Trading strategies based on economics blogs sentiment

Vincenzo Farina, Albana Nako, Antonio Parisi, Ugo Pomante

University of Rome Tor Vergata, Department of Economics and Finance,

Via Columbia 2, 00133, Rome, Italy

E-mail: vincenzo.farina@uniroma2.it; albana.nako@uniroma2.it, antonio.parisi@uniroma2.it; ugo.pomante@uniroma2.it

Abstract

This paper analyses the effects of a particular type of mass media – economics blogs - on investors' behaviors. Using a large amount of linguistic data regarding about 900,000 posts during a five years time period (from 2008 to 2013), we propose the Moodec Index as new measure of investor sentiment and define a simple trading strategy to test its effectiveness. At the end of the period, our Moodec Index based strategy outperforms the "buy and hold" one, with a higher performance of 13,75%.

Keywords: Investor sentiment, stock market, text analysis

JEL classification: E32, G12, G14

1. Introduction

One of the most important research streams in finance aims at understanding the determinants of stock market dynamics. According to the theory of efficient financial markets (Fama, 1970), stock prices should reflect all available information. However, the evidence of autocorrelation of stock returns in the short term (Jegadeesh and Titman, 1993; Moskowitz and Grinblatt 1999; Hong, Touros, and Valkanov, 2007) suggests that stock prices do not fully adjust to new information.

Market sentiment does not play any role in the classic approach to finance. On the contrary, the behavioral approach suggests that waves of irrational sentiment, i.e. times of overly optimistic or pessimistic expectations, can persist and affect asset prices for significant periods of time (De Long, Shleifer, Summers and Waldmann, 1990). Barberis, Shleifer and Vishny (1998) find a psychological foundation of mispricing and show that investors place more emphasis on the strength of the information than on its statistical weight, relative to a rational Bayesian.

In this behavioral approach investor sentiment is a priced factor in market equilibrium and relevant questions are how to measure sentiment and how to quantify its effects (Baker and Wurgler, 2007).

In this paper we propose a new measure of investor sentiment based on mass media content and define a simple trading strategy to test its effectiveness.

We use the daily posts of all economics blogs indexed by Palgrave Econolog to reveal market-level sentiment. By considering the ratio daily posts referring to some negative economics words (e.g. bankruptcy, crisis, recession, unemployment, etc.) to total posts of the day, we construct the Moodec Index as a new measure of (pessimistic) sentiment.

From a methodologically point of view, we consider economics blogs sentiment as a measure of investor expectation about future dynamics of the Standard & Poor's 500 Index. We suggest that economics blogs sentiment does not only reflect the current state of the stock markets but may also be able to anticipate certain future (medium and long-term) trends.

Our findings are consistent with the intriguing proposal that movements in the financial market are preceded by a notable change in the slope of the Moodec Index. Therefore, the rationale of the trading strategy that we defined is very simple: i) when Moodec Index shows a growing trend, stay out of the market or sell the market index; ii) when Moodec Index shows a decreasing trend, buy the market index.

These results suggest that, following this logic, during the period 2008 to 2013 the analysis of economics blogs sentiment could have been used in the construction of a profitable trading strategy. The inversion signals of our sentiment index timed the transition periods of the stock market very well.

Our main contributions are: i) the evidence that investors' behavior is influenced by the content of economics blogs news; ii) a large-scale textual analysis of data from economics blogs (we examined about 900,000 blogs' posts during a five years time period) in order to build consistent sentiment time series; iii) the definition of an effective trading strategy based on the sentiment indicator defined.

This paper is organized as follows. First we review related work. We then describe the blogs and financial data we applied. The next part is dedicated to the trading strategy implemented using the smoothing spline method. In the final part of the paper we show the results and conclude.

2. Literature Review

Mass media (newspapers, television, Web 1.0) represent a fundamental tool in the production and the widespread of the information. Beside these traditional channels of information sharing, the birth of Web 2.0 platform (Twitter, Facebook, Google+, LinkedIn, etc.) is transforming the way we seek information and communicate with each other (Raacke, Bonds-Raacke, 2008; Nielsen-Wire, 2010; Mor Naaman, 2010).

How information is originated and disseminated by the traditional media (such as: newspapers, radios or television channels) is pointed out in many papers (Shoemaker, Vos, 2009; Salcito, 2009; Deephouse, 2000). Moreover, cognitive studies (Baumeister et al., 2001; Rozin, Royzman, 2001; Fiske, Taylor, 1991; Brief, Motowidlo, 1986), demonstrate that positive and negative news have different impacts on people' perceptions, and negative news also cause a stronger impact than positive news. Other works demonstrate media affect the behavior of investors to the extent that they are able to attract attention by making some titles more familiar than others (Kang, Stulz, 1997; Huberman, 2001; Barber, Odean, 2003; Li, 2004; Hong, Kubik, Stein, 2004; Shiller, 2005; Pollock, Rindova, Maggitti, 2008; Barber, Odean, 2008; Lehavy, Sloan, 2008; Da, Engelberg, Gao, 2011).

Some economic studies concentrate on the examination of the language used by the media when reporting economic and financial events by demonstrating the impact financial information exerts on the retail investor's behavior as a consequence the stock market (e.g., Ferguson, 2012; Carretta et al., 2011; Tetlock et al., 2008; Tetlock, 2007).

Ferguson et. al., (2012) examine whether tone (positive/negative) of firm specific news media content are able to predict stock returns. Findings demonstrate that information embedded in positive as well as negative words in news stories predicts next day returns, with pronounced impact from news about fundamentals. The predictive power of tone is stronger among low visibility (small) firms. The findings suggest that firm-specific news media content incorporates valuable information that predicts asset returns.

The paper of Carretta et. al., (2011) considers corporate governance news and its interaction with the corporate economic and financial standing. Results show that stock returns tend to increase after ownership news if the firm involved is not profitable at the time of the news release. Otherwise, the overall effect on stock returns is substantially negative. This would indicate that investors dislike ownership related news for profitable firms, so that they tend to sell their stocks. Another finding is that investors are influenced by rumors about corporate governance and that, before publication, they are simply able to assess the type of corporate governance event. However, after publication, investors' behavior is influenced by the content and tone of the news.

Tetlock, Saar-Tsechansky and Mackassy (2008) find that some firms' fundamentals are a function of the percentage of negative words in the news. Some news exerts an effect in a relatively short period and other news in the medium and long term (for example, news regarding core aspects of firm management).

The interactions between media content and stock market activity is investigated also by Tetlock, (2007); the author took under investigation the news published in the section "Abreast of the market" in the Wall Street Journal. The paper points out that high values of the pessimistic indicator induce downward pressure on market prices; unusually high or low values of pessimism lead to temporarily high market trading volume. It also shows that the price impact of pessimism appears especially large and slow to reverse in small stocks.

The birth of Internet first and the development of the new technologies, Web 2.0 platform, are changing radically the way we obtain and share information in the eWOM (electronic word of mouth) form. According to a survey conducted by Pew Research (2010), the Internet has overtaken newspapers in terms of popularity as a news platform and ranks just behind television. Recent researches evidence that several blogs, social networks etc. focus on opinion and information sharing through micro blogging (Kwak et al., 2010; Mor Naaman, 2010; Johnson, Yang, 2009; Pfitzner, 2012) rather than on reciprocal social interaction (Huberman et al., 2009) and that is one of the main reasons why the Web 2.0 platform becomes an interesting venue to put under a microscope.

One important value of Twitter and other public social media platforms is that they allow us to listen (public) conversations and to infer people's opinions or moods based on their public statements (Pfitzner, 2012). Examples are mood surveys of communication on Myspace (Thelwall, Wilkinson, Uppal, 2009), and Twitter (Thelwall et al., 2010). Milstein et al., (2008) reviews general background on Twitter and microblogging. Ebner, Schiefner (2008) and Grosseck, Holotescu (2008) examine microblogging in an educational setting.

The financial information disseminated on this new platform is attractive also for financial researchers. In particular, Bollen et al., (2011) link the mood of tweets with the fluctuations of the stock market. The study highlights that the sentiment of Twitter is able to predict, with a high accuracy, the daily changes (up and down) of the closing prices of the DJIA. Different from other studies Timm et al., (2013) investigate the microblogging messages published on Twitter that express an explicit content referred to the stock market by restricting the amount of tweets available as an output decreasing the noise related with the big amount of data. The study shows that sentiment embedded in the tweets affect not only the stock market returns, but also the trading volume and volatility. The study also demonstrate that users who come with investment advices are more popular in terms of followers they have and retweets of their massages, as expression of their power to be "heard" on the specific blog. Mizrach (2009) analyses the trading activity in an Internet chat room over a 4-year period and he finds that these traders are more skilled than retail investors analyzed in other studies. Chen et al., (2013) investigates the extent to which investor opinions transmitted through social media predict financial market outcome variables through the textual analysis conducted on articles published on one of the most popular social media venues for investors in the US Market. The paper evidences that the views expressed in articles and commentaries each substantially contribute in predicting future stock returns and earnings surprises. Gilbert, Karahalios (2010) show how the mood of millions in a large online community, even one that primarily discusses daily life, can affect the stock market.

Overall, mass media seem to be have an important impact on stock market dynamics. In this paper we analyze the effects of a particular type of mass media – economics blogs - on investors' behaviors. Therefore, our hypothesis is that the tone of information disseminated through economics blogs may affect the stock market.

3. Methods

This section illustrates the research methods used in the paper. To describe the nature of the interactions between blogs sentiment and stock market reaction we consider daily data in a specific time period (from March 1, 2008 to December 16, 2013) and construct trading strategies basing only on the pessimistic indicator defined.

3.1 Data and variables

For the measurement of media pessimism we considered the degree of pessimism expressed by internationally recognized economic blogs indexed on Palgrave Econolog for each day of the specified period. In the observation period we analyzed about 900,000 blog posts written in the English in order to calculate the daily measure of financial pessimism, here and after named Moodec (economic mood indicator).

Due to the nature of data and in order to allow the stability, the reproducibility and the accuracy of the measure of pessimism, we apply a content analysis methodology (Stone, Dunphy, Smith and Ogilvie, 1966) using a specifically defined linguistic dictionary of negative financial words (e.g. bankruptcy, crisis, recession, unemployment, etc.). In order to classify a post as pessimist, we simply observe if the post contains words that falls within the negative category.

Therefore, the Moodec Index is defined as the ration between the number of posts containing (in the title or in the message) a word with negative meaning on the overall number of posts available in that specific day on the blogs' platform.

As a measure of returns we consider the daily log-differences of the Standard & Poor's 500 index levels in two contiguous days. We obtain daily time series of returns from Datastream.

Thus we obtain a trading strategy which is based only on the sentiment index and which outperforms the "buy and hold" strategy.

3.2 Model

In order to decrease the frequency of the trading signals, we decided to filter the signal brought by the Moodec Index using a smoothing spline.

A smoothing spline (see, for example, Green and Silverman, 1994 and Hastie, Tibshirani, and Friedman, 2001) is a method that solves the knot selection problem and controls the complexity of the fit by regularization.

Let (xi, yi), with x1 < x2 < ... < xn, be a sequence of observations. The smoothing spline estimate of the function f(x) is defined as the minimizer (over the class of twice differentiable functions) of the penalized residual sum of squares where $\lambda \in (0,\infty)$ is a smoothing parameter. The second term penalizes the roughness of the function f: If $\lambda = 0$, f converges to the interpolating spline; if $\lambda = \infty$, f represents the least squares line fit. We will select the optimal value of λ using a forecasting approach.

3.3 The trading strategy

Let y_t , t = 1, ..., T denote the Moodec Index over the period defined previously. Our trading strategy relies on the assumption of an existing relationship between the yt and the future behavior of a relevant index. Given that the Moodec has been calculated using posts from international blogs, a relevant index is given by the Standard & Poor's 500 index, here and after named SPY, denoted as x_t. As the index yt is quite noisy, we will use a obtain a smoothing spline, say z_t, using observations (t, y_t).

Figure 1: Level of the SPY index (right axis) and the complement to 1 of the smoothed Moodec index (left axis).

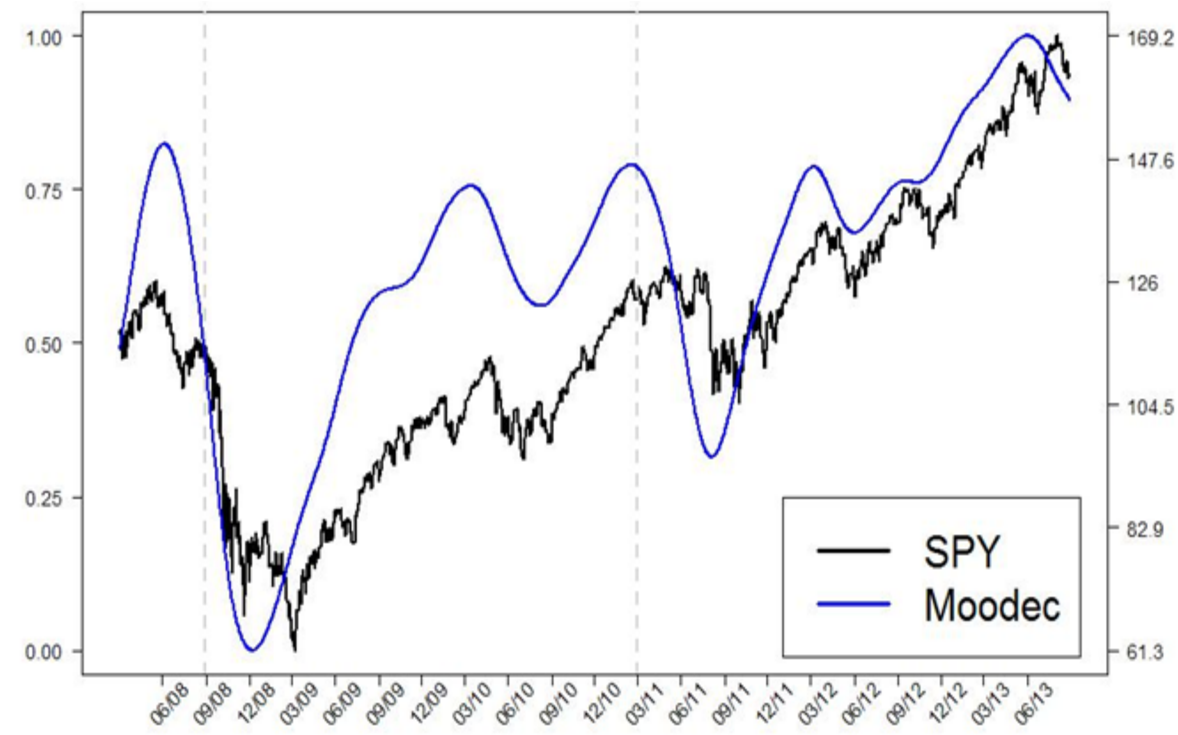

Figure 1 gives an a posteriori descriptive representation of the relationship between the filtered series z_t (obtained using a value for λ equal to 20) and the index xt. Notice that the figure shows the complement to 1 of z_t, as the Moodec Index is a pessimism index. It is possible to define different trading strategies based on the index: what we propose here is to check for changes of the concavity of z_t, as these changes should forewarn about turning points in the related price index. However, the estimate of the entire spline changes at every time period, and the maximum variability concerns the last days in the sample: while the spline is stable over observations in the far past, it shows considerable changes for recent observations. As a consequence, even if z_t shows a change in the concavity at time t, this change could not be confirmed in the subsequent days. To avoid these contradictory signals we propose to delay the validation of the trading signal after a sufficient number of time periods. The time lost for this delay is partly compensated by the fact that the Moodec index is available for all calendar dates, also in weekends and holidays.

As a result, we propose a strategy which is based on smoothing splines, in which we have to optimize two parameters: the smoothing parameter λ and the delay d. Both of them are related to the timeliness of the strategy and to the frequency of trading signals. We aim at obtaining a combination of parameters which determines a fast response to substantial changes in the pessimism index and a moderate frequency of trading signals. In order to find such a combination, they are chosen on the basis of their predictive performance.

Results

We obtained the value of y_t from March 1, 2008 to December 16, 2013 (2115 observations). For the same period, we obtained the values of the SPY index (1460 observations).

The observation period has been divided in three parts, as depicted in figure 1 by the gray dashed lines: the first one (March 3, 2008 to August 31, 2008, that is 180 days) will serve as a starting period, in the second one (September 1, 2008 to February 28, 2011, that is about three years from the starting date) the performances of the different combinations of λ and d are evaluated, in the last part (March 1, 2011 to August 31, 2013, that is 30 months) we evaluate the performance of the strategy using the best combination of parameters.

It turns out that the best values of the parameters are given by λ = 50 and d = 12. Figure 2 depicts the results of such an optimization as a comparison between the "buy and hold" strategy against the strategy suggested by our index, on the third period.

Figure 2: Comparison between strategies

The black line is the value of the SPY index, while the blue one is the performance of the trading strategy: red lines represent sell signals, while green lines represent buy signals. Gray lines indicate a standard deviation of the cumulative return of a random strategy: maintaining the number of the transactions we obtained with our strategy, the random strategy draws the dates of the transactions from a hypergeometric distribution. The expected value and the standard deviation are obtained by simulation of 10,000 independent realization of the random investment strategy.

Summarizing, the picture shows that our strategy overperforms the “buy and hold” strategy and that this spread on the performance is not a casual result. Notice that in our investment strategy, since determining transaction fees is quite difficult, we neglected them at all. In any case, because of the low number of transactions, their impact on the performance would be marginal.

More in details, our strategy requires that, when a “buy” signal is confirmed, the value of the Moodec strategy follows the same dynamic as the underlying index, while if a “sell” signal is confirmed, its value stays fixed until the arrival of the following signal.

It is possible to notice that the Moodec strategy prevents many, even if not all, the falls of the SPY index. On the contrary, it takes advantage of the majority of the positive changes of the SPY. At the end of the period, our Moodec Index based strategy outperforms the “buy and hold” one, with an higher performance of 13,75%.

In 912 trading days, we registered 48 trading signals (24 buy plus 24 sell signals), that is more or less one signal each month, even if they concentrate in periods where market uncertainty arise.

Conclusion

This paper proposes a new measure of investor sentiment based on mass media content and defines a simple trading strategy to test its effectiveness.

Our results suggest that during the period we investigate, Moodec Index does not only reflect aspects of the current state of the economy, but may have also pro- vided some insight into (medium and long term) future trends of the Standard & Poor’s 500 index.

The contribution of this study is manifold.

Our results support the general argument that the characteristics of information provided by mass media influence investor choices. In this perspective, we evidence that mass media information is important not only for its novelty, but also for its effects on investor sentiment. As one could expect, media sentiment, measured as semantic content of the blog posts, influences the investors' preferences and therefore stock market returns.

During the period 2008 to 2013 the analysis of economics blogs sentiment could have been used in the construction of a profitable trading strategy. At the end of the period, our Moodec Index based strategy outperforms the "buy and hold" one, with a higher performance of 13,75%. In general, the inversion signals of the Moodec Index timed the transition periods of the stock market very well.

References

Baker M., Wurgler J., 2006, Investor sentiment and the cross-section of stock returns, Journal of Finance 61, 1645–1680.

Barber, B. M., Odean, T., 'Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors', Journal of Finance, 2003, Vol. 55, pp. 773-806.

Barber, Brad M. and Odean, Terrance, All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors (November 2006). EFA 2005 Moscow Meetings Paper.

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307-343.

Baumeister R.F., Bratslavsky E., Finkenauer C., Vohs K.D., Bad is stronger than good, Review of General Psychology 5: 323–370, 2001.

Bollen, J., Mao, H., Zeng, X., 'Twitter mood predicts the stock market', Journal of Computational Science, 2010, Vol. 2, No. 1, pp. 1-8.

Brief, A. P., & Motowidlo, S. J. 1986. Prosocial organizational behaviors. Academy of Management Review, 11: 710–725.

Carretta, A., Farina, V., Fiordelisi, F., Martelli, D., Schwizer, P., 'The impact of corporate governance press news on stock market returns', European Financial Management, 2011, Vol. 17, No. 1, pp. 100-119.

Chen, H., P. De, J. Hu, and B.-H. Hwang, 2013, "Customers as advisors: The role of social media in financial markets," Working Paper.

Da, Z., J. Engelberg, and P. Gao, 2012, In search of attention, Journal of Finance.

De Long J., B.; Shleifer A.; Summers H. L.; Waldmann R. J., "Noise Trader Risk in Financial Markets", The Journal of Political Economy, Vol. 98, No. 4. 1990), pp. 703-738.

Deephouse, D. L. (2000), Media reputation as a strategic resource: An integration of mass communication and resource-based theories, Journal of Management, 26: 1091-1112.

Ebner, M.,& Schiefner, M. (2008). Microblogging—more than fun? In Inmaculada Arnedillo Sánchez & Pedro Isaías (Eds.), Proceedings of IADIS Mobile Learning Conference 2008 (pp. 155–159). Portugal. Grosseck, G., and Holotescu, C. (2008, April) "Can we use Twitter for Educational Activities?" Paper presented at the 4th International Scientific Conference, eLearning and Software for Education, Bucharest, Romania. Timm O. S., Andranik T., Philipp G. S., Isabell M. W., "Tweets and Trades:

the Information Content of Stock Microblogs", European Financial Management, Vol. 9999, No. 9999, 2013, 1–32

Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. Journal of Finance 25, 383-417.

Ferguson, N., Guo, J. Lam, H. and D. Philip. 2011. Media sentiment and UK stock returns. Working Paper (University of Durham).

Fiske, S.T., Taylor, S.E., Social cognition, New York: McGraw-Hill, 1991.

Gilbert, E. & Karahalios, K. (2010). Widespread Worry and the Stock Market. 4th International AAAI Conference on Weblogs and Social Media (ICWSM), 2010.

Green P.J., Silverman B. W., "Nonparametric Regression and Generalized Linear Models: A roughness penalty approach", Chapman & Hall/CRC – 1994.

Hong, H., Kubik, J. D., Stein, J. C., 'Social Interaction and Stock-Market Participation', Journal of Finance, 2004, Vol. 59, pp. 137-163.

Hong, H., Touros W., Valkanov R., 2007, "Do industries lead stock markets?", Journal of Financial Economics 83, 367-396.

Huberman, B. A., Romero, D. M., Wu, F., 'Social networks that matter: Twitter under the micro-scope', 2009, Available at SSRN: http://ssrn.com/abstract=1313405 14(1):8.

Huberman, G., 'Familiarity breeds investment', Review of Financial Studies, Vol. 14, 2001, pp. 659–80.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. Journal of Finance 48, 65–91.

Johnson, P. R., Yang, S.. Uses and gratifications of Twitter: An examination of user motives and satisfaction of Twitter use. Paper presented at the Communication

Technology Division of the annual convention of the Association for Education in Journalism and Mass Communication in Boston, MA., (2009).

Kang, J.-K., Stulz, R.M., 1997. Why is there a home bias? An analysis of foreign portfolio equity ownership in Japan. Journal of Financial Economics 46, 2}28.

Kwak, H., Lee, Changhyun, H., Moon, S., 'What is Twitter, a Social Network or a News Media?' In Proceedings of the 19th World-Wide Web (WWW) Conference, Raleigh, North Carolina, 2010.

Milstein, S., Chowdhury, A., Hochmuth, G., Lorica, B., & Magoulas, R. (2008). Twitter and the micro-messaging revolution: Communication, connections, and immediacy—140 characters at a time. Sebastopol, CA: O'Reilly Media.

Mizrach B., Weerts S., "Experts online: An analysis of trading activity in a public Internet chat room", Journal of Economic Behavior & Organization Volume 70, Issues 1–2, May 2009, Pages 266–281

Mor Naaman, Jeffrey B., C., "Is it all About Me? User Content in Social Awareness Streams." In Proceedings of the 2010 ACM conference on Computer Supported Cooperative Work., 2010.

Moskowitz, T.J., Grinblatt, M., 1999. Do industries explain momentum? Journal of Finance 54, 1249–1290.

Nielsen-Wire., "Led by Facebook, Twitter, Global Time Spent on Social Media Sites up 82% Year over Year". http://blog.nielsen.com/nielsenwire/global/ledby- facebook-twitter-global-time-spent-on-social-media-sites-up-82-year-overyear/

Pfitzner´, R., Garas, A., Schweitzer, F., "Emotional Divergence Influences Information Spreading in Twitter", Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, 2012.

Raacke, J., Bonds-Raacke, J., (2008). MySpace and Facebook: Applying the uses and gratifications theory to exploring friend-networking sites. Cyberpsychology & Behavior, 11(2), 169–174.

Rozin, P., Royzman, E., Negativity bias, negativity dominance, and contagion, Personality and Social Psychology Review, 5: 296-320, 2001.

Salcito, K., Online journalism ethics: "Gatekeeping". <http://www.journalismethics.ca/online_journalism_ethics/gatekeeping.htm> (2009).

Shiller, R. J., 'Irrational Exuberance', Second Edition, Princeton, N.J., Princeton University Press, 2005.

Shoemaker, P. J., Vos, T. P., "Media gatekeeping". In D. W. Stacks & M. B. Salwen (Eds.), An integrated approach to communication theory and research (2nd ed., pp. 75–89). New York: Routledge. 2009.

Stone, P.J., Dunphy, D.C., Smith M.S. and Ogilvie, D.M., 1966. The general inquirer: a computer approach to content analysis. MIT studies in comparative politics, MIT Press, Cambridge.

Tetlock, P. C., 'Giving content to investor sentiment: the role of media in the stock market', Journal of Finance, 2007, Vol. 62, No. 3, pp. 1139-68.

Tetlock, P. C., Saar-Tsechansky M., Mackassy, S., 'More than words: quantifying language to measure firms' fundamentals', Journal of Finance, 2008, Vol. 63, No. 3, pp. 1437-67.

Thelwall, M., Buckley, K., Paltoglou, G., 'Sentiment in Twitter Events', Journal of the American Society for Information Science, 2011, Vol. 62, No. 2, pp. 406-418.

Thelwall, M.; Wilkinson, D.; and Uppal, S. 2009. Data mining emotion in social network communication: Gender differences in MySpace. Journal of the American Society for Information Science and Technology In press.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Springer Series in Statistics Springer New York Inc., New York, NY, USA, (2001).