EB5204 NEW MEDIA AND SENTIMENT MINING

SOCIAL MEDIA AND STOCK PERFORMANCE

Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media
Hailiang Chen, Prabuddha De, Yu (Jeffrey) Hu and Byoung-Hyoun Hwang
The Review of Financial Studies, Vol. 27, No. 5 (May 2014), pp. 1367-1403

SUBMISSION DATE: 05 NOVEMBER 2018

GROUP 04
SIDDHARTH PANDEY
PRANSHU RANJAN SINGH
NYON YAN ZHENG
TAN KOK KENG

INSTITUTE OF SYSTEMS SCIENCE
NATIONAL UNIVERSITY OF SINGAPORE

Summary

Issac Newton's remarks "I can calculate the motion of heavenly bodies, but not the madness of people" after he lost nearly all his life savings by being swept by the wild enthusiasm of the crowd, is an excellent example of the influence crowd's opinion has on a callow investor. Fast forward to the 21st century, with the growing rise of social media platforms, it is now easier than ever to disseminate ideas and utilize public opinions through commentaries and forum discussions. In the early 2000s, as chat boards rose to popularity many retail investors could discuss their strategies and share opinions about the market conditions. But the utility of such casual and unstructured discussions and if these chatters actually captured any market hunch still needs to be established. Driven by the aforementioned question of news or noise, Tumarkin & Whitelaw (2001), analyzed forum threads of the website Raging Bull. The study revealed a lack of association between message board activity and industry-adjusted return or abnormal trading volume. Similar investigations performed by Antweiler and Frank (2004) and Das and Chen (2007), also suggest that the opinions transmitted through social media outlets do not predict stock returns. In the last few years, a number of people are turning towards more sophisticated platforms like Seeking Alpha (SA), Ayondo and StockTwits which follow pyramid structure of knowledge dissemination. Opinions submitted by gurus often have rigorous market analysis to strengthen them, and the aftermath discussions are more focused in contrast to message board chatters. The changing landscape undermines the conclusions of the previous researches. The paper "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media" by Chen Hailiang et al. aims to find answers to the same questions and reassess the utility of social media for providing reliable financial advice.

The findings in this paper suggest that stock return predictability is not only significant but also economically meaningful. This contrasts with the previous findings and points towards a shift in the way investment related social media outlets influence the decision making. The influence of control variables explored in the study is statistically significant in predicting stock returns and earnings surprises. The authors suggest that this influence can be realized either through a predictability channel or clout channel. Though not entirely confident, the tone of the article and interpretation of findings hint towards predictability channel. The correcting behaviour of SA readers and decrease in popularity and article views of an inconsistent author raises apprehension over the existence of a significant influence of the clout channel. Ensuing from this discussion, it can be suggested that the SA articles and SA commentaries contain pieces of value-relevant information, which, as of the article publication date, are not fully factored into the price. Finally, in the end, the findings reported in the paper establishes the usefulness and value relevance of peer-based advice in the investment domain.

Background

The rise of social media has seen the impact of peer opinions in business such as restaurants, hotels, movies, e-commerce websites. These businesses are heavily influenced by peer reviews. Such reviews capture the sentiment of the crowd and clout the decision of peers. However, the impact of peer opinions on financial markets is still unknown and evolving. Influence of peer-based advice from user-generated content is increasing, while the influence of traditional advice sources is decreasing, peer opinions have begun to play a greater role in financial markets (one in four in the USA).

The aim of the paper was to investigate the extent to which investor opinions are transmitted through social media and its effect on the prediction of future stock abnormal returns and earnings surprises. The analysis is performed mainly over published articles and commentaries on SA.

Some previous works had focused on analysis to find some associations between conversations on Internet message boards and stock returns. Tumarkin and Whitelaw (2001) detect no association between the two. Antweiler and Frank (2004 found a statistically significant, yet economically meaningless association. Das and Chen (2007) found no strong relationship from sentiment to stock prices on average across the individual stocks. This paper, in contrast, detects that there is both statistically significant and economically meaningful association between peer stock opinions and stock return predictability. The differences in findings with other works can be attributed to narrower sample size, short messages on message boards and still evolving social media in the early 2000s.

Data Description

The data sources for this study include SA articles, SA commentaries, Dow Jones News Service (DJNS) articles. Compustat financial statement data and Center for Research in Security and Prices (CRSP) financial market data. The sample period for the collected data was from 2005 to 2012.

A total of 97,070 single-ticker (focusses on only one stock) articles were scraped from SA website and extracted information was stored in a MySQL database. The extracted information included article ID, article title, main text, date of publication, author name and stock ticker. A total of 459,679 commentaries written in first two days of article publication were also scraped and extracted information such as article ID, comment ID, main text, date of the comment, and author name was stored in the database. The list of negative words was compiled by Loughran and McDonald (2011) was used for this analysis.

The DJNS articles were accessed based on the stocks covered by SA articles. A search query was formulated to find the matched news articles for each stock. The Factiva coding of company names, obtained from the CRSP database, were matched with the first 50 words of DJNS news articles. A total of 322,046 articles were obtained from the matching.

Methodology and Main Findings

To gain a complete 360 view of association among social media and stock performance, various control variables were extracted from the dataset. The association (correlation) among these control variables and stock returns was obtained using a Linear Regression analysis. The analysis can be broadly classified into multiple parts. The first part attempts to find **whether peer opinions on social media can be used to predict stock returns**. The authors also investigated whether the prediction of future stock market performance was due to value or naïve investor reaction by conducting a regression to see the association with subsequent earnings surprises. The second part investigates **whether such peer opinions can be trusted as a source of investment advice**.

The fraction of negative words in SA articles and fraction of negative words in SA comments over two days upon article publication were computed for all articles. The stock returns after three months of publication of each article were measured. Upon linear regression analysis, it was observed that the stock returns lowered by 0.332% and 0.194% with every 1% increase in the fraction of negative words in articles and comments respectively. Thus, it was observed that the fraction of negative words in SA articles and comments, both negatively predict stock returns over the ensuing three months.

To investigate the effectiveness of correlation among the fraction of negative words in articles and comments with stock returns, the analysis was extended for different holding periods of stocks. It was the predictability of stock return, using both articles and comments, increases with the length of holding period. The reason for such observation can be attributed to under-reaction for articles on social media. Financial analyst received more attention from investors, hence their opinions are incorporated into the market at a faster pace.

To investigate the impact of the magnitude of negative words transmitted through SA articles and comments, all the SA articles and comments were clustered in quintiles based on the fraction of negative words. The difference between the three months stock returns from the highest and lowest fraction of negative words quintiles were measured. The observed difference in returns were 2.6 bps and 2.2 bps for articles and comments respectively. A similar analysis was performed using quartiles instead of quintiles and the results were consistent with each other. Additionally, the authors formed two hypothetical portfolios, one each for signals from articles and comments, to long and short stocks in the bottom and top quintiles (and quartiles) respectively over 6 years. The returns were 50% for articles (approximately 7% p.a.) and 40% for comments (approximately 5.7% p.a.) were observed.

Furthermore, the SA articles were split into terciles based on the number of comments posted for that article. The average number of comments in high, medium and low tercile were 20.63, 4.18 and 1.39 respectively. For every 1% increase in the fraction of negative words in comments, the stock returns were 0.512%, 0.316% and 0.120% lower for high, medium and low terciles respectively. Thus, articles with a higher number of comments have higher predictive power on future returns.

The earnings surprise is measured as the difference between the reported earnings per share (EPS) and the average of financial analysts' EPS forecasts issued within 30 days prior to the earnings announcement. A linear regression analysis was performed to investigate the correlation between the fraction of negative words in SA articles and comments and the earnings surprise. The earnings surprises were between 0.232% and 0.266% further below market expectations when the fraction of negative words in SA articles was 1% higher. It suggests that the fraction of negative words in articles and comments strongly predict subsequent earnings surprises. The earnings surprises for the articles containing the word "earn" were between 0.267% and 0.306% below market expectations while earning surprises for articles that didn't contain the word "earn" were between 0.193% and 0.209% below market expectations, when the fraction of negative words in the articles was 1% higher. It suggests that the articles containing the word "earn" have even stronger predictive power while predicting subsequent earnings surprises. Therefore, the authors concluded that opinions transmitted via SA impart value-relevant information that provides a meaningful platform for people to help each other make more informed investment decisions.

Opinions on social media platforms such as SA can be trusted as a valuable source of investment advice. Such platforms allow the SA contributor incentives in from earnings ($10 per 1000-page views). The articles posted on the SA website pass through a review panel before being published. Furthermore, other peers can post comments on the article which would discourage malignant articles from SA contributors.

A bullish article has the fraction of negative words below the median of its overall distribution, whereas a bearish article has the fraction of negative words above the median. An article is considered "consistent" if a bullish (bearish) article is followed by positive (negative) abnormal returns. Page-views measures the number of article visits by peers across all articles written by the author. Read-to-ends measures the number of articles visits where the reader read the entire article. A linear regression analysis was performed to check for correlation between the

fraction of consistent articles with page-views and read-to-ends. When the fraction of consistent articles increased by 10%, the page-views increased by 4915, whereas read-to-ends increased by 2775, suggesting both page-views and read-to-ends increases with the author's historical level of consistency. Thus, it can be observed that the intelligent followers can differentiate between authors that offer historically good versus bad advice and the popularity of these authors changes accordingly.

To investigate whether the followers (readers) are able to distinguish between consistent and inconsistent authors, the correlation between author-track record and follower disagreement is evaluated. Author track record can be measured using the consistency of articles written, article length, etc. whereas follower disagreement is measured as the absolute difference between the fraction of negative words in the article and across all comments. The linear regression analysis shows that such a correlation exists and is significant. This suggests that followers are more prone to challenge the author's viewpoint when the author writes a more negative article. Followers disagree with authors more when their articles have been inconsistent. Longer articles face less disagreement from followers as such articles tend to be detailed and are able to convince the followers. Another observation was that, for an author with a poor track record, the tone of the comments reliably predicts the stock market performance.

Conclusion

The authors find that the opinions revealed on SA strongly predict stock returns and earnings surprises. As trust and influence of new media grow, the effect will be even more substantial, with a greater number of people turning towards peer-based advice platforms for decision making. The authors predict 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 change the way information is produced, evaluated and disseminated.

Suggestions for improvement to topic

The research methodology followed considers the impact of two articles with an equal number of negative words as equal. This might not necessarily be the case, as the impact of an article will vary with its tone and also depend upon the consistency and reputation of the author. A weighted measure should be employed to account for this variation.

Sources from other popular outlets Ayondo, StockTwits should also be assimilated to strengthen the findings of the paper. An investigative study on predictability power of the different social media outlets can also be undertaken.