2012

Will Twitter Make You a Better Investor? A Look at Sentiment, User Reputation and Their Effect on the Stock Market

Eric D. Brown
Dakota State University, edbrown@dsu.edu

Follow this and additional works at: http://aisel.aisnet.org/sais2012

WILL TWITTER MAKE YOU A BETTER INVESTOR? A LOOK AT SENTIMENT, USER REPUTATION AND THEIR EFFECT ON THE STOCK MARKET

Eric D. Brown
Dakota State University
edbrown@dsu.edu

ABSTRACT

The use of social networks like Twitter and Facebook has grown exponentially over the last few years. Twitter, which was founded in 2006, had an estimated 200 million users on January 1 2011 with more than 95 million tweets sent per day. With this rapid growth and significant adoption, Twitter has become an important tool for businesses and individuals to communicate and share information. In addition, Twitter has rapidly grown as a medium to share ideas and thoughts on investing decisions. This research builds on prior published research and attempts to determine whether there is correlation between twitter and the stock market by studying sentiment, message volume, price movement and stock volume as well as the affect that a twitter user's reputation may have on sentiment and the stock market.

Keywords

Sentiment analysis, decision support, knowledge sharing, Twitter

INTRODUCTION

The use of social networks like Twitter and Facebook has grown exponentially over the last few years. Twitter, which was founded in 2006, had an estimated 200 million users on January 1 2011 with more than 95 million tweets being sent per day (Chiang, 2011). With this rapid growth and significant adoption, Twitter has become an important tool for businesses and individuals to communicate and share information. Although a good portion of the messages shared on Twitter might be considered as 'noise' to many, there are communities within Twitter that have begun to use this system to share information relative to their interests. One of these communities has grown up around the area of the stock market where investors and traders share their thoughts on trade and investing ideas.

Recent research involving data mining and sentiment analysis of Twitter messages has produced interesting results. Researchers have been studying Twitter messages to attempt to gauge whether there is any measurable and useful sentiment and to gather opinions of Twitter users toward companies, products and services (Bifet and Frank, 2010; Bollen, Mao and Zeng, 2010; Pak and Paroubek, 2010; Romero, Meeder and Klienberg, 2010; Sprenger and Welpe, 2010; Vincent and Armstrong, 2010).

This research project expands upon previous research on sentiment analysis and attempts to measure whether there is any correlation between Twitter sentiment and the stock market; and more importantly, if that correlation exists what might be causing that correlation. More specifically, this study will attempt to objectively determine whether an analysis of messages found in publicly available Twitter streams can be mined and analyzed for sentiment. This sentiment will then be used to as an input into various modeling techniques to determine if sentiment can be used within a model to assist with the decision making process for investing decisions.

This research builds upon previous research projects using sentiment analysis on discussion forums as well as Twitter messages. For example, Bollen et al. (2010) used Twitter data mining and sentiment analysis to determine the mood of the Twitter population to predict the movement of the Dow Jones Industrial Average (DJIA) with a claimed 87.6% accuracy. Sprenger and Welpe (2010) have used sentiment analysis of tweets gathered for the top 100 stocks of the Standard & Poor's index (S&P 100) and were able to show a consistent correlation between Twitter sentiment and stock market returns and between Twitter message volume and stock market volume.

This research will attempt to further understand whether Twitter messages are leading or lagging the movement in the stock market and what affect the volume of Twitter messages has on a stock of a specific firm or an entire sector. Rather than focus on the macro Twitter environment and public mood like Bollen et al. (2010) or the S&P 100 in a similar approach to Sprenger and Welpe (2010), this research will focus on a specific sector of the stock market by tracking an Exchange Traded Fund (ETF) and the stocks that make up that sector upon which the ETF is based.

More formally, this research will attempt to: 1) determine whether the sentiment for a sector as a whole matches the weighted sentiment of the companies that make up the sector; 2) determine if there are times or days that provide more useful sentiment gathered from Twitter messages; 3) determine whether a poster's reputation or number of Twitter followers affects the contribution of that user's sentiment towards a stock or sector ; and 4) determine whether a 'Retweet' (someone re-sending a previously sent tweet) has a significant effect on sentiment scores for a stock or sector and, if so, whether these Retweets result in a form of 'artificial' sentiment for a stock or sector by (i.e., can someone use the idea of Retweets to artificially increase / decrease the sentiment of a stock?).

LITERATURE REVIEW

The automated use of natural language processing and computational linguistics has grown in popularity in recent years with researchers studying sentiment analysis techniques (Abbasi, Chen and Salem, 2008; Berger, Della Pietra and Della Pietra, 1996; Boiy and Moens, 2009; Choi, Kim and Myaeng, 2009; Lin and He 2009; Narayanan, Liu and Choudhary, 2009; Pang, Lee and Vaithyanathan, 2002; Whitelaw, Garg and Argamon, 2005) and the application of those techniques to various domains including movie reviews (Pang et al., 2002; Thet, Na, Khoo and Shakthikumar, 2009), general opinion mining (Choi et al., 2009; Hui and Gregory, 2010; Hursman, 2010; Pak et al., 2010; Romero et al., 2010; Tucker, 2010) and even attempts to predict the movement of the stock market (Antweiler and Frank, 2004; Bollen et al., 2010; Chua, Milosavljevic and Curran, 2009; Delort, Arunasalam, Milosavljevic and Leung, 2009; Gu, Konana, Liu, Rajagopalan and Ghosh, 2006; Sprenger et al., 2010; Tumarkin and Whitelaw, 2001; Wysocki, 1998; Zhang, 2009).

Wysoki (1998) conducted a research project to study stock market message boards for over 3,000 stocks to determine if a correlation between discussion board message volume and message quality had any effect on the volume or price for a stock. A key contribution of this research shows a strong positive correlation between the volume of messages posted on the discussion boards during the hours that the stock market is closed (between 4:01 PM and 8:29 AM weekdays) and the next trading day's volume and stock returns. The researchers report that a tenfold increase in message postings in the overnight hours led to an average increase in the next day's stock grade volume of approximately 15.6% and a 0.7% increase in next day stock returns (Wysocki 1998).

Tumarkin and Whitelaw (2001) studied the effect of Internet stock forum postings for use in predicting stock returns or stock trading volume. For this project, the researchers used a popular website at the time to study the Internet Service sector. They analyzed the messages posted for specific companies to determine if any predictive features could be found by looking at the volume of messages on these boards as well as running sentiment analysis techniques of messages posted.   Their research showed that there was no conclusive predictive capabilities found within message board activity (Tumarkin et al., 2001).

In a similar research project, Antweiler and Frank (2004) studied messages posted on stock-related message boards and how those messages might affect movements in the stock market. The researchers examined approximately 1.5 million messages posted on two message boards for 45 companies and performed text classification and sentiment analysis to understand the sentiment of each message. The researchers were able to show a strongly positive correlation between message board posts, trading volume, trading volatility and a minor correlation between messages and price activity on the following day (Antweiler et al., 2004).

Das and Chen (2007) added to the research in the field by investigating a more formal approach to the use of sentiment analysis techniques when applied to internet stock message boards. Previous research had used manual classification techniques or simple text classification algorithms to assign a 'buy', 'sell' or 'hold/neutral' signal to messages. While these approaches delivered acceptable results, the researchers set out to find a more automated and more robust classification technique to classify message as 'bullish', 'bearish' or 'neutral'. While the classification technique is interesting and worth further study, a key outcome of this study shows no significant correlation between sentiment and individual stock price movements. Additionally, the researchers were able to show a positive correlation between the aggregate sentiment of a set of stocks and the aggregate movement of those stocks (Das et al., 2007).

Research by Gu et al. (2006) as well as Zhang (2009) takes a slightly different approach by focusing on the reputation of the message poster rather than purely on the message sentiment or message volume. Both of these studies reported that a users' reputation can be helpful when determining whether to use their comments in formulas or algorithms when determining sentiment (Gu et al., 2006; Zhang, 2009).

While these previous research projects used stock message discussion board, there have been recent attempts to use other forms of media. Researchers have used blogs, Twitter and other social media outlets to determine sentiment of a stock, sector and or the market as a whole. For example, O'Hare et al. (2009) used sentiment analysis techniques to classify text within financial blogs. Bollen, Mao and Zeng (2010) have conducted a study using sentiment analysis of Twitter messages to determine public mood and Sprenger and Welpe (2010) use sentiment analysis of Twitter messages to determine sentiment

towards individual stocks. The methodologies used by Bollen et al. (2010) and Sprenger and Welpe (2010) will be applied to the proposed research.

Bollen et al. (2010) report on the use of sentiment analysis of a large corpus of Twitter messages to determine the 'mood' of the Twitter population on a given day. This 'mood' is then used as an input into a neural network prediction engine to predict the movement of the stock market on the following day with a reported 87.6% accuracy of prediction of the Dow Jones Industrial Average (Bollen et al. 2010). While this research project shows a correlation between sentiment gathered via Twitter and market movements, the researchers used massive amounts of data from the entire Twitter population in an attempt to understand the overall sentiment of the Twitter population, rather than the sentiment of the population specifically directed toward the stock market.

Sprenger and Welpe (2010) have taken a more targeted approach by focusing on the Standard & Poor's top 100 stocks, known as the S&P 100, and gathering tweets corresponding to these . These Twitter messages are analyzed to study whether sentiment of a company on Twitter has any correlation to the movement in price or volume. The researchers took a novel approach to reduce the large amount of 'noise' on Twitter by following the dollar symbol ('$') preceding the stock market symbol nomenclature popularized by the Stocktwits.com website and its users. This nomenclature allowed the researchers to focus on tweets that had been created and shared by only those people with an interest in the stock market. The outcome of the research shows that the sentiment of a company on Twitter closely follows market movements and that message volume on Twitter is positively correlated to the trading volume for that stock (Sprenger et al. 2010).

THEORETICAL FOUNDATION

While the research methods of this project are grounded in data mining, sentiment analysis and statistical computation, a theoretical foundation to explain the underlying behavioral finance foundation of the use of Twitter messages and the sentiment contained within these messages.

The Adaptive Market Hypothesis (AMH) presented by Andrew Lo (2004) presents an interesting framework for the foundation of this research. The AMH provides a modified approach to the Efficient Market Hypothesis (EMH) by asserting that competition, learning and a process of evolutionary selection provides pressure on the market to drive prices to 'efficient' levels (Neely, Weller and Ulrich, 2009). In addition, the AMH provides for the existence of profit opportunities in the marketplace and provides a mechanism for learning and competition to gradually remove these opportunities from the market (Neely et al., 2009).

By using the AMH as a theoretical foundation, this research hopes to be able to show that Twitter is contributing to the competition and learning process that drives prices.

RESEARCH METHODS

To conduct this research, it is necessary to have access to the Twitter 'stream' in order to access, download and store tweets pertaining to the sectors and companies that will be studied. Thankfully, Twitter provides an application programming interface (API) that can be used to access many aspects of the Twitter system, including user information and individual tweets along with timestamps for those tweets (Twitter 2011a). The "track" method of the Twitter Streaming API will be used to track stock symbols for the sectors and companies that will be studied (Twitter 2011b). The Twitter collection prototype has proven to work well for any number of symbols or companies.

For this research project, a select number of companies within a given sector will be tracked and studied to try to gauge sentiment. Rather than determine sentiment of the larger Twitter universe, this research is more interested in the sentiment of active investors and how these investors share this sentiment on Twitter. Therefore, taking an approach similar to Sprenger and Welpe (2010), only tweets that use the "$" nomenclature made popular by the Stocktwits.com website and widely used within the Twitter investing and trading community will be tracked. For example, in order to track Apple's stock symbol on Twitter, a "$" would be added to Apple's stock symbol, AAPL, to get $AAPL, which would be used to track Apple on Twitter.

Once tweets have been collected and stored, a semantic analysis technique will be applied using the Natural Language Processing Toolkit available in the Python programming language determine sentiment using a Naïve Bayesian text classification algorithm. In order to use this algorithm, a training set will need to be created to 'train' the filter. This dataset will be developed by manually classifying a corpus of twitter message selected randomly from the captured data. These tweets will be classified as bullish (i.e., positive), neutral or bearish (i.e., negative) and will then be used to classify the remaining dataset as bullish, neutral or bearish

Preliminary research has shown that the approach to capturing and analyzing tweets is fairly straightforward. All aspects of each tweet can be analyzed to determine time of day, user name of the person who tweeted, the number of followers of that user, the type of tweet sent (original tweet versus a Retweet) and the contents of the tweet itself.

By studying the timeline of tweets an analysis can be performed to determine whether the time of day a tweet is sent has an effect sentiment and whether there are any days of the week or month that provide more positive or negative correlation with the movement in the stock market. Lastly, an analysis of the social graph sender of a tweet will be performed to determine whether the number of followers of a user has any correlation to how that user's tweet(s) influence the sentiment of a sector or stock mentioned in the tweet.

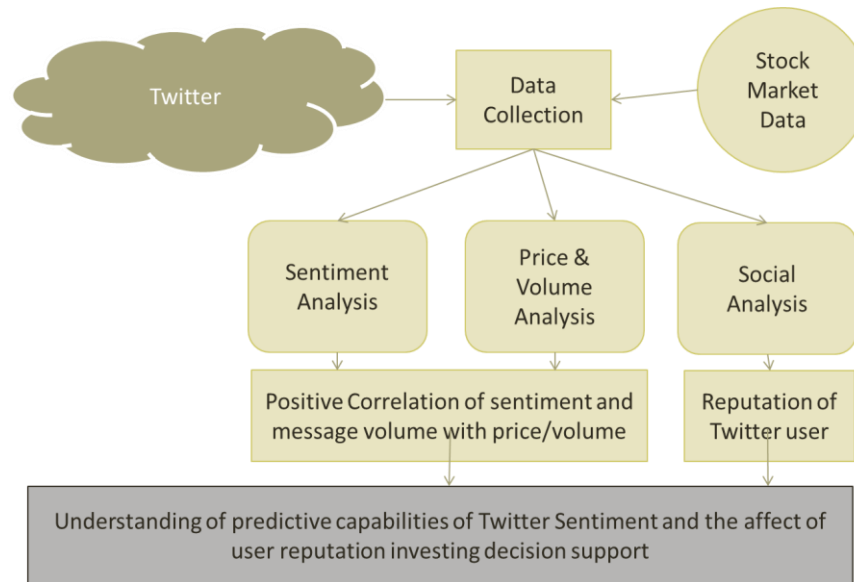A visual representation of the approach to data collection and analysis is given in Figure 1 below.

Figure 1. Twitter Data Collection and Analysis

PRELIMINARY RESEARCH AND RESULTS

Using a Twitter collection prototype to collect data, a test collection run was initiated on May 2, 2011 and ended on May 11. During this test run approximately 13,000 tweets were collected covering the XLE and XLP ETF's and the companies that make up these ETF's.

A review of the timeframe of this test (May 1 to May 12 2011) shows a period of volatility in the stock market, especially in the energy sector. During this period, the XLE ETF saw a little over an eight-point drop from a high of $80.80 on May 1, 2011 to a low of $72.78 on May 11, 2011 (StockCharts.com, 2011).

The captured tweets were analyzed using a Naïve Bayes Classification algorithm to assign a sentiment using Hu and Liu's (2004) polarity dataset as the training dataset. While this training dataset isn't ideal, it was a good place to start to begin to understand sentiment and with the twitter dataset.

For the XLE ETF, the average sentiment over the test time period was 0.115 (with 1 being bullish, 0 being neutral and -1 being Bearish). A Comparison of the XLE price movement compared to the classified sentiment over the test period can be seen in

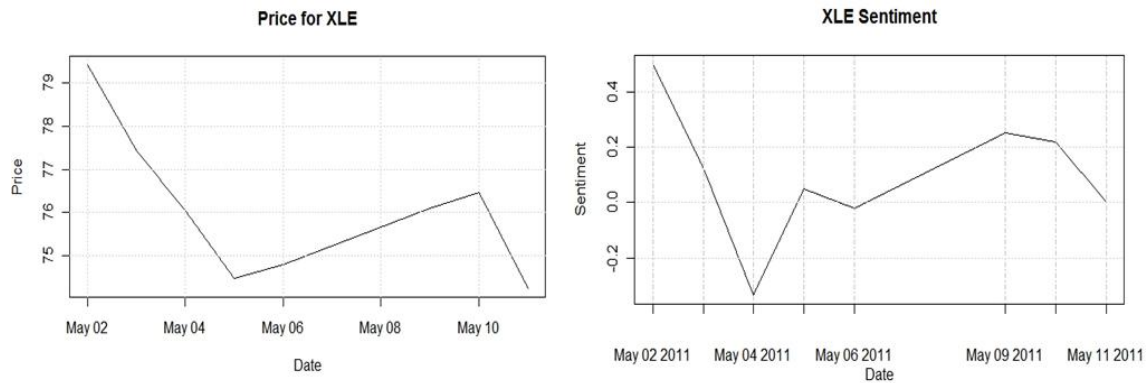Figure 2. XLE Price vs. Sentiment for May 2 to May 11

.

Figure 2. XLE Price vs. Sentiment for May 2 to May 11

Next Steps

Data Collection continues with tweets that mention XLE, XLP and the 82 companies comprising those ETF'S being collected and analyzed. In addition, other ETF's, including the S&P 500 ETF with the symbol SPY, are being captured for additional statistical analysis and study. Analysis is ongoing to determine whether correlation does exist between the stock market and twitter sentiment and/or volume as well as user reputation on the sentiment of an ETF and/or stock.

In addition, a random sampling of captured tweets is currently being manually categorized to use as a training dataset for the Naïve Bayes Classification algorithm. Additional research will be performed to determine if this training dataset provides a more accurate / robust dataset than other pre-existing datasets (e.g., Hu and Liu (2004), Pang and Lee (2008; 2002)).

CONCLUSIONS, DISCUSSION, AND SUGGESTIONS FOR FUTURE RESEARCH

Based on the preliminary research and literature review, future research into how Twitter sentiment might be used to predict movements of a stock or sector may yield promising insights into potential practical applications. This could, potentially, provide a decision support mechanism for investors and traders to use while attempting to determine whether to invest into a particular stock or sector.

By undertaking the research proposed, a better understanding of whether Twitter sentiment is driven by the internal happenings of the stock markets or whether Twitter sentiment is driving movement of the stock market and can be used as a predictive tool for decision support for investing decisions.

The literature shows conflicting results in some areas of sentiment analysis of message board postings. Some researchers report that message board sentiment has no predictive capabilities (Tumarkin et al., 2001) while other researchers have reported either strong or weak predictive capabilities of sentiment analysis of messages (Antweiler et al., 2004; Bollen et al., 2010; Sprenger et al., 2010). The evidence from this study may lend support to one side or the other in this disagreement within the area of previous work.

The research outlined in this proposal could lead to future research projects by extending the analysis of sentiment to determine and report sentiment in near real-time to allow investors and traders to decide which stocks or sectors should be invested in throughout the trading day. Additionally, aspects of this research may be extended into non-investing areas by attempting to understand whether a user's Twitter reputation has any relationship with how followers that user has as well as how many times that user has their tweets shared by their followers.

ACKNOWLEDGMENTS

REFERENCES

1. Abbasi, A., Chen, H. and Salem, A. (2008) Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums, *ACM Trans. Inf. Syst.* (26:3) 1-34.
2. Antweiler, W. and Frank, M. Z. (2004) Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards, *Journal of Finance* (59:3) 1259-1294.

3. Berger, A. L., Della Pietra, S. A. and Della Pietra, V. J. (1996) A Maximum Entropy Approach to Natural Language Processing, *Computational Linguistics* (22:1)39-71.

4. Bifet, A. and Frank, E. (2010) Sentiment Knowledge Discovery in Twitter Streaming Data, in: *Proceedings of the 13th international conference on Discovery science*, Springer-Verlag, Canberra, Australia, pp. 1-15.

5. Boiy, E. and Moens, M. (2009) A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts, *Information Retrieval* (12:5)526.

6. Bollen, J., Mao, H. and Zeng, X.-J. (2010) Twitter Mood Predicts the Stock Market.

7. Chiang, O. (2011) Twitter Hits Nearly 200M Accounts, 110M Tweets Per Day, Focuses On Global Expansion, Forbes.

8. Choi, Y., Kim, Y. and Myaeng, S.-H. (2009) Domain-Specific Sentiment Analysis Using Contextual Feature Generation, in: *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, ACM, Hong Kong, China, pp. 37-44.

9. Chua, C. C., Milosavljevic, M. and Curran, J. R. (2009) A Sentiment Detection Engine for Internet Stock Message Boards.

10. Das, S. R. and Chen, M. Y. (2007) Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web, *Journal of Management Science* (54:9)1375-1388.

11. Delort, J.-Y., Arunasalam, B., Milosavljevic, M. and Leung, H. (2009) The Impact of Manipulation in Internet Stock Message Boards, *SSRN eLibrary*).

12. Gu, B., Konana, P., Liu, A., Rajagopalan, B. and Ghosh, J. (2006) Identifying Information in Stock Message Boards and Its Implications for Stock Market Efficiency.

13. Hu, M. and Liu, B. (2004) Mining and Summarizing Customer Reviews, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA.

14. Hui, P. and Gregory, M. (2010) Quantifying sentiment and influence in blogspaces, in: *Proceedings of the First Workshop on Social Media Analytics*, ACM, Washington D.C., District of Columbia, pp. 53-61.

15. Hursman, A. (2010) Is Your Online Reputation Doomed?, *Information Management* (20:3)48.

16. Lin, C. and He, Y. (2009) Joint Sentiment/Topic Model for Sentiment Analysis, in: *Proceeding of the 18th ACM conference on Information and knowledge management*, ACM, Hong Kong, China, pp. 375-384.

17. Lo, A. W. (2004) The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective, *Journal of Portfolio Management*:30th Anniversary Issue)15-29.

18. Narayanan, R., Liu, B. and Choudhary, A. (2009) Sentiment analysis of conditional sentences, in: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, Association for Computational Linguistics, Singapore, pp. 180-189.

19. Neely, C. J., Weller, P. A. and Ulrich, J. M. (2009) The Adaptive Markets Hypothesis: Evidence from the Foreign Exchange Market, *Journal of Financial and Quantitative Analysis* (44:2)467-488.

20. O'Hare, N., Davy, M., Bermingham, A., Ferguson, P., Sheridan, P., Gurrin, C. and Smeaton, A. F. (2009) Topic-Dependent Sentiment Analysis of Financial Blogs, in: *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, ACM, Hong Kong, China, pp. 9-16.

21. Pak, A. and Paroubek, P. (2010) Twitter as a Corpus for Sentiment analysis and Opinion Mining, in: *Language Resources and Evaluation (LREC)* LREC 2010 Proceedings, Malta.

22. Pang, B. and Lee, L. (2008) Opinion Mining and Sentiment Analysis, *Found. Trends Inf. Retr.* (2:1-2)1-135.

23. Pang, B., Lee, L. and Vaithyanathan, S. (2002) Thumbs up?: sentiment classification using machine learning techniques, in: *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, Association for Computational Linguistics, pp. 79-86.

24. Romero, D. M., Meeder, B. and Klienberg, J. (2010) Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter, Cornell.

25. Sprenger, T. O. and Welpe, I. M. (2010) Tweets and Trades: The Information Content of Stock Microblogs, in: *Working Paper Series* Technische Universität München (TUM), p. 89.

26. StockCharts.com (2011) XLE Energy Sector Daily Chart, StockCharts.com.

27. Thet, T. T., Na, J.-C., Khoo, C. S. G. and Shakthikumar, S. (2009) Sentiment Analysis of Movie Reviews on Discussion Boards Using a Linguistic Approach, in: *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, ACM, Hong Kong, China, pp. 81-84.

28. Tucker, P. (2010) Pop Music as an Economic Indicator, *The Futurist* (44:2)10.

29. Tumarkin, R. and Whitelaw, R. F. (2001) News or Noise? Internet Postings and Stock Prices, *Financial Analysts Journal* (57:3)41-51.

30. Twitter (2011a) Twitter API, http://dev.twitter.com.

31. Twitter (2011b) Twitter Streaming API Methods, p. Description of the Various Methods Found in the Twitter Streaming API.

32. Vincent, A. and Armstrong, M. (2010) Predicting Break-Points in Trading Strategies with Twitter, *SSRN eLibrary*).
33. Whitelaw, C., Garg, N. and Argamon, S. (2005) Using Appraisal Groups for Sentiment Analysis, in: *Proceedings of the 14th ACM international conference on Information and knowledge management*, ACM, Bremen, Germany, pp. 625-631.
34. Wysocki, P. D. (1998) Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards, *SSRN eLibrary*).
35. Zhang, Y. (2009) Determinants of Poster Reputation on Internet Stock Message Boards, *American Journal of Economics and Business Administration* (1:2)114-121.