DATA MINING TWITTER TO PREDICT STOCK MARKET MOVEMENTS

Maxim PECIONCHIN[1], PhD candidate
University of International Business and Economics, Beijing, China
Muhammad USMAN[2], PhD candidate
University of International Business and Economics, Beijing, China

In this paper we apply sentiment analysis of Twitter data from July through December, 2013 to find correlation between users' sentiments and NASDAQ closing price and trading volume. Our analysis is based on the Affective Norms for English Words (ANEW). We propose a novel way of determining weighted mood level based on PageRank algorithm. We find that sentiment data is Granger-causal to financial market performance with high degree of significance. "Happy" and "sad" sentiment variables' lags are strongly correlated with closing price and "excited" and "calm" lags are strongly correlated with trading volume.

Key words: *sentiment analysis, opinion mining, financial market, trading volume.*

În această lucrare am aplicat analiza sentimentelor de date Twitter din iulie până în decembrie anul 2013, pentru a găsi corelația între sentimentele utilizatorilor şi prețul de inchidere NASDAQ şi volumul tranzacțiilor. Analiza noastră se bazează pe Norme Afective pentru Cuvinte Engleze (NACE). Noi propunem un mod nou de determinare a nivelului stării de spirit ponderat bazat pe algoritmul PageRank. Considerăm, că datele sentimentelor sunt legate prin relatia cauza-effect după Granger cu performanța pieței financiare cu grad ridicat de importanță. Lag-uri "Fericit" și "trist" sunt corelate puternic cu prețul de închidere și lag-uri "excitat" şi "calm" sunt puternic corelate cu volumul de tranzacționare.

Cuvinte cheie: *analiza sentimentelor, extragerea opiniilor, piețe financiare, volumul tranzacțiilor.*

В этой статье мы применяем анализ тональности текста данных платформы Twitter в период с июля по декабрь 2013 года, чтобы изучить корреляцию между настроением пользователей и ценой закрытия и объемом торгов биржи NASDAQ. Наш анализ основан на Аффективных Нормах Английских Слов (ANEW). Мы предлагаем новый способ определения взвешенного уровня настроения на основе алгоритма PageRank. Мы обнаружили, что настроение пользователей является каузальным по Грейнджеру по отношению к показателям финансового рынка с высокой степенью статистической значимости. Лаги переменных настроения "счастливый" и "грустный" статистически значимо коррелируют с ценой закрытия, а лаги переменных "возбужденный" и "спокойный" коррелирует с объемом торгов.

Ключевые слова: *анализ тональности текста, экстракция мнения, финансовые рынки, объем торгов.*

JEL Classification: C81; G12; F13; F10

Introduction. This paper offers insight in the way in which the mood of Internet users is correlated to the financial market performance. The question being addressed in this paper is whether a person's mood affects his investment decisions and whether there exists a lag between the mood of a person and the resulting investment decisions of investors as represented by NASDAQ index and its trading volume.

There is no direct way to determine population's mood level at each moment in time, but there are ways to measure it indirectly. In this paper, we use sentiment analysis techniques to analyze more than 275 million "tweets" that correspond to 6 months of Twitter data stream from July to December 2013. The six sentiment variables, representing different dimensions of mood are the proxy of population's mood level.

We then proceed to test whether the sentiment variables obtained from Twitter have causal effect on NASDAQ hourly price, as well as its trading volume. The result is that there exists statistically significant

[1] ***© Maxim PECIONCHIN, maxim.pechyonkin@gmail.com***
[2] ***© Muhammad USMAN, usman.bhutta@yahoo.com***

evidence exhibiting correlation between NASDAQ indicators and preceding Twitter users' sentiment values. We find that sentiment variables representing happy and sad emotions affect NASDAQ price the most, and the other sentiment variables such as excitement, calmness, dominance and submissiveness have somewhat of less effect.

The contribution of this paper is that the data used for analysis is hourly data, whereas the previous research was focused on daily data. The hourly data allows understanding the dynamics of sentiments affecting the stock market at a finer scale.

In addition, we employ a weighting algorithm that allows to better estimate the effect of sentiments on NASDAQ market. The motivation behind the weighting is to allow the more popular users of Twitter to have more effect on total sentiment value than the less popular users. We find that this weighting procedure results in more statistically significant results.

In the following section, we are going to present current research in this area. In section “The Data”, we describe the data set. Then in section “Methodology”, we describe basic assumptions that we are making that may be driving our results, the work we have done to measure the sentiment level of Twitter stream and then the econometric methodology we used to infer causality of sentiment variables on NASDAQ performance. Finally, we present our empirical results in section “Results”.

Literature review. Research in behavioral economics tells us that emotions of an investor can affect its investment decisions. Being unable to directly observe investor's emotions, we try to find an appropriate proxy to person's emotional state. Social media can be treated as such proxy. We assume that what people post on Twitter represents how they think and what they feel.

Bollen et al. [3] use more than 9 million tweets to explore the possibility of predicting the stock market movements. They use tools called OpinionFinder, which is a system that processes documents and automatically identifies subjective sentences as well as various aspects of subjectivity within sentences, including agents who are sources of opinion, direct subjective expressions and speech events, and sentiment expressions. They also build GPOMS (Google-profile of Mood States) to range the mood level of tweets measured by six dimensions: calm, alert, sure, vital, kind and happy. Then they proceed to use Granger's causality test to determine whether mood metrics have predictive information on the stock price. They find that only GPOMS Calm metric has predictive power on the stock market. They also fit a nonlinear model, where mood causes stock price change nonlinearly. They use self-organizing fuzzy neural network for this purpose. The result is claimed 87.6% accuracy in predicting up and down movements of DJIA over the period February 28 – December 20, 2008.

Gilbert et al. [9] in their paper explore whether the data from more than 20 million LiveJournal posts can be used to predict movements in the S&P 500 index. They construct the so called Anxiety Index and test it using Granger's causality test. They find that the Anxiety Index has some novel predictive information and anxiety is negatively correlated with future S&P 500 index value.

Mittal et al. [14] base their work on Bollen's [3] methodology. Although they use simplified approach based on POMS to construct sentiment index in four dimensions: calm, happy, alert and kind, they have much larger dataset of 476 million tweets from July to December 2009. They apply Granger's causality test, the results are that calm and happy dimensions have the strongest predictive power. After the authors proceed to assess the effectiveness of predictions of stock price of four tools, namely, linear regression, logistic regression, support vector machine (SVM) and self-organizing fuzzy neural network (SOFNN). They find that logistic regression and SVM are the least accurate, whereas SOFNN is the most accurate tool with the level of price directional prediction accuracy as high as 75%. Then the authors implement simple trading algorithm that shows its effectiveness over the period of 40 consecutive days demonstrating the profit of more than 520 Dow Points.

Ruiz et al. [16] were the first to use graph theoretic approach to analyze correlation between Twitter data and S&P 500 companies' financial time series during the first half of 2010. They build a graph of all the tweets relating to 150 randomly chosen S&P 500 companies and analyze this graph's features. Then they regress these features to the stock's price and trading volume using lagged cross-correlation coefficient. They find that certain graph features are closely correlated with trading volume and correlated weakly with the stock's price. Namely, the number of tweets about a particular company, the number of different users who tweeted about that company and the number of connected components in the company's graph are the most important features. Despite the weak correlation between these features and the company’s stock price, the authors demonstrate that this information can be successfully used to outperform a number of baseline trading strategies. They compare randomized, fixed profit margin, autoregression, their Twitter-augmented regression strategies and simply buying DJIA index. The result is that their strategy is the only one that demonstrated positive average return, even when all the other strategies and DJIA were declining between March and June 2010.

Zhang et al. [19] using rather simple methodology by simply counting the number of keywords appearing in tweets collected from March through September 2009 demonstrate that "negative" words have negative correlation with stock market returns, whereas "positive" words have positive correlation.

The data. We use a large dataset from social microblogging platform Twitter comprised of more than 275 million tweets, which corresponds to a period from July through December 2013. The data was gathered by ArchiveTeam research group and is publicly available on the Internet [20]. The data feed represents what Twitter calls "Firehose" feed, which is a portion of actual stream of all the tweets in real time, representing approximately 5 percent of all the tweets. Each tweet is presented as a "tweet object" that not only packs text of the tweet, but also provides a list of metadata, such as the time of posting, language, number of people, following the author of the tweet etc. All these metadata provide abundant resource for analysis.

After gathering, the data were processed to leave only tweets in English with metadata used in later analysis, such as the text of a tweet, time of posting and the number of followers of the author of the tweet.

Financial data for NASDAQ index from July through December 2013 was obtained from the Russian "Finam" investment holding company [21]. The stock market was open 8 hours every day, excluding national holidays and days-off, for 127 days. In total, we have 1012 hourly data points during these 6 months. The variables we get from the data set are time series for hourly closing price (CLOSE) and hourly trading volume (VOL).

Methodology. In this paper, we make the following theoretical assumptions. First, we assume that a person's mood can affect their investment decisions. Second, that social and informational environment can affect person's mood. This leads to the third assumption that social and informational environment can affect person's investment decisions.

We use sentiment analysis of Twitter data to measure the mood level of social environment. Then we use Granger's causality test to investigate whether the mood state of people has any predictive power on investment decisions of investors in the NASDAQ stock market.

Sentiment Analysis. We use sentiment analysis technique based on Affective Norms for English Words (ANEW) dictionary. It was developed "to provide a set of normative emotional ratings for a large number of words in the English language" [4]. ANEW is a list of 1,034 words rated among three dimensions, according to Osgood, Suci and Tannenbaum's theory of emotions [15]. The first is the valence of the emotions invoked by the word, going from unhappy to happy. The second is the level of arousal evoked by the word, going from calm to excited. The third dimension refers to the dominance (power) of the word, going from weak (submissive) to strong (dominant). The original results presented in [4] are normalized to take values in range from -4 to 4.

This research analyzes every tweet on a 6-dimensional sentiment scale:

1. happy (valence > 0);
2. sad (valence < 0);
3. excited (arousal > 0);
4. calm (arousal < 0);
5. dominant (dominance > 0);
6. submissive (dominance < 0).

Each tweet is separated into a list of words, and then all the words are converted to lower case. Any word in a tweet that is on the ANEW list is then mapped back to the corresponding sentiment value, hence forming a 6-dimensional sentiment vector. Then all the word vectors for a given tweet are added together to form the tweet sentiment vector $\boldsymbol{m_{tweet}}$.

The mood vector $\boldsymbol{m_t}$ of length *k* for a given period *t* can be computed by adding all the tweet vectors $\boldsymbol{m_{tweet}}$ in the set of all tweets $\boldsymbol{T_t}$ for the given period:

$$m_t = \sum_{\forall t \subset T_t} m_{tweet}$$

PageRank. Novel contribution of this paper is to use the PageRank as a weight function to determine total current mood state of the Twitter community. Google founders Sergey Brin and Larry

Page [5] introduced PageRank in 1999 in an attempt to improve web search and make it more relevant. Each tweet can mention other user, or it can be retweeted. In the first case we count it as an outbound link, in the latter it is an inbound link. The idea of PageRank is quite simple. If a large number of users mentions some user that means that this user has high rank. On the other hand, if this user mentions some other user and suppose that user is only mentioned once, that user still will have a rather high rank because a user with such a high rank mentioned him.

However, as was shown by Kwak et al. [13], it is possible to avoid a complex and cumbersome computation using PageRank on a user directed graph with more than 50 million nodes and more than 2 billion links. He showed that ranking users by follower count bears the same results as ranging users by PageRank. They used Kendall's tau rank correlation coefficient and found it to be no lower than 0.6, which implies significant similarity between two rankings. This allows achieving the same results, but avoiding resource demanding and timing consuming computation. Hence, in this paper we use the number of followers as proxy of PageRank weight.

We are going to test two types of weight coefficients: one that is equal to the number of followers of a user, and another one equal to the natural logarithm of the number of followers of a user. The motivation behind taking logarithm of the number of followers is that in this way the sentiments of a small number of extremely popular users will not be able to dominate overall sentiment level of all Twitter users. This logic is supported by the fact that all social networks and Twitter in particular exhibit large-scale properties. In other words, there is a small number of users that have the number of followers exceeding to many degrees of magnitude the median number of user followers [2].

Granger's causality test. The test was developed by Granger in 1969 [10]. It is based on the assumption that if a random variable X causes Y, then changes in X will systematically occur before changes in Y [3]. If the mood state values time series contains predictive information about the stock market, then we might expect lagged values of the mood series to show statistically significant correlation with the financial indicators of stock market.

The test compares a pair of models, where the second model adds variables of interest, up to *n* lags back (better description of models here). So *M0* is the null model, it is nested in *Mx*, which is the alternative model. We are going to test several alternative models incorporating different combinations of sentiment lag values and then perform the likelihood-value Wald test of the two models.

$$\boldsymbol{M0}: M_t = \alpha + \sum_{i=1}^{n} \beta_i M_{t-i} + \varepsilon_t$$

$$\boldsymbol{Mx}: M_t = \alpha + \sum_{i=1}^{n} \beta_i M_{t-i} + \sum_{k=1}^{6} \sum_{i=1}^{n} \lambda_{k,i} m_{k,t-i} + \varepsilon_t$$

M_t is the first difference of the dependent variable CLOSE or VOL. In the case of closing price, we use the first-difference in log-returns, as it is the standard solution to achieve sufficient stationarity of a time series, at least asymptotically [9]. M_{t-i} is the i-th lag of dependent variable and $m_{k,t-i}$ represents k-th element of the lag mood vector $\boldsymbol{m_{t-i}}$. Note that the base model M0 is nested in all models Mx.

Prior to evaluating the models above, we tested all variables for stationarity using Augmented Dickey-Fuller test. The null hypothesis is of existence of the unit root (non-stationary process); the alternative hypothesis is that there is no unit root (stationary process). The results show that neither of the variables is non-stationary.

We decided to limit our analysis to 120 hourly lags, which corresponds to 5 days of lead-time of public sentiment affecting investment decisions in the financial market. For each data point in the 2 dependent variables: NASDAQ hourly close price (CLOSE) and trading volume (VOL) we computed 120 corresponding lags for all six sentiment variables as well as dependent variables themselves. However, including all the 120 hourly lags for dependent variables as well as for each of six sentiment time series means estimating a model with 720 parameters. This leads to over-specification problem. In order to solve this issue, we only include statistically significant lags after estimation of cross-correlation functions for CLOSE and VOL. The resulting cross-correlograms can be observed in Figures 1 and 2. Dark colored bars represent lags at 99% significance level; grey bars represent insignificant lags.

Results. All the models were evaluated using GLM. We evaluated 24 models for both CLOSE and VOL, which tested different combinations of sentiment variables and statistical significance of their predictive power on dependent variables.

We use Wald test of nested models to test for goodness of fit of models that use sentiment variables. The null hypothesis is that M0 model is true model and adding sentiment lags will make the model fit worse. For all 48 models with high level of significance (<1%) we can reject the null hypothesis. This means that including significant sentiment lags are Granger-causal for our data.

To further investigate the quality of the 48 models, we ranked them based on the Akaike information criterion (AIC). Models with the AIC less than that of *M0* are good in the sense of balance between goodness of fit and complexity of a model. If AIC of a model is higher than AIC of the base model, the interpretation is that some information was lost by including additional variables into the model and its quality is lower.

The main result is that log-weighted models in all cases are superior to the base models *M0*, and no plain-weighted models exhibit higher quality than the base models.

We find that for NASDAQ closing price (CLOSE) two models provide the best predictors. One model includes only lags corresponding to sentiment variables HAPPY and SAD, and the second model includes only lags corresponding to negative sentiment variables SAD, CALM and SUBMISSIVE. For NASDAQ trading volume (VOL) the best model includes only lags for sentiment variables EXCITED and CALM. Details on these models can be found in the table below.

Upon examining significant models, we can see that closing price is positively correlated with 48-th and 55-th “happy” and “calm” lags, and negatively correlated with 72-nd “sad” lag, all at 5% significance level. This means that happy or calm Twitter mood will likely cause closing price of NASDAQ go up 48 to 55 hours later, and if the mood is sad, this will lead closing prices down 72 hours later.

As for trading volume, we see that there is a very significant ($p < 0.1\%$) positive correlation with 62-nd “excited” lag and very significant ($p < 0.1\%$) negative correlation with 62-nd “calm” lag. This implies that if Twitter mood is very excited, then trading volume of NASDAQ will likely increase 62 hours later, and the opposite is true if Twitter mood is very calm.

Table 1

Regression results for NASDAQ closing price

Variable name	M0			M8			M12		
	Coef.	SD	Sign.	Coef.	SD	Sign.	Coef.	SD	Sign.
AIC		**6,068.0**			**6,035.3**			**6,039.8**	
Wald test p-value		**-**			**< 0.001**			**< 0.001**	
intercept	0.697	(0.309)	*	0.588	(0.310)	'	0.590	(0.310)	'
close lag 39	-0.081	(0.033)	*	-0.091	(0.033)	**	-0.085	(0.033)	*
close lag 72	-0.088	(0.034)	**	-0.098	(0.035)	**	-0.104	(0.035)	**
close lag 119	0.094	(0.034)	**	0.095	(0.034)	**	0.093	(0.034)	**
lgw happy lag 16				0.883	(0.924)				
lgw happy lag 48				0.595	(0.254)	*			
lgw happy lag 55				0.491	(0.227)	*			
lgw happy lag 83				0.120	(0.669)				
lgw happy lag 120				-0.076	(0.812)				
lgw sad lag 16				0.170	(0.778)		1.692	(1.867)	
lgw sad lag 72				-0.547	(0.238)	*	-0.499	(0.239)	*
lgw sad lag 83				0.790	(0.623)		-0.164	(1.502)	
lgw sad lag 120				0.678	(0.735)		0.520	(2.367)	
lgw calm lag 16							-1.479	(0.855)	'
lgw calm lag 48							0.481	(0.242)	*
lgw calm lag 55							0.468	(0.213)	*
lgw calm lag 57							0.279	(0.205)	
lgw calm lag 83							-0.003	(0.956)	

lgw calm lag 120	-0.747	(1.366)
lgw submissive lag 16	0.561	(2.116)
lgw submissive lag 83	1.013	(1.690)
lgw submissive lag 120	0.814	(2.613)

Note: 'p<0.1; *p<0.05; **p<0.01; ***p<0.001

Source: Authors.

Table 2

Regression results for NASDAQ trading volume (source: authors)

Variable name	**M0**			**M9**		
	Coef.	**SD**	**Sign.**	**Coef.**	**SD**	**Sign.**
AIC	**6,482.6**			**2,241.5**		
Wald test p-value	**-**			**< 0.001**		
intercept	0.001	(0.033)		0.002	(0.032)	
vol lag 1	-0.525	(0.032)	***	-0.499	(0.033)	***
vol lag 2	-0.300	(0.031)	***	-0.289	(0.031)	***
vol lag 8	0.075	(0.028)	**	0.078	(0.029)	**
vol lag 48	0.398	(0.044)	***	0.355	(0.044)	***
vol lag 49	0.104	(0.046)	*	0.104	(0.045)	*
vol lag 120	-0.165	(0.041)	***	-0.141	(0.041)	***
lgw excited lag 33				-0.006	(0.024)	
lgw excited lag 37				0.034	(0.027)	
lgw excited lag 50				-0.033	(0.026)	
lgw excited lag 59				-0.070	(0.069)	
lgw excited lag 62				0.365	(0.086)	***
lgw excited lag 77				-0.140	(0.060)	*
lgw excited lag 82				0.075	(0.026)	**
lgw excited lag 105				0.039	(0.026)	
lgw excited lag 107				-0.003	(0.027)	
lgw excited lag 111				-0.143	(0.113)	
lgw calm lag 7				-0.047	(0.023)	*
lgw calm lag 15				0.017	(0.024)	
lgw calm lag 40				0.077	(0.028)	**
lgw calm lag 41				0.063	(0.030)	*
lgw calm lag 59				-0.001	(0.060)	
lgw calm lag 62				-0.302	(0.075)	***
lgw calm lag 77				0.094	(0.059)	
lgw calm lag 86				0.004	(0.026)	
lgw calm lag 88				0.053	(0.027)	*
lgw calm lag 111				0.165	(0.097)	'

Note: 'p<0.1; *p<0.05; **p<0.01; ***p<0.001

Source: Authors.

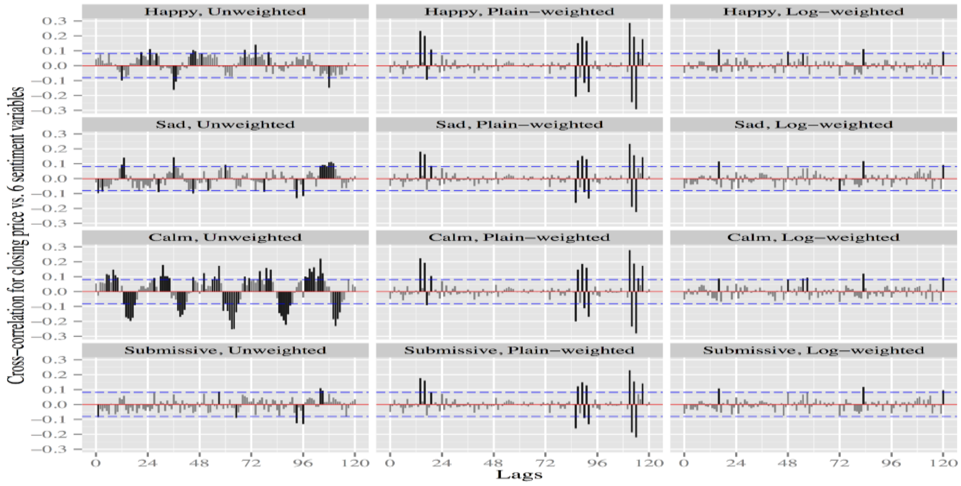

Fig. 1. Cross-correlations for closing price and six sentiment variables' lags
Source: Authors.

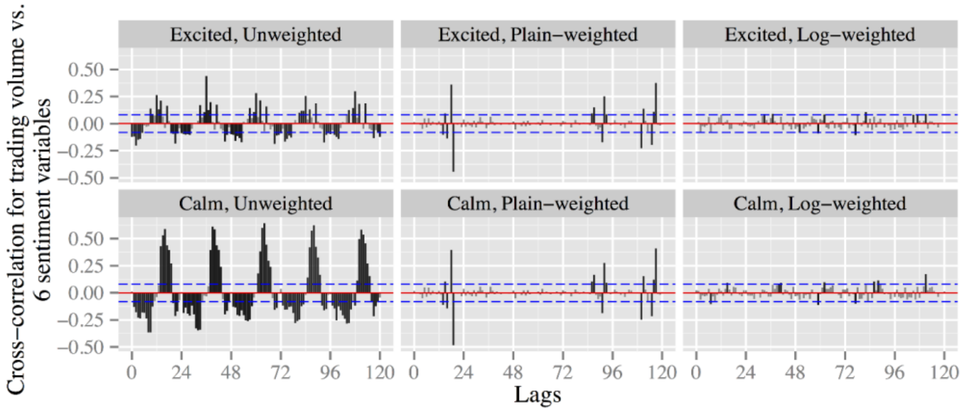

Fig. 2 Cross-correlations for trading volume and six sentiment variables' lags
Source: Authors.

REFERENCES

1. ARNAUD, V., ARMSTRONG, M. Predicting break points in trading strategies with Twitter. 2010, october [accesat 16 ianuarie 2015]. Disponibil: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1685150
2. BARABASI, A., BONABEAU, E. Scale-Free Networks. In: Scientific American. 2003, may, pp. 60-69.
3. BOLLEN, J, MAO, H., ZENG, X-J. Twitter mood predicts the stock market. In: Journal of Computational Science. 2011, vol. 2, issue 1, pp. 1-8.
4. BRADLEY M.M., LANG P.J. Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings. 1999 [accesat 12 ianuarie 2015]. Disponibil: http://a.parsons.edu/~spani621/thesis/context/ANEW.pdf
5. BRIN, S., PAGE, L. The PageRank Citation Ranking: Bringing Order to the Web. 1999 [accesat 12 februarie 2015]. Disponibil: http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf
6. CHA, M., HADDADI, H. et al. Measuring User Influence in Twitter: The Million-Follower Fallacy. In: Proceedings of the Fourth International AAAI Conference on WEblogs and Social Media, may 23-26 2010. Washington, 2010, pp. 10-17.
7. CHEN, A-S., LEUNG, M. T., DAOUK, H. Application of Neural Networks to an Emerging Financial Market: Forecasting and Trading the Taiwan Stock Index. In: Computers&Operations Research. 2003, vol. 30, issue 6, pp. 901-923.
8. CHOUDHURY, M. D., SUNDARAM, H. et al. Can Blog Communication Dynamics be correlated with Stock Market Activity? 2008 [accesat 24 ianuarie 2015]. Disponibil: http://www.munmund.net/pubs/ht_2_08.pdf

9. GILBERT, E., KARAHALIOS, K. Widespread Worry and the Stock Market. 2009 [accesat 5 februarie 2015]. Disponibil: http://social.cs.uiuc.edu/people/gilbert/pub/icwsm10.worry.gilbert.pdf
10. GRANGER, C. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. In: Econometrica. 1969, vol. 37, no. 3, pp. 424-438.
11. GRANGER, C., NEWBOLD, P. Spurious Regressions in Econometrics. In: Journal of Econometrics. 1974, no. 2, pp. 111-120.
12. HUBERMAN, B.A., ROMERO, D.M., WU, F. Social Networks That Matter: Twitter under microscope. 2008, december [accesat 24 ianuarie 2015]. Disponibil: http://arxiv.org/pdf/0812.1045.pdf
13. KWAK, H., LEE, C., PARK, H., MOON, S. What is Twitter, a social network or a news media? 2010 [accesat 24 ianuarie 2015]. Disponibil: http://www.eecs.wsu.edu/~assefaw/CptS580-06/papers/2010-www-twitter.pdf
14. MITTAL, A., GOEL, A. Stock Prediction Using Twitter Sentiment Analysis. 2010 [accesat 24 ianuarie 2015]. Disponibil: http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf
15. OSGOOD, C.E., SUCI, G.J., TANNENBAUM, P. *The Measurement of Meaning*. Champaign: University of Illinois Press, 1957. 342 p.
16. RUIZ, E. J., CASTILLO, C. et al. Correlating Financial Time Series with Micro-Blogging Activity. In: Web Search and Data Mining: international conference, february 8-12 2012. Seattle, 2012 [accesat 24 ianuarie 2015]. Disponibil: http://www.cs.ucr.edu/~vagelis/publications/wsdm2012-microblog-financial.pdf
17. RUSSEL, J.A. A Circumplex Model of Affect. In: Journal of Personality and Social Psychology. 1980, vol. 39, no. 6, pp. 1161-1178.
18. SHARMA, J., VYAS, A. Twitter sentiment analysis. 2010 [accesat 24 ianuarie 2015]. Disponibil: http://www.cse.iitk.ac.in/users/cs365/2012/submissions/jaysha/cs365/projects/
19. ZHANG, X., FUEHRES, H., GLOOR, P.A. Predicting Stock Market Indicators Through Twitter: “I hope it is not as bad as I fear”. 2011 [accesat 16 ianuarie 2015]. Disponibil: http://www.ickn.org/documents/COINs2010_Twitter4.pdf
20. Archive Team: The Twitter Stream Grab [accesat 9 ianuarie 2015]. Disponibil: https://archive.org/details/twitterstream
21. Финам [accesat 24 ianuarie 2015]. Disponibil: http://www.finam.ru
22. Yahoo Finance [accesat 24 ianuarie 2015]. Disponibil: http://finance.yahoo.com

Recommended for publication: 23.02.2015