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

Hailiang Chen, Prabuddha De, Yu (Jeffrey) Hu, and Byoung-Hyoun Hwang[1]

June 2013

Social media has become a popular venue for individuals to share the results of their own security analysis. We conduct textual analysis of articles published on the most popular social-media platform in the United States; we also consider the readers' perspective as inferred via commentaries written in response to these articles. We find that the views expressed in articles and commentaries each substantially contribute in predicting stock returns and earnings surprises, and that the predictability strengthens with the number of commentaries over which readers' views are measured. Together, these findings point to the usefulness of investor-turned-advisor opinions in financial markets.

Keywords: Investor-turned-advisors, Social media, Financial markets.

[1] Chen is from the Department of Information Systems, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong. De and Hwang are from the Krannert School of Management, Purdue University, 403 W State Street, West Lafayette, IN 47907. Hu is from the Scheller College of Business, Georgia Institute of Technology, 800 West Peachtree St. NW, Atlanta, GA 30308. Email: hailchen@cityu.edu.hk, pde@purdue.edu, jeffrey.hu@scheller.gatech.edu, and bhwang@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-Quantitative Group, Korea University, University of Notre Dame, SAC-Quantitative Group, University of Sydney, University of Technology – Sydney, the 3rd Annual Behavioral Finance Conference at Queen's 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 … became commoditized by the use of new tools or access to information, the era of DIY [do-it-yourself] financial analysis is dawning*."[1]

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

"Peer" opinions also have begun to play a greater role in financial markets. Traditionally the domain of professional market participants (e.g., sell-side analysts), financial analysis is increasingly being performed and broadcast by investors themselves. As of 2008, nearly one in four adults in the US report to directly rely on investment advice transmitted via social media outlets (Cogent Research 2008) and regulators conclude that "*social media is landscape-shifting*," with its relevance to financial markets growing continuously (SEC 2012). But do peer opinions actually impart value-relevant news? Or do they merely constitute "random chatter" in a task best left to professional sell-side analysts? The goal of this study is to assess the performance of investor-turned-advisors and to test whether investors can turn to their peers for genuine, useful investment advice.

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

This setting is interesting for several reasons. A product's quality is difficult to capture objectively, which renders an assessment of the value relevance of peer opinions challenging. Prior work examines the related but separate question of how peer opinions predict a product's sales (e.g., Chevalier and Mayzlin 2006; Liu 2006; Zhu and Zhang 2010), but this link could be one of imitation-based herding into a popular, but possibly mediocre product rather than one of sorting based on the product's actual quality. The appealing feature of our setting is that we have a relatively objective measure of performance, in the form of stock returns and earnings surprises, against which users' views can be compared. Earnings, in particular, are not affected by users' opinions and, thus, are highly unlikely to be endogenously determined. A second interesting feature of our analysis is that financial securities are complicated products. Evidence that the collection of opinions posted on a social media outlet can identify "good" or "bad" stocks would, therefore, be particularly impressive.

To examine the role of peer-based advice, we extract user-generated opinions from Seeking Alpha (hereafter, SA; http://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*" (Seeking Alpha 2012). The channels through which investors can voice their opinions and exchange investment ideas are twofold: (1) Users can submit opinion articles to SA, which are reviewed by a panel and subject to editorial changes. If deemed of adequate quality, these articles are then published on the SA website. (2) In response to these articles, any interested potential investor can write a commentary, sharing his or her own view, which might agree or disagree with the author's view on the company in question. SA articles and SA commentaries, over our sample period, are written by around 5,000 and 130,000 different users, respectively.

To quantify the views disseminated through SA, we employ textual analysis. Specifically, we build on prior literature suggesting that the frequency of negative words in an article captures its tone (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 and commentaries.

To preview our results, 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 and SA comments negatively predict subsequent stock returns. The association is statistically significant and economically meaningful. Authors' views, as inferred via articles, and readers' views, as inferred via commentaries, are frequently not aligned, and we find that articles and comments, as a system, predict future stock returns much more strongly than articles alone. In other words, our results imply that both articles and comments are value-relevant and that when authors miss out on an important piece of information, it is frequently picked up by the readers. This pattern is particularly evident when the number of comments is relatively high, which is consistent with the argument that, while the signal by any one individual is relatively noisy, when aggregated, the collective signal becomes informative. The return effect we observe does not revert, and our results hold when augmenting our regression equation with analyst recommendation upgrades/downgrades, positive/negative earnings surprises, and the average fraction of negative words in Dow Jones News Services (DJNS) articles.

To further explore whether SA-views contain value-relevant news or merely incite 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 the reported earnings-per-share (EPS) and the average (or median) of financial analysts' EPS forecasts issued/updated within thirty days prior to the earnings announcement. If opinions expressed through SA are unrelated to firms' fundamentals, or if the information is already fully incorporated by financial analysts into their reported EPS forecasts, then no association should be observed between subsequent scaled earnings surprise and our measure of peer-based advice. In contrast, we find that the fraction of negative words in articles and comments prior to the earnings announcement strongly predict subsequent scaled earnings surprises. These findings suggest that the opinions expressed in SA articles and comments not only affect investor behavior, but also provide value-relevant information beyond that provided by financial analysts.

Together, our findings point to the usefulness and value-relevance of peer-based advice and hint at the possibility that social-media outlets specializing in financial markets may eventually mirror the development of other "bottom-up knowledge generators" such as Wikipedia and the way they have changed how information is produced, evaluated, and disseminated (Tyckoson et al. 2011). The popular press has broached this issue when reporting that social media outlets, through their growing influence among the investor population, are already creating a rivalry with traditional advice sources, such as professional sell-side analysts[2], with far-reaching implications for financial market participants (SEC 2012).

Our study adds to several lines of research. First, by providing evidence that the views reflected in SA help predict future stock returns and earnings surprises, we contribute to the interdisciplinary research on the usefulness of peer-based advice (e.g., Chevalier and Mayzlin 2006; Liu 2006; Chen and Xie 2008; Zhu and Zhang 2010). Our study also relates to the

[2] Examples include: "*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.

literature analyzing the media's effect on the stock market (e.g., Barber and Loeffler 1993; Huberman and Regev 2001; Busse and Green 2002; Tetlock 2007; Engelberg 2008; Tetlock et al. 2008; Fang and Peress 2009; Engelberg and Parsons 2011; Dougal et al 2012, Gurun and Butler 2012, Solomon 2012). In particular, we are the first to examine not only the tone of an article itself, but also how readers actually respond to these articles (via comments). Put differently, while prior literature focuses on journalists and the *production* of information and opinions, here, we also consider how readers *consume* and process these articles, and, in turn, generate their own insights. We provide evidence that the views expressed both in articles and in comments each substantially contribute in predicting stock returns and earnings surprises. Finally, 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. Social media sites make information shared among investors accessible to researchers and, as such, pose an interesting setting to conduct research on social interactions, information exchange/diffusion, and implications for financial markets.

2. Data

2.1 Seeking Alpha (SA)

Articles submitted to SA are reviewed by a panel and subject to editorial changes. The review process is intended to improve the quality of published articles, without interfering with the author's original opinion. Authors are required to disclose their identity and, as of 2010, have to report their holdings on the stocks they discuss. Many authors also maintain their own subscriber-based financial blogs and, as such, have a genuine incentive to produce high-quality research reports, which increase their network of clients and paying subscribers.

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

SA allows any interested investors to not only write and read articles, but also to post commentaries in response to an article. We download all commentaries written in response to the 79,142 single-ticker articles in our sample. 60% of the commentaries are posted on the day of article publication, an additional 20% are posted on the ensuing calendar day, and the remaining 20% are posted sporadically over the ensuing three weeks. In our analysis, we focus on the 256,619 commentaries written in the first two days of article publication. The information we collect about each commentary includes the following items: article id, comment id, main text, date the comment is made, and author name.

To extract authors' opinions, we use the negative words list compiled by Loughran and McDonald (2011), which they designed specifically for use in studies on financial markets. $NegSA_{i,t}$ is the average fraction of negative words across all articles published on SA about company i on day t. $NegSA\text{-}Comment_{i,t}$ is the average fraction of negative words across all SA comments posted over days t to t+1 in response to SA articles about company i on day t, if there were any such comments, and zero otherwise. In our regression analysis, we include *NegSA-*

$Comment_{i,t}$, as well as an indicator variable, $I(NegSA\text{-}Comment_{i,t})$, denoting whether there were any comments posted in response to SA articles discussing company i on day t.

2.2 Dow Jones News Service (DJNS)

To explore whether SA-views have an effect above and beyond news released in more traditional media outlets, we also construct a measure of information revealed by articles published in the DJNS. We access DJNS articles for the stocks covered by single-ticker SA articles via the Factiva database. Since DJNS articles are not tagged by company name or stock ticker, we formulate a search query to find matched news articles for each stock from 2005 to 2011. We start with each company's name as it appears in the CRSP database and require the CRSP company name to show up at least once in the first 50 words of the DJNS news article.[3] The information collected about each DJNS article includes: article title, main text, and date of publication. The DJNS-variable, $NegDJNS_{i,t}$, is the average fraction of negative words across all articles published in the DJNS about company i on day t, if there were any such articles, and zero otherwise. In our regression analysis, we include $NegDJNS_{i,t}$, as well as an indicator variable, $I(DJNS_{i,t})$, denoting whether there were articles published in the DJNS about company i on day t.

Table 1 illustrates a few features of our data. The average length of an SA article is 491 words, which is longer than the average length of a DJNS article (335 words). The average length of comments posted in response to SA articles is 85 words. This length is meaningful and significantly longer than that of messages posted on Internet message boards studied by prior literature, which, according to Antweiler and Frank (2004), "*is most frequently between 20 and 50 [words]*." (p. 1263). The average fraction of negative words used in SA articles is 1.61%; and the average fraction of negative words used in SA comments is 1.82%. In comparison, the average fraction of negative words used in DJNS articles is 1.47%. The fraction of negative

[3] We observe similar results using 25-word and 100-word cutoff points.

words used fluctuates over time, with the fraction being highest in 2008 and 2009 when the stock market performed poorly. The correlation between *NegSA* and *NegSA-Comment* within the subset of observations with comments posted to an SA article is 0.203. In section 4, we examine what factors determine the magnitude of the correlation between *NegSA* and *NegSA-Comment* and the degree to which readers adopt the author's viewpoint on the company in question.

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 and volatility. 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, book-to-market ratio and past returns (Daniel et al. 1997). Because our main variable of interest, $NegSA\text{-}Comment_{i,t}$, is the average fraction of negative words across SA comments posted over days t to t+1, we compute three-months holding-period returns from trading day t+3 to t+60, $ARet_{i,t+3,t+60}$. If the SA article is published on a non-trading day, we move the beginning of the three-months holding-period forward to ensure that the days over which $NegSA_{i,t}$ and $NegSA\text{-}Comment_{i,t}$ are computed do not overlap. Other variables include: $Volatility_{i,t}$, which is the sum of squared daily returns in the calendar month prior to day t; and $ARet_{i,t}$, $ARet_{i,t-1}$, $ARet_{i,t-2}$, and $ARet_{i,t-60,t-3}$, which are abnormal returns on day t, day t-1, day t-2 and cumulative abnormal returns over the three calendar months prior to day t, respectively.

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

Table 2 presents the descriptive statistics of the main variables used in this study. The mean and the median of our abnormal return measure are slightly negative, which is the result of larger firms outperforming smaller firms during our sample period and the abnormal return measure being defined as the difference between raw returns and the *value-weighted* return of a portfolio consisting of stocks with similar size/book-to-market ratio/past returns.

Table 3 reports characteristics of the firms in our sample. The average market capitalization is \$9.2 billion, the average book-to-market ratio is 0.686, the average one-year holding period return is 12%, the average analyst coverage is 9, and the average retail holdings are 28%. In comparison, the average firm in the full CRSP/Compustat sample from 2005 to 2011 has a market capitalization of \$3.4 billion, a book-to-market ratio of 0.832, one-year holding period returns of 8%, an analyst coverage of 5, and retail holdings of 44%. Compared to the average CRSP/Compustat firm, our average sample firm is, therefore, larger, has a higher market-to-book ratio and has higher past stock returns.

3. Main Results

We organize our analysis around the following regression specification:

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

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

Our independent variables are: $NegSA_{i,t}$, which is the average fraction of negative words across all articles published on SA about company i on day t. $NegSA\text{-}Comment_{i,t}$, which is the average fraction of negative words across SA comments posted over days t to t+1 in response to the SA articles, if there were any such comments, and zero otherwise. X includes the following variables as described in Section 2: $NegDJNS_{i,t}$, $I(SA\text{-}Comment_{i,t})$, $I(DJNS_{i,t})$, $Upgrade_{i,t}$, $Downgrade_{i,t}$, $PosES_{i,t}$, $NegES_{i,t}$, $Volatility_{i,t}$, $ARet_{i,t}$, $ARet_{i,t-1}$, $ARet_{i,t-2}$, and $ARet_{i,t-60,t-3}$. 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 linked to firm size, book-to-market ratio and past returns. *T*-statistics are computed using Newey and West (1987) standard errors with sixty lags 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 predict future stock returns. The coefficient estimate on $NegSA_{i,t}$, by itself, in column (1) is -0.287 (*t*-statistic=-2.08), which indicates that future abnormal returns are 0.287% lower when the fraction of negative words in SA articles increases by 1%. When including $NegSA\text{-}Comment_{i,t}$ in the regression equation, the coefficient estimate on $NegSA_{i,t}$ turns to -0.241 (*t*-statistic=-1.77). The coefficient estimate on $NegSA\text{-}Comment_{i,t}$ equals -0.254 (*t*-statistic=-3.05), which implies that future abnormal returns decrease by 0.254% when the fraction of negative words in SA comments increases by 1%. The finding that both $NegSA_{i,t}$ and $NegSA\text{-}Comment_{i,t}$ each substantially contribute in predicting stock returns implies that both sources are value-relevant and that when one source misses out on an important piece of information, it is frequently picked up by the other. In other words, articles and comments, as a system, contain more information and predict returns more accurately than either one considered by itself.

Our results hold whether we control for company news reported in DJNS articles (column 3) or not (column 2). Including leads and lags of the average fraction of negative words in DJNS articles does not alter this observation. The results are also robust to controlling for earnings surprises and analyst upgrades/downgrades. Together, these findings suggest that the association between future abnormal returns and the views expressed in SA does not merely reflect the market's (slow) reaction to company-specific news and/or analyst upgrades/downgrades (e.g., Ball and Brown 1968; Womack 1996).

Our results are also similar when altering the return window. For instance, the coefficient estimate on $NegSA\text{-}Comment_{i,t}$ turns to -0.087 (*t*-statistic=-1.77) when examining how comments predict abnormal returns over the ensuing one month, and to -0.483 (*t*-statistic=-2.22) when examining how comments predict abnormal returns over the ensuing three years. For reference, as reported in Table 4, when examining how comments predict abnormal returns over the ensuing three months, the estimate equals -0.254 (*t*-statistic=-3.05).

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

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

3.1 Number of Comments

In an attempt to better understand the underlying mechanisms, we explore whether the effect here depends on the number of comments over which aggregate views on a stock are computed. While the signal by any one individual is relatively noisy, the wisdom-of-the-crowd effect (e.g., Surowiecki 2005) implies that, when aggregated, the common component in the collective signal becomes valuable. We, therefore, expect $NegSA\text{-}Comment_{i,t}$ to be more informative when computed over a relatively high number of comments. To test our conjecture, we focus on the subset of observations with SA comments and we assign each observation its tercile rank based on the number of comments over which $NegSA\text{-}Comment_{i,t}$ is computed. We then re-estimate our main regression with the addition of this new tercile-rank variable and its interaction term with $NegSA\text{-}Comment_{i,t}$. The average number of comments across the bottom-tercile observations is 1; the average number of comments across the medium-tercile observations is 2.74; and the average number of comments across the top-tercile observations is 16.49.

As reported in Table 5, the regression produces a strong negative slope on the interaction term, suggesting that the predictive power of $NegSA\text{-}Comment_{i,t}$ for future stock-returns is stronger when $NegSA\text{-}Comment_{i,t}$ is computed over many comments. The coefficient estimates on $NegSA\text{-}Comment_{i,t}$ and its interaction term are -0.137 (*t*-statistic=-1.42) and -0.284 (*t*-statistic=-2.11), respectively. These numbers imply that when the average fraction of negative words in SA comments increases by 1%, future abnormal returns for firms in the top tercile are 0.705% lower,[4] whereas future abnormal returns for firms in the bottom tercile are "merely" 0.137% lower.[5] These findings suggest that the number of comments has a strong moderating effect on the predictability of future stock returns.

3.2 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. Table 6 provides evidence on this matter. In particular, we estimate a regression of price-scaled earnings surprise on the fraction of negative words in SA articles and comments. Earnings surprise is the difference between the reported quarterly EPS and the average EPS forecast across all analysts issuing estimates. We do not consider "stale" forecasts issued more than 30 days prior to the earnings announcement. SA-views and earnings consensus forecasts are, thus, computed over the same horizon. We winsorize the absolute value of scaled earnings surprise at the 99th percentile to mitigate the influence of outliers on our results.

Motivated by Tetlock et al. (2008), our independent variables include: (a) the average fraction of negative words across all SA articles about company *i* from thirty days to three days prior to the earnings announcement; (b) the average fraction of negative words across comments

[4] Calculation for firms in the top tercile, i.e., $Rank(\#SA\text{-}Comment_{i,t})$ = 2: (0.137+0.284×2)*1%) = 0.705%.
[5] Calculation for firms in the bottom tercile, i.e., $Rank(\#SA\text{-}Comment_{i,t})$ = 0: (0.137+0.284×0)*1%) = 0.137%.

posted in response to these SA articles – if there are any such comments – and zero otherwise; (c) the average fraction of negative words across all DJNS articles about company *i* from thirty days to three days prior to the earnings announcement – if there are any such articles – and zero otherwise; (d) indicator variables denoting whether any comments were posted in response to SA articles and whether any DJNS article appeared about company *i* from thirty days to three days prior to the earnings announcement; (e) lagged scaled earnings surprise; (f) price-scaled standard deviation of analysts' earnings-per-share forecasts; (g) the logarithm of market capitalization; (h) the logarithm of book-to-market ratio as of December of the calendar year prior to the earnings announcement; and (i) cumulative abnormal return from thirty to three calendar days prior to the earnings announcement. We also include year fixed effects.

If opinions expressed through SA were unrelated to firms' fundamentals, or if information was spurious and fully incorporated by financial analysts into their reported EPS forecasts, no association should be observed between scaled earnings surprise and our social-media-sentiment measure. As reported in Table 6, the coefficient estimate on $NegSA_{i,t\text{-}30,t\text{-}3}$ ranges from -0.208 (*t*-statistic=-2.46) to -0.273 (*t*-statistic=-2.39) depending on the set of control variables, suggesting that when the fraction of negative words in SA articles increases by 1%, subsequent scaled earnings are between 0.208% and 0.273% below "market expectations", as measured by financial analysts' forecasts. For reference, the mean scaled earnings surprise is -0.136% and the median is 0.054%. The coefficient estimate on $NegSA\text{-}Comment_{i,t\text{-}30,t\text{-}3}$ ranges from -0.112 (*t*-statistic=-2.32) to -0.108 (*t*-statistic=-2.34). Together, these findings point to the usefulness of peer-based advice over traditional advice sources.

3.3 Discussion

The results suggest that conversations on SA collectively provide value-relevant information to investors. A natural question that arises from this interpretation is why SA users would be willing to share such value-relevant information with others. Traditional theories predict that an investor 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 frameworks (e.g., Dow and Gorton 1994; Stein 2008; Gray and Kern 2011) propose that investors share their 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 our setting. Writing articles and posting comments can help users confirm, disprove or refine their investment ideas. Also, SA users may lack the capital to push market prices to their perceived true fundamental values. Having established a full position in the asset, SA users can, thus, write an article and post comments to publicize their investment ideas, 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).

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. This study examines how views expressed on a popular social-media site for investors pertain to security prices. We find that the opinions revealed on this site strongly predict future stock returns and earnings surprises, even after controlling for the effect of traditional advice sources, such as financial analysts and newspaper articles. Together, our findings point to the usefulness of peer-based advice in financial markets, and the value-relevance of peer opinions in general.

References

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

Ball, R. and P. Brown, 1968, “An Empirical Evaluation of Accounting Income Numbers,” *Journal of Accounting Research*, 6(2), 159-178.

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

Barber, B. M. and D. Loeffler, 1993, “The ’dartboard’ column: Second-hand information and price pressure,” *Journal of Financial and Quantitative Analysis*, 28(2), 273-284.

Benartzi, S., 2001, “Excessive extrapolation and the allocation of 401(k) accounts to company stock,” *Journal of Finance,* 56(5), 1747-1764.

Busse, J. A. and T. C. Green, 2002, “Market efficiency in real time,” *Journal of Financial Economics*, 65(3), 415-437.

Chevalier J. A. and D. Mayzlin, 2006, “The effect of Word of Mouth on sales: Online book reviews,” *Journal of Marketing Research*, 43(3), 345-354.

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

Cogent Research, 2008, “Social Media’s Impact on Personal Finance & Investing.” Available from: http://www.cogentresearch.com.

Coval, J. D. and T. Shumway, 2005, “Do Behavioral Biases Affect Prices?” *Journal of Finance*, 60(1), 1-34.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, “Measuring Mutual Fund Performance with Characteristic-Based Benchmarks,” *Journal of Finance*, 52(3), 1035-1058.

Das, S. R. and M. Y. Chen, 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., J. M. Piger, and L. M. Sedor, 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 G. Gorton, 1994, “Arbitrage chains,” *Journal of Finance*, 49(3), 819-849.

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

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

Fang, L. and J. Peress, 2009, “Media coverage and the cross-section of stock returns,” *Journal of Finance*, 64(5), 2023-2052.

Friedman, M., 1953, *The case for flexible exchange rates*. University of Chicago Press, Chicago.

Gartner (2010). User survey analysis: Consumer marketing using social network analysis, worldwide.

Gray, W. R. and A. E. Kern, 2011, “Talking your book: Social networks and price discovery,” SSRN eLibrary.

Griffin, J. M., J. Harris, T. Shu, and S. Topaloglu, 2011, “Who Drove and Burst the Tech Bubble?” *Journal of Finance,* 66(4), 1251-1290

Gurun, U. G. and A. W. Butler, 2012, “Don’t Believe the Hype: Local Media Slant, Local Advertising, and Firm Value”, *Journal of Finance,* forthcoming.

Huberman, G. and T. Regev, 2001, “Contagious speculation and a cure for Cancer: A non-event that made stock prices soar,” *Journal of Finance*, 56(1), 387-396.

Kelley, E. K. and P. C. Tetlock, 2012, “How Wise Are Crowds? Insights from Retail Orders and Stock Returns,” *Journal of Finance*, forthcoming.

Kaniel, R., G. Saar, and S. Titman, 2008, “Individual Investor Trading and Stock Returns,” *Journal of Finance*, 63(1), 273-310.

Kaniel, R., S. Liu, G. Saar, and S. Titman, 2012, “Individual Investor Trading and Return Patterns around Earnings and Announcements,” *Journal of Finance,* 67(2), 639-680.

Li, F., 2008, “Annual report readability, current earnings, and earnings persistence,” *Journal of Accounting and Economics*, 45(2-3), 221-247.

Liu, Y., 2006, “Word of Mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, 70(3), 74-89.

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

Newey, W. K. and K. D. West, 1987, “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica*, 55(3), 703-708.

Odean, T., 1998, “Are Investors Reluctant to Realize Their Losses?” *Journal of Finance*, 53(5), 1775-1798.

SEC, 2012, “Investment adviser use of social media,” *National Examination Risk Alert*, 2(1), 1-7.

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

Solomon, David H., 2012, “Selective Publicity and Stock Prices”, *Journal of Finance*, forthcoming.

Stein, J. C., 2008, “Conversations among competitors,” *American Economic Review*, 98(5), 2150-2162.

Surowiecki, J., 2005, *The Wisdom of Crowds*. New York: Anchor Books.

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

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

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

Womack, K., L., 1996, “Do Brokerage Analysts’ Recommendations Have Investment Value?” *Journal of Finance*, 51(1), 137-167.

Zhu, F. and X. M. Zhang, 2010, Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2), 133-148.

Table 1. Descriptive Statistics of Seeking Alpha and Dow Jones News Service Articles

This table reports summary statistics for single-ticker *Seeking Alpha* (SA) articles, SA comments written in response to single-ticker SA articles and *Dow Jones News Service* (DJNS) articles. The sample period is 2005-2011.

	2005	2006	2007	2008	2009	2010	2011	All Years
Panel A: Seeking Alpha (SA) Articles								
Total # Stock tickers	479	1,725	2,851	2,748	2,674	2,841	2,925	6,772
Total # Articles	1,611	5,725	13,830	13,047	15,692	13,903	15,334	79,142
Avg. # Words per article	305	414	424	458	470	512	628	491
Avg. % Negative words	1.08%	1.44%	1.47%	1.73%	1.81%	1.59%	1.58%	1.61%
StDev % Negative words	1.15%	1.22%	1.20%	1.28%	1.32%	1.26%	1.16%	1.25%
Panel B: Seeking Alpha (SA) Comments								
Total # Stock tickers	26	302	860	1,593	1,747	1,819	2,079	4,205
Total # Comments	67	917	7,361	43,299	61,807	56,262	86,906	256,619
Avg. # Words per comment	104	95	95	89	89	81	80	85
Avg. % Negative words	2.00%	1.61%	1.69%	1.80%	1.95%	1.83%	1.74%	1.82%
StDev % Negative words	4.50%	2.20%	2.50%	2.70%	2.80%	2.72%	2.76%	2.75%
Panel C: Dow Jones News Service (DJNS) Articles								
Total # Stock tickers	2,462	2,926	3,101	2,928	1,961	2,387	2,066	3,887
Fraction of stock tickers covered by SA, but NOT by the DJNS	53.03%	66.73%	63.14%	58.15%	45.06%	56.60%	43.80%	54.77%
Total # Articles	33,355	51,620	49,162	35,987	27,739	38,052	30,810	266,725
Avg. # Words per article	357	267	260	329	404	400	411	335
Avg. % Negative words	1.37%	1.25%	1.27%	1.54%	1.76%	1.65%	1.68%	1.47%
StDev % Negative words	1.41%	1.44%	1.46%	1.50%	1.46%	1.51%	1.52%	1.48%

Table 2. Summary Statistics – Firm/Trading Day Level

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

	N	Mean	Std. Dev	25th Pctl	50th Pctl	75th Pctl
$ARet_{i,t+2,t+60}$	33,800	-0.004	0.190	-0.089	-0.009	0.076
$ARet_{i,t}$	33,800	0.001	0.061	-0.012	0.000	0.012
$ARet_{i,t-1}$	33,800	0.001	0.056	-0.011	0.000	0.012
$ARet_{i,t-2}$	33,800	0.001	0.044	-0.011	0.000	0.010
$ARet_{i,t-60,t-3}$	33,800	0.007	0.284	-0.096	-0.009	0.083
$Upgrade_{i,t}$	33,800	0.030	0.169	0.000	0.000	0.000
$Downgrade_{i,t}$	33,800	0.034	0.179	0.000	0.000	0.000
$PosES_{i,t}$	33,800	0.057	0.231	0.000	0.000	0.000
$NegES_{i,t}$	33,800	0.018	0.133	0.000	0.000	0.000
$Volatility_{i,t}$	33,800	0.029	0.362	0.004	0.009	0.020

Table 3. Summary Statistics – Firm/Calendar Year Level

This table reports summary statistics of various firm characteristics. The observations are on a firm/year level. Every year *t* (from 2005 to 2011), we compile a list of firms in our sample with a *Seeking Alpha* article in year *t*. We then compute the respective firms' characteristics as of December. *Size* is the firm's market capitalization in millions. *BM* is the firm's book-to-market ratio. *Past Return* is the firm's cumulative one-year return. *Coverage* is the firm's analyst coverage, which is set equal to zero if the firm is not covered by any analysts in that year. *Retail Holdings* is one minus the fraction of shares held by institutional investors.

	N	Mean	Std. Dev	25^{th} Pctl	50^{th} Pctl	75^{th} Pctl
Size	7,478	9,201	28,102	325	1,374	5,717
BM	7,478	0.686	1.181	0.267	0.468	0.785
Past Return	7,478	0.120	1.364	-0.282	0.023	0.314
Coverage	7,478	9.001	7.870	2.000	7.000	14.000
Retail Holdings	7,478	0.284	0.277	0.095	0.226	0.432

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 and comments. The sample period is 2005-2011. Abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market/past return-characteristics from t+3 to t+60, where t is the day of article appearance or the ensuing trading day if the article is published on a non-trading day. $NegSA_{i,t}$ is the average fraction of negative words across all articles published on SA about company i on day t. $NegSA\text{-}Comment_{i,t}$ is the average fraction of negative words across SA comments posted over days t to t+1 in response to the SA articles, if there were any such comments, and zero otherwise. $NegDJNS_{i,t}$ is the average fraction of negative words across all articles published in the DJNS about company i on day t, if there were any such articles, and zero otherwise. $I(SA\text{-}Comment_{i,t})$ and $I(DJNS_{i,t})$ are indicator variables denoting whether there were comments posted on SA articles and whether there were articles published in the DJNS. We include year-month fixed effects. Other independent variables are as described in Table 2. *T*-statistics are computed using Newey-West (1987) standard errors with 60 lags and are reported in parentheses.

	(1)	(2)	(3)
$NegSA_{i,t}$	-0.287	-0.241	-0.218
	(-2.08)	(-1.77)	(-1.64)
$NegSA\text{-}Comment_{i,t}$		-0.254	-0.252
		(-3.05)	(-3.02)
$I(SA\text{-}Comment_{i,t})$		0.000	0.000
		(0.14)	(0.14)
$NegDJNS_{i,t}$			-0.354
			(-1.37)
$I(DJNS_{i,t})$			0.006
			(1.16)
$Upgrade_{i,t}$	0.003	0.003	0.003
	(0.51)	(0.48)	(0.47)
$Downgrade_{i,t}$	-0.003	-0.003	-0.003
	(-0.58)	(-0.55)	(-0.50)
$PosES_{i,t}$	0.008	0.007	0.007
	(2.19)	(2.08)	(1.57)
$NegES_{i,t}$	-0.011	-0.011	-0.010
	(-1.43)	(-1.48)	(-1.26)
$Volatility_{i,t}$	-0.011	-0.011	-0.011
	(-1.59)	(-1.60)	(-1.59)
$ARet_{i,t}$	0.011	0.010	0.009
	(0.26)	(0.25)	(0.23)
$ARet_{i,t-1}$	-0.066	-0.066	-0.066
	(-1.90)	(-1.91)	(-1.91)
$ARet_{i,t-2}$	0.018	0.017	0.017
	(0.38)	(0.36)	(0.35)
$ARet_{i,t-60,t-3}$	-0.016	-0.016	-0.016
	(-1.65)	(-1.67)	(-1.69)
# Obs.	33,800	33,800	33,800
Adj. R^2	1.28%	1.32%	1.32%

Table 5. Seeking Alpha, Abnormal Returns and Number of SA Comments

This table reports coefficient estimates from regressions of abnormal returns on measures of the views reflected in *Seeking Alpha* (SA) articles and comments. The sample period is 2005-2011. The sample consists of observations with SA comments. Abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market/past return-characteristics from t+3 to t+60, where t is the day of article appearance or the ensuing trading day if the article is published on a non-trading day. $NegSA_{i,t}$ is the average fraction of negative words across all articles published on SA about company i on day t. $NegSA\text{-}Comment_{i,t}$ is the average fraction of negative words across SA comments posted over days t to t+1 in response to the SA articles, if there were any such comments, and zero otherwise. $Rank(\#SA\text{-}Comment_{i,t})$ is the tercile rank of number of SA-comments posted. $NegDJNS_{i,t}$ is the average fraction of negative words across all articles published in the DJNS about company i on day t, if there were any such articles, and zero otherwise. $I(SA\text{-}Comment_{i,t})$ and $I(DJNS_{i,t})$ are indicator variables denoting whether there were comments posted on SA articles and whether there were articles published in the DJNS. Other independent variables are as described in Table 2. *T*-statistics are computed using Newey-West (1987) standard errors with 60 lags and are reported in parentheses.

	(1)	(2)
$NegSA_{i,t}$	-0.253	-0.223
	(-1.26)	(-1.17)
$NegSA\text{-}Comment_{i,t}$	-0.137	-0.138
	(-1.42)	(-1.44)
$NegSA\text{-}Comment_{i,t} * Rank(\#SA\text{-}Comment_{i,t})$	-0.284	-0.275
	(-2.11)	(-2.07)
$Rank(\#SA\text{-}Comment_{i,t})$	0.007	0.007
	(2.21)	(2.23)
$NegDJNS_{i,t}$		-0.357
		(-0.85)
$I(DJNS_{i,t})$		0.003
		(0.45)
$Upgrade_{i,t}$	0.010	0.011
	(1.03)	(1.07)
$Downgrade_{i,t}$	-0.002	-0.001
	(-0.20)	(-0.08)
$PosES_{i,t}$	0.012	0.014
	(1.98)	(1.89)
$NegES_{i,t}$	-0.010	-0.007
	(-0.62)	(-0.41)
$Volatility_{i,t}$	-0.016	-0.016
	(-1.88)	(-1.88)
$ARet_{i,t}$	0.005	0.005
	(0.10)	(0.09)
$ARet_{i,t-1}$	-0.067	-0.068
	(-1.31)	(-1.32)
$ARet_{i,t-2}$	-0.014	-0.015
	(-0.24)	(-0.24)
$ARet_{i,t-60,t-3}$	-0.028	-0.028
	(-1.95)	(-1.99)
# Obs.	14,863	14,863
Adj. R^2	2.29%	2.29%

Table 6. Seeking Alpha and Earnings Surprises

We estimate a regression of price-scaled earnings surprise on measures of the views reflected in *Seeking Alpha* (SA). The sample period is 2005-2011. Earnings surprise is the difference between reported quarterly EPS and the consensus EPS forecast across all analysts issuing earnings estimates from thirty to three calendar days prior to the earnings announcement. $NegSA_{i,t-30,t-3}$ is the fraction of negative words in SA articles about company *i* from thirty to three days prior to the earnings announcement on day *t*. $NegSA\text{-}Comment_{i,t-30,t-3}$ and $NegDJNS_{i,t-30,t-3}$ are the fraction of negative words in SA comments in response to the SA articles and the fraction of negative words in *Dow Jones News Service* articles about company *i* from thirty to three days prior to the earnings announcement; if there were no comments or no DJNS articles, these variables equal zero. $I(SA\text{-}Comment_{i,t-30,t-3})$ and $I(DJNS_{i,t-30,t-3})$ are indicator variables denoting whether there were any SA comments posted and whether there were any DJNS articles written about company *i* thirty to three days prior to the earnings announcement. $Lagged(DependentVar_{i,t})$ is the price-scaled earnings surprise (our dependent variable) from the previous quarter. $ForecastDispersion_{i,t}$ is the price-scaled standard deviation of analysts' EPS forecasts. $ln(MarketCapital_{i,lagged})$ is the logarithm of the market capitalization as of the quarterly earning's corresponding fiscal quarter end. $ln(Book/Market_{i,lagged})$ is the logarithm of the book-to-market ratio as of the most recent fiscal year end. $PastReturn_{i,t-30,t-3}$ is the cumulative stock market performance from thirty to three calendar days prior to the earnings announcement. We include year fixed effects. *T*-statistics are reported in parentheses and clustered by year.

	(1)	(2)	(3)
$NegSA_{i,t-30,t-3}$	-0.273	-0.245	-0.208
	(-2.39)	(-2.55)	(-2.46)
$NegSA\text{-}Comment_{i,t-30,t-3}$		-0.112	-0.108
		(-2.32)	(-2.34)
$I(SA\text{-}Comment_{i,t-30,t-3})$		-0.004	-0.004
		(-1.80)	(-1.77)
$NegDJNS_{i,t-30,t-3}$			-0.167
			(-1.00)
$I(DJNS_{i,t-30,t-3})$			-0.001
			(-0.39)
$Lagged(DependentVar_{i,t})$	0.194	0.195	0.194
	(1.45)	(1.48)	(1.48)
$ForecastDispersion_{i,t}$	0.408	0.417	0.421
	(1.17)	(1.20)	(1.20)
$ln(MarketCapital_{i,lagged})$	0.001	0.002	0.002
	(2.62)	(2.63)	(2.48)
$ln(Book/Market_{i,lagged})$	0.004	0.004	0.004
	(1.58)	(1.64)	(1.67)
$PastReturn_{i,t-30,t-3}$	0.007	0.007	0.006
	(1.81)	(1.81)	(1.75)
# Obs.	2,681	2,681	2,681
Adj. R^2	8.18%	8.58%	8.78%