International Business Research; Vol. 7, No. 5; 2014
ISSN 1913-9004 E-ISSN 1913-9012
Published by Canadian Center of Science and Education

Panic Indicator for Measurements of Pessimistic Sentiments from Business News

Rodion Remorov[1]

[1] Consultant, www.newsreturn.com, Canada

Correspondence: Rodion Remorov, 54 Southvale Dr., Vaughan, Ontario, L6A0Y6, Canada. E-mail: rodionremorov@gmail.com

Received: February 6, 2014 Accepted: March 3, 2014 Online Published: April 24, 2014

doi: 10.5539/ibr.v7n5p103 URL: http://dx.doi.org/10.5539/ibr.v7n5p103

Abstract

Computational semantic analysis was applied for measurements of pessimistic sentiments from business news. Using semantic tree method, the number of pessimistic and optimistic news was estimated for the analysis of the bearish and bullish stock markets. It was found that the number of pessimistic and optimistic macroeconomic news is very sensitive to the sharp changes of the market indices at the extreme market conditions. A new sentiment indicator was constructed for quantitative measurements of the pessimistic investor sentiments. The proposed sentiment indicator is called a Panic Indicator. We found that the Panic Indicator is appropriate for the explanation of the relationship between the negative public information and stock index declines (S&P-TSX), as well as the sharp changes of VIX index. The proposed Panic Indicator would be useful for stock price modeling, the quantified description of the pessimistic opinions, and computational trading algorithms.

Keywords: investor sentiments, sentiment indicator, semantic analysis, bearish market, market crash, computational linguistics

1. Introduction

Since the financial crisis on the US market in 2007–2008, there has been an increasing interest in the study of the role of the investor sentiments in the stock market. The pessimistic or optimistic sentiments can be defined as the herd behaviour of investors, who follow the economic, political, or social factors with respect to the future stock price returns. The basic concept of herd behavior was stated by Tarde in 1903 in the idea of the imitation process, for which people learn from one another through the process of imitation. The market crowd behavior was studied by Soros (1987) using the concept of reflexivity of the market participants. The concept of reflexivity may explain the violation of the Random Walk Hypothesis during boom and bust market through disequilibrium process in the financial market. The negative expectations or pessimistic sentiments can lead to the downward direction of the stock market.

The difficulty for measuring herding effects lies in the lack of the effective model of the disequilibrium market as well as the methods, which could measure the activity of rational and irrational investors in the speculative market. There are different models for the analysis of the investor sentiments. Existing quantitative approaches of the sentiment measurements have mainly focused on the empirical investigation of the time-series of the price returns, trading volume, the dividend premium, IPO volume, volatility (see, for example, Baker & Urgler, 2007; Brown & Cliff, 2004). DeLong, Shliefer, Summers and Waldmann (1990) proposed the behavioural model to explain the risk created by the unpredictability of the unsophisticated investor's opinions. The model is based on the abstractive qualitative analysis of two types of investors: rational arbitrageurs (sentiment-free) and irrational investors. They showed that the market risk could be generated by the activity of the unsophisticated investors. The trading activity of the unsophisticated investors could limit the applications of the various arbitrage strategies applied by rational arbitrageurs. Remorov (2014) showed that the disequilibrium market can be modelled through cash-flow exchange between market-makers at the stock market crashes through behavioural process. Graham (1999) investigated the impact of public and private information on the herd behaviour among investment analysts. Graham proposed the qualitative reputation herding model to describe the relation between information signals with herd behaviour of smart analysts. Hong et al. (2000) developed a behavioural model, which explains the sharp decline of profitability of momentum strategies with the firm size. Most interesting research works are related to the information-based herding model (Welch, 1992), compensation based herding model (Brennan, 1993), and reputation based herding model (Scharfstein & Stein, 1990).

The study of the information-based herding requires the advanced semantic methods for the computational analysis of the public information impact. Beth (1954) invented the method of semantic tableaux (semantic tree method). A "Bag of Words" method was developed by Harris (1954) and is commonly used methods of the document classification, which consist of the distributional methodology for linguistics. Using the concept of the similarity of the members for each basis classes, Harris (1954) grouped the members according to the distributional relations between members of each basis classes. Some approaches were developed using the "Bag of Words" methods (see, for example, a Noun Phrase approach reported by Tolle & Chen, 2000). The goal of the automated application the textual analysis is closely related to the machine learning algorithms, which are intensively developed with the web-semantic analysis. The machine learning algorithm uses different statistical techniques to optimize the model parameters with forecasting the stock price movement. The Genetic Algorithm was applied by Thomas and Sycara (2002) to investigate the stock price movement with the number of postings and the number of posted words on a daily basis. The Naïve Bayesian technique uses the weighted vector of keywords for the stock rice prediction applying textual classification; see Seo, Giampapa et al. (2002).

Public business news is primary information source for the both rational and irrational investors. There is a limited range of semantic techniques for the analysis of the public business news with respect to the pessimistic/optimistic investor sentiments. Significant work was undertaken to develop the computational linguistic technique for stock price forecast. Mitchell and Mulherin (1994) studied the relation between the number of news announcements reported daily by Dow Jones and Company and aggregate measures of securities market activity including trading volume and market returns. Tetlock et al. (2008) showed that the linguistic media content may capture the information of the company fundamentals, and; therefore, investors may quickly incorporate this information in the different strategies. Ormos and Vazsonyi (2011) investigated the relation of the nouns and adjectives within business sentences with positive and negative returns of the S&P 500 stocks. Xie et al. (2012) developed the semantic approach for the analysis the business news with respect to the stock price movement and the company performance. They showed that their developed algorithm of the semantic analysis gives the significantly better prediction performance of the price return. Birza and Lott (2011) empirically showed that business news about GDP and unemployment does affect the stock returns. Zhang et al. (2011) investigated the textual information from Twitter and found that emotional tweet percentage significantly negatively correlated with Dow Jones, NASDAQ and S&P 500, but displayed significant positive correlation to VIX.

The range of the potential applications of the semantic analysis is extremely broad. In the current report, the new methodology of the pessimistic sentiment measurements was set out for computational linguistic analysis with the focus on the extreme market conditions. Our motivation of the sentiments measurement methodology is related to the construction of the computational linguistic analysis of the impact of public news on equity market for replication of the typical behaviour of financial analysts. The new sentiment indicator is presented for the measurement of the panic herd behaviour during the extreme market conditions. We investigate how shocking market conditions are related to the issues of business news and investor sentiments. The current article demonstrates first results of the study of the relations between stock price and investor sentiment during market crashes.

2. Method

2.1 Measurements of the Pessimistic Sentiments from Business News

Our sentiment technique is mainly based on the semantic tree method with subjective classification of the selected sentences. Some principles of the subjective classification and semantic tree methods are described in the academic literature (Beth, 1955; Bondecka-Krzykowska, 2005; Liu, 2010). A large group of the sentences were selected from three types of the textual information sources: business sentences from the economic analytical articles, titles from analytical articles, sentences from educational courses, (for example, CFA courses) with positive or negative impact on the equity market. The textual database exceeded 65,000 sentences (connectives). The textual information was used as a reference benchmark database for the application for the semantic analysis of new business news. The selected sentences included the information from economics, accounting, derivative instruments, technical analysis for the stock market, political news, social news, and so on. Although some topics, such as the technical analysis from CFA courses, don't have scientific support, we used the application of technical analysis for the stock market as an important part of the herd behavioral processes in the financial industry.

Our developed algorithm monitors business analytical news on a daily basis directly from web-publishers, such as Thomson Reuters, Bloomberg, The Wall Street Journal, and The New York Times Company. The number of issued negative and positive news was calculated for the weekly period from different news agencies, however, in the current report we present the results of our analysis obtained from Thomson Reuters. The articles from Thomson Reuters were collected since 2007 year; the total number of articles exceeded 6 mln articles. Textual information from newspaper titles and news abstracts was used for the semantic analysis; because the titles and abstracts better describe the news announcement in general. Similar approach was used by Birza and Lott (2011).

Initially, new business news was separated by 6 topic categories with the following hierarchy: macroeconomic news, government news, stock market news, real estate news, company specific news, and commodity news using the key-word classification (for example, the key-words "oil" and "NYSE" define commodity news and stock market news, respectively).

The logic tree method was applied for the selection of the documents as optimistic, pessimistic, and neutral information with respect to the growth of the US stock market. The method of the verification consists of finding the best coincidences of the benchmark sentences with the tested sentences. Each sentence from the benchmark sentence database (65,000 sentences) was compared with the title or abstract of the article from the web-publisher. Then, the number of optimistic and pessimistic news was counted for each day. The results of the sentiment analysis are published and updated weekly in web-site *www.newsreturn.com*.

3. Results

In the current article, the sentiments measurements rely upon on the extreme panic market conditions with two key objectives. The first is the development of the application of the semantic approach for the computational analysis of public business news to determine the analyst opinions. The second is the application of the sentiment indicator for the measurement of the investor sentiments with respect to their expectations of the future stock price return. We consider the usefulness of the application of the news sentiment indicator from two perspectives: as a quantitative measure of the pessimistic investor sentiments from business news and the impact of negative news on the equity market. Many of the study in the literature rely on the relationship of public information and stock price return. Nevertheless, the impact of public news can be strictly criticized for number reasons. For example, the issue of public news can cause the stock price changes and vice versa. As pointed out by Mitchell and Mulherin (1994), some newspaper articles might be written after the closing of the stock market on the day of the economic releases. For this reason, we also investigated the relation between the issued articles with the past events in stock market. We found that issued news, which is specified as stock market news, had the highest correlation with past events (stock price movement) and, therefore, did not have the predictive power for modeling. In contrast to stock market news, macroeconomic news demonstrated short and long-term correlation with future stock price indices. For this reason, macroeconomic news was taken for our investigation for the detection of the news impact on the investor sentiments.

We also tried to use the statistical methods to count the bad and good words from the articles, but we found that the justification of optimistic or pessimistic documents from the statistical methods (bad-good word counting) did not work.

The sample tests of randomly selected news showed that the accuracy of the sentiment news prediction exceeded ~90%. The subjective sentiment classification is most critical part in our computational linguistic technique, because this approach replicates the behaviour of the equity analyst for the short/long term price return perspective.

Obviously, the proposed method is very sensitive to the subjective classification of the original benchmark sentences for affective judgment of the optimistic or pessimistic documents. However, we consider that the proposed method is simple and robust for the semantic analysis without contribution of the possible modeling risk from the statistical methods. Moreover, the proposed method can be easily extended and modified for optimization with the maximum news impact for forecasting purposes.

We analyzed the time-series of various indicators, which include optimistic and pessimistic news from different selected topics: macroeconomic news, political news, monetary policy, stock market, real estate, company specific news, and commodities. Since the number of combinations of the possible indicators is extremely broad, we focused on the sensitivity of the indicators to the extreme market conditions only. We observed from the analysis of the pessimistic sentiments that the most appropriate textual information can be found from macroeconomic news. The textual information from stock market news, commodities news, or company specific news reflected the past event in the stock market, and, therefore, we didn't use these articles for our analysis.

To define whether the number of macroeconomic news is a good measure of the investor sentiments, we examine the historical dependencies of different variables for conformance with our assumptions about the relationship of negative news with market indices. Our empirical analysis showed that the ratio of pessimistic to optimistic macroeconomic news is the most sensitive to the negative stock price declines during stock market crisis. This ratio is called a pessimistic rate, Equation (1):

$$k_p = \frac{number\ of\ pessimistic\ macroeconomic\ news}{number\ of\ optimistic\ macroeconomic\ news} \tag{1}$$

where k_P is the pessimistic rate (PR). Figure 1 summarizes the calculated pessimistic rates for a period from 1/1/2007 to 1/1/2014. As can be seen, the pessimistic rates demonstrate a positive trend with jumps during bearish market period (2007–2008).

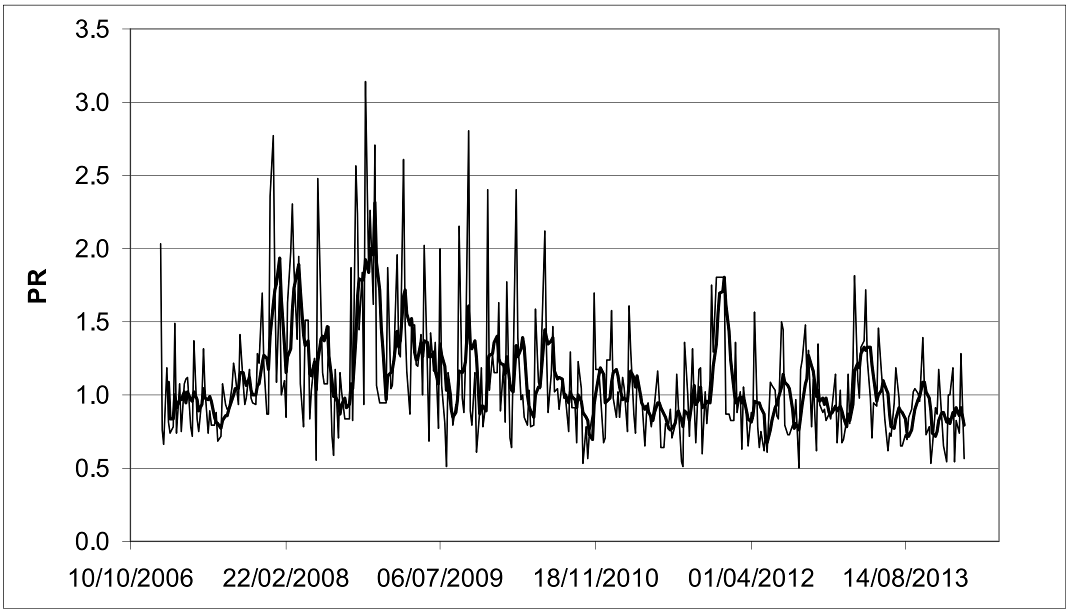
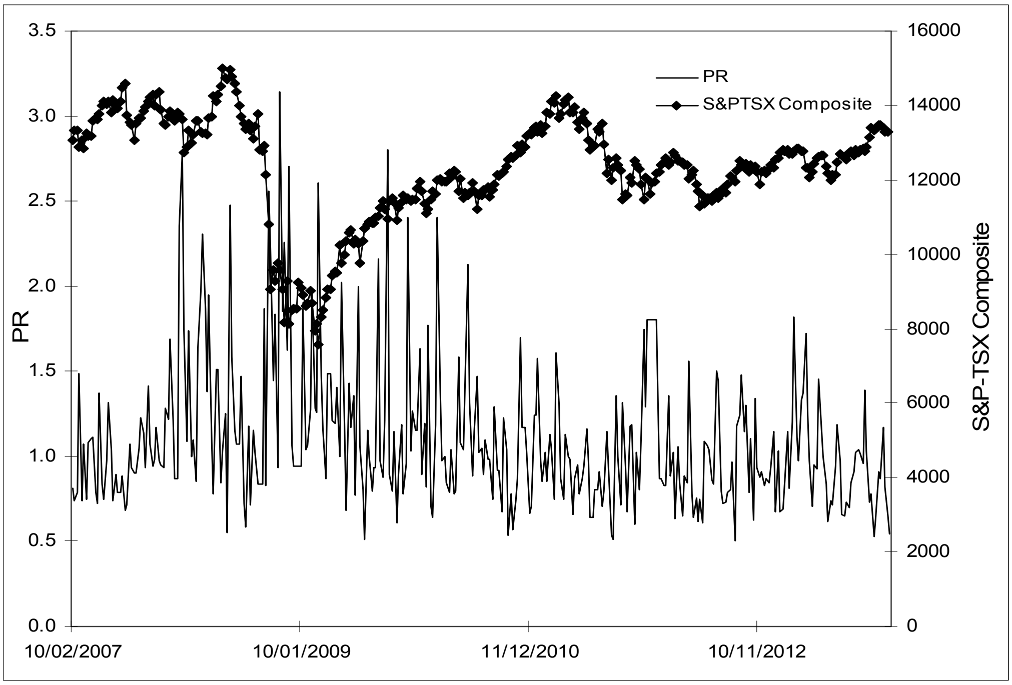


Figure 1. The time-series of the ratio of the pessimistic to optimistic macroeconomic news

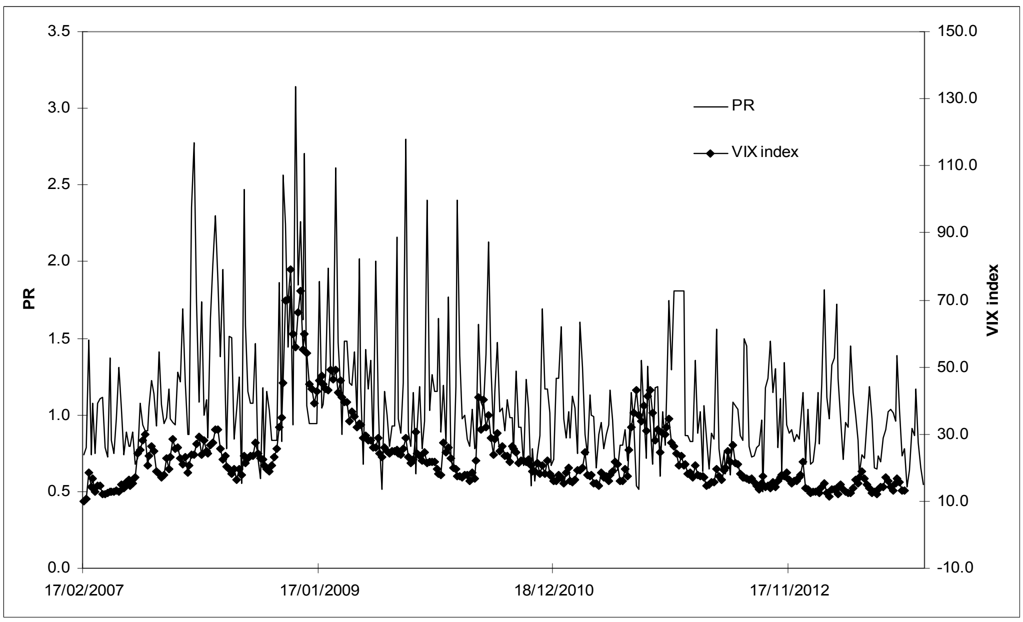
News source: Thomson Reuters.

4. Discussion

It is interesting to observe how negative information was issued during the financial crisis. The sharp changes of PR are characterized by the sentiment overreaction to the market through negative news. The pessimistic rates exhibit the considerable correlation with market index movement during the shocking market conditions. Figures 2 and 3 show the comparison of the stock index (S&P-TSX Composite index) and VIX index with the pessimistic rates, respectively. As can be seen, the density of negative news reaches maximum during the shocking market conditions. The Pearson correlation (6-month period) of the pessimistic rate with weekly return (S&P-TSX) reached 60% and statistically significant during the market crashes, Figure 3. The PR values differentiate the period of a stable market and shocking market conditions. Moreover, for the high volatile market, the PR values deviate considerably. Our result supports the study of Zhang et al. (2011), who showed that the emotional textual information from Twitter significantly negatively correlated with Dow Jones, NASDAQ and S&P 500, but displayed significant positive correlation to VIX.



(a)



(b)

Figure 2. Comparison of the S&P-TSX Composite index (a) and VIX index (b) with the pessimistic rate (PR) for period from 1/1/2007 to 1/1/2014

Data source: Yahoo Finance.

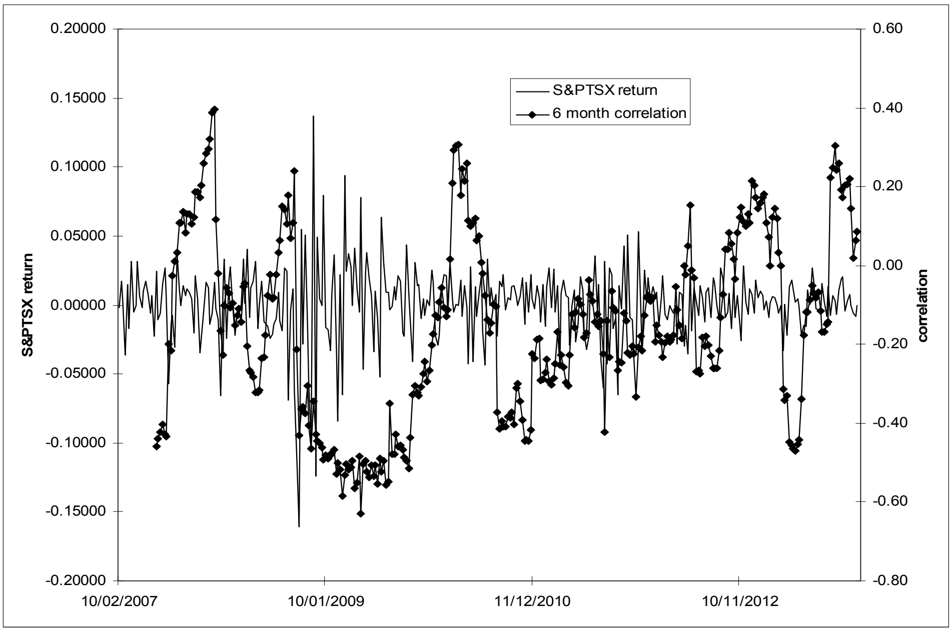


Figure 3. Correlation between the pessimistic rate with weekly return (S&P-TSX) time frame is 6 month

Data source: Yahoo Finance.

The similar sentiment indicator, which can be constructed from the pessimistic rate, is calculated from the difference of optimistic and pessimistic news to the total number of macroeconomic news. The proposed indicator is called a Panic Indicator (PI), Equation (2):

$$PI = \frac{optimistic\ macroeconomic\ news - pessimistic\ macroeconomic\ news}{optimistic\ macroeconomic\ news + pessimistic\ macroeconomic\ news} \quad (2)$$

The panic indicator can be expressed through the pessimistic rate:

$$PI = \frac{1-k_p}{1+k_p} \quad (3)$$

The range of the Panic Indicator is-1 to +1, i.e., the minimum value (-1) corresponds to the extreme case, when all published macroeconomic articles are negative; neutral market corresponds to the PI values close to zero, bullish market conditions correspond to the high positive PI. Since the pessimistic rates change from the large range from 0.5 to 3.14, see the historical distribution of PR, Figure 4, we consider that the panic indicator is more suitable for the quantitative analysis of stock price return.

The historical distribution of the panic indicator is shown in Figure 5. It is interesting to observe that the distribution of the Panic Indicator exhibits two picks near the neutral PI and negative skewness. Comparison of PI values with S&PTSX index shows that if PI values is lower than (-0.2), the market reaches the bearish market conditions.

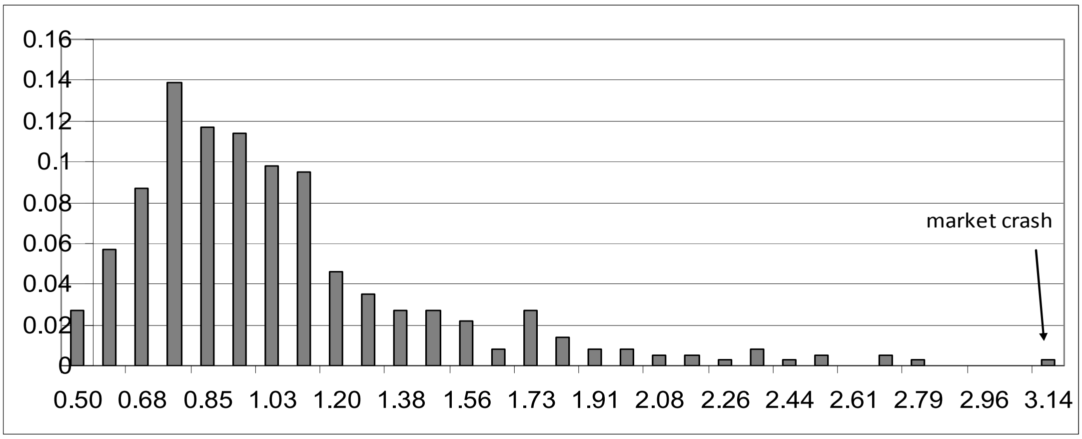
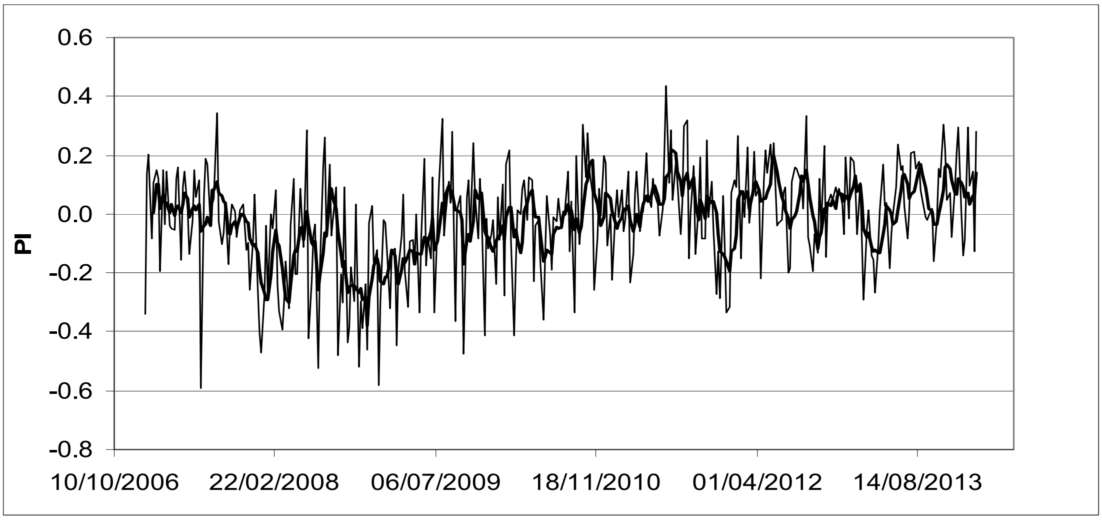
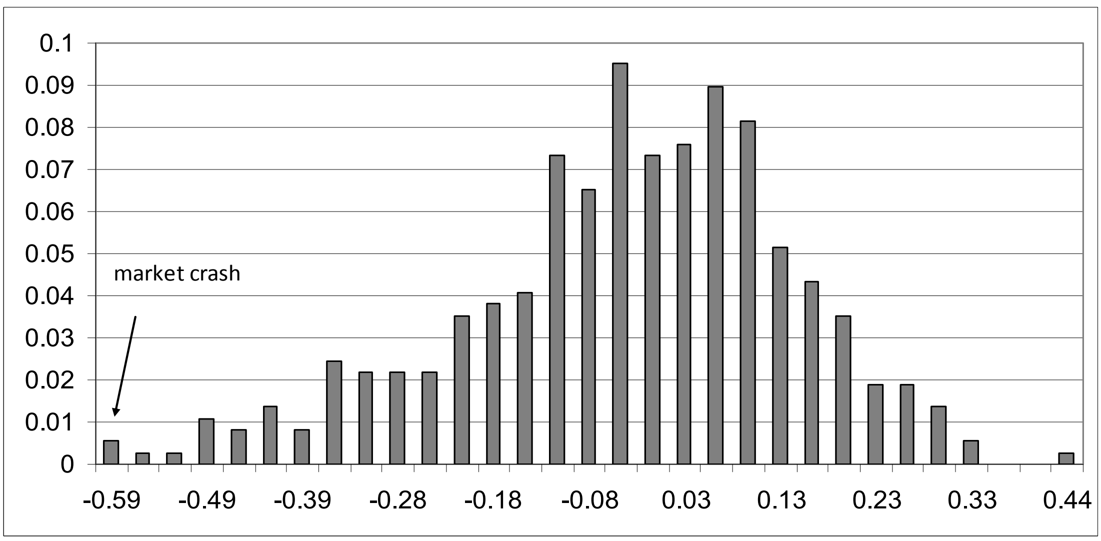


Figure 4. The historical distribution of the pessimistic rates for the period from 1/1/2007 to 1/1/2014



(a)



(b)

Figure 5. (a) Time-series of the Panic Indicator calculated from Equation (2)

(b) The historical distribution of the Panic Indicator

The proposed panic indicator is very straightforward, using the total number of positive and negative news, we may classify the bearish or bullish market. Moreover, the extreme events are clearly observed in both in analytical news and market indices, since the significant investor pessimism can be observed from business news, which is clearly confirmed by the current report. Although we may construct many different sentiment indicators from business news, we consider that the proposed sentiment indicator can be easily explained and replicated for the stock price forecast.

Obviously, the proposed indicator can be easily applied for the regression analysis of the stock price return. The application of the panic indicator could work well for the various herding strategies, when the traditional macroeconomic and fundamental factors don't have predictive power. For example, from the observed correlation between sentiment indicator and S&P-TSX index returns (Figure 3) during financial crisis, we can apply a simple regression formula for the analysis of the stock price forecast:

$$r(t_0,t_1) = a_0 + a_1 PI(t_{-1},t_0) + a_2 PI(t_{-2},t_{-1}) + a_3 PI(t_{-3},t_{-2}) + \dots \tag{4}$$

where $r(t_0,t_1)$ is the stock price return for future period (t_0,t_1), $PI(t_{-1},t_0)$ is the panic indicator calculated for the "past" period (t_{-1},t_0), $a_0, a_1, a_2, \dots$ are the regression coefficients.

The results of the sentiment measurements demonstrated a good performance of the panic indicator during the financial crisis. The proposed panic indicator would be useful for stock price modeling, description of the pessimistic opinions, and computational trading algorithms. The proposed sentiment indicator (Panic Indicator) has following advantages. (1) The application of the Panic Indicator can provide alternative information for the analysis of the bullish or bearish market conditions. (2) The Panic Indicator can be an additional source of the market and liquidity risk, when the stock price decline is controlled by the negative information flows. (3) The proposed indicator exhibits the predictive power for the extreme market conditions.

References

Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives, 21*(2), 129–151. http://dx.doi.org/10.1257/jep.21.2.129

Beth, E. W. (1995). *Semantic Entailment and Formal Derivability* (*Mededelingen der Koninklijke Nederlandse Akademie van wetenschapen, afd. Letterkunde*).

Birza, G., & Lott, J. R. (2011). The effect of macroeconomic news on stock returns: New evidence from newspaper coverage. *Journal of Banking & Finance, 35*(11), 2791–2800. http://dx.doi.org/10.1016/j.jbankfin.2011.03.006

Bondecka-Krzykowska, I. (2005). Semantic tree method—historical perspective and applications. *Annales UMCS Informatica AI, 3*(2005), 15–25

Brennan, M. J. (1993). *Agency and Asset Prices*. Finance Working Paper No. 6–93, UCLA.

Brown, G. W., & Cliff, M. T. (2004). Investor Sentiment and the Near-Term Stock Market. *Journal of Empirical Finance, 11*(1), 1–27. http://dx.doi.org/10.1016/j.jempfin.2002.12.001

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy, 98*(4), 703–738. http://dx.doi.org/10.1086/261703

Graham, J. R. (1999). Herding among Investment Newsletters: Theory and Evidence. *The Journal of Finance, LIV*(1), 237–268. http://dx.doi.org/10.1111/0022-1082.00103

Harris, Z. (1954). Distributional Structure. *Word, 10*(2/3), 146–162.

Hong, H., Lim T., & Stein, J. (2000). Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategy. *The Journal of Finance, LV*(1), 265–295. http://dx.doi.org/10.1111/0022-1082.00206

Liu, B. (2010). *Sentiment Analysis and Subjectivity, Invited Chapter for the Handbook of Natural Language Processing* (2nd ed.). March.

Mitchell, M. L., & Mulherin, J. H. (1994). The Impact of Public Information on the Stock Market. *Journal of Finance, 49*, 923–950. http://dx.doi.org/10.1111/j.1540-6261.1994.tb00083.x

Ormos, M., & Vazsonyi, M. (2011). Impact of Public News on Stock Market Prices: Evidence from S&P 500. *Interdisciplinary Journal of Research in Business, 1*(2), 1–17.

Remorov, R. (2014). Stock Price and Trading Volume at the Market Panic Conditions. *International Journal of Marketing Studies, 6*(1), 22–30. http://dx.doi.org/10.5539/ijms.v6n1p21

Sekine, S., & Nobata, C. (2003). Definition, dictionaries and tagger for Extended Named Entity Hierarchy. *Proceedings of the LREC.*

Seo, Y. W., Giampapa, J., & Sycara, K. (2002). *Text Classification for Intelligent Portfolio Management*. Robotics Institute, Canegie Mellon University.

Scharfstein, D. S., & Stein, J. C. (1990). Herd Behavior and Investment. *The American Economic Review, 80*(3), 465–479.

Soros, G. (1987, 2003). *The Alchemy of Finance*. Wiley & Sons, Incorporated, John.

Tarde, G., & Parsons, E. W. C. (1903). *The Laws of Imitation* (pp. 393). New York: H. Holt and Company.

Thomas, J. D., & Sycara, K. (2002). *Integrating Genetic Algorithms and Text Learning for Financial Prediction.* Genetic and Evolutionary Computation Conference (GECCO), Las Vegas, NV.

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firm's Fundamentals. *Journal of Finance, LXIII*(3), 1437–1467. http://dx.doi.org/10.1111/j.1540-6261.2008.01362.x

Tolle, K. M., & Chen, H. (2000). Comparing Noun Phrasing Techniques for Use with Medical Digital Library Tools. *Journal of the American Society for Information Science, 51*(4), 352–370. http://dx.doi.org/10.1002/(SICI)1097-4571(2000)51:4<352::AID-ASI5>3.0.CO;2-8

Welch, I. (1992). Sequential Sales, Learning and Cascades. *Journal of Finance, 47*, 695–732. http://dx.doi.org/10.1111/j.1540-6261.1992.tb04406.x

Xie, B., Passonneau, J., Wu, L., & Creamer, G. (2013). *Semantic Frames to Predict Stock Price Movement* (pp. 873–883). Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics. Sofia, Bulgaria, August 4–9.

Zhang, X., Fuehres, H., & Gloor, P. (2011). Predicting Stock Market Indicators through Twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences, 26*, 55–62. http://dx.doi.org/10.1016/j.sbspro.2011.10.562