NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Alexander V. Porshnev, Valeria V. Lakshina
Ilya E. Red'kin

COULD EMOTIONAL MARKERS IN TWITTER POSTS ADD INFORMATION TO THE STOCK MARKET ARMAX-GARCH MODEL

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: FINANCIAL ECONOMICS
WP BRP 54/FE/2016

SERIES: *FINANCIAL ECONOMICS*

Alexander V. Porshnev, [1] *Valeria V. Lakshina* [2]
Ilya E. Red'kin [3]

COULD EMOTIONAL MARKERS IN TWITTER POSTS ADD INFORMATION TO THE STOCK MARKET ARMAX-GARCH MODEL

Abstract

In our paper, we analyze the possibility of improving the prediction of stock market indicators by adding information about public mood derived from Twitter posts. To estimate public mood, we analyzed the frequencies of 175 emotional markers — words, emoticons, acronyms and abbreviations — in more than two billion tweets collected via Twitter API over the period from 13.02.2013 to 22.04.2015. We found that, from 17 emotional markers frequencies with established Granger causality, six provide additional information for the baseline ARMAX-GARCH model according to Bayesian information criteria for the in-sample period of 421 days, and two emotional markers improve directional accuracy and a decrease in the mean-squared error of the model. Our analysis reveals several groups of emotional markers, such as general and specific, direct and indirect, which relate differently to the dynamics of returns.

JEL Classification: G17, G10, G14.

Keywords: Twitter, mood, emotional markers, stock market, volatility.

[1]National Research University Higher School of Economics. Department of General and Strategic Management. Associate Professor. E-mail: aporshnev@hse.ru. Nizhniy Novgorod, Russia.

[2]National Research University Higher School of Economics. Department of Mathematical Economics. Lecturer. E-mail: vlakshina@hse.ru. Nizhniy Novgorod, Russia.

[3]Freelance Data Analyst, Nizhny Novgorod, Russia. E-mail: ilya-redkin@yandex.ru

1 Introduction

Mood, emotions and decision-making are closely connected. For example, Johnson and Tversky (1983) reported that psychological states invoked by reading reports about tragic events can affect evaluations of risk level. Positive moods lead individuals to make more optimistic choices and, conversely, negative moods lead to pessimistic choices (Isen & Patrick, 1983; Mayer & Gaschke, 1988; Mayer, Gaschke, Braverman, & Evans, 1992). In a positive mood people usually underestimate risks and focus on positive consequences of their decision, while in a sad mood they often concentrate on negative consequences (Loewenstein, 2000).

Individuals in a positive mood tend to spend less time on decision making by referring more rarely to alternatives that have already been reviewed and by ignoring information they believe is irrelevant, according to Isen and Means (1983).

The role of moods and emotions in decision-making grows in situations of uncertainty intrinsic to the stock market. Behavioral economists have found that a trader's decision-making process regularly demonstrates a wide set of human cognitive biases and is highly influenced by emotional factors (Falato, 2009; Hirshleifer & Shumway, 2003; Lepori, 2015a; Saunders, 1993).

Following Nofsinger (2005), we regard the economy not as a physical system, but as a complex system of human interactions, where moods and irrationalities can play a significant role. This point can be supported by observing the informational cascade phenomenon in the stock market (R. T. Zhou & Lai, 2009).

Publicly-expressed emotions in Facebook and Twitter have attracted the attention of many researchers (Bollen, Mao, & Zeng, 2011; Nofer & Hinz, 2015; Rao & Srivastava, 2012). A relation between Facebook's Gross National Happiness Index and 20 international markets is shown in Siganos, Vagenas-Nanos, and Verwijmeren (2014), who demonstrated that negative sentiments are associated with increases in trading volumes and return volatility. According to Kearney and Liu (2014), most of the 38 studies run in this area in 2004–2013 were devoted to the usage of news articles, annual financial reports, earnings press releases or other financial-related information, and only one dealt with the Internet messages. While the first approach appears to be more relevant, it may fail to recognize faked or republished historical news (Prezioso, 2013; Selyukh, 2013).

We propose a simple alternative measure of sentiment based on posts published by Twitter users. It is noteworthy that people use not only words to express their emotional state in Twitter, but also abbreviations and emoticons. We term all these signs "emotional markers".

In our view, information about Twitter users' sentiment is significant for the dynamics of financial time series, and emotional markers can therefore increase the explanatory power of financial time series models. This power is measured by several goodness of fit criteria, including Akaike, Bayesian and Hannan-Quinn information criteria (Javed & Mantalos, 2013).

Nofsinger (2005) supposes that the stock market itself can be a direct measure of social mood. Taking the stock market and Twitter as two possible indicators of social mood, we can assume their correlation and the possibility of using moods expressed in tweets to increase the explanatory and predictive accuracy of financial time series models. Although current research reports that Twitter mood could be used to enhance the quality of stock market forecasts, the validity of the conclusions reached in these works remains doubtful (Bollen et al., 2011; Chen & Lazer, 2013; Porshnev, Redkin, & Shevchenko, 2013; Zhang, Fuehres, & Gloor, 2011). As a matter of fact, early research either used a short (less than 40 days) out-of-sample period, failed to account for autocorrelations while modeling the returns, or even reported only correlations. In addition, in their detailed review of methods and models applied in textual sentiment analysis in the financial field, Kearney and Liu (2014) note that volatility models have rarely been used. For example, in Nofer and Hinz (2015), returns are modeled by linear regression without taking into account autoregressive and conditional heteroskedasticity effects.

Our research is aimed at filling this gap. We use ARMAX-GARCH for returns modeling, one of the most widespread and effective models in time series analysis, which allows us to capture the main distinctive features of financial time series (Horv & Kokoszka, 2003).

Assuming that some words, abbreviations and emoticons may be more related to returns, we test the hypothesis that emotional marker frequencies can add information and increase the explanatory power of a dynamic returns model.

The rest of the article is organized as follows: Section 2 explains the methodology employed in the research, Section 3 describes the data and data preprocessing, empirical results, and discusses findings, and Section 4 concludes.

2 Methodology

2.1 Emotional markers

One of the simplest and most intuitive methods for sentiment analysis is word counting, so we use the frequencies of words from a specially drawn-up list instead of combining them into one or several mood indexes[4]. Early in our research we tried to establish indexes of moods and sentiments of Twitter users in a way similar to that suggested by Bollen et al. (2011), but soon we realized that the application of machine learning techniques made this procedure non-transparent and dependent on the preliminary choice of relevant words. We therefore discarded the idea of applying complex sentiment analysis methods, and analyze frequencies of specially chosen emotional markers as a more transparent measure.

We compiled our list of emotional markers using a Brief Mood Introspection Scale with 8 scales and 2 adjectives representing each mood as the starting point in creating dictionaries (Mayer & Gaschke, 1988). We have augmented this list with the synonyms of the adjectives selected from the WordNet dictionary (Miller, 1995), SentiWordNet (Esuli & Sebastiani,

[4] A similar approach is used by Zhang et al. (2011). They analyze frequencies of several words (e.g. "worry", "hope", "fear" etc.) and find high correlation between the frequencies of emotional posts and S&P500, DJI, and VIX indexes.

2006) and Word Associations Network Project (Rotmistrov, 2016). For example, we measure the presence of an energetic state in tweets by the occurrence of the following words: animate, animated, athletic, brisk, chipper, emphatic, enterprising, exuberant, fresh, lusty, passionate, robust, sprightly, spry, strenuous, strong, tireless, trenchant, warming party, honor, and vote. Besides adding synonyms from the Brief Mood Introspection Scale, we have augmented our list with words which are associated with emotions, for instance, “cancer”, “hell”, “hang over”, etc. At a later stage in our analysis, we distinguish these words from direct emotional markers expressing emotions, like the word “fear”.

We also include recognizable derived words, such as “happyyy” or “happpppppyyyyyyy”, and counted them using regular expressions (see, for example, Friedl (2002)). Although it might be possible to treat the emotional states expressed by “I'm happy” and “I'm happpppppyyyyyyy” differently, we regard these posts as equal contributions to the frequency of the emotional marker “happy”.

We do not include negations, because after analyzing a testing sample of 9000 tweets we found that negations were not common. For example, a sample with fifty one instances of the word “happy” contained the negation “not happy” only once. The same is the case with “but” and sentences expressing desires, e. g. “wanna be happy”. The probable reason for the low frequency of negations is the small number of characters allowed for a Twitter message (140 characters).

Boia, Faltings, Musat, and Pu (2013) show that emoticons have a high degree of classificatory power, and that the accuracy of emoticon-based sentiment classification exceeds 90% for tweets with emoticons. Impressed by this result, we have augmented our list with all emoticons from Schnoebelen (2012), and distinguished different types of smile emoticons. For example, “:)”, “:-)”, and “:D” are not treated as synonyms.

Importantly, the Twitter lexicon contains numerous abbreviations and slang words, such as “LOL”[5] and “WTH”[6]. In our analysis, we have added abbreviations expressing emotional states from “The Online Slang Dictionary” (2016).

Our list of emotional markers contains 175 items in total. We count the number of posts per day which include each emotional marker, and consider this number an emotional marker frequency. During analysis, tweets are transferred to all-lower case format. The frequencies are included in the ARMAX-GARCH model as additional regressors in the mean equation.

2.2 Granger causality

As shown above, logarithmic returns of financial assets and emotions are correlated with each other. While there are evidently situations in which emotions cause returns, there are also cases in which the dynamics of the returns cause changes in people's moods. Clive Granger, in his seminal paper (1969), proposed a procedure which enables differentiation between these two situations statistically. The idea of Granger test is that, if one time series

[5] “laughing out loud”
[6] “what the hell”

precedes another time series, then the former most likely causes the latter. This kind of relation between time series has come to be called Granger causality.

Since we are interested in the situation in which the changes in returns' dynamics are caused by emotions, we employ a Granger causality test by estimating Eq. (1) in a way similar to S.-H. Kim and Kim (2014).

$$
\begin{aligned}
R_t &= a_0 + \sum_{i=1}^{L} \alpha_i R_{t-i} + \sum_{j=1}^{L} \beta_j X_{t-j} + \varepsilon_t, \\
X_t &= \widetilde{a}_0 + \sum_{i=1}^{L} \widetilde{\alpha}_i X_{t-i} + \sum_{j=1}^{L} \widetilde{\beta}_j R_{t-j} + \widetilde{\varepsilon}_t,
\end{aligned}
\tag{1}
$$

where R_t is asset's returns, X_t — is an emotional marker, a_0, α_i, β_j and their tilde counterparts are parameters, which help to identify the direction of Granger causality relations, ϵ_t and $\widetilde{\epsilon_t}$ are uncorrelated error terms, and i and j are summation indexes. In the Granger test, if β_j is significant at some chosen level (usually 5%), then emotional marker X_t indeed causes the log returns R_t with a lag over L periods; if $\widetilde{\beta}_j$ is significant, then the emotional marker is caused by log returns. In order to avoid inclusion of the reverse causation, we kept only those emotional markers for which β_j is significant and $\widetilde{\beta}_j$ is insignificant at the 5% level according to F-test. Thereby we avoid the reverse causality problem described in Brown and Cliff (2004). We call the resulting set of emotional markers Granger-valid.

We also found the optimal lag of each sentiment series X_t by varying the L parameter from 1 to 30, whereas in other research the lags for Granger tests do not commonly exceed 7 days (Bollen et al., 2011; Zhang et al., 2011).

2.3 ARMAX-GARCH model

To examine the impact of Twitter moods on the returns of stocks and stock market indexes, and to check if emotional markers can add new information to the model, this study uses the well-known ARMAX(p,q)-GARCH(r,m) model (see, for example, Francq and Zakoian (2004)). The model can be represented as in Eq. (2).

$$
\begin{aligned}
y_t &= R_t - E(R_t|\mathcal{F}_{t-1}), \\
y_t &= \sigma_t \cdot \eta_t,\ \eta_t \sim f(\theta),
\end{aligned}
\tag{2}
$$

where $E(R_t|\mathcal{F}_{t-1})$ is a conditional mean of daily returns R_t at time t conditional on all information available at $t-1$, F_{t-1}; y_t represents GARCH innovations; σ_t^2 represents volatility and η_t represents an error term, distributed according to some distribution f with parameter set θ. Returns R_t are calculated as a logarithm of today's price divided by yesterday's price: $R_t = log(\frac{p_t}{p_{t-1}})$. Conditional mean $E(R_t|\mathcal{F}_{t-1})$ is modeled as ARMAX(p,q), as in Eq. (3).

$$
E(R_t|\mathcal{F}_{t-1}) = a_0 + \sum_{i=1}^{p} \alpha_i R_{t-i} + \sum_{j=1}^{q} \beta_j y_{t-j} + \sum_{k=1}^{n} \gamma_k X_{k,t-L}, \tag{3}
$$

where parameter α_i and β_j are the ith-order autoregressive (AR) and jth-order moving average (MA) terms respectively, parameter measures the impact of additional regressor $X_{k,t-L}$ on the index return, p and q are the orders of ARMAX model, and n is the number of regressors $X_{k,t-L}$, L is lag obtained from Eq. (1). In our research, emotional marker frequencies play the role of additional regressors $X_{k,t-L}$.

Conditional variance σ_t^2 is modelled as GARCH(r,m), as in Eq. (4).

$$\sigma_t^2 = c_0 + \sum_{i=1}^{r} \kappa_i y_{t-i}^2 + \sum_{j=1}^{m} \mu_j \sigma_{t-j}^2, \tag{4}$$

where parameters κ_i and μ_j account for the ARCH and GARCH effects of ith- and jth-orders respectively, and r, m are the orders of GARCH model. It is also possible to add Twitter mood $X_{k,t-L}$ to the GARCH equation in order to measure the influence of Twitter mood on volatility.

Traditional specifications of ARMAX-GARCH imply normal or Student's t-distributions of the error term η_t. One obvious disadvantage of these distributions is that they cannot capture asymmetry in the distribution of returns. In order to eliminate this drawback, we implemented skewed normal and skewed Student's t-distributions for the error term. We also estimate ARMAX-GARCH with normal errors as a benchmark.

We choose the parameters p, q, r and m by means of Bayesian information criteria (BIC), with the best specification corresponding to the lowest BIC. Estimation is carried out by means of *rugarch* package by Ghalanos (2014). We employ the Vuong test (Vuong, 1989) in order to compare models. This test can be used for non-nested models, in contrast to the traditional likelihood ratio test. The null hypothesis implies equal goodness of fit for the comparing models. Since observations in financial time series are not typically independent, we use heteroskedasticity and an autocorrelation-consistent version of the Vuong test developed in Calvet and Fisher (2004).

We use mean squared error (MSE) and directional accuracy (DAC) as measures of out-of-sample performance. The latter shows the percentage of matches between returns and their forecast.

2.4 Testing process

Firstly, we evaluate the causality relations between emotional markers and returns by the Granger test, as explained in Section 2.2.

Secondly we define three groups of assets: emotion-sensitive stocks, emotion-insensitive stocks, and indexes. A distinction is made between the first two groups on the basis of the idea that the stocks of companies which are permanently visible for large number of people are affected by people's moods more than others. We suggest that for such companies, emotional markers will demonstrate a substantial and significant effect on stock returns. On the contrary, emotions should have a smaller or even insignificant impact on returns for emotion-insensitive stocks. Indexes are included in our analysis to determine the influence of emotional markers on the financial market generally. The groups we selected include Apple, Facebook and Google (emotion-sensitive); JP Morgan Chase, Pfizer and Exxon Mobil (emotion-insensitive); and the Standard and Poor's 500 and Dow Jones Industrial Average (indexes)[7].

Thirdly, the whole data set was divided into two subsamples for in-sample and out-of-sample estimation. We use a 100-day period for out-of-sample testing, which amounts to approximately 20% of the sample.

For each group, two ARMAX-GARCH models are estimated: a sentiment model, which contains an emotional marker in Eq. (3), and a baseline model without an additional regressor in this equation. Parameters p, q, r and m vary from zero to three (except r, which cannot be smaller than one) and are estimated using BIC. Each asset has as many baseline models as it has emotional markers, since each emotional marker has its own optimal L parameter in Eq. (1), which forms a unique subsample.

Lastly, we compare the explanatory power of sentiment and baseline models by calculating three information criteria: Akaike, Bayesian and Hannan-Quinn. If the criterion value of the sentiment model is less than the value of the baseline model, and if parameter γ from Eq. (3) is significant, then the emotional marker included in this sentiment model is considered to have added information to the return model. The significance of the additional information is verified by the Vuong test.

We also compare the predictive power of estimated models using the DAC and MSE.

3 Empirical results

3.1 Data description

The data set includes eight assets: the S&P500 and DJI indexes, and Apple, Facebook, Google, JP Morgan Chase, Pfizer and Exxon Mobil stocks, information on all of which is obtained from "Yahoo! Finance" (2015). The period under consideration spanned 521 trading days and lasted from February 13, 2013 to April 22, 2015. The descriptive statistics of the assets' logarithmic returns is presented in Appendix B.

Via Twitter API, we downloaded 2,349,036,300 tweets from the period under consideration, an average of 3,098,992 tweets per day. The only restriction placed on posts that were downloaded is that they were published by people located in US. All tweets were sorted by date and analyzed automatically by a JAVA application of our creation. For each day we calculated the frequencies of posts with each item from the emotional markers list, as described in Section 2.1, and normalized them by the number of tweets downloaded on each day.

Most of the frequencies exhibited non-stationary behavior which is apparent on the line

[7]Below we will sometimes use tickers instead of assets' names. The tickers and corresponding names are in Appendix A.

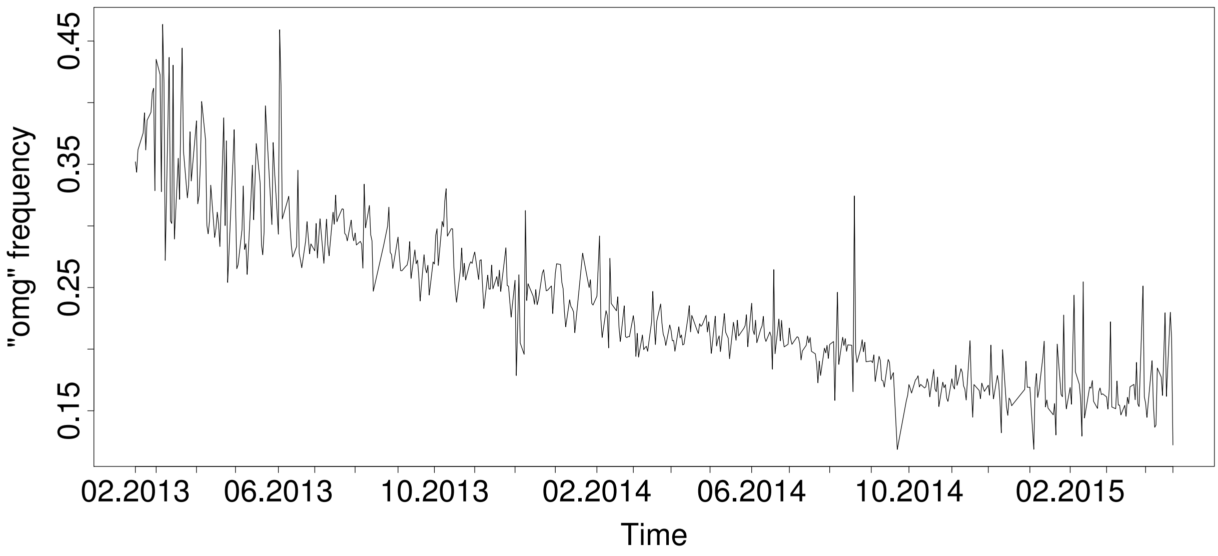


Figure 1: Downward trend in “omg” emotional marker.

plots. For example, “omg”[8] displays a downward trend, Fig. 1.

On the other hand, some frequencies (approximately 10% out of the total) are difference stationary, i. e. have a unit root, confirmed by the augmented Dickey-Fuller test (Dickey & Fuller, 1979). For example, the emoticons “:))” and “:((” demonstrate the absence of stationarity, Fig. 2.

If non-stationary emotional markers are included in the ARMAX-GARCH model, then conventional statistical measures, such as t-statistics or R-squared, are inapplicable (Z. Zhou & Shao, 2013). The emotional markers listed in Appendix C are initially stationary at the 5% level according to the ADF test. The other are brought to stationarity by means of either de-trending (for trend-stationary series) or by taking the first difference (for difference stationary series). The repeated ADF test rejects non-stationarity in all cases.

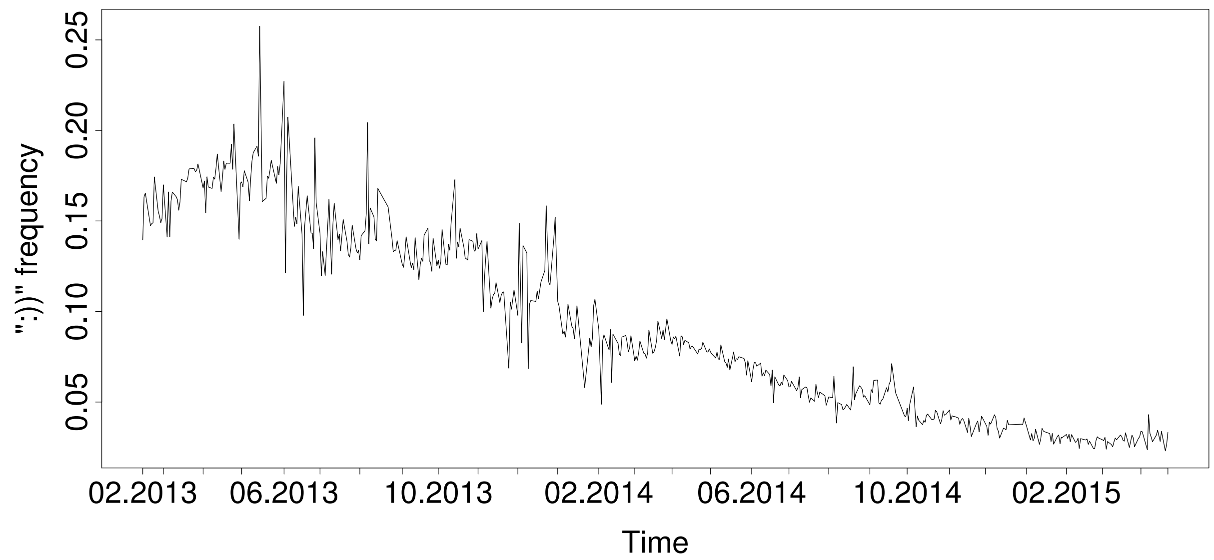
There are also some emotional markers which appear very seldom in tweets, for example the emoticons “:-c”, “>:o” and “:-t”, and the abbreviations “icnbi”[9] and “glhf”[10]. We have excluded these rare markers from our emotional markers list.

3.2 Granger-valid emotional markers

The estimation of Eq. (1) results in 17 emotional markers, listed in Appendix D for each asset. Some emotional markers appear more than once and the number of each Granger-valid emotional marker is presented on Fig. 3.

The fact that more than 90% of emotional markers are not Granger-valid indicates that many emotional markers are driven by returns according to the Granger test. In other words, changes in returns precede an increase or decrease in the frequency of emotional markers.

[8] “Oh my god”
[9] “I cannot believe it”
[10] “Good luck have fun”



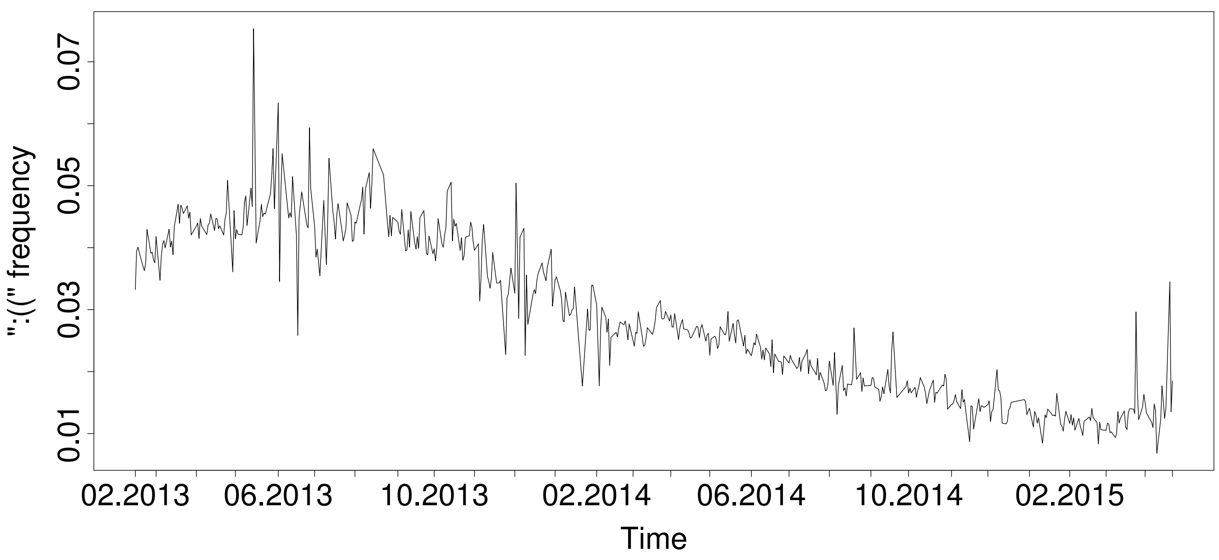


Figure 2: Difference stationary series of ":))" and ":((" emotional markers.

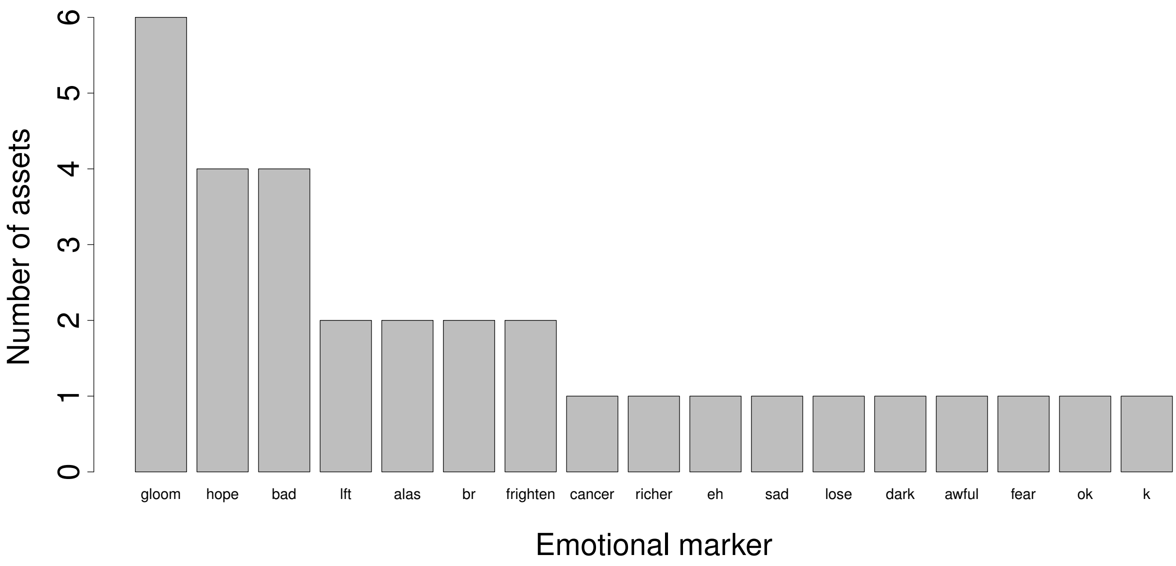


Figure 3: Number of assets influenced by Granger-valid emotional markers

Moreover, different words or emotional markers have different senses in various contexts. For example, the acronym "omg" could be an expression of fear, surprise or happiness. If we calculate the frequency of "omg", it would represent an average of the occurrence of all three emotions. On the other hand, "gloom" expresses a more narrow range of emotions. In subsequent research we plan to include context and disambiguation algorithms in sentiment analysis.

The marker *gloom* is a Granger-valid cause for changes in returns for almost each asset, with the exceptions of JP Morgan and Pfizer. *Hope* and *bad* are Granger-valid for half of the assets, but not for emotion-sensitive assets. The other markers influence one or two assets and can be considered specific to those assets. For example, *cancer* is a Granger-valid cause for change in Pfizer returns, and *richer* for Exxon Mobil returns.

3.3 In-sample

The optimal specifications of the ARMAX(p,q)-GARCH(r,m) selected via BIC are presented in Appendix E.

The group of indexes consisting of the S&P500 and DJI has three common emotional markers: *hope*, *bad* and *gloom*. The dynamics of the DJI is also affected by *alas*. However, *alas* provides additional information according to AIC and HQIC, but not according to BIC. The emotional marker *bad* adds information to models with normal and skewed Student's t-distributions. Specifications which provide better fits, according to BIC, in most cases have a skewed normal distribution for the error term. Indexes show no GARCH-effects for these criteria, because the m parameter is equal to zero for optimized models.

The coefficients in optimal models are significant at the 5% level. While the emotional marker *gloom* has a positive effect on the dynamics of both indexes, its effect on the DJI is almost ten times greater than its effect on the S&P500 (see Table 1). The marker *alas* has a substantial negative impact on DJI log returns.

Table 1: Summary for sentiment models which significantly outperformed baseline models

	Asset	DJI	SNP	AAPL	JPM
Baseline model parameters	Distribution	snorm	sstd	sstd	sstd
	p,q,r,m	0,0,3,0	2,1,1,1	0,0,3,0	0,0,1,0
Sentiment model parameters	Distribution	snorm	sstd	sstd	sstd
	p,q,r,m	2,3,1,1	2,2,1,1	3,2,1,0	3,2,1,0
	Emotional marker	*gloom*	*gloom*	*ok*	*hope*
	Coefficient	0.634*	0.065*	-0.016	0.053*
	Lag	8	8	29	8
BIC	Baseline	-7.309	-7.192	-5.750	-6.030
	Sentiment	-7.314	-7.203	-5.764	-6.031
AIC	Baseline	-7.397	-7.280	-5.811	-6.079
	Sentiment	-7.421	-7.310	-5.875	-6.138
Vuong test		1.042*	0.407*	0.885*	0.053*

* means significant on 1% level.

p,q,r,m are corresponding parameters in Eq. (3) and Eq. (4).

The next group of assets, which we call emotion-sensitive stocks, includes Apple, Facebook and Google. There are more Granger-valid emotional markers for this group than for the previous group. These markers include *awful*, *fear*, and *frighten*, as well as *bad* and *gloom*.

For emotion-sensitive stocks, sentiment models with skewed normal distribution again perform better than baseline models. A normal distribution is also present among optimal specifications. A skewed Student's t-distribution is included among optimal specifications only for the *ok* marker for Apple stock returns. The Vuong test supports the alternative hypothesis that sentiment models have a better fit than do baseline models (see Table 1).

Emotional markers associated with fear, i.e. *fear* and *frighten*, exhibit a strong negative impact on returns. The same behavior is demonstrated by *awful*. Interestingly, *ok* also has negative effect, but its magnitude is much smaller than for the *fear* marker. It worth mentioning that, because of the transformation of all text in our analysis into lowercase letters, we could not distinguish between the expressions “Ok”, “OK”, and “ok”, and did not try to compensate for the ambiguity created by the acronym “ok” (“only kidding” or “Okay” by context). We plan to incorporate this in our subsequent research.

The *looking forward to* marker has a substantial positive impact on Apple and Google stock returns. The marker *bad* that is likewise among Apple and Google's Granger-valid markers has a minor positive effect on returns in specifications with normal and skewed normal errors. The marker *gloom*, which is Granger-valid for all stocks in the group under consideration, is insignificant at the 5% level for optimal specifications.

For the last group of emotion-insensitive stocks, normal and skewed normal distributions similarly demonstrate a better fit, verified by the Vuong test. The markers *hope* and *frighten*, already mentioned, have significant positive and negative impact correspondingly. For Exxon Mobil *hope* turns out to be insignificant at the 5% level.

We found that emotion-insensitive stocks have specific emotional markers, discovered by the Granger test in Eq. (1). These markers are *cancer* for Pfizer, *br* for JP Morgan and *richer* for Exxon Mobil. Although the *cancer* marker is statistically insignificant, it helps to improve predictive power compared to the baseline model. JP Morgan's specific marker results in a slight decrease in returns. The marker *richer*, on the other hand, is one of the strongest predictor for growth in Exxon Mobil's returns. The other important emotional markers for Exxon Mobil are *gloom* and *dark*, which affect the stock positively, and *sad*, which affects it negatively.

Our hypothesis is that the emotion-sensitive group of stocks is more affected by emotional markers than the emotion-insensitive group. The results demonstrate that stocks in both groups are influenced by emotional markers. Although the tweets we analyzed are not restricted in their focus to the economy, business climate, world affairs, or specific businesses, and, for example, include tweets by teenagers talking about themes or events from everyday life, we found that the sentiment measurements we suggested do add information to the ARMAX-GARCH model. One of the possible ways to continue research in this area would be to change the way downloaded messages are filtered to measure the sentiments of a more restrictive group, based on the context of their posts (i.e., whether they are business- or economics-related).

Our analysis shows that we can divide emotional markers into two categories: direct and indirect. Direct emotional markers consist of expressions synonymous with their eponymous emotions, for example the emotion "fear", signaled by the words "fear", "scared", "frighten" etc. Indirect emotional markers express emotions by indicating a context, for example "cancer", "hang over" etc. We found that negative direct emotional markers have negative influence on returns (*awful*, *frighten*, *fear*, *sad*, *alas*) and positive direct markers have positive influence (*hope*). Indirect emotional markers demonstrated opposite behavior: negative markers have a positive influence (*bad*, *gloom*, *cancer*, *lose*, *dark*), while positive markers (*br*, *ok*) have a negative influence. These results are in line with the findings from Hirshleifer and Shumway (2003) and Siganos et al. (2014), who demonstrated the direct influence of emotions on returns.

3.4 Out-of-sample

100 observations were retained to evaluate the out-of-sample performance of emotional markers. We calculated MSE and DAC (refer to Section 2.3 for details) as measures of emotional

markers' predictive power. The results are in Appendix F.

The optimal models for the index group demonstrate less successful predictive performance when compared to baseline models, with the exception of the marker *hope*, which provides a smaller MSE for both indexes. In addition, the markers *hope* and *bad* with normal and skewed Student's t-distribution errors increased directional accuracy of DJI and S&P500 returns to 58% and 54%, correspondingly.

Directional accuracy for the emotion-sensitive group of assets is higher than for the index group, even in BIC-selected models. The DAC obtained for optimal models starts from 50% and peaks at 63% for the marker *frighten*.

It should be noted that "k"[11], which is insignificant at the 5% level, yields outstanding out-of-sample results (57-58%) with MSE and DAC. This confirms our suggestion that emotional markers which provide poor in-sample performance can be successfully used in prediction models.

The directional accuracy of the models for the last group of emotion-insensitive stocks varies from 47% to 56%. The same distributions, namely normal and skewed normal, provide enhanced prediction compared to baseline models. Emotional markers which contribute most to out-of-sample performance are *br* for JP Morgan, *cancer* for Pfizer and *richer*, *hope* and *dark* for Exxon Mobil.

Models which exhibit poorer performance in-sample demonstrate promising out-of-sample results. We consider this as motivation to find optimal specifications by some predictive criterion, such as MSE or DAC, to obtain models with increased predictive power.

The inclusion of emotional markers in the volatility equation Eq. (4) is insignificant at any reasonable level of significance. We also controlled the mean equation for date effects and found no evidence of their presence or impact on predictive ability.

3.5 Discussion

According to neoclassical economic theory, the market is inhabited by rational agents, whose behavior and decisions are not influenced by emotions and emotional states. In other words, the price of assets should not react to changes in social moods, but numerous studies challenge this assumption (see Falato (2009), Hirshleifer and Shumway (2003), Lepori (2015a, 2015b)).

According to De Long, Shleifer, Summers, and Waldmann (1990), rational economic agents operate side-by-side with noise traders, who display irrational behavior in the market. Noise traders tend to be affected by emotional impulses, and hold beliefs which are stochastic and unpredictable. One of the strongest characteristics of noise traders is overreaction (Daniel, Hirshleifer, & Subrahmanyam, 1998). Overreaction is related to the well-known cognitive bias of overconfidence, which implies that people tend constantly to overestimate the subjective accuracy of their own judgments compared to objective accuracy (Pallier et al., 2002). As Daniel et al. (1998) show, overreaction results in positive autocorrelation in

[11] "k" is short for "ok". It is worth mentioning that we calculate frequencies of standalone "k" (accompanied by spaces, commas or other symbols), not the letter "k" associated with any word.

returns, followed by a correction.

We suggest that the phenomenon of overreaction is closely related to the results we obtained for indirect emotional markers. When social moods become negative (which in our research corresponds to the increased frequency of the emotional markers *gloom* or *bad* in tweets), noise traders are led by emotion to react accordingly and drive prices down. These processes take place simultaneously, because noise traders and Twitters users who post their moods are likely to be the same people. After some time their confidence about future price levels declines and the price returns to its fundamental value. This price movement is usually called the mean-reverting component of returns. The reaction to social mood changes can be divided into two phases — a mood-following phase and compensation (De Long et al., 1990).

We found emotional markers which significantly correlate with both immediate changes in returns and with changes that lag behind L periods (see Eq. (1)). The valuable feature of indirect emotional markers is that they reveal a statistically significant connection with the returns precisely during the second phase of mean-reverting price movement. Direct emotional markers, on the other hand, correlate with the returns in the first phase. In addition, the L parameter can be interpreted as a half of the mean-reverting period for a given asset.

4 Conclusion

Twitter is microblogging network which allows millions of people to express their sentiments and feelings. Our research starts with a question: can Twitter data bring additional information to the financial time series model? To find the answer we collected and pre-processed Twitter posts, developed a list of emotional markers, and examined the relation of their frequencies in Twitter posts to logarithmic returns of two indexes and six stocks. We formed three groups of assets, namely, indexes, emotion-sensitive stocks, and emotion-insensitive stocks.

To make the textual analysis stage both transparent and simple, we used a parsimonious word count technique to create emotional markers, which subsequently were used as the determinants of the logarithmic return dynamics in the ARMAX-GARCH model with different error term distributions. To avoid the reverse causality problem, we retained only those emotional markers which are Granger-valid causes of logarithmic returns and, at the same time, are not caused by the logarithmic returns.

In our analysis of the explanatory power our models, we show that parsimonious emotional markers demonstrate smaller BIC and provide significant positive improvements to the likelihood function, verified by the Vuong test. In order to capture higher-order effects of returns, such as skewness associated with the third moment of the distribution of returns, we applied skewed versions of the normal and Student's t-distributions for errors. Our study revealed that a skewed normal distribution deals adequately with the problem of modeling the asymmetry of returns. In the cases of some stocks, including Facebook, Google and JP Morgan, the third moment effects turn out to be insignificant, so a normal distribution also works well for these stocks.

The analysis of indexes and of sensitive and insensitive groups shows that emotional markers could add information according to BIC. We called "general" any emotional markers that influenced a large number of analyzed assets (e.g. *hope*, *bad* and *gloom*). General emotional markers do not, however, increase the predictive ability of the ARMAX-GARCH model we tested.

Our analysis of the emotion-insensitive group revealed the existence of specific emotional markers for members of this group. Although these markers had negligible explanatory power, they increased the predictive ability of the ARMAX-GARCH model substantially.

By dividing emotional markers into the two categories of direct (expressions of emotions) and indirect (facts and context) markers, we were able to better understand the influence of these markers. We observed that direct emotional markers have a direct impact on the dynamics of both stocks and indexes, in line with the results obtained by Hirshleifer and Shumway (2003) and Siganos et al. (2014). Indirect emotional markers, such as *cancer* and *hang over*, demonstrate opposite behavior, and can be a useful tool for estimating the determinants of mean-reverting behavior of returns produced by the irrational decisions of noise traders (De Long et al., 1990).

In summary, our study provides new evidence that emotions expressed in Twitter contain valuable information and can enhance explanations of the dynamics of indexes and stocks. We will continue research on the predictive power of Twitter moods. We expect that the relationship between emotional markers and returns will be found to change over time, depending on certain fundamental factors. In a period of financial stability, for example, emotions may play a smaller role than during a downturn, and market response to similar financial news may differ depending on the mood prevailing in society at different times. In subsequent research we plan to move in two directions. Firstly, we will distinguish periods when the stock market is emotionally-driven and when it is news-driven. And secondly, we will monitor Twitter posts to see if there will be significant changes in emotional marker frequencies, which could be a sign of manipulation.

Acknowledgements

We gratefully acknowledge the helpful comments received from our colleagues at the International Laboratory of Intangible-driven Economy[12] and from anonymous reviewers.

[12]National Research University Higher School of Economics, Perm, Russia, 614070, 38 Studencheskaya Ulitsa. E-mail: info@hse.perm.ru

References

Boia, M., Faltings, B., Musat, C.-C., & Pu, P. (2013). A:) is worth a thousand words: how people attach sentiment to emoticons and words in tweets. In *Social computing (socialcom), 2013 international conference on* (pp. 345–350). IEEE.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. doi:10.1016/j.jocs.2010.12.007.

Brown, G. W. & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, *11*(1), 1–27. doi:10.1016/j.jempfin.2002.12.001.

Calvet, L. E. & Fisher, A. J. (2004). How to forecast long-run volatility: regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics*, *2*(1), 49–83. doi:10.1093/jjfinec/nbh003.

Chen, R. & Lazer, M. (2013). *Sentiment analysis of twitter feeds for the prediction of stock market movement*. Stanford.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *the Journal of Finance*, *53*(6), 1839–1885.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, 703–738.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427–431. doi:10.1080/01621459.1979.10482531.

Esuli, A. & Sebastiani, F. (2006). Sentiwordnet: a publicly available lexical resource for opinion mining. In *Proceedings of lrec* (Vol. 6, pp. 417–422). Citeseer.

Falato, A. (2009). Happiness maintenance and asset prices. *Journal of Economic Dynamics and Control*, *33*(6), 1247–1262.

Francq, C., Zakoian, J.-M. et al. (2004). Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, *10*(4), 605–637. http://www.jstor.org/stable/3318818.

Friedl, J. E. (2002). *Mastering regular expressions*. "O'Reilly Media, Inc."

Ghalanos, A. (2014). *rugarch: univariate garch models*. R package version 1.3-5. https://cran.r-project.org/web/packages/rugarch/rugarch.pdf.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424–438.

Hirshleifer, D. & Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *The Journal of Finance*, *58*(3), 1009–1032. doi:10.2469/dig.v33.n4.1369.

Horv, L. & Kokoszka, P. (2003). Garch processes: structure and estimation. *Bernoulli*, *9*(2), 201–227. doi:10.3150%2fbj%2f1068128975.

Isen, A. M. & Means, B. (1983). The influence of positive affect on decision-making strategy. *Social cognition*, *2*(1), 18–31. doi:10.1521/soco.1983.2.1.18.

Isen, A. M. & Patrick, R. (1983). The effect of positive feelings on risk taking: When the chips are down. *Organizational Behavior and Human Performance*, *31*(2), 194–202. doi:10.1016/0030-5073(83)90120-4.

Javed, F. & Mantalos, P. (2013). Garch-type models and performance of information criteria. *Communications in Statistics-Simulation and Computation*, *42*(8), 1917–1933. doi:10.1080/03610918.2012.683924.

Johnson, E. J. & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of personality and social psychology*, *45*(1), 20. doi:10.1037/0022-3514.45.1.20.

Kearney, C. & Liu, S. (2014, May). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, *33*, 171–185. doi:10.1016/j.irfa.2014.02.006.

Kim, S.-H. & Kim, D. (2014). Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization*, *107*, 708–729. doi:10.1016/j.jebo.2014.04.015.

Lepori, G. M. (2015a). Investor mood and demand for stocks: evidence from popular tv series finales. *Journal of Economic Psychology*, *48*, 33–47.

Lepori, G. M. (2015b). Positive mood and investment decisions: evidence from comedy movie attendance in the us. *Research in International Business and Finance*, *34*, 142–163.

Loewenstein, G. (2000). Emotions in Economic Theory and Economic Behavior. *The American Economic Review*, *90*(2), 426–432. ArticleType: research-article / Issue Title: Papers and Proceedings of the One Hundred Twelfth Annual Meeting of the American Economic Association / Full publication date: May, 2000 / Copyright 2000 American Economic Association. doi:10.2307/117263.

Mayer, J. D. & Gaschke, Y. N. (1988). The experience and meta-experience of mood. *Journal of personality and social psychology*, *55*(1), 102. doi:10.1037/0022-3514.55.1.102.

Mayer, J. D., Gaschke, Y. N., Braverman, D. L., & Evans, T. W. (1992). Mood-congruent judgment is a general effect. *Journal of Personality and Social Psychology*, *63*(1), 119. doi:10.1037/0022-3514.63.1.119.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, *38*(11), 39–41. doi:10.1145/219717.219748.

Nofer, D. M. & Hinz, P. D. O. (2015, June). Using Twitter to Predict the Stock Market. *Business & Information Systems Engineering*, *57*(4), 229–242. doi:10.1007/s12599-015-0390-4.

Nofsinger, J. R. (2005, September). Social Mood and Financial Economics. *Journal of Behavioral Finance*, *6*(3), 144–160. doi:10.1207/s15427579jpfm0603_4.

Yahoo! Finance. (2015). Retrieved January 22, 2014, from http://finance.yahoo.com

The Online Slang Dictionary — Urban Thesaurus. (2016, August). Dictionary. http://onlineslangdictionary.com/thesaurus/.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *The Journal of General Psychology*, *129*(3), 257–299.

Porshnev, A., Redkin, I., & Shevchenko, A. (2013, December). *Improving Prediction of Stock Market Indices by Analyzing the Psychological States of Twitter Users* (SSRN Scholarly Paper No. ID 2368151). Social Science Research Network. Rochester, NY. doi:10.2139/ssrn.2368151.

Prezioso, J. (2013, October). Yom Kippur War Tweet Prompts Higher Oil Prices. Retrieved January 22, 2014, from http://www.huffingtonpost.com/2013/10/10/yom-kippur-war-tweet-oil-prices-traders_n_4079634.html

Rao, T. & Srivastava, S. (2012). *Using twitter sentiments and search volumes index to predict oil, gold, forex and markets indices*. https://repository.iiitd.edu.in/jspui/handle/123456789/31.

Rotmistrov, Y. A. (2016). Word associations network. Retrieved from http://wordassociations.net/

Saunders, E. M., Jr. (1993, December). Stock Prices and Wall Street Weather. *The American Economic Review*, *83*(5), 1337–1345. http://www.jstor.org/stable/2117565.

Schnoebelen, T. (2012). *Do you smile with your nose? Stylistic variation in Twitter emoticons*. University of Pennsylvania Working Papers in Linguistics. http://repository.upenn.edu/pwpl/vol18/iss2/14/.

Selyukh, A. (2013). Hackers send fake market-moving AP tweet on White House explosions — Reuters. Retrieved September 17, 2013, from http://www.reuters.com/article/2013/04/23/net-us-usa-whitehouse-ap-idUSBRE93M12Y20130423

Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2014, November). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*. Empirical Behavioral Finance, *107, Part B*, 730–743. doi:10.1016/j.jebo.2014.06.004.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307–333. http://www.jstor.org/stable/1912557.

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear". *The 2nd Collaborative Innovation Networks Conference - COINs2010*, *26*, 55–62. doi:10.1016/j.sbspro.2011.10.562.

Zhou, R. T. & Lai, R. N. (2009). Herding and information based trading. *Journal of Empirical Finance*, *16*(3), 388–393.

Zhou, Z. & Shao, X. (2013). Inference for linear models with dependent errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(2), 323–343. doi:10.1111/j.1467-9868.2012.01044.x.

Appendix A. Tickers

DJI	Dow Jones Industrial Average
SNP	Standard and Poor's 500
AAPL	Apple
FB	Facebook
GOOG	Google
JPM	JP Morgan Chase
PFE	Pfizer
XOM	Exxon Mobil

Appendix B. Descriptive statistics of log returns

	AAPL	DJI	FB	GOOG	SNP	JPM	PFE	XOM
1	-0.833	-0.210	-0.719	-0.470	-0.232	-0.433	-0.046	-0.043
2	-0.072	-0.030	-0.109	-0.072	-0.034	-0.006	-0.005	-0.053
3	0.012	0.005	0.017	0.006	0.001	0.001	0.007	0.001
4	0.011	0.007	0.011	0.000	0.001	0.012	0.003	-0.004
5	0.100	0.046	0.013	0.008	0.050	0.078	0.062	0.064
6	0.788	0.024	0.259	1.292	0.237	0.438	0.411	0.298
7	0.015	0.007	0.240	0.135	0.072	0.122	0.101	0.099
8	-0.107	-2.078	25.951	16.431	-2.978	-1.492	-0.967	-0.296
9	71.080	35.870	294.266	187.103	36.528	37.952	4.689	4.895

1 — Min.·10, 2 — 1st Quantile·10, 3 — Mean·10, 4 — Median·10, 5 — 3rd Quantile·10, 6 — Max.·10, 7 — St.dev.·10, 8 — Skewness, 9 — Kurtosis

Appendix C. Emotional markers with unit root

:'(	:(	:((	:(((	h.o.	ok	:-|	k	:)
0.074	0.507	0.066	0.252	0.531	0.164	0.078	0.087	0.192
:-)	kk	lhh	oh	seriously?	o_o	wth	gtfoh	
0.580	0.440	0.646	0.061	0.614	0.266	0.092	0.064	

Emotional markers for which augmented Dickey-Fuller test null hypothesis is not rejected on 5% level.

Appendix D. Granger-valid emotional markers

DJI	SNP	AAPL	FB	GOOG	JPM	PFE	XOM
hope	hope	awful	fear	lft	hope	frighten	richer
alas	bad	frighten	gloom	bad	br	cancer	hope
bad	gloom	lft		gloom		eh	alas
gloom		bad				lose	gloom
		gloom					sad
		ok					dark
		k					br

Acronyms

lft - looking forward to

br - best regards

ok - only kidding

k - OK

alas - expression of regret, sorrow

Appendix E. Optimal ARMAX-GARCH specifications

Numbers of the form "p,q,r,m" are ARMAX(p,q)-GARCH(r,m) parameters, see Eq. (3) and Eq. (4). Each model is estimated with three different error distributions: normal, skewed normal and skewed Student's, which corresponds to norm, snorm and sstd in the table.

	Sentiment model			Baseline model		
DJI	norm	snorm	sstd	norm	snorm	sstd
alas	2, 2, 3, 0	2, 2, 1, 1	2, 3, 2, 1	0, 0, 3, 0	2, 1, 3, 0	2, 1, 3, 0
BIC	-7.281	-7.307	-7.296	-7.274	-7.309	-7.297
AIC	-7.379	-7.404	-7.423	-7.323	-7.397	-7.394
HQIC	-7.340	-7.366	-7.373	-7.303	-7.362	-7.355
V	0.815	0.124	0.441			
prob.	0.000	0.000	0.000			
coef.	-0.106	-0.222	-0.075			
prob.	0.000	0.000	0.000			
bad	1, 2, 3, 0	1, 2, 3, 0	1, 2, 3, 0	1, 2, 3, 0	1, 2, 3, 0	1, 2, 3, 0
BIC	-7.240	-7.285	-7.272	-7.253	-7.281	-7.264
AIC	-7.331	-7.387	-7.383	-7.334	-7.373	-7.366
HQIC	-7.295	-7.346	-7.339	-7.302	-7.336	-7.325
V	0.021	0.189	0.225			
prob.	0.000	0.000	0.000			
coef.	0.008	0.013	0.012			
prob.	0.736	0.000	0.000			
gloom	2, 3, 3, 1	2, 3, 1, 1	2, 1, 3, 0	0, 0, 3, 0	2, 1, 3, 0	2, 1, 3, 0
BIC	-7.274	-7.314	-7.283	-7.274	-7.309	-7.297
AIC	-7.391	-7.421	-7.390	-7.323	-7.397	-7.394
HQIC	-7.345	-7.379	-7.348	-7.303	-7.362	-7.355
V	1.042	0.347	0.010			
prob.	0.000	0.000	0.000			
coef.	0.468	0.634	0.342			
prob.	0.000	0.000	0.604			

	Sentiment model			Baseline model		
SNP	norm	snorm	sstd	norm	snorm	sstd
hope	1, 2, 3, 0	3, 2, 3, 0	3, 2, 3, 0	1, 0, 3, 0	3, 3, 1, 1	3, 3, 1, 1
BIC	-7.139	-7.191	-7.196	-7.139	-7.171	-7.186
AIC	-7.226	-7.308	-7.321	-7.198	-7.278	-7.302
HQIC	-7.192	-7.262	-7.272	-7.175	-7.236	-7.256
V	0.438	0.353	0.248			
prob.	0.000	0.000	0.000			
coef.	0.027	0.013	0.018			
prob.	0.000	0.000	0.000			
bad	3, 2, 3, 1	1, 2, 2, 0	3, 3, 1, 1	0, 0, 3, 0	3, 3, 1, 1	1, 1, 1, 1
BIC	-7.141	-7.181	-7.173	-7.145	-7.178	-7.182
AIC	-7.257	-7.268	-7.298	-7.193	-7.284	-7.259
HQIC	-7.211	-7.234	-7.249	-7.174	-7.242	-7.229
V	0.994	-0.259	0.644			
prob.	0.000	0.000	0.000			
coef.	0.004	0.008	0.004			
prob.	0.000	0.000	0.014			
gloom	2, 3, 3, 0	2, 3, 3, 0	2, 2, 1, 1	2, 1, 3, 0	2, 1, 3, 0	2, 1, 1, 1
BIC	-7.155	-7.199	-7.203	-7.165	-7.195	-7.192
AIC	-7.263	-7.316	-7.310	-7.243	-7.283	-7.280
HQIC	-7.220	-7.270	-7.268	-7.212	-7.248	-7.245
V	0.348	0.481	0.407			
prob.	0.000	0.000	0.000			
coef.	0.710	0.235	0.065			
prob.	0.000	0.000	0.000			

	Sentiment model			Baseline model		
AAPL	norm	snorm	sstd	norm	snorm	sstd
frighten	3, 3, 3, 2	3, 3, 3, 0	2, 3, 1, 0	0, 0, 3, 0	3, 3, 3, 0	0, 0, 1, 0
BIC	-5.640	-5.654	-5.716	-5.601	-5.591	-5.716
AIC	-5.776	-5.780	-5.822	-5.649	-5.708	-5.765
HQIC	-5.722	-5.730	-5.780	-5.630	-5.662	-5.745
V	1.728	0.789	0.882			
prob.	0.000	0.000	0.000			
coef.	-0.491	-0.253	2.040			
prob.	0.000	0.000	0.000			
bad	2, 3, 2, 0	3, 3, 3, 0	0, 0, 1, 0	0, 0, 3, 0	0, 0, 3, 0	0, 0, 1, 1
BIC	-5.639	-5.602	-5.733	-5.613	-5.599	-5.750
AIC	-5.741	-5.734	-5.794	-5.664	-5.659	-5.811
HQIC	-5.701	-5.682	-5.770	-5.644	-5.635	-5.787
V	1.010	1.092	-0.167			
prob.	0.000	0.000	0.000			
coef.	0.076	-0.025	0.012			
prob.	0.000	0.000	0.621			
gloom	3, 3, 1, 2	3, 2, 3, 0	0, 0, 1, 0	0, 0, 3, 0	0, 0, 3, 0	0, 0, 1, 0
BIC	-5.629	-5.625	-5.729	-5.614	-5.599	-5.740
AIC	-5.747	-5.744	-5.788	-5.663	-5.658	-5.789
HQIC	-5.701	-5.697	-5.764	-5.643	-5.635	-5.770
V	1.197	1.158	0.036			
prob.	0.000	0.000	0.000			
coef.	-0.674	-0.597	1.583			
prob.	0.420	0.187	0.239			
ok	2, 2, 3, 0	2, 2, 3, 0	3, 2, 1, 0	0, 0, 3, 0	0, 0, 3, 0	0, 0, 1, 1
BIC	-5.605	-5.611	-5.764	-5.613	-5.599	-5.750
AIC	-5.706	-5.723	-5.875	-5.664	-5.659	-5.811
HQIC	-5.666	-5.679	-5.831	-5.644	-5.635	-5.787
V	0.670	0.881	0.885			
prob.	0.000	0.000	0.000			
coef.	-0.009	-0.016	-0.016			
prob.	0.632	0.000	0.857			
k	2, 2, 3, 0	2, 2, 3, 0	0, 0, 1, 0	0, 0, 3, 0	0, 0, 3, 0	0, 0, 1, 0
BIC	-5.597	-5.586	-5.698	-5.599	-5.585	-5.711
AIC	-5.693	-5.692	-5.756	-5.647	-5.643	-5.759
HQIC	-5.655	-5.650	-5.733	-5.628	-5.620	-5.740
V	0.712	0.746	0.013			
prob.	0.000	0.000	0.000			
coef.	-0.002	-0.002	-0.001			
prob.	0.303	0.163	0.557			

	Sentiment model			Baseline model		
FB	norm	snorm	sstd	norm	snorm	sstd
fear	2, 2, 2, 1	3, 3, 3, 1	1, 0, 1, 1	0, 0, 1, 0	0, 0, 1, 0	0, 0, 1, 1
BIC	-4.437	-4.461	-4.682	-4.420	-4.472	-4.703
AIC	-4.536	-4.599	-4.762	-4.450	-4.511	-4.763
HQIC	-4.497	-4.544	-4.730	-4.438	-4.496	-4.739
V	1.211	1.379	0.087			
prob.	0.000	0.000	0.000			
coef.	-1.188	-1.275	-0.419			
prob.	0.000	0.000	0.085			

	Sentiment model			Baseline model		
GOOG	norm	snorm	sstd	norm	snorm	sstd
lft	1, 2, 1, 0	1, 2, 1, 0	0, 0, 1, 0	0, 0, 2, 0	0, 0, 3, 0	0, 0, 1, 0
BIC	-5.796	-5.791	-5.912	-5.783	-5.780	-5.927
AIC	-5.865	-5.870	-5.972	-5.823	-5.839	-5.977
HQIC	-5.838	-5.839	-5.948	-5.807	-5.816	-5.957
V	0.573	0.410	0.001			
prob.	0.000	0.000	0.010			
coef.	0.309	-0.066	0.116			
prob.	0.031	0.735	0.796			
bad	3, 2, 2, 0	3, 2, 1, 1	0, 0, 1, 0	1, 2, 1, 1	3, 2, 3, 0	0, 0, 1, 0
BIC	-5.790	-5.819	-5.919	-5.810	-5.801	-5.933
AIC	-5.886	-5.925	-5.977	-5.878	-5.906	-5.981
HQIC	-5.848	-5.883	-5.954	-5.851	-5.865	-5.962
V	0.229	0.186	0.002			
prob.	0.000	0.000	0.002			
coef.	0.002	0.013	0.006			
prob.	0.150	0.000	0.756			

	Sentiment model			Baseline model		
JPM	norm	snorm	sstd	norm	snorm	sstd
hope	0, 0, 1, 0	2, 3, 1, 1	3, 2, 1, 0	0, 0, 1, 0	0, 0, 1, 1	0, 0, 1, 0
BIC	-6.009	-6.000	-6.031	-6.022	-5.996	-6.030
AIC	-6.048	-6.107	-6.138	-6.052	-6.045	-6.079
HQIC	-6.033	-6.065	-6.095	-6.040	-6.026	-6.059
V	0.013	0.930	0.898			
prob.	0.000	0.000	0.000			
coef.	0.022	0.043	0.053			
prob.	0.619	0.000	0.000			
br	0, 0, 1, 0	1, 2, 1, 0	0, 0, 1, 0	0, 1, 1, 0	3, 2, 2, 0	0, 0, 1, 0
BIC	-6.022	-5.999	-6.027	-6.017	-6.004	-6.038
AIC	-6.062	-6.080	-6.087	-6.057	-6.104	-6.088
HQIC	-6.046	-6.048	-6.063	-6.041	-6.064	-6.068
V	0.046	-0.345	0.042			
prob.	0.000	0.000	0.000			
coef.	-0.007	-0.007	-0.005			
prob.	0.106	0.000	0.289			

	Sentiment model			Baseline model		
PFE	norm	snorm	sstd	norm	snorm	sstd
frighten	2, 3, 2, 1	3, 2, 1, 1	3, 2, 1, 1	0, 0, 1, 1	0, 0, 1, 1	0, 0, 1, 1
BIC	-6.331	-6.297	-6.338	-6.323	-6.308	-6.343
AIC	-6.441	-6.408	-6.458	-6.363	-6.358	-6.403
HQIC	-6.397	-6.364	-6.410	-6.347	-6.338	-6.379
V	1.130	0.798	0.850			
prob.	0.000	0.000	0.000			
coef.	-0.701	-0.442	-0.408			
prob.	0.000	0.000	0.000			
eh	3, 2, 2, 1	3, 2, 1, 1	3, 2, 2, 0	0, 0, 1, 1	0, 0, 1, 1	0, 0, 1, 1
BIC	-6.317	-6.315	-6.332	-6.320	-6.305	-6.343
AIC	-6.428	-6.426	-6.454	-6.361	-6.356	-6.404
HQIC	-6.384	-6.382	-6.405	-6.345	-6.336	-6.380
V	1.021	0.997	0.797			
prob.	0.000	0.000	0.000			
coef.	-0.011	0.010	-0.007			
prob.	0.000	0.000	0.000			

	Sentiment model			Baseline model		
XOM	norm	snorm	sstd	norm	snorm	sstd
richer	0, 0, 1, 0	1, 2, 1, 0	0, 0, 1, 0	0, 0, 1, 0	0, 0, 2, 0	0, 0, 1, 0
BIC	-6.589	-6.585	-6.606	-6.592	-6.576	-6.608
AIC	-6.627	-6.662	-6.663	-6.620	-6.625	-6.656
HQIC	-6.612	-6.631	-6.641	-6.609	-6.606	-6.637
V	0.117	0.529	0.120			
prob.	0.000	0.000	0.000			
coef.	2.658	2.225	2.418			
prob.	0.005	0.000	0.006			
hope	2, 2, 2, 0	2, 2, 2, 1	2, 2, 1, 1	0, 0, 2, 0	3, 2, 1, 1	2, 2, 1, 1
BIC	-6.589	-6.569	-6.595	-6.586	-6.600	-6.608
AIC	-6.676	-6.675	-6.701	-6.625	-6.697	-6.705
HQIC	-6.642	-6.633	-6.659	-6.609	-6.659	-6.666
V	0.772	-0.173	0.015			
prob.	0.000	0.000	0.000			
coef.	0.021	0.019	0.018			
prob.	0.566	0.518	0.543			
gloom	2, 3, 1, 1	3, 2, 2, 1	2, 2, 1, 0	0, 0, 1, 0	2, 3, 2, 2	2, 3, 1, 1
BIC	-6.593	-6.589	-6.616	-6.574	-6.566	-6.615
AIC	-6.693	-6.709	-6.716	-6.604	-6.686	-6.725
HQIC	-6.653	-6.662	-6.676	-6.592	-6.638	-6.681
V	1.238	0.232	-0.141			
prob.	0.000	0.000	0.000			
coef.	0.926	-0.330	1.003			
prob.	0.000	0.000	0.000			
dark	3, 2, 2, 0	3, 2, 1, 1	3, 2, 2, 0	0, 0, 1, 0	0, 0, 2, 0	0, 0, 1, 0
BIC	-6.638	-6.585	-6.613	-6.592	-6.576	-6.608
AIC	-6.734	-6.691	-6.729	-6.620	-6.625	-6.656
HQIC	-6.696	-6.649	-6.683	-6.609	-6.606	-6.637
V	1.510	0.973	1.084			
prob.	0.000	0.000	0.000			
coef.	0.132	0.171	0.198			
prob.	0.000	0.000	0.000			

Appendix F. Out-of-sample

MSE is multiplied by 1000.

DJI		Sentiment model norm	snorm	sstd	Baseline model norm	snorm	sstd
alas	MSE	0.219	0.160	0.140	0.078	0.082	0.082
	DAC	0.470	0.480	0.490	0.490	0.490	0.490
bad	MSE	0.224	0.179	0.219	0.168	0.180	0.203
	DAC	0.490	0.480	0.480	0.490	0.480	0.490
gloom	MSE	0.145	0.276	0.083	0.078	0.082	0.082
	DAC	0.490	0.470	0.490	0.490	0.490	0.490

SNP		Sentiment model norm	snorm	sstd	Baseline model norm	snorm	sstd
hope	MSE	0.187	0.126	0.237	0.073	0.167	0.165
	DAC	0.460	0.460	0.470	0.490	0.460	0.470
bad	MSE	0.168	0.253	0.071	0.072	0.072	0.072
	DAC	0.500	0.460	0.540	0.460	0.530	0.470
gloom	MSE	0.197	0.203	0.196	0.085	0.086	0.075
	DAC	0.460	0.460	0.450	0.450	0.450	0.460

AAPL		Sentiment model norm	snorm	sstd	Baseline model norm	snorm	sstd
frighten	MSE	0.339	0.321	0.271	0.258	0.302	0.258
	DAC	0.600	0.630	0.500	0.510	0.470	0.510
bad	MSE	0.546	0.711	0.258	0.258	0.258	0.258
	DAC	0.500	0.400	0.510	0.510	0.510	0.510
gloom	MSE	0.280	0.280	0.261	0.259	0.259	0.258
	DAC	0.550	0.580	0.510	0.510	0.510	0.510
ok	MSE	0.253	0.316	0.258	0.258	0.258	0.258
	DAC	0.560	0.390	0.530	0.510	0.510	0.510
k	MSE	0.255	0.256	0.259	0.258	0.258	0.258
	DAC	0.580	0.570	0.510	0.510	0.510	0.510

FB		Sentiment model norm	snorm	sstd	Baseline model norm	snorm	sstd
fear	MSE	1.270	0.215	0.206	0.199	0.203	0.199
	DAC	0.450	0.540	0.550	0.540	0.540	0.540

GOOG		Sentiment model norm	Sentiment model snorm	Sentiment model sstd	Baseline model norm	Baseline model snorm	Baseline model sstd
lft	MSE	1.331	0.550	0.184	0.183	0.183	0.184
	DAC	0.500	0.500	0.500	0.500	0.500	0.500
bad	MSE	0.334	0.383	0.184	1.586	0.258	0.184
	DAC	0.440	0.500	0.500	0.500	0.440	0.500

JPM		Sentiment model norm	Sentiment model snorm	Sentiment model sstd	Baseline model norm	Baseline model snorm	Baseline model sstd
hope	MSE	0.194	1.297	0.504	0.192	0.192	0.192
	DAC	0.540	0.510	0.510	0.540	0.540	0.540
br	MSE	0.194	0.310	0.193	0.194	0.490	0.192
	DAC	0.540	0.540	0.530	0.500	0.530	0.540

PFE		Sentiment model norm	Sentiment model snorm	Sentiment model sstd	Baseline model norm	Baseline model snorm	Baseline model sstd
frighten	MSE	0.342	0.281	0.140	0.103	0.103	0.103
	DAC	0.530	0.470	0.510	0.480	0.480	0.480
eh	MSE	0.516	0.364	0.401	0.103	0.103	0.103
	DAC	0.530	0.470	0.550	0.480	0.480	0.480

XOM		Sentiment model norm	Sentiment model snorm	Sentiment model sstd	Baseline model norm	Baseline model snorm	Baseline model sstd
richer	MSE	0.189	0.470	0.189	0.189	0.189	0.189
	DAC	0.560	0.410	0.550	0.400	0.400	0.400
hope	MSE	0.243	0.250	0.245	0.189	0.841	0.249
	DAC	0.530	0.530	0.490	0.400	0.430	0.510
gloom	MSE	1.548	2.008	0.844	0.189	0.711	0.779
	DAC	0.500	0.500	0.500	0.400	0.510	0.500
dark	MSE	0.223	0.365	0.258	0.189	0.189	0.189
	DAC	0.540	0.540	0.470	0.400	0.400	0.400

Alexander V. Porshnev

National Research University Higher School of Economics. Department of General and Strategic Management. Associate Professor. Nizhniy Novgorod, Russia. E-mail: aporshnev@hse.ru

Valeria V. Lakshina

National Research University Higher School of Economics. Department of Mathematical Economics. Senior Lecturer. Nizhniy Novgorod, Russia. E-mail: vlakshina@hse.ru

Ilya E. Red'kin

Freelance Data Analyst. Nizhny Novgorod. Russia. E-mail: ilya-redkin@yandex.ru