i,t})$                  | 0.001<br>(1.04)     | 0.001<br>(1.22)         | 0.000<br>(0.27)       | 0.001<br>(0.58)           |
| $\ln(1+Comment_{i,t})*NegSA_{i,t}$      | -0.091<br>(-2.23)   | -0.068<br>(-1.12)       | 0.037<br>(1.30)       | 0.033<br>(0.49)           |
| $\ln(1+PriorArticle_{i,t})$             | 0.000<br>(0.75)     | 0.000<br>(0.04)         | -0.000<br>(-0.89)     | -0.000<br>(-0.71)         |
| $\ln(1+PriorArticle_{i,t})*NegSA_{i,t}$ | -0.052<br>(-3.03)   | -0.045<br>(-2.50)       | 0.008<br>(0.71)       | 0.014<br>(0.81)           |
| $NegSA_{i,t}$                           | 0.037<br>(0.38)     | -0.090<br>(-0.92)       | -0.136<br>(-2.03)     | -0.191<br>(-2.08)         |
| $NegWSJ_{i,t}$                          | -0.206<br>(-2.84)   | -0.142<br>(-1.65)       | 0.052<br>(1.03)       | 0.030<br>(0.45)           |
| $DummyWSJ_{i,t}$                        | -0.004<br>(-2.19)   | -0.004<br>(-1.93)       | 0.000<br>(0.36)       | 0.000<br>(0.00)           |
| $Upgrade_{i,t}$                         | 0.033<br>(19.70)    | 0.035<br>(6.78)         | 0.002<br>(1.29)       | 0.002<br>(1.64)           |
| $Downgrade_{i,t}$                       | -0.042<br>(-30.66)  | -0.045<br>(-14.12)      | -0.004<br>(-4.13)     | -0.005<br>(-2.51)         |
| $PosES_{i,t}$                           | 0.005<br>(2.62)     | 0.008<br>(3.74)         | 0.003<br>(2.61)       | 0.003<br>(1.69)           |
| $NegES_{i,t}$                           | -0.019<br>(-6.02)   | -0.036<br>(-6.83)       | -0.017<br>(-7.98)     | -0.022<br>(-5.45)         |
| $\ln(Turnover)_{i,t}$                   | -0.001<br>(-1.23)   | -0.001<br>(-0.95)       | -0.000<br>(-1.09)     | -0.001<br>(-1.54)         |
| $Volatility_{i,t}$                      | -0.002<br>(-1.31)   | -0.001<br>(-0.30)       | -0.001<br>(-0.61)     | 0.003<br>(1.74)           |
| $PastReturn_{i,t}$                      | -0.000<br>(-0.18)   | -0.002<br>(-0.33)       | -0.000<br>(-0.06)     | -0.010<br>(-1.82)         |
| # Obs.                                  | 30,255              | 30,236                  | 30,236                | 30,212                    |
| Adj. $R^2$                              | 4.93%               | 4.27%                   | 0.62%                 | 0.92%                     |

Table 7. Seeking Alpha and Earnings Surprises

We estimate a regression of price-scaled earnings surprise on measures of the views reflected in *Seeking Alpha* (SA) articles. Earnings surprise is the difference between reported quarterly EPS and the average EPS forecast across all analysts issuing estimates for company  $i$ 's EPS reported at time  $t$ .  $NegSA_{i,t-30,t-3}$  is the fraction of negative words from thirty to three days prior to the earnings announcement. Other independent variables include: the fraction of negative words in *Wall Street Journal* (WSJ) articles on company  $i$  thirty to three days prior to the earnings announcement, if there were any such articles, and zero otherwise; an indicator variable denoting whether no WSJ article(s) was (were) written on company  $i$  thirty to three days prior to the earnings announcement; lagged price-scaled earnings surprise (our dependent variable); forecast dispersion, which equals the price-scaled standard deviation of analysts' EPS forecasts, or zero if a company is only covered by one analyst; logarithm of scaled advertising expenditure; the logarithm of market capitalization; the logarithm of the book-to-market ratio; and cumulative abnormal returns from thirty to three calendar days prior to the earnings announcement. We include calendar year/quarter fixed effects. T-statistics are reported in parentheses.

|   | (1)               | (2)               | (3)                | (4)               |
|---|-------------------|-------------------|--------------------|-------------------|
| $NegSA_{i,t-30,t-3}$                      | -0.185<br>(-4.57) | -0.172<br>(-4.19) | -0.137<br>(-3.37)  | 0.032<br>(0.49)   |
| $NegWSJ_{i,t-30,t-3}$                     |                   | -0.075<br>(-1.26) | -0.061<br>(-1.04)  | -0.055<br>(-0.76) |
| $DummyWSJ_{i,t-30,t-3}$                   |                   | -0.000<br>(-0.08) | -0.001<br>(-0.37)  | -0.002<br>(-0.97) |
| $Lag(Dependent Var_{i,t})$                | 0.272<br>(13.87)  | 0.271<br>(13.82)  | 0.226<br>(11.43)   | 0.110<br>(3.79)   |
| $ForecastDispersion_{i,t}$                |                   |                   | -0.018<br>(-10.50) | -0.034<br>(-8.64) |
| $NegSA_{i,t-30,t-3} * \ln(1+Advertising)$ |                   |                   |                    | -2.376<br>(-2.30) |
| $\ln(1+Advertising)_{i,t}$                |                   |                   |                    | 0.034<br>(1.83)   |
| $\ln(MarketCapital_{i,t})$                | 0.001<br>(3.74)   | 0.001<br>(3.85)   | 0.001<br>(3.82)    | 0.001<br>(3.00)   |
| $\ln(Book/Market_{i,t})$                  | -0.002<br>(-3.42) | -0.002<br>(-3.32) | -0.001<br>(-2.06)  | -0.000<br>(-0.46) |
| $PastReturn_{i,t}$                        | 0.006<br>(1.97)   | 0.006<br>(1.92)   | 0.004<br>(1.35)    | 0.011<br>(2.52)   |
| # Obs.                                    | 3,212             | 3,212             | 3,212              | 1,559             |
| Adj. $R^2$                                | 8.67%             | 8.70%             | 11.72%             | 9.21%             |

Table 8. Trading Strategy

This table reports returns on trading strategies based on measures of the views reflected in *Seeking Alpha* (SA) articles. The sample period is 2006-2010. We exclude stocks with a price below \$5 and market capitalization below \$2 billion. In the basic trading strategy (Panel A), at the end of each trading day  $t$ , we assign stocks into tercile portfolios based on the average fraction of negative words across all articles published on SA about company  $i$  on day  $t$ . In the “isolated” trading strategy (Panel B), at the end of each trading day  $t$ , we again assign stocks into tercile portfolios based on the average fraction of negative words across all articles published on SA about company  $i$  on day  $t$ , but to isolate the effect of SA, the top tercile portfolio based on  $NegSA_{i,t}$  now does not contain stocks that are simultaneously in the top tercile based on  $NegWSJ_{i,t}$  on day  $t$ . The same applies to the bottom-tercile portfolios. For both strategies, stocks are held in the portfolio for one day, two days, or five days. The portfolio of securities in the bottom  $NegSA_{i,t}$ -tercile is referred to as “Bullish”; the portfolio of securities in the top  $NegSA_{i,t}$ -tercile is referred to as “Bearish.” We compute value-weighted portfolio returns. For the “bullish” (“bearish”) portfolio, returns are computed from the closing ask (bid) price on day  $t$  to the closing bid (ask) price on day  $t+1$ , day  $t+2$  or day  $t+5$  to account for the bid-ask spread. T-statistics are computed using Newey-West (1987) standard errors and are reported in parentheses.

|                                      | Tercile            |                    |                    |
|--------------------------------------|--------------------|--------------------|--------------------|
|                                      | 1-day              | 2-day              | 5-day              |
| Panel A: Basic Trading Strategy      |                    |                    |                    |
| “Bullish”                            | 0.069%<br>(1.17)   | 0.147%<br>(1.64)   | 0.181%<br>(1.00)   |
| “Bearish”                            | -0.066%<br>(-0.86) | -0.059%<br>(-0.52) | -0.057%<br>(-0.23) |
| Spread                               | 0.135%<br>(2.05)   | 0.206%<br>(2.27)   | 0.238%<br>(1.47)   |
| Panel B: “Isolated” Trading Strategy |                    |                    |                    |
| “Bullish”                            | 0.060%<br>(1.02)   | 0.142%<br>(1.58)   | 0.207%<br>(1.14)   |
| “Bearish”                            | -0.063%<br>(-0.83) | -0.056%<br>(-0.51) | -0.046%<br>(-0.21) |
| Spread                               | 0.123%<br>(1.92)   | 0.198%<br>(2.21)   | 0.253%<br>(1.77)   |