

# The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume and survey sentiment indices

Nuno Oliveira<sup>a,\*</sup>, Paulo Cortez<sup>a</sup>, Nelson Areal<sup>b</sup>

<sup>a</sup>*ALGORITMI Centre, Department of Information Systems, University of Minho,  
4804-533 Guimarães, Portugal*

<sup>b</sup>*School of Economics and Management, Department of Management, University of Minho, 4710-057 Braga,  
Portugal*

---

## Abstract

In this paper, we propose a robust evaluation of the information content of microblogging data to forecast useful stock market variables: returns, volatility and trading volume of diverse dataset of indices and portfolios. We analyze a large Twitter dataset, from December 2012 to October 2015, with about 31 million messages related with 3,800 stocks traded in US markets. Also, we apply a sound prediction procedure (e.g., rolling window evaluation, four regression methods) along with a statistical test of predictive accuracy. Furthermore, we explore the diversity of traditional sentiment indicators and assess their complementarity value with microblogging sentiment. A Kalman Filter (KF) procedure is applied to create an unique daily sentiment indicator from a Twitter indicator and four other sentiment indicators (created from surveys). We also predicted two popular survey sentiment indicators using microblogging data.

We found that Twitter sentiment and posting volume were particularly important for the forecasting of returns of S&P 500 index, portfolios of lower market capitalization and some industries. Additionally, KF sentiment was informative for the forecasting of returns. Furthermore, Twitter and KF sentiment indicators were useful for the prediction of some AAI and II survey sentiment indicators. These results show that microblogging data are relevant to forecast stock market behavior and can provide a valuable alternative for existing measures (e.g., survey sentiment) with various advantages (e.g., fast and cheap creation, daily frequency).

*Keywords:* Predictive sentiment analysis; Stock market; Twitter; Data and text mining;  
Regression; [Big data analysis](#)

---

\*Corresponding author at: ALGORITMI Centre, Department of Information Systems, University of Minho, 4804-533 Guimarães, Portugal Tel. +351253510319.

*Email addresses:* [nunomroliveira@gmail.com](mailto:nunomroliveira@gmail.com) (Nuno Oliveira), [pcortez@dsi.uminho.pt](mailto:pcortez@dsi.uminho.pt) (Paulo Cortez), [nareal@eeg.uminho.pt](mailto:nareal@eeg.uminho.pt) (Nelson Areal)

---

## 1. Introduction

Nowadays, very large volumes of data are supplied from diverse sources (e.g., sensors, mobile phones, social media, online shopping) at different speeds and periodicities [15, 50, 12, 13, 59]. Some sources deliver data continuously in real time [59] and it is estimated that the volume of business data duplicates every 1.2 years [12]. Doug Laney classifies these high volume (large volume), variety (diversity of data types and sources) and velocity (fast generation) data as Big Data [42]. The analysis of big data permits the extraction of valuable information that may improve decision making and organizational performance [50]. The identification of new customers and markets, the creation of new products, a better customer service and more operational efficiency are some potential advantages obtained by big data analytics [12].

Social media is an important source of big data. For instance, Facebook has 1.65 billion monthly active users<sup>1</sup>, there are 1 billion visits monthly to sites with embedded Tweets<sup>2</sup> and Weibo has 236 million monthly active users<sup>3</sup>. Users spend a significant part of their time on social media services [27]. Thus, they disclose substantial information about diverse aspects of their lives in these sites. The analysis of these social media data may allow a deeper understanding of users' behavior such as the assessment of their sentiment (e.g., [9]), the identification of their interests (e.g., [34]) or the measurement of users influence (e.g., [72]).

Social media mining can be utilized for various purposes. For instance, it can be used to predict sales [27]; to manage brand reputation [33, 55]; to create recommendation systems [33]; to identify and predict crime [13]; to improve advertisement by finding influential users to promote their products and by evaluating the reactions to an ad-campaign [13, 55, 33]; to identify pandemics and to assess mental health [55] or to measure users' sentiment towards political parties [55].

Financial services have been exploring social media data. For instance, the financial application Thomson Reuters Eikon includes sentiment analysis of tweets<sup>4</sup> and the financial information platform Bloomberg provides social media indicators for their subscribers<sup>5</sup>. Moreover, the evaluation

---

<sup>1</sup><https://investor.fb.com/investor-news/press-release-details/2016/Facebook-Reports-First-Quarter-2016-Results-and-Default.aspx>

<sup>2</sup><https://about.twitter.com/company>

<sup>3</sup><http://ir.weibo.com/phoenix.zhtml?c=253076&p=irol-newsArticle&ID=2145407>

<sup>4</sup><http://thomsonreuters.com/en/press-releases/2014/thomson-reuters-adds-unique-twitter-and-news-sentiment-analysis.html>

<sup>5</sup><http://www.bloomberg.com/company/announcements/trending-on-twitter-social-sentiment-analytics/>

of the predictive information of social media data for stock market behavior is an active research topic. Several studies have extracted investor sentiment and attention indicators from social media contents and have analyzed their impact on stock market (e.g., [2, 21, 9, 70, 52]).

Indeed, there is a strand of the finance literature (behavioral finance) that argues that sentiment may affect financial prices [65]. De Long et al. (1990) [47] argue that the unpredictability of irrational traders (noise traders) creates an additional risk on prices that prevents rational investors from vigorously betting against them. Shleifer and Vishny (1997) [66] refers that in some extreme circumstances, specialized professional arbitrageurs may not correct security prices effectively for various reasons, such as capital limitations or agency costs. In addition, some institutional investors are forbidden by their charters from taking short positions [71]. Therefore, security prices may diverge from fundamental values because arbitrage may fail to eliminate mispricing caused by investor sentiment. If it happens, future prices may be predictable and sentiment indicators may have predictive information.

Different proxies for investor sentiment have been applied and most of them are not related to social media. Some papers use indirect measures such as economic and financial variables to infer the emotional state of investors. For instance, the Baker and Wurgler (BW) monthly index [6] was adopted in several studies [7, 41, 82, 16, 38, 71, 17] and it uses six different proxies for sentiment: the closed-end fund discount, New York Stock Exchange (NYSE) share turnover, the number and average first-day returns on Initial Public Offerings (IPOs), the equity share in new issues, and the dividend premium. Other papers applied direct sentiment measures derived from surveys of investors' feelings about the stock market. For instance, the American Association of Individual Investors (AAII) provides weekly values of the votes of their members to a poll questioning their sentiment (bullish, bearish, neutral) on the stock market for the next six months. Research works such as [28, 10, 40, 79, 78] used AAII index as a sentiment measure. Also, the University of Michigan Surveys of Consumers (UMSC) is a monthly sentiment index constructed from a consumer confidence survey answered by a random group of five hundred continental US households. Despite being considered less related to investor sentiment, UMSC index has been applied in a considerable number of stock market studies (e.g., [28, 45, 58, 37, 62, 86, 16, 71]). Moreover, Sentix ([www.sentix.de](http://www.sentix.de)) creates sentiment indices for various stock markets (e.g., US, Japan, Germany, Euro zone) from surveys answered by more than 3,500 participants. Some research works used Sentix indices such as [61, 17, 36]. Another example is the Investors Intelligence (II) weekly index based on over a hundred independent market newsletters, with each newsletter being

categorized as bullish, bearish or correction. II measures may be more correlated to institutional sentiment than AAI, because many of the newsletters' authors are market professionals [10]. The II index is widely applied in behavior finance literature (e.g., [69, 28, 44, 10, 11, 40, 37, 41, 79, 78]).

Recently, various studies used investor sentiment and attention indicators automatically extracted from textual contents. Diverse data sources were used to create these indicators, such as newspapers (e.g., New York Times, Wall Street Journal) [74, 24, 32, 63, 29, 83, 30], message boards (e.g., thelion.com, ragingbull.com) [77, 2, 22, 21, 60, 83, 46], or microblogs (e.g., Twitter, StockTwits) [68, 83, 84, 67, 48, 14, 9, 70, 52, 53, 54]. The indicators extracted from texts (e.g., Twitter) have many advantages when compared with survey sentiment indices (e.g., AAI, UMSC, II). The creation of text based sentiment indicators is faster and cheaper, permits greater periodicities (e.g., daily) and may be targeted to a more restrict set of stocks (e.g., stock market indices or individual stocks). Also, several studies have used these sentiment indicators to make trading decisions [63, 14, 48, 30] or predict useful stock market variables, such as: daily or intraday values of stock prices [74, 9, 84, 85], price directions [32, 52, 67, 46, 68], returns [70, 9, 24, 60, 29], volatility [2, 60, 84] and trading volume [74, 2, 60].

For lower periodicities (e.g., monthly), some studies identified a relationship between sentiment and future aggregate stock market returns (e.g., [7, 28]), volatility (e.g., [44]) and crisis on stock market (e.g., [86]). However, a considerable body of research has focused on the analysis of the influence of sentiment on portfolios formed on diverse characteristics (e.g., firm dimension, age, volatility, book-to-market ratio). Sentiment seems to have more effect on returns of some portfolios having extreme values on these characteristics [43, 51, 44, 6, 7, 8, 16, 71, 17]. This cross-sectional variation may be explained by a different vulnerability to speculation and difficulty to arbitrage in some categories of stocks [6]. Some stocks are more unsafe and difficult to value and they require riskier and costlier arbitrage actions [1, 23, 80]. There is a higher concentration of irrational investors in such stocks because more rational investors prefer safer investments. These studies usually rely on indirect and survey sentiment measures and low frequencies (e.g., monthly, annual).

Posting volume on social media services (e.g., microblogs and message boards) has also been applied in the prediction of stock market behavior. Some studies found that the number of posts on financial message boards has impact on future returns (e.g., [81, 2]), trading volume (e.g., [81]) and volatility (e.g., [2, 21]). Moreover, other works applied the number of microblogging messages as proxy for attention and found some predictive information for trading volume (e.g., [70, 53]). Weekly Google search volumes were also used in [49] to predict stock prices.

Nevertheless, these findings are not consensual. [The efficient market hypothesis, proposes that investors act as rational agents and all available information is reflected immediately in stock prices, which implies that assets are traded at their fair value \[26\].](#) Indeed, there are studies that find scarce evidence of the predictive power of sentiment for stock prices or returns [77, 2, 22, 53]. Timmermann [75] defends that there is in general scant evidence of return predictability. Tumarkin and Whitelaw (2001) [77] refer that posting volume of financial message boards do not add information for the forecasting of trading volume and returns.

Moreover, most forecasting studies that use text extracted sentiment indicators present limitations in terms of lack of a robust evaluation and usage of very simple sentiment analysis methods. For instance, no out of sample evaluation was executed in [43, 51, 44, 2, 6, 7, 74, 70, 32, 60, 84, 8, 71, 17, 29, 67, 68, 85]. Other works applied very short test sets: 8 predictions [52]; 19 forecasts [9]; 23 predictions [63]; 20 and 30 forecasts [49]; and 25 and 35 predictions [14]. Also, as recently shown in [54], generic domain independent lexicons (e.g., General Inquirer, MPQA, SentiWordNet) are ineffective for assessing the sentiment of stock market related messages. For instance, the term “explosive” is often negative in generic contexts but can be positive within the financial domain (“explosive rise”). Yet, several works are based on such lexicons (e.g., General Inquirer was adopted in [74], SentiWordNet was used in [24, 14], MPQA was adopted in [9, 63]). Other works [49, 53] adopted a very simplistic lexicon with just two terms (“bullish” and “bearish”) and that only allows the sentiment extraction of a very small fraction of the available messages. In [52], a manual effort was used to extract the sentiment of messages but such manual labeling is costly and impractical when applied to large datasets.

The main goal of this paper is to evaluate the information content of microblogging data to the forecasting of stock market variables by conducting a large set of experiments. We extend and improve our previous work [53], leading to the following main contributions:

- i) we generate [new](#) Twitter sentiment indicators by adopting a recently proposed lexicon [54] that contains more than 20,000 entries and that is specifically adjusted to financial microblogs. [This is the first paper that applies sentiment indicators created by using specialized financial microblogging lexicons. These indicators were extracted from a recent and very large Twitter based dataset, containing around 31 million messages from December 2012 to October 2015 related with all stocks traded in US markets \(about 3,800 stocks\);](#)
- ii) [we propose a novel, more complete and less noisy sentiment indicator based on a Kalman Filter \(KF\) that combines measures of different periodicities, namely daily Twitter indicators with](#)

weekly AAI and II values and monthly UMSC and Sentix indices. We believe this is the first paper of this topic that uses a KF procedure to aggregate diverse sentiment values. We also explore diverse sentiment aggregation formulas (e.g., bullish ratio, variation, agreement)

- iii) we predict several daily stock market variables (returns, trading volume and volatility) of diverse indices, such as Standard & Poor's 500 (SP500), Russell 2000 (RSL), Dow Jones Industrial Average (DJIA), Nasdaq 100 (NDQ), and portfolios, such as formed on size and industries (we note that such portfolio analysis has been scarcely executed for social media data and for daily frequencies, most works apply monthly indicators derived from surveys, such as AAI or II, or financial variables, such as BW);
- iv) we also predict two popular survey sentiment indicators (AAI and II) using Twitter and KF sentiment indicators, which may permit a satisfactory anticipation of AAI and II values or a decent alternative whenever they are unavailable (to the best of our knowledge, this is the first attempt of such prediction);
- v) we adopt a realistic and robust evaluation that uses a rolling window scheme (with 300 training days for stock market variables and 50 days for survey indicators), the Diebold-Mariano (DM) test for predictive accuracy [25] and four regression models, Multiple Regression (MR), Neural Network (NN), Support Vector Machine (SVM) and Random Forest (RF), under two main strategies: baseline model (without microblog features) and microblog based (with such features).

Moreover, we consider that this study may support further advances in big data analytics applications. For instance, real time financial decision support systems can be enhanced by the utilization of predictive models applying microblogging data or the display of different sentiment indicators according to user needs (e.g., various periodicities, related to specific stocks, from diverse types of social media users).

This paper is structured as follows. Section 2 presents the microblogging and stock market data, survey sentiment indicators, methods for the creation of microblogging sentiment and attention indicators, KF process, the forecasting procedure and respective evaluation. Section 3 shows and analyzes the results of the prediction of returns, volatility, trading volume and survey sentiment values. Finally, conclusions are presented in Section 4.

## 2. Material and Methods

### 2.1. Microblogging Data

Microblogging data have characteristics that may indicate potential informative value to the forecasting of stock market behavior. Services such as Twitter and StockTwits have large communities of investors sharing information about stock market. These users frequently interact during the day and react readily to events. Messages are usually very objective due to the character limit. Microblogging users usually apply cashtags in stock market conversations to refer to the involved stocks. Cashtags are composed by a “\$” character and the respective ticker (e.g., \$AAPL) and its presence means that the message is related with that stock. The utilization of cashtags permits an easy and less noisy selection of messages related to specific stocks. The extraction of attention and sentiment from microblogging is faster and cheaper than from traditional sources (e.g., surveys) because data is promptly available at very low cost.

The sentiment and attention indicators created in this study were extracted from Twitter, which is a large microblog platform with more than 300 million active users<sup>6</sup>. Using Twitter REST API (<https://dev.twitter.com/docs/api>), we collected all messages (around 31 million) from 22<sup>nd</sup> of December 2012 to 29<sup>th</sup> of October 2015 holding cashtags of all stocks traded in US markets (nearly 3,800 stocks). R tool (<http://www.r-project.org>) was used in all processing tasks and MongoDB (<https://www.mongodb.org>) was applied to store Twitter data.

### 2.2. Microblogging Sentiment and Attention Indicators

In this paper, we assessed the information content of investor sentiment and attention indicators extracted from Twitter data to predict stock market behavior. The number of tweets were applied to produce the attention indicators. We opted to use the first difference of the posting volume because there is a visible growing number of tweets during the analyzed time period. To create the investor sentiment indicators, we used the sentiment scores produced by sentiment analysis of all tweets. The sentiment analysis applies a recently proposed lexicon [54] that is properly adapted to microblogging conversations about stock market and publicly available at [https://github.com/nunomroliveira/stock\\_market\\_lexicon](https://github.com/nunomroliveira/stock_market_lexicon). This lexicon was automatically created using data from June, 2010 to March, 2013. It is an up-to-date lexicon because its training data ends nearly one year before the first prediction days of this study. It contains approximately 7000

---

<sup>6</sup><http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>

unigrams, 13000 bigrams and the respective sentiment scores for affirmative and negated contexts. For instance, a negative score indicates a bearish word and a positive sentiment value indicates a bullish word. Negated contexts are text segments starting with a negation expression while the affirmative contexts are all other segments. To identify affirmative and negated segments, we applied the same approach used in the lexicon creation. Negated segments begin with a negation item included in the Christopher Potts’ sentiment tutorial (<http://sentiment.christopherpotts.net/lingstruc.html>) and end with one of the following punctuation marks: ”, ”. ”, ”: ”, ”; ”, ”! ”, ”? ”. The sentiment score of each tweet corresponds to the sum of the sentiment score of all lexicon items present in the message. When lexicon bigrams are identified in the text, we only account the score of the bigrams and do not consider the score of their individual components. **In our opinion, bigrams scores are more precise than unigram scores because bigrams have a more defined context. For instance, the bigram “debt free” is usually bullish while its individual components may have distinct sentiment orientation (e.g., “greek debt”, “more debt”, “free fall”).** To adequately verify the presence of lexicon elements, we executed the preprocessing tasks:

- replace all cashtags by the tag ”tkr”; all numbers by the tag ”NUM”; all mentions by ”@user”; all URL addresses by ”URL”;
- execute tokenization, Part of Speech (POS) tagging and lemmatization by applying Stanford CoreNLP [76].
- identify the affirmative and negated segments in order to apply the adequate score.

The sentiment indicators are created using the scores produced by the sentiment analysis. We created two major types of investor sentiment indicators: general and sectorial. The general indicators represent the sentiment of the whole investor community. Thus, we used all tweets in the construction of these indicators. The sectorial indicators measure the sentiment regarding specific sectors (e.g., industries). In the creation of these indicators, we applied tweets enclosing cashtags of stocks belonging to the respective sector. We selected the cashtags composing each sector based on their Standard Industrial Classification (SIC) code.

We also experimented diverse forms to calculate the daily sentiment values:

- Bullish Ratio (BullR) (e.g., [53, 67, 68])

$$BullR_t = \frac{NBull_t}{NBull_t + NBear_t} \quad (1)$$



- Bearish Ratio (BearR)

$$BearR_t = \frac{NBear_t}{NBull_t + NBear_t} \quad (2)$$

- Bullishness Index (BI) (e.g., [2, 70, 52, 53])

$$BI_t = \ln \frac{NBull_t + 1}{NBear_t + 1} \quad (3)$$

- Variation (VA) (e.g., [53, 67, 68])

$$VA_t = BullR_t - BullR_{t-1} \quad (4)$$

- Agreement (AG) (e.g., [2, 70])

$$AG_t = 1 - \sqrt{1 - \left( \frac{NBull_t - NBear_t}{NBull_t + NBear_t} \right)^2} \quad (5)$$

where  $NBull_t$  and  $NBear_t$  are the bullish and bearish score of day  $t$ . We did not find any paper applying a sentiment measure computed exactly like BearR, however there are papers using similar measures such as [74, 29]. The tested sentiment aggregation measures are distinct formulas applied in diverse studies in this research topic. In this study, we applied BullR, BI and VA indicators in the prediction of returns, volatility, trading volume and survey sentiment values. Since dispersion of expectations is considered to be related with trading volume [64, 2] and volatility ([64]), we added AG indicators in the forecasting of these stock market variables. Additionally, we applied BearR indicators in the prediction of negative values of survey sentiment values, because we used Twitter indicators counterparts in the forecasting of each survey sentiment value (e.g., BullR indicators applied in the prediction of AAI positive values or BearR indicators in the forecasting of AAI negative values).

### 2.3. Survey Sentiment Indicators

Survey sentiment indicators are frequently applied in studies about the analysis of the sentiment impact on stock market behavior. Investors indicate their expectations about stock market by answering surveys or by writing their opinions (e.g., newsletters). For instance, the American Association of Individual Investors (AAII) calculates the weekly percentage of bullish, bearish and neutral answers of their members to a survey about the stock market. It is used by several studies (e.g., [28, 10, 40, 79, 78]). Investors Intelligence (II) computes three weekly percentages (bullish,

bearish and correction) of investor sentiment by analyzing independent market newsletters. Diverse studies (e.g., [69, 28, 44, 10, 11, 40, 79, 37, 78, 41]) used II indicators as proxy for investor sentiment.

This paper tests the prediction of these weekly sentiment indicators by using Twitter sentiment measures. This procedure may allow an acceptable anticipation of AAI and II values or a satisfactory alternative whenever they are unavailable. We did not experimented the forecasting of monthly sentiment indicators because there is not sufficient data to perform a feasible analysis. However, we used two additional monthly sentiment measures in the creation of a sentiment indicator using the Kalman Filter measure described in the next subsection. The University of Michigan Surveys of Consumers (UMSC) creates a monthly sentiment index by processing a survey answered by five hundred US households. This index is used in diverse studies about stock market (e.g., [28, 45, 58, 37, 62, 86, 16, 71]). Sentix ([www.sentix.de](http://www.sentix.de)) produces sentiment indices for different stock markets (e.g., US, Germany, Euro Zone, Japan) by delivering surveys to more than 3,500 participants. Some studies apply Sentix indices such as [61, 17, 36]. In this study we used the monthly sentiment indicators for US. AAI, II, UMSC and Sentix values were obtained from Thompson Reuters Datastream (<http://online.thomsonreuters.com/datastream/>).

#### *2.4. Sentiment Indicators created by Kalman Filter procedure*

Kalman Filter (KF) permits the combination of diverse observed variables in order to extract a latent variable. Existing investor sentiment indicators are measured by different approaches at different frequencies (e.g., surveys or social media interactions) and they usually produce distinct values. These observed values are related to sentiment but they contain some noise. The utilization of KF may permit the creation of a sentiment indicator that is more representative and less noisy than their individual components. Moreover, KF allows the usage of indicators with distinct frequencies (e.g., daily, weekly, monthly). Thus, we can produce a daily sentiment indicator from the combination of daily, weekly and monthly values. The linear dynamical model can be represented as follows:

$$\begin{aligned} Y_t &= F_t \theta_t + v_t, & v_t &\sim N(0, V_t) \\ \theta_t &= G_t \theta_{t-1} + w_t, & w_t &\sim N(0, W_t) \end{aligned} \tag{6}$$

where the first equation describes the observed variables (e.g., survey or social media sentiment indicators), the second equation represents the latent variable and  $V, W$  are parameter matrices.  $\theta_0$  is assumed to be normally distributed with mean 0 and variance  $1e7$ . The model is estimated by maximum likelihood allowing the observation noises,  $v_t$ , to be cross-correlated. To reduce the complexity of the optimization problem and ensure that the variance-covariance matrix is

positive semi-definite we use the approach suggested by [57] and followed by [56], parametrizing the covariance matrix,  $V$ , in terms of the elements of its log-Cholesky decomposition and the system variance,  $W$ , using its log.

We created a daily sentiment indicator by applying the KF procedure to five different sentiment indicators: AAIL, II, UMSC, Sentix and the daily Twitter sentiment indicator (TWT) created in this work. AAIL, II and TWT values correspond to the Bull-Bear spread (e.g., [10, 78]):

$$SP_t = \frac{NBull_t - NBear_t}{NBull_t + NBear_t} \quad (7)$$

All indicators were normalized by calculating their standard score. The model parameters were estimated using the first training rolling window also applied in the forecasting of stock market variables (i.e., first 300 days). Then, we created the sentiment indicators for the entire time period by filtering the series using the estimated model.

KF values were tested as sentiment indicators in forecasting models to assess eventual improvements compared to the utilization of TWT as proxy for sentiment. As an example, Figure 1 shows the overall (for all analyzed stocks) KF sentiment indicator values and its individual components (AAIL, II, UMSC, Sentix, TWT) when considering the period that ranges from 1st January of 2014 to 30th June of 2014. The plot confirms that KF indicator presents smoother values than TWT.

## 2.5. Stock Market Data

Various studies have analyzed the influence of sentiment on portfolios formed on different characteristics (e.g., market capitalization, book-to-market ratio, industries). For example, some papers refer that sentiment has more impact on returns of stocks with lower market capitalization (e.g, [43, 6, 8]). However, few of these studies apply sentiment measures extracted from social media and higher frequencies than the monthly periodicity. Therefore, in this study we explored a comprehensive set of stocks and portfolios having distinct characteristics:

- Standard & Poor’s 500 (SP500): index composed by 500 large companies.
- Russell 2000 (RSL): index of the smallest 2000 companies belonging to Russell 3000 index.
- Dow Jones Industrial Average (DJIA): constituted by 30 large companies listed on NYSE and NASDAQ.
- Nasdaq 100 (NDQ): includes the 100 of the largest non-financial stocks traded on NASDAQ.

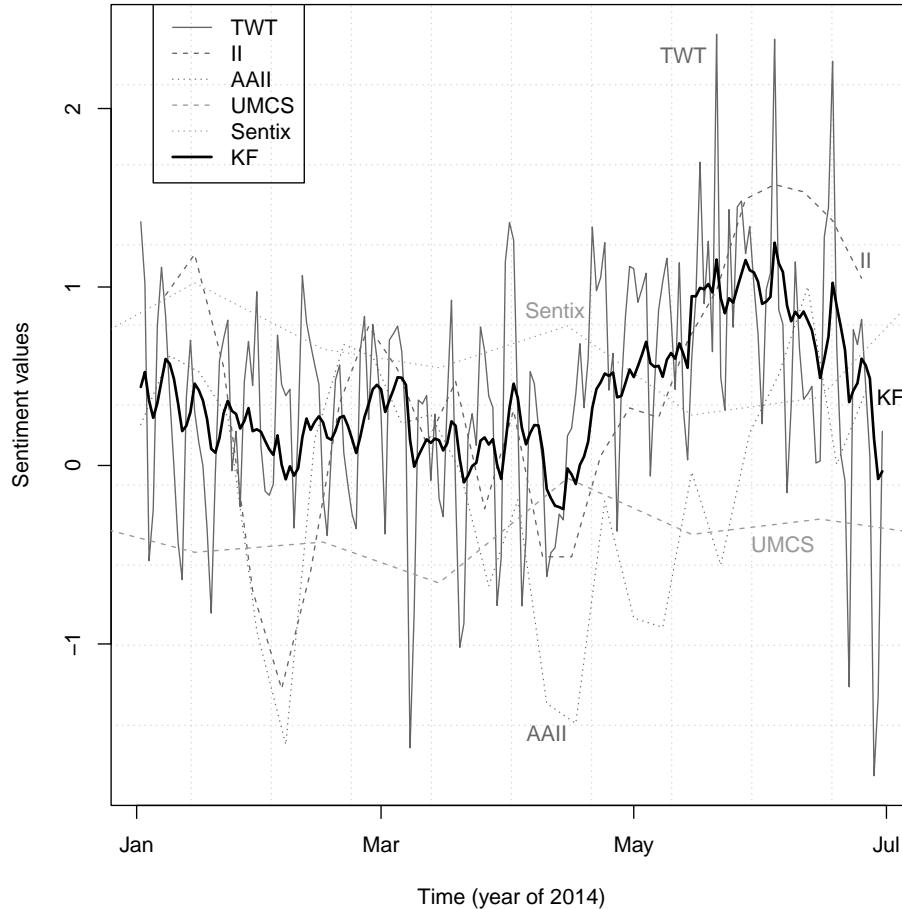


Figure 1: Sentiment indicator values (KF, AAIL, II, UMCS, Sentix and TWT)

- Excess return on the market (RMRF): return of the market minus the risk-free return rate. The return of the market corresponds to the value-weight return of all CRSP companies integrated in the US and traded on the NYSE, AMEX, or NASDAQ while the risk-free return rate is the one-month Treasury bill rate.
- Small Minus Big (SMB): a Fama and French factor corresponding to the difference in returns between small and large firms.
- High Minus Low (HML): a Fama and French factor that is equal to the difference in returns between value (i.e., high book-to-market ratios) and growth (i.e., low book-to-market ratios) stocks.
- Momentum Factor (MOM): spread in returns between high prior return portfolios and low prior return portfolios.
- CBOE Volatility Index (VIX): measures the implied volatility of SP500 index options. It is of-

ten considered the market’s “fear gauge” because it intends to represent investors expectation of future 30 days volatility.

- Portfolios formed on size (PSize): contains return values of portfolios constructed on the market capitalization, namely bottom 30% (Lo30), middle 40% (Mid40), top 30% (Hi30) and quintiles (Lo20, Qnt2, Qnt3, Qnt4, Hi20).
- 10 Industry Portfolios (PInd): returns of ten different industries, namely Consumer Non-Durables (NoDur), Consumer Durables (Durbl), Manufacturing (Manuf), Energy (Enrgy), Business Equipment (HiTec), Telecommunications (Telcm), Shops (Shops), Health (Hlth), Utilities (Utils) and Other (Other).

Thus, we analyze the effect of sentiment and attention on stocks having distinct size (e.g., PSize, SMB), industries (e.g., PInd), momentum (e.g., MOM) and book-to-market ratios (e.g., HML).

Several studies defend that investor sentiment and attention may influence stock market variables. For instance, sentiment may have predictive information for returns (e.g., [70, 9, 24, 60, 29]), volatility (e.g., [2, 60, 84]) and trading volume (e.g., [2, 74, 60]). Posting volume on social media (e.g., microblogs and message boards) may also add information for the forecasting of returns (e.g., [81, 2]), trading volume (e.g., [81, 70, 53]) and volatility (e.g., [2, 21]). We have focused on the prediction of these three different stock market variables:

- Daily returns of SP500, RSL, DJIA, NDQ, RMRF, SMB, HML, MOM, PSize and PInd. Returns measure changes in the asset value. We calculated the returns of SP500, RSL, DJIA and NDQ using the total return index (RI datatype) retrieved from Datastream as follows:

$$r_t = \frac{RI_t - RI_{t-1}}{RI_{t-1}} * 100 \quad (8)$$

where  $RI_t$  and  $RI_{t-1}$  are the total return index values of day  $t$  and  $t - 1$ . The returns of the remaining stocks were directly collected from Professor Kenneth French webpage ([http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)). All values are in percentage.

- Daily trading volume of SP500 and DJIA collected from Datastream. Trading volume is the number of shares traded in a given period of time (values in thousands).
- Daily volatility is measured using the model free estimate given the VIX index, and also the realized volatility measure given by a realized kernel for SP500, RSL, DJIA and NDQ. Since

the realized kernel values are very small, we converted them into annualized realized volatility as [3]:

$$av_t = \sqrt{rk_t} * \sqrt{252} * 100 \quad (9)$$

where  $av_t$  is the annualized realized volatility and  $rk_t$  is the realized kernel value. Volatility provides a measures of total risk associated with an investment. VIX data is available at CBOE webpage (<http://www.cboe.com/micro/vix/historical.aspx>) and the realized kernel of the referred indices were collected from Oxford-Man Institute of Quantitative Finance (<http://realized.oxford-man.ox.ac.uk/data/download>).

## 2.6. Models

In this work, we tested four different regression methods [35] to predict stock market variables: Multiple Regression (MR), Neural Network (NN), Support Vector Machine (SVM) and Random Forest (RF). We applied a MR model because it is very fast to learn, easy to interpret, extensively applied in finance and has less tendency to overfit. However, it has limited capacity to extract nonlinear associations. Therefore, we experimented three other methods that are more suitable to deal with nonlinearity: NN, SVM and RF. Often, these nonlinear methods lead to complex data-driven models. However, these models can still be understood by humans by using a sensitivity analysis and visualization techniques [19].

NN is a system of interconnected neurons whose functioning is inspired by biological neural networks. The numeric weights linking the neurons are calibrated during the learning process. In this study we applied the multilayer perceptron having one hidden layer with logistic functions. The output node applies a linear function for regression (e.g., returns, volatility, trading volume). The final output is dependent of the selection of initial weights. To address this problem, we apply an ensemble of NNs and calculate the average of the individual predictions [35]. The number of nodes in the hidden layer ( $H$ ) was the only hyperparameter we had to tune in this work.

The SVM model was initially proposed to perform classification tasks and then extended to regression by adopting an  $\epsilon$ -insensitive loss function. SVM can execute a nonlinear mapping by projecting the inputs into high-dimensional space using kernel functions. When compared with multilayer perceptron, the SVM algorithm has the advantage of always converging to the optimal set of weights. In this work, we utilize the popular gaussian kernel and tuned  $\gamma$ ,  $C$  and  $\epsilon$  hyperparameters.

RF is an ensemble model that generates a larger number of unpruned decision trees during the

training process. The individual trees are based on a random feature selection, using bootstrap training samples. The final RF predictions are built by averaging the outputs of the individual trees.

In order to evaluate the relevance of microblogging data to predict stock market behavior, we test all four regression models under two main strategies: with and without microblogging data. The baseline model is the Auto-Regressive (AR) model of order  $p$  (past time lags) of the predicted target, while the microblog based models are described next. When comparing the baseline and microblog based models, we select the best regression method (among MR, NN, SVM and RF). Given the extensive number of experiments conducted in this paper, we fixed the number of adopted time lags to  $p = 5$  for both baseline, i.e., AR(5), and microblog based models. We note that several other related works also used a similar short and fixed number of time lags. For instance, the same five past trading days were used in [74, 24, 52, 29].

All predictive experiments were conducted using the R tool and rminer package, which facilitates the application of data mining methods in real-world tasks [18]. The same methods (MR, NN, SVM and RF) were executed for both baseline and microblog based strategies. The default parameters (e.g., 500 trees for RF, ensemble with 3 multilayer perceptrons for NN, tolerance termination criterion of 0.001 for SVM) were used for the all methods except for the hyperparameters ( $H$  for NN and  $\gamma$ ,  $C$  or  $\epsilon$  for SVM). These hyperparameters were set using a grid search and internal 10-fold cross-validation considering the first training window (e.g., first 300 trading days for the daily stock market variables). The grid search values were set to  $H \in \{1, 2, 3, 4\}$ ,  $\gamma \in \{2^{-9}, 2^{-7}, \dots, 2^3\}$ ,  $C \in \{2^{-3}, 2^{-1}, \dots, 2^9\}$  and  $\epsilon \in \{0, 0.01, 0.1, 0.2, 0.4, 0.6\}$ .

The tested models to predict returns were:

$$\begin{aligned}
\hat{R}_t &= f(R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}) && \text{(MRet1, baseline)} \\
\hat{R}_t &= f(S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}) && \text{(MRet2)} \\
\hat{R}_t &= f(R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}) && \text{(MRet3)} \\
\hat{R}_t &= f(S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}, Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MRet4)} \\
\hat{R}_t &= f(R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}, \\
&Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MRet5)} \\
\hat{R}_t &= f(Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MRet6)} \\
\hat{R}_t &= f(R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MRet7)}
\end{aligned} \tag{10}$$

where  $S_t$  refers to the sentiment value of day  $t$  (VA, BI, BR or KF indicators) and  $Nt_t$  refers to

the first difference of posting volume of day  $t$ .

To forecast volatility, we experimented the following models:

$$\begin{aligned}
\hat{V}t_t &= f(Vt_{t-1}, Vt_{t-2}, Vt_{t-3}, Vt_{t-4}, Vt_{t-5}) && \text{(MVt1, baseline)} \\
\hat{V}t_t &= f(S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}) && \text{(MVt2)} \\
\hat{V}t_t &= f(Vt_{t-1}, Vt_{t-2}, Vt_{t-3}, Vt_{t-4}, Vt_{t-5}, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}) && \text{(MVt3)} \\
\hat{V}t_t &= f(S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}, Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MVt4)} \\
\hat{V}t_t &= f(Vt_{t-1}, Vt_{t-2}, Vt_{t-3}, Vt_{t-4}, Vt_{t-5}, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}, \\
&Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MVt5)} \\
\hat{V}t_t &= f(Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MVt6)} \\
\hat{V}t_t &= f(Vt_{t-1}, Vt_{t-2}, Vt_{t-3}, Vt_{t-4}, Vt_{t-5}, Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MVt7)}
\end{aligned} \tag{11}$$

where  $Vt_t$  is the volatility value of day  $t$ .

In the prediction of trading volume, we tested the following models:

$$\begin{aligned}
\hat{V}o_t &= f(Vo_{t-1}, Vo_{t-2}, Vo_{t-3}, Vo_{t-4}, Vo_{t-5}) && \text{(MVol1, baseline)} \\
\hat{V}o_t &= f(S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}) && \text{(MVol2)} \\
\hat{V}o_t &= f(Vo_{t-1}, Vo_{t-2}, Vo_{t-3}, Vo_{t-4}, Vo_{t-5}, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}) && \text{(MVol3)} \\
\hat{V}o_t &= f(S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}, Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MVol4)} \\
\hat{V}o_t &= f(Vo_{t-1}, Vo_{t-2}, Vo_{t-3}, Vo_{t-4}, Vo_{t-5}, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}, \\
&Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MVol5)} \\
\hat{V}o_t &= f(Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MVol6)} \\
\hat{V}o_t &= f(Vo_{t-1}, Vo_{t-2}, Vo_{t-3}, Vo_{t-4}, Vo_{t-5}, Nt_{t-1}, Nt_{t-2}, Nt_{t-3}, Nt_{t-4}, Nt_{t-5}) && \text{(MVol7)}
\end{aligned} \tag{12}$$

where  $Vo_t$  is the trading volume of day  $t$ .

In this study, we also used Twitter and KF sentiment indicators to predict survey sentiment indicators. AAI and II are popular survey sentiment indicators already used in diverse research studies about stock market behavior (e.g., [28, 10, 40, 78, 79, 44, 69, 11, 41, 37]). These indicators have distinct publication days. AAI values are released each Thursday comprising data from previous Thursday until Wednesday. II indicators are published each Wednesday and they are related to analysis performed from previous Wednesday to last Tuesday. To properly compare Twitter and KF indicators to each survey measure, we created different weekly microblogging indicators corresponding to the time interval used by each survey indicator. Additionally, we tested the utilization of the previous 7 daily (one week) Twitter and KF indicators.



We predicted four different values for each survey indicator: VA, BI, negative and positive percentage. In the forecasting of each survey value, we used the equivalent Twitter indicator (i.e., VA, BI, BearR and BullR). To compute the weekly Twitter values, we also experimented two different approaches. The first approach (AA) calculates the weekly value using the total number of positive and/or negative messages of the week while the second approach (MA) computes the average of the seven daily indicators that compose the week. The production of the weekly KF values applies the second approach.

We applied seven different models by exploring five lags of the target survey indicator, five lags of the weekly microblogging indicators and seven lags (one week) of the daily microblogging indicators. The explored forecasting of survey sentiment models are:

$$\begin{aligned}
\hat{Sv}_t &= f(Sv_{t-1}, Sv_{t-2}, Sv_{t-3}, Sv_{t-4}, Sv_{t-5}) && \text{(MSv1, baseline)} \\
\hat{Sw}_t &= f(Sw_{t-1}, Sw_{t-2}, Sw_{t-3}, Sw_{t-4}, Sw_{t-5}) && \text{(MSv2)} \\
\hat{Sd}_t &= f(Sd_{t-1}, Sd_{t-2}, Sd_{t-3}, Sd_{t-4}, Sd_{t-5}, Sd_{t-6}, Sd_{t-7}) && \text{(MSv3)} \\
\hat{Sv}_t &= f(Sw_{t-1}, Sw_{t-2}, Sw_{t-3}, Sw_{t-4}, Sw_{t-5}, \\
&Sd_{t-1}, Sd_{t-2}, Sd_{t-3}, Sd_{t-4}, Sd_{t-5}, Sd_{t-6}, Sd_{t-7}) && \text{(MSv4)} \\
\hat{Sv}_t &= f(Sv_{t-1}, Sv_{t-2}, Sv_{t-3}, Sv_{t-4}, Sv_{t-5}, \\
&Sw_{t-1}, Sw_{t-2}, Sw_{t-3}, Sw_{t-4}, Sw_{t-5}) && \text{(MSv5)} \\
\hat{Sv}_t &= f(Sv_{t-1}, Sv_{t-2}, Sv_{t-3}, Sv_{t-4}, Sv_{t-5}, \\
&Sd_{t-1}, Sd_{t-2}, Sd_{t-3}, Sd_{t-4}, Sd_{t-5}, Sd_{t-6}, Sd_{t-7}) && \text{(MSv6)} \\
\hat{Sv}_t &= f(Sv_{t-1}, Sv_{t-2}, Sv_{t-3}, Sv_{t-4}, Sv_{t-5}, \\
&Sw_{t-1}, Sw_{t-2}, Sw_{t-3}, Sw_{t-4}, Sw_{t-5}, \\
&Sd_{t-1}, Sd_{t-2}, Sd_{t-3}, Sd_{t-4}, Sd_{t-5}, Sd_{t-6}, Sd_{t-7}) && \text{(MSv7)}
\end{aligned} \tag{13}$$

where  $Sv_t$  corresponds to the weekly survey sentiment values (AAII or II) of day  $t$ ,  $Sw_t$  corresponds to the weekly microblogging sentiment values (BI, VA, BullR, BearR or KF) of day  $t$  and  $Sd_t$  corresponds to the daily microblogging values (AAII or II) of day  $t$ .

## 2.7. Evaluation

There is empirical evidence that good forecasting methods provide consistent results across multiple metrics [20]. Given the large number of experiments conducted in this work, we opted for a single error metric when evaluating the quality of the predictions. In this work, we selected an absolute error based metric, which is a common approach in the forecasting domain [39]. For instance, in [5, 4] it is argued that squared error metrics, such as Root Mean Square Error (RMSE),

are not reliable due to their sensitivity to outliers and should be replaced by absolute error metrics when comparing across time series. We note that other related works also have adopted absolute error metrics, such as: Mean Absolute Error (MAE) [24, 85] and Mean Absolute Percentage Error (MAPE) [9, 24, 49].

Using any absolute error measure should lead to the same ranking differences when comparing distinct forecasting models, thus the particular choice of such measure affects mostly its range of values and interpretation. In this paper, we selected the Normalized Mean Absolute Error (NMAE) metric that is calculated as [31]:

$$\begin{aligned} MAE &= \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \\ NMAE &= \frac{MAE}{y_H - y_L} \end{aligned} \tag{14}$$

where  $y_i$  is the target value for time  $i$ ,  $y_H$  is the highest target value,  $y_L$  is the lowest target value,  $\hat{y}_i$  is the predicted value and  $N$  corresponds to the number of predictions considered. When compared with other absolute based metrics, the NMAE presents several advantages. First, it is easier to interpret than MAE, since it expresses the error as a percentage of the full target scale. The lower the NMAE values, the better are the forecasts. Second, it is scale independent, which is particularly useful in this work since we predict variables with distinct scales. Third, it does not contain the limitations of other scale independent measures. For instance, MAPE can produce infinity values when the denominator (target values) is zero [39], which might occur in several of the predicted variables (e.g., returns can have near zero values). While we only consider NMAE values, in the tables with predictive results we also show the target range value ( $y_H - y_L$ ), thus the MAE values can be easily obtained by computing the inverse of Equation 14.

To measure the generalization capability of the predictive models, we applied a fixed-size rolling windows scheme [73]. For each prediction (e.g., day  $t$ ), the model is trained using a window of the previous  $W$  consecutive samples (e.g., from day  $t - W$  to day  $t - 1$ ) and used to predict the next value (time  $t$ ). Then, the training window is slid by discarding its oldest element and adding the value of  $t$  in order to retrain the model and predict the value at time  $t + 1$ , and so on. Therefore, a dataset of length  $L$  will produce  $L - W$  model trainings and their respective predictions. This rolling windows validation is realistic since it mimics the way a predictive model would be used in a real-environment, trained with a large number of past data and used to predict the next daily/weekly values. And it is robust, since it allows the training and testing of a large number of models. In this work, we applied a window size of  $W=300$  days for the prediction of the daily stock market variables and  $W=50$  days for the forecasting of the weekly survey sentiment indicators.

The number of predictions range from 392 to 439 for the daily variables and 92 to 93 for the weekly survey indices. We note that these numbers are much higher than most state of the art works (e.g., 8 [52] and 35 [14] predictions).

The prediction ability of the models was evaluated by the Diebold-Mariano (DM) test for predictive accuracy [25], under a pairwise comparison between the baseline and microblog based models. Thus, we assume that the microblogging data has predictive content if the respective model has a statistical significant DM test.

### 3. Results and Discussion

#### 3.1. Prediction of Returns

This work analyzed the prediction of returns of diverse indices as well as portfolios formed on size and industries. We tested sentiment indicators created by three different formulas (BullR, BI and VA) and KF procedure. Table 1 presents the NMAE results produced by the four regression methods (MR, RF, NN and SVM) using sentiment indicators produced by BullR approach and the p-value calculated by DM test for predictive accuracy [25] for SP500, RSL, DJIA, NDQ, RMRF, SMB, HML and MOM. In the table, the baseline results are underlined, while the lowest NMAE value is in **bold**. The first column also shows in brackets the number of predictions ( $L - W$ ) and the target range ( $y_H - y_L$ ).

We summarize the results for BullR, BI, VA and KF indicators in Table 2. For each index, Table 2 identifies the baseline model, the lowest NMAE model and models generating statistically significant results in the DM test. Ten SVM models significantly improve the results of baseline models for the forecasting of SP500, DJIA, MOM, SMB and RMRF. Two of these models obtain p-value inferior to 5%, both for the forecasting of SP500. There are no microblogging models having significantly higher predictive accuracy than baseline models for the remaining indices (i.e., HML, NDQ and RSL). However, the lowest NMAE values are produced by models containing microblogging features for all items, except for RSL. The utilization of KF sentiment indicators lowered the NMAE results for four indices: DJIA, MOM, RMRF and SP500. Furthermore, models applying KF indicators are significantly more accurate than baseline for SP500, DJIA, MOM and RMRF, while models using TWT indicators obtain this statistical significant results only for DJIA, SMB and SP500. Moreover, SVM MRet3 model using KF indicators produces statistically better forecasts of the SP500 than the baseline. The posting volume features were important for some indices. The lowest NMAE values of the prediction of NDQ and SMB were produced by models

Table 1: Predictive results for returns of SP500, RSL, DJIA, NDQ, RMRF, SMB, HML and MOM. Utilization of general sentiment indicators calculated using BullR formula and first difference of the posting volume. For each index, the baseline model is underlined and the lowest NMAE value is in **bold** (\* – p-value < 10%, \*\* – p-value < 5%, \*\*\* – p-value < 1%, NMAE in %).

	Mtd	MRet1	MRet2	MRet3	MRet4	MRet5	MRet6	MRet7
(number of predictions: 414; returns range: 7.53)	DJIA MR	<u>8.12</u>	8.11	8.24	8.20	8.33	8.11	8.25
	RF	8.38	8.41	8.43	8.44	8.42	8.38	8.45
	SVM	8.19	8.09	<b>8.01*</b>	8.07	8.07	8.05	8.05
	NN	8.28	8.42	8.16	8.29	8.44	8.56	8.26
(number of predictions: 392; returns range: 3.36)	HML MR	10.49	10.37	10.54	10.36	10.54	10.34	10.52
	RF	10.42	10.56	10.39	10.49	10.49	10.62	10.56
	SVM	<u>10.29</u>	10.31	<b>10.24</b>	10.26	10.25	10.26	10.33
	NN	10.92	10.84	10.66	10.62	10.73	10.99	10.47
(number of predictions: 392; returns range: 4.63)	MOM MR	10.91	10.80	10.94	11.02	11.18	11.03	11.11
	RF	11.00	11.14	10.93	11.18	11.17	11.31	11.21
	SVM	<u>10.78</u>	10.79	<b>10.73</b>	10.84	10.78	10.88	10.81
	NN	11.18	11.01	11.99	11.33	11.32	11.13	11.11
(number of predictions: 439; returns range: 9.35)	NDQ MR	7.68	7.70	7.82	7.86	7.99	7.74	7.85
	RF	7.85	7.90	7.88	7.97	7.96	7.88	7.93
	SVM	<u>7.61</u>	<b>7.59</b>	7.61	7.78	7.62	7.64	7.69
	NN	7.99	8.38	7.94	8.93	7.81	7.97	8.04
(number of predictions: 392; returns range: 7.58)	RMRF MR	8.31	8.32	8.45	8.42	8.58	8.36	8.44
	RF	8.54	8.59	8.61	8.68	8.81	8.56	8.74
	SVM	<u>8.27</u>	8.35	8.28	<b>8.24</b>	<b>8.24</b>	<b>8.24</b>	8.26
	NN	8.95	8.46	8.47	8.65	8.54	9.57	8.61
(number of predictions: 439; returns range: 7.02)	RSL MR	<b>11.10</b>	11.31	11.35	11.46	11.49	11.28	11.30
	RF	11.17	11.47	11.19	11.58	11.40	11.44	11.33
	SVM	11.15	11.71	11.20	11.24	11.19	11.13	11.13
	NN	11.21	11.60	11.21	11.46	11.64	12.57	13.85
(number of predictions: 392; returns range: 3.36)	SMB MR	<u>12.44</u>	12.44	12.59	12.50	12.76	12.40	12.59
	RF	12.46	12.72	12.54	12.74	12.50	12.72	12.42
	SVM	<u>12.44</u>	12.41	12.40	<b>12.27*</b>	12.46	12.48	12.44
	NN	13.49	13.09	13.05	13.41	12.70	12.55	12.89
(number of predictions: 413; returns range: 7.85)	SP500 MR	7.90	7.88	8.01	7.97	8.11	7.92	8.01
	RF	8.32	8.20	8.32	8.23	8.35	8.13	8.32
	SVM	<u>7.87</u>	7.83	7.88	7.92	7.86	7.83	<b>7.80**</b>
	NN	8.59	8.17	7.95	8.00	8.02	7.92	8.40

using the first difference of the number of tweets. Additionally, there are diverse models applying these features significantly more accurate than baseline for SP500 and SMB. The majority of the most accurate models applies the SVM method.

Table 2: Predictive results for returns of SP500, RSL, DJIA, NDQ, RMRF, SMB, HML and MOM. Utilization of general sentiment indicators (BullR, BI and VA approaches), KF indicators and first difference of the posting volume. NMAE values of baseline model, lowest NMAE model and models producing statistical significant results in DM test (\* - p-value < 10%, \*\* - p-value < 5%, \*\*\* - p-value < 1%, sentiment indicators in parenthesis, NMAE in %, NMAE values lower than baseline are in **bold**).

Index	Baseline	Lowest NMAE	Statistical significant results
DJIA (n° predictions: 414; returns range: 7.53)	MR: 8.12	SVM MRet2 (KF): <b>7.98*</b>	SVM MRet3 (BR): <b>8.01*</b> SVM MRet2 (KF): <b>7.98*</b>
HML (n° predictions: 392; returns range: 3.36)	SVM: 10.29	SVM MRet3 (BR): <b>10.24</b>	
MOM (n° predictions: 392; returns range: 4.63)	SVM: 10.78	SVM MRet2 (KF): <b>10.69*</b>	SVM MRet2 (KF): <b>10.69*</b>
NDQ (n° predictions: 439; returns range: 9.35)	SVM: 7.61	SVM MRet7: <b>7.58</b>	
RMRF (n° predictions: 392; returns range: 7.58)	SVM: 8.27	SVM MRet3 (KF): <b>8.19*</b>	SVM MRet3 (KF): <b>8.19*</b>
RSL (n° predictions: 439; returns range: 7.02)	MR: 11.10	MR MRet1: 11.10	
SMB (n° predictions: 392; returns range: 3.36)	MR: 12.44 SVM: 12.44	SVM MRet4 (BR): <b>12.27*</b>	SVM MRet4 (BR): <b>12.27*</b>
SP500 (n° predictions: 439; returns range: 7.85)	SVM: 7.87	SVM MRet3 (KF): <b>7.79**</b>	SVM MRet7: <b>7.80**</b> SVM MRet6: <b>7.81*</b> SVM MRet4 (VA): <b>7.81*</b> SVM MRet5 (VA): <b>7.81*</b> SVM MRet3 (KF): <b>7.79**</b>

Some studies refer that there is a distinct sentiment effect in prices of stocks with large and small market capitalization (e.g., [6, 8, 51, 44, 43]). To analyze the influence of sentiment on stocks with different size, we predicted returns of portfolios formed on size. Table 3 shows a summary of all tested models for the prediction of portfolios formed on size. There are fifteen models significantly more accurate than baseline in the forecasting of portfolios of lower market capitalization (i.e., Lo20 and Lo30). The prediction of Lo20 returns has one model obtaining p-value less than 1% in the pairwise DM test and five other models generating p-value inferior to 5%. These results may indicate that sentiment is more informative to the prediction of stocks of smaller capitalization, which is consistent with previous findings (e.g., [6, 8, 43]). The most accurate microblogging models are SVM. Posting volume seems informative for the prediction of these portfolios. The majority of the models obtaining statistical significant results in the pairwise DM test and the lowest NMAE values apply the first difference of posting volume. KF indicators seem less useful for the forecasting of portfolios formed on size than for the prediction of indices. Models applying

KF indicators outperformed models using Twitter indicators only for Qnt4. Nevertheless, in the prediction of Lo20 there is one model applying KF indicators that produced p-value less than 5% in the DM test and another model obtained p-value less than 10%.

Table 3: Predictive results for returns of portfolios formed on size. Utilization of general sentiment indicators (BullR, BI and VA approaches), KF indicators and first difference of the posting volume. NMAE values of baseline model, lowest NMAE model and models producing statistical significant results in DM test (\* – p-value < 10%, \*\* – p-value < 5%, \*\*\* – p-value < 1%, sentiment indicators in parenthesis, NMAE in %, NMAE values lower than baseline are in **bold**, 392 predictions).

Index	Baseline	Lowest NMAE	Statistical significant results
Lo30 (returns range: 5.96)	MR: 11.75	SVM MRet4 (VA): <b>11.57*</b>	SVM MRet4 (BR): <b>11.59*</b> SVM MRet4 (BI): <b>11.59*</b> SVM MRet4 (VA): <b>11.57*</b> SVM MRet5 (VA): <b>11.60*</b>
Med40 (returns range: 6.87)	MR: 11.08 SVM: 11.08	SVM MRet4 (BR): <b>11.01</b> SVM MRet5 (BR): <b>11.01</b> SVM MRet7: <b>11.01</b> SVM MRet5 (BI): <b>11.01</b> SVM MRet5 (KF): <b>11.01</b>	
Hi30 (returns range: 7.34)	SVM: 8.71	SVM MRet7 (BI): <b>8.68</b> SVM MRet4 (VA): <b>8.68</b> SVM MRet5 (VA): <b>8.68</b> SVM MRet6: <b>8.68</b>	
Lo20 (returns range: 5.64)	MR: 11.94	SVM MRet4 (VA): <b>11.66***</b>	SVM MRet4 (BR): <b>11.74*</b> SVM MRet4 (BI): <b>11.75*</b> SVM MRet5 (BI): <b>11.70**</b> SVM MRet7: <b>11.75**</b> SVM MRet2 (VA): <b>11.75*</b> SVM MRet3 (VA): <b>11.78*</b> SVM MRet4 (VA): <b>11.66***</b> SVM MRet5 (VA): <b>11.74**</b> SVM MRet6: <b>11.71**</b> SVM MRet2 (KF): <b>11.74*</b> SVM MRet4 (KF): <b>11.72**</b>
Qnt2 (returns range: 7.31)	SVM: 11.58	SVM MRet6: <b>11.53</b>	
Qnt3 (returns range: 6.94)	MR: 10.83	SVM MRet5 (BR): <b>10.77</b>	
Qnt4 (returns range: 6.72)	SVM: 10.31	SVM MRet2 (KF): <b>10.28</b>	
Hi20 (returns range: 7.33)	SVM: 8.63	SVM MRet6: <b>8.57</b> SVM MRet7: <b>8.57</b>	

We also tested the prediction of portfolios of 10 different industries. The respective evaluation results are summarized in Table 4. There are several microblogging models significantly more accurate than baseline for Enrgy, HiTec and Other. Seven of these models obtain p-value less than 5% in the pairwise DM test for the prediction of Enrgy and three models have p-value less than 5% for the forecasting of HiTec. Therefore, microblogging features are particularly informative for Enrgy and HiTec. A possible explanation for these results is the unusual high number of microblogging messages related to stocks of these sectors, mainly for HiTec. Thus, the general

sentiment indicators may be biased toward these industries. For demonstration purposes, the forecasted returns of Enrgy sector by the NN MRet2 (VA) are presented in the Figure 2. The utilization of posting volume was also important. The majority of models obtaining statistical significant results in the pairwise DM test use posting volume features. Moreover, the utilization of KF indicators is beneficial in some situations. KF indicators permit lower NMAE values than TWT indicators for the HiTec and Utils sectors, and they are the most applied sentiment indicator in models which are significantly more accurate than baselines.

Table 4: Predictive results for returns of portfolios formed on industries. Utilization of general sentiment indicators (BullR, BI and VA approaches), KF indicators and first difference of the posting volume. NMAE values of baseline model, lowest NMAE model and models producing statistical significant results in DM test (\* – p-value < 10%, \*\* – p-value < 5%, \*\*\* – p-value < 1%, sentiment indicators in parenthesis, NMAE in %, NMAE values lower than baseline are in **bold**, 350 predictions).

Index	Baseline	Lowest NMAE	Statistical significant results
Durbl (returns range: 5.71)	MR: 13.90	SVM MRet6: <b>13.74</b>	
Enrgy (returns range: 21.12)	SVM: 7.18	NN MRet2 (VA): <b>7.07**</b>	SVM MRet3 (BI): <b>7.10**</b> SVM MRet5 (BI): <b>7.10**</b> SVM MRet6: <b>7.13*</b> SVM MRet7: <b>7.12*</b> SVM MRet2 (VA): <b>7.08**</b> NN MRet2 (VA): <b>7.07**</b> SVM MRet3 (VA): <b>7.11*</b> SVM MRet5 (VA): <b>7.10**</b> SVM MRet2 (KF): <b>7.10**</b> SVM MRet3 (KF): <b>7.10**</b> SVM MRet4 (KF): <b>7.11*</b> SVM MRet5 (KF): <b>7.12*</b>
HiTec (returns range: 5.90)	MR: 12.79	SVM MRet5 (KF): <b>12.55**</b>	SVM MRet3 (BR): <b>12.64*</b> SVM MRet4 (BR): <b>12.56**</b> SVM MRet4 (BI): <b>12.56**</b> SVM MRet2 (KF): <b>12.58*</b> SVM MRet3 (KF): <b>12.64*</b> SVM MRet5 (KF): <b>12.55**</b>
Hlth (returns range: 7.10)	NN: 12.94	SVM MRet6: <b>12.85</b>	
Manuf (returns range: 5.17)	SVM: 13.66	SVM MRet7: <b>13.62</b>	
NoDur (returns range: 4.55)	SVM: 12.77	SVM MRet3 (BR): <b>12.72</b>	
Other (returns range: 4.04)	SVM: 13.51	SVM MRet6: <b>13.34*</b> SVM MRet4 (VA): <b>13.34*</b> SVM MRet4 (KF): <b>13.34*</b>	SVM MRet6: <b>13.34*</b> SVM MRet3 (VA): <b>13.37*</b> SVM MRet4 (VA): <b>13.34*</b> SVM MRet4 (KF): <b>13.34*</b>
Shops (returns range: 4.76)	SVM: 13.85	SVM MRet4 (BI): <b>13.67*</b>	SVM MRet4 (BI): <b>13.67*</b>
Telcm (returns range: 5.52)	SVM: 14.15	SVM MRet2 (BR): <b>13.98*</b>	SVM MRet2 (BR): <b>13.98*</b>
Utils (returns range: 5.77)	SVM: 11.18	RF MRet5 (KF): <b>11.04</b>	

The application of sectorial sentiment indicators may allow a better analysis of the effect of sentiment in each sector. We had to exclude the “Other” sector from this analysis because there was no information about its SIC codes, so we were unable to create its sectorial indicator. Additionally,

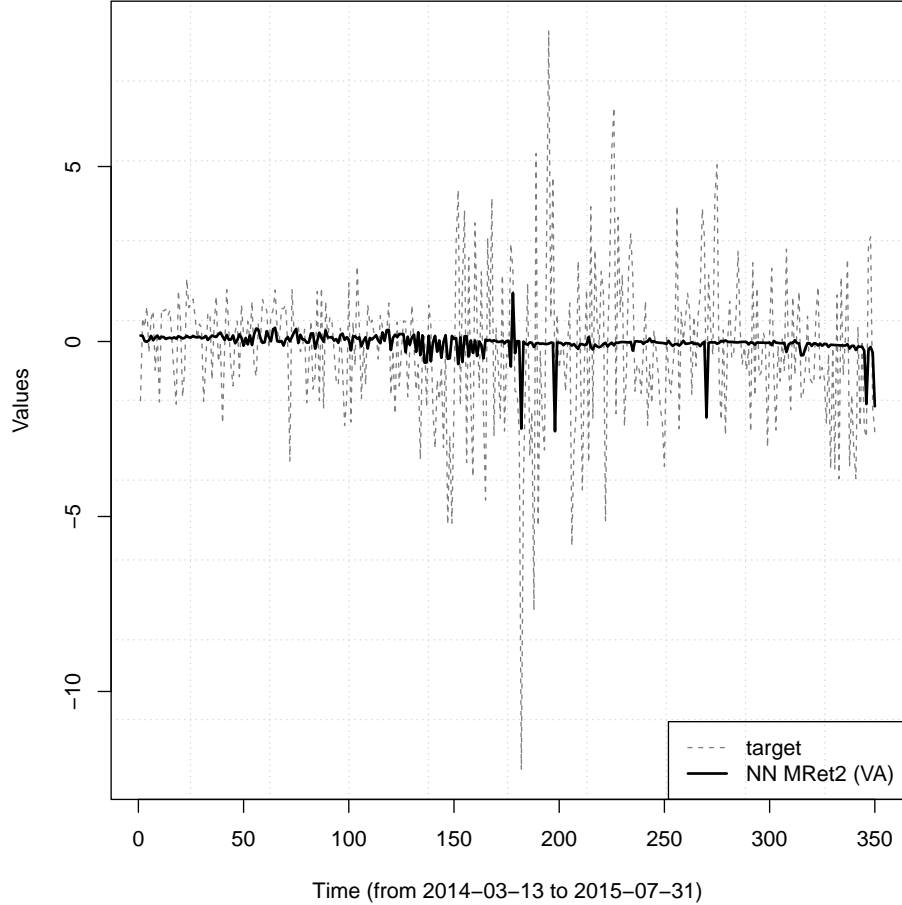


Figure 2: Predicted and real values for Enrgy

there are no tweets for Durbl sector for 37 days, so we excluded those days from the analysis. Table 5 present the results for these models.

Sectorial indicators are particularly useful for Shops and Telcm industries. There are more models significantly more accurate than baseline using sectorial indicators than applying general indicators in these sectors. The SVM model using previous values of returns, posting volume and BI sentiment (SVM MRet5 (BI)) obtains a p-value less than 5% for the prediction of Telcm. Moreover, the lowest NMAE values for Telcm and Shops are obtained by models utilizing sectorial indicators.

In summary, microblogging sentiment and attention indicators were particularly informative for the forecasting of returns of SP500, portfolios of smaller market capitalization (Lo20 and Lo30) and some sectors such as HiTec, Enrgy and Telcm. In the prediction of the returns of the mentioned stocks, there are diverse microblogging models obtaining p-value less than 5% in the pairwise DM test. Many of these models apply both sentiment and posting volume indicators. These results may suggest that sentiment and attention have more impact on future returns of stocks of inferior



Table 5: Predictive results for returns of portfolios formed on industries. Utilization of sectorial sentiment indicators (BullR, BI and VA approaches) and first difference of the posting volume. NMAE values of baseline model, lowest NMAE model and models producing statistical significant results in DM test (\* – p-value < 10%, \*\* – p-value < 5%, \*\*\* – p-value < 1%, sentiment indicators in parenthesis, NMAE in %, NMAE values lower than baseline are in **bold**, 350 predictions).

Index	Baseline	Lowest NMAE	Statistical significant results
Durbl (returns range: 5.71)	MR: 13.90	SVM MRet2 (BI): <b>13.75</b>	
Enrgy (returns range: 21.12)	SVM: 7.18	SVM MRet3 (BR): <b>7.10*</b> SVM MRet3 (BI): <b>7.10**</b> SVM MRet5 (BI): <b>7.10**</b>	SVM MRet3 (BR): <b>7.10*</b> SVM MRet4 (BR): <b>7.11*</b> SVM MRet3 (BI): <b>7.10**</b> SVM MRet5 (BI): <b>7.10**</b>
HiTec (returns range: 5.90)	MR: 12.79	SVM MRet6: <b>12.56**</b> SVM MRet4 (VA): <b>12.56**</b>	SVM MRet2 (BR): <b>12.57**</b> SVM MRet4 (BR): <b>12.58*</b> SVM MRet6: <b>12.56**</b> SVM MRet2 (BI): <b>12.61*</b> SVM MRet4 (BI): <b>12.62*</b> SVM MRet4 (VA): <b>12.56**</b>
Hlth (returns range: 7.10)	NN: 12.94	NN MRet1: 12.94	
Manuf (returns range: 5.17)	SVM: 13.66	SVM MRet5 (VA): <b>13.60</b>	
NoDur (returns range: 4.55)	SVM: 12.77	MR MRet2 (VA): <b>12.72</b>	
Shops (returns range: 4.76)	SVM: 13.85	SVM MRet5 (BI): <b>13.63*</b>	SVM MRet5 (BR): <b>13.64*</b> SVM MRet6: <b>13.67*</b> SVM MRet5 (BI): <b>13.63*</b> SVM MRet4 (VA): <b>13.65*</b>
Telcm (returns range: 5.52)	SVM: 14.15	SVM MRet4 (BR): <b>13.92*</b>	SVM MRet4 (BR): <b>13.92*</b> SVM MRet5 (BR): <b>13.95*</b> SVM MRet5 (BI): <b>13.93**</b>
Utils (returns range: 5.77)	SVM: 11.18	SVM MRet6: <b>11.14</b>	

capitalization and technology related companies. The impact of sentiment on these type of stocks may be explained by a higher concentration of irrational investors. Small stocks are considered to be mostly held by individual investors [43] and less attractive to rational investors [6, 8]. Additionally, a considerable part of technology companies are young and have a small track record. So, they are less appealing to professional investors that prefer easier to value stocks [6]. The unpredictability of irrational traders adds risk on prices and makes arbitrage strategies more difficult to implement [47]. Thus, stock prices may differ from fundamental values because arbitrage may be insufficient to instantly eliminate mispricing generated by investor sentiment. In these situations, there is margin to predict future prices and sentiment indicators can be informative.

The utilization of KF indicators was important in some situations. For instance, models applying KF indicators have p-value less than 5% for the prediction of SP500, Lo20, HiTec and Enrgy. Moreover, the usage of KF sentiment indicators decreased the NMAE values for DJIA, MOM, RMRF, SP500, Qnt4, HiTec and Utils. The sectorial indicators were useful for the prediction of returns of Shops and Telcm industries.

Regarding the tested methods, SVM was clearly the most effective. Therefore, the relationship between sentiment, attention and returns may be nonlinear.

### 3.2. Prediction of Volatility

In this paper, we forecasted VIX and the annualized realized volatility of SP500, RSL, DJIA and NDQ. Table 6 outlines the results produced by models applying posting volume and general sentiment indicators computed by BullR, BI, VA, AG and KF approaches.

Table 6: Predictive results for VIX and annualized realized volatility of SP500, RSL, DJIA and NDQ. Utilization of general sentiment indicators (BullR, BI, VA and AG approaches), KF indicators and first difference of the posting volume. NMAE values of baseline model, lowest NMAE model and models producing statistical significant results in DM test (\* – p-value < 10%, \*\* – p-value < 5%, \*\*\* – p-value < 1%, sentiment indicators in parenthesis, NMAE in %, NMAE values lower than baseline are in **bold**).

Index	Baseline	Lowest NMAE	Statistical significant results
DJIA (n° predictions: 413; realized volatility range: 92.41)	MR: 2.91	SVM MVlt3 (AG): <b>2.79</b>	MR MVlt3 (KF): <b>2.85*</b>
NDQ (n° predictions: 413; realized volatility range: 57.18)	MR: 3.89	SVM MVlt3 (VA): <b>3.87</b>	
RSL (n° predictions: 412; realized volatility range: 39.56)	MR: 5.71	MR MVlt1: 5.71	
SP500 (n° predictions: 413; realized volatility range: 67.90)	MR: 3.40	MR MVlt3 (KF): <b>3.36</b>	
VIX (n° predictions: 413; VIX range: 30.42)	MR: 3.27	SVM MVlt3 (BR): <b>3.25</b>	

The utilization of KF sentiment values and previous volatility values significantly improves the forecasting of DJIA realized volatility compared to the AR(5) model. However, the lowest p-value of DM test is not inferior to 5%. KF indicators were also used in the lowest NMAE model for the prediction of SP500 realized volatility. AG indicators were applied in the model that obtained the lowest NMAE value for DJIA. However, this model is not significantly more accurate than baseline. The inclusion of the number of tweets do not seems to benefit the forecasting of volatility. There are no models utilizing posting volume that significantly outperform the respective baseline models. Also, the most accurate models using posting volume features do not improve the results of models without posting volume information. MR and SVM are the most effective methods.

### 3.3. Prediction of Trading Volume

In this paper, we predicted the trading volume of SP500 and DJIA. Table 7 digests the results for models using general sentiment indicators, KF indicators and first difference of the number of tweets.

Table 7: Predictive results for trading volume of SP500 and DJIA. Utilization of general sentiment indicators (BullR, BI, VA and AG approaches), KF indicators and first difference of the posting volume. NMAE values of baseline model, lowest NMAE model and models producing statistical significant results in DM test (\* – p-value < 10%, \*\* – p-value < 5%, \*\*\* – p-value < 1%, sentiment indicators in parenthesis, NMAE in %, NMAE values lower than baseline are in **bold**).

Index	Baseline	Lowest NMAE	Statistical significant results
DJIA (n° predictions: 414; volume range: 310804)	SVM: 6.00	SVM MVlt5 (BR): <b>5.84*</b> SVM MVlt5 (BI): <b>5.85*</b>	SVM MVlt5 (BR): <b>5.84*</b>
SP500 (n° predictions: 413; volume range: 1636036)	SVM: 4.98	SVM MVlt1: 4.98	

The application of microblogging sentiment values and the first difference of the number of tweets significantly improves baseline results for the prediction of DJIA trading volume in two situations: SVM MVlt5 (BR) and SVM MVlt5 (BI). However, the lowest NMAE values for SP500 are obtained by the AR(5) model. These results add some evidence that posting volume and sentiment may have predictive content for forecasting of trading volume (e.g., [60, 2, 74, 81, 53, 70]). The utilization of AG indicators do not result in statistical significant DM tests. Therefore, we do not have evidence that disagreement is associated with trading volume as found in some studies (e.g., [2, 64]). The most accurate models also did not applied KF indicators. The SVM method is clearly the best performing approach for the forecasting of trading volume.

### 3.4. Prediction of survey sentiment indicators using Twitter and KF sentiment indicators

In this subsection, we apply Twitter and KF indicators to forecast AAI and II indicators. The applied procedure is similar to the applied in the forecasting of stock market variables. However, we reduced the size of the rolling windows to 50 because the weekly periodicity downsized the data set. We tested the prediction of four different survey values: VA, BI, negative and positive values. Table 8 presents the evaluation results of the forecasting of AAI and II indicators using weekly Twitter indicators produced by AA approach (aggregating positive and negative messages of the week). For each prediction, the table also shows the number of predictions and target range ( $y_L - y_H$ ).

The utilization of Twitter indicators produced by AA approach was important for the prediction of VA and negative AAI values. The combined utilization of weekly Twitter indicators and previous AAI values (MSv5) produced lower NMAE results than baseline for VA values. However, the relevance of Twitter indicators is higher for AAI negative values because there are seven models producing lower NMAE results than baseline, four of which use only Twitter indicators. These seven models apply daily Twitter indicators. Furthermore, SVM MSv7 is significantly more accurate than

Table 8: Prediction of weekly AAI and II values using Twitter sentiment indicators calculated by AA approach. For each survey sentiment indicator, the baseline model is underlined and NMAE values lower than baseline are in **bold** (\* - p-value < 10%, \*\* - p-value < 5% and \*\*\* - p-value < 1%, NMAE values in %).

Mtd	MSv1	MSv2	MSv3	MSv4	MSv5	MSv6	MSv7	Mtd	MSv1	MSv2	MSv3	MSv4	MSv5	MSv6	MSv7
Panel A: Prediction of AAI values calculated by BI formula (93 predictions; range: 58.21)															
MR	14.35	18.10	17.72	19.42	15.52	16.19	16.98	RF	14.61	17.59	17.92	16.95	14.71	15.18	14.69
SVM	<u>14.30</u>	16.92	16.84	16.56	14.41	16.42	14.73	NN	16.11	21.58	21.71	20.61	17.15	17.81	17.39
Panel B: Prediction of AAI values calculated by VA formula (92 predictions; range: 25.2)															
MR	<u>18.74</u>	19.36	21.7	22.42	20.12	21.96	24.01	RF	19.11	19.35	19.67	19.74	19.11	19.77	19.43
SVM	18.97	21.24	21.54	19.20	<b>18.60</b>	18.93	18.96	NN	19.76	27.84	23.57	22.77	24.14	21.08	23.79
Panel C: Prediction of AAI positive values (93 predictions; range: 37.89)															
MR	<u>12.52</u>	19.26	19.43	20.18	13.76	14.52	15.83	RF	13.52	18.50	19.61	18.58	13.03	13.74	13.52
SVM	14.54	18.31	19.49	18.29	14.19	13.55	14.36	NN	13.41	18.63	21.80	21.39	14.50	17.32	19.07
Panel D: Prediction of AAI negative values (93 predictions; range: 25.65)															
MR	<u>16.42</u>	18.26	17.17	18.77	17.62	16.56	18.07	RF	16.64	17.86	<b>16.23</b>	<b>15.90</b>	16.74	<b>15.57</b>	<b>15.44*</b>
SVM	16.77	18.30	<b>16.33</b>	<b>15.88</b>	17.18	17.09	<b>16.37</b>	NN	18.05	22.61	18.34	18.05	17.59	19.81	22.63
Panel A: Prediction of II values calculated by BI formula (92 predictions; range: 55.8)															
MR	<u>5.85</u>	12.56	12.12	12.43	6.35	6.49	7.29	RF	8.60	12.14	11.18	11.06	8.37	7.78	8.13
SVM	6.33	14.64	13.10	13.57	6.80	6.79	7.50	NN	8.43	13.18	11.83	12.50	10.95	7.82	9.19
Panel B: Prediction of II values calculated by VA formula (91 predictions; range: 19.4)															
MR	<u>16.20</u>	17.48	17.69	19.00	18.12	17.43	19.17	RF	16.7	17.76	16.44	16.78	17.06	16.23	16.59
SVM	<u>16.20</u>	20.27	17.08	16.82	17.63	17.23	16.58	NN	18.16	22.27	<b>15.91</b>	16.95	20.61	17.83	18.97
Panel C: Prediction of II positive values (92 predictions; range: 37.9)															
MR	<u>8.31</u>	14.61	12.69	13.31	8.76	8.48	9.11	RF	11.12	14.31	12.74	12.49	10.78	10.24	10.43
SVM	8.76	19.64	13.97	14.04	11.10	8.56	10.39	NN	8.73	16.89	14.11	13.89	12.73	9.67	11.84
Panel D: Prediction of II negative values (92 predictions; range: 21.8)															
MR	<u>4.70</u>	11.97	11.93	12.12	5.05	5.53	5.80	RF	6.17	11.31	11.18	10.86	6.23	6.12	6.47
SVM	6.38	13.86	12.26	11.59	9.25	7.00	7.70	NN	6.07	11.42	10.85	11.58	7.39	7.26	8.08

baseline according to the pairwise DM test. In the prediction of II indicators, there is only one model more accurate than AR(5) model. The utilization of daily Twitter indicators (i.e., MSv3) produces lower NMAE values than baseline for the forecasting of II values calculated by VA formula. For demonstration purposes, the prediction of negative values of AAI by the SVM MSv4 model (using only Twitter indicators) is present in the left of Figure 3.

Table 9 shows the results of the prediction of survey indicators using Twitter indicators created by MA approach (average of the daily indicators). The most accurate models using Twitter indicators computed by MA approach produce slightly lower NMAE values than models applying Twitter indicators calculated by AA approach for the forecasting of AAI BI and VA values. However, there are no models significantly more accurate than baseline for the prediction of AAI negative values. Moreover, models utilizing these Twitter indicators do not outperform AR(5) models for the forecasting of any II value.

Table 10 presents the forecasting of survey values using KF indicators. The usage of KF indicators produced worse results than Twitter indicators in the forecasting of AAI indicators.

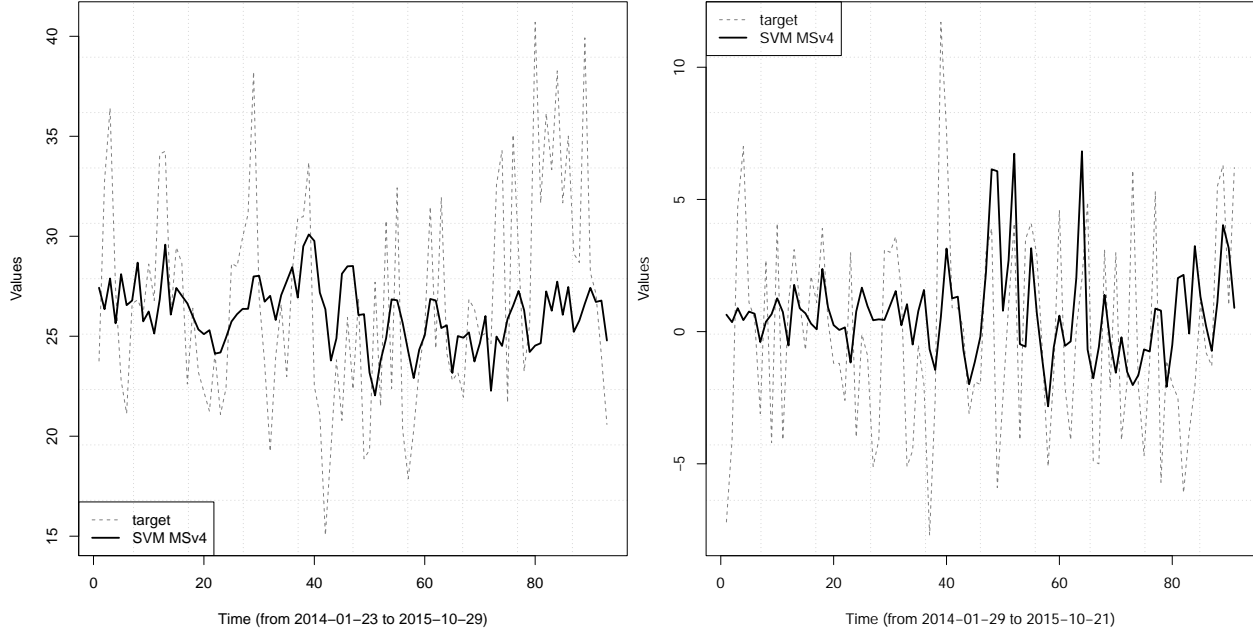


Figure 3: Predicted results for negative values of AAI (left) and values of II calculated by VA formula using KF indicators (right)

Yet, KF indicators were more informative for the prediction of II VA values. While there is only one model applying Twitter indicators outperforming the baseline model for II values, there are eight models using KF values more accurate than the best AR(5) model for II values calculated by VA approach. These II forecasts are presented in the right of Figure 3.

Overall, Twitter sentiment indicators proved to be useful for the prediction of negative values of AAI but less important for the forecasting of II values. Nevertheless, KF indicators are informative for the prediction of II computed by VA formula. Contrary to the forecasting of stock market variables, SVM is not the dominant regression model and the most accurate AAI and II prediction models are obtained by distinct learning models (e.g., RF is the best model for negative AAI predictions; NN obtains the lowest NMAE values for the prediction of II calculated by VA formula).

#### 4. Conclusions

Diverse studies refer that sentiment may affect future returns (e.g., [7, 28, 60, 70, 9]), volatility (e.g., [60, 2, 84, 44]) and trading volume (e.g., [60, 2, 74]). Also, that posting volume on social media can be applied in the forecasting of returns (e.g., [81, 2]), trading volume (e.g., [81, 53, 70]) and volatility (e.g., [2, 21]). However, many of these studies use very simple sentiment analysis methods (e.g., use of generic lexicons [74, 9]) and do not execute a solid evaluation (e.g., with no out of sample [2, 7, 60, 84] or usage of very small test sets [9]). Other studies find scarce evidence of

Table 9: Prediction of weekly AAI and II values using Twitter sentiment indicators calculated by MA approach. For each survey sentiment indicator, the baseline model is underlined and NMAE values lower than baseline are in **bold** (\* - p-value < 10%, \*\* - p-value < 5% and \*\*\* - p-value < 1%, NMAE values in %).

Mtd	MSv1	MSv2	MSv3	MSv4	MSv5	MSv6	MSv7	Mtd	MSv1	MSv2	MSv3	MSv4	MSv5	MSv6	MSv7
Panel A: Prediction of AAI values calculated by BI formula (93 predictions; range: 58.21)															
MR	14.35	18.81	17.72	19.86	15.93	16.19	17.05	RF	14.61	17.80	17.96	17.71	<b>14.26</b>	15.14	14.85
SVM	<u>14.30</u>	16.44	16.84	17.35	14.40	16.42	15.36	NN	16.11	17.35	17.92	22.71	16.23	17.87	17.58
Panel B: Prediction of AAI values calculated by VA formula (92 predictions; range: 25.2)															
MR	<u>18.74</u>	19.65	21.70	23.09	19.89	21.96	23.12	RF	19.11	19.30	19.70	19.58	19.33	19.87	19.34
SVM	18.97	18.92	20.96	19.49	19.11	<b>18.44</b>	18.97	NN	19.76	20.21	23.10	21.53	25.50	22.37	22.15
Panel C: Prediction of AAI positive values (93 predictions; range: 37.89)															
MR	<u>12.52</u>	20.65	19.43	20.98	14.15	14.52	16.19	RF	13.52	19.41	19.59	19.67	13.32	13.71	14.01
SVM	14.54	19.38	18.64	19.39	14.07	14.36	14.60	NN	13.41	21.28	20.71	21.11	16.37	15.34	18.72
Panel D: Prediction of AAI negative values (93 predictions; range: 25.65)															
MR	<u>16.42</u>	18.51	17.17	18.86	18.11	16.56	18.06	RF	16.64	16.97	<b>16.35</b>	<b>16.06</b>	16.56	<b>15.45</b>	<b>15.60</b>
SVM	16.77	17.19	<b>16.33</b>	<b>16.30</b>	17.36	17.07	<b>16.19</b>	NN	18.05	18.26	17.80	18.79	17.70	18.70	17.73
Panel A: Prediction of II values calculated by BI formula (92 predictions; range: 55.8)															
MR	<u>5.85</u>	12.82	12.12	12.65	6.26	6.49	7.31	RF	8.60	11.87	11.16	10.74	8.14	7.66	8.08
SVM	6.33	14.80	13.32	12.27	10.37	7.11	7.82	NN	8.43	14.61	12.51	12.84	7.37	7.19	10.07
Panel B: Prediction of II values calculated by VA formula (91 predictions; range: 19.4)															
MR	<u>16.20</u>	18.14	17.69	18.52	17.96	17.43	17.97	RF	16.70	17.73	16.38	16.92	17.00	16.39	16.58
SVM	<u>16.20</u>	17.78	17.37	19.29	16.85	17.23	18.70	NN	18.16	18.90	16.27	19.17	19.43	20.04	19.48
Panel C: Prediction of II positive values (92 predictions; range: 37.9)															
MR	<u>8.31</u>	14.44	12.69	13.45	8.84	8.48	9.25	RF	11.12	14.41	12.80	13.14	10.83	10.12	10.77
SVM	8.76	20.15	14.54	14.24	11.44	8.68	10.51	NN	8.73	15.21	14.42	14.83	9.78	9.94	11.73
Panel D: Prediction of II negative values (92 predictions; range: 21.8)															
MR	<u>4.70</u>	11.98	11.93	12.05	4.88	5.53	5.82	RF	6.17	11.11	11.06	10.74	6.12	6.07	6.43
SVM	6.38	12.91	12.25	11.47	9.13	5.47	7.79	NN	6.07	11.43	10.90	11.51	6.39	6.97	8.38

the predictive content of sentiment and posting volume on stock market variables such as returns or trading volume.

In this work, we performed a robust evaluation of information content of microblogging data to the prediction of valuable stock market variables. A very recent and large Twitter dataset was collected and adopted, with around 31 million tweets from December 2012 to October 2015, related with around 3,800 stocks traded in US markets. The Twitter sentiment indicators were extracted by considering a recent lexicon specifically adjusted to financial microblogging data [54]. [This is the first paper of this research topic that uses sentiment indicators created by applying specialized financial microblogging lexicons.](#) Moreover, it uses a much larger data period than the majority of studies using Twitter data to predict stock market behavior. Furthermore, we created a novel daily sentiment indicator from the combination of a daily Twitter indicator, weekly American Association of Individual Investors (AAII) and Investors Intelligence (II) values and monthly University of Michigan Surveys of Consumers (UMSC) and Sentix values. Existing sentiment indicators (e.g., survey, social media) have distinct characteristics and may contain different informative value.

Table 10: Prediction of weekly AAI and II values using KF sentiment indicators. For each survey sentiment indicator, the baseline model is underlined and NMAE values lower than baseline are in **bold** (\* – p-value < 10%, \*\* – p-value < 5% and \*\*\* – p-value < 1%, NMAE values in %).

Mtd	MSv1	MSv2	MSv3	MSv4	MSv5	MSv6	MSv7	Mtd	MSv1	MSv2	MSv3	MSv4	MSv5	MSv6	MSv7
Panel A: Prediction of AAI values calculated by BI formula (93 predictions; range: 58.21)															
MR	14.35	18.96	17.77	19.1	15.22	15.69	17.73	RF	14.61	17.34	18.73	17.68	14.36	15.84	15.27
SVM	<u>14.3</u>	18.12	16.36	17.69	14.58	14.88	15.32	NN	16.11	20.77	17.11	19.14	17.11	17.38	19.64
Panel B: Prediction of AAI values calculated by VA formula (92 predictions; range: 25.2)															
MR	<u>18.74</u>	20.48	20.82	23.76	20.55	20.86	23.55	RF	19.11	19.67	20.75	20.59	18.95	20.2	19.87
SVM	18.97	20.93	21.79	26.9	19.27	21.25	20.07	NN	19.76	21.2	21.12	24.34	23.24	19.61	21.89
Panel C: Prediction of AAI positive values (93 predictions; range: 37.89)															
MR	<u>12.52</u>	20.47	19.94	21.87	13.6	14.16	15.94	RF	13.52	19.41	20.35	20.14	12.8	14.06	13.83
SVM	14.54	18.79	21	20.15	13.42	15.43	14.38	NN	13.41	23.29	20.46	25.15	15.23	15.86	15
Panel D: Prediction of AAI negative values (93 predictions; range: 25.65)															
MR	<u>16.42</u>	18.63	16.76	17.67	18.11	<b>16.39</b>	18.13	RF	16.64	17.34	17.04	16.45	16.55	16.48	<b>16.33</b>
SVM	16.77	18.88	<b>16.13</b>	16.56	17.85	18.41	16.59	NN	18.05	18.94	16.86	19.99	20.05	20.21	20.17
Panel A: Prediction of II values calculated by BI formula (92 predictions; range: 55.8)															
MR	<u>5.85</u>	11.74	11.83	11.44	6.12	6.44	6.81	RF	8.6	10.95	10.55	10.13	8.06	7.71	7.88
SVM	6.33	14.05	11.56	10.88	8.61	6.55	7.02	NN	8.43	13.09	11.62	12.41	6.81	7.28	7.93
Panel B: Prediction of II values calculated by VA formula (91 predictions; range: 19.4)															
MR	16.2	<b>15.70</b>	16.72	17.19	<b>15.99</b>	17.44	17.99	RF	16.7	<b>15.67</b>	16.77	<b>16.18</b>	<b>15.75</b>	16.55	16.27
SVM	<u>16.2</u>	16.36	16.57	<b>15.43</b>	16.58	<b>15.74</b>	<b>16.00</b>	NN	18.16	17.36	19.4	16.46	16.73	17.4	17.87
Panel C: Prediction of II positive values (92 predictions; range: 37.9)															
MR	<u>8.31</u>	14.54	13.39	13.41	8.65	8.71	9.36	RF	11.12	13.85	12.99	12.45	11.2	10.34	10.7
SVM	8.76	16.19	13.39	12.78	9.77	8.96	9.02	NN	8.73	16.13	14.6	13.81	10.08	9.85	11.21
Panel D: Prediction of II negative values (92 predictions; range: 21.8)															
MR	<u>4.7</u>	10.81	10.95	10.58	5.09	5.21	5.45	RF	6.17	10.16	9.59	9.41	6.12	6.03	6.18
SVM	6.38	11.33	10.91	10.47	5.74	7.31	7.99	NN	6.07	13.07	9.22	10.6	5.9	6.12	6.62

Kalman Filter (KF) procedure allows the production of a daily sentiment indicator by combining various sentiment measures of diverse frequencies. The resulting indicator may be more representative of general investor sentiment. The predictive content of this KF indicator was compared with the content of Twitter indicator. We also explore different sentiment aggregation formulas, such as bullish ratio, variation and agreement. We predict daily returns, trading volume and volatility of diverse indices, such as Standard & Poor’s 500 (SP500), Russell 2000 (RSL), Dow Jones Industrial Average (DJIA) and Nasdaq 100 (NDQ), and portfolios (e.g., formed on size and industries). A fixed-sized rolling window of 300 training days was applied to predict the next day, allowing to perform a large number of model trainings and predictions (ranging from 392 to 439). Also, we explored four different regression methods: Multiple Regression (MR), Neural Network (NN), Support Vector Machine (SVM) and Random Forest (RF). To analyze the predictive value of microblogging features, we considered the Normalized Mean Absolute Error (NMAE) and applied the Diebold-Mariano (DM) test between the most accurate AR(5) model (baseline) and similar models using microblogging extracted variables. [In our opinion, this is a more robust evaluation scheme](#)

when compared with most state of the art works.

Additionally, some studies have analyzed the influence of sentiment on portfolios formed on diverse characteristics (e.g., market capitalization, book-to-market ratio). Many of these works refer that the effect of sentiment is more evident on returns of some portfolios having extreme values (e.g., small market capitalization [6, 8, 43]). However, most of these studies apply low frequencies (e.g., monthly, annual) and do not use sentiment indicators extracted from social media. Therefore, we analyzed in this paper the predictive content of daily sentiment indicators extracted from Twitter on returns of portfolios formed on size and industries.

Considering that AAI and II are widely used by academics (e.g., [44, 69, 28, 10, 11, 40, 41, 37, 78, 79]) and practitioners, we also created a forecasting exercise for these sentiment measures using Twitter and KF sentiment indicators. To the best of our knowledge, this is the first study that addresses the forecasting of such survey sentiment indicators.

We found that microblogging sentiment and attention indicators were particularly useful for the prediction of returns of S&P 500 index, portfolios of lower market capitalization and some sectors such as High Technology, Energy and Telecommunications. In these situations, there are models obtaining p-value less than 5% in the pairwise DM test with the baseline model. The application of microblogging features were less convincing for the forecasting of trading volume and volatility. These results add some evidence about the predictive content of social media sentiment and posting volume for returns (e.g., [60, 70, 9]). Additionally, microblogging sentiment indicators have various advantages when compared to traditional sentiment measures (e.g., surveys). For instance, their creation is faster and cheaper, allows greater frequencies (e.g., daily) and may be targeted to a more limited group of stocks (e.g., individual stocks or indices). The obtained results also corroborate previous findings that sentiment has more impact on smaller stocks (e.g., [6, 8, 43]). We note that daily microblogging features were used in this paper while the majority of previous studies apply monthly sentiment indicators extracted from economic variables or surveys.

The utilization of sentiment indicators produced by Kalman Filter procedure were informative for the prediction of returns of some portfolios and indices. There are models using KF indicators having p-value less than 5% in the pairwise test for SP500, Lo20, HiTec and Enrgy. Also, the application of KF indicators decreased NMAE values for diverse indices and portfolios. However, KF indicators were less effective for the forecasting of trading volume and volatility.

Twitter sentiment values were specially informative for negative values of AAI. In this case, there were seven models applying Twitter indicators producing lower NMAE results than the most



accurate AR(5) model. One of these models is significantly more accurate than baseline according to the pairwise DM test, and four of them apply exclusively Twitter indicators. KF indicators were particularly important for the prediction of II values calculated by VA formula. In this situation, there were eight models more accurate than the baseline model. These results show that Twitter and KF indicators can be valuable to forecast AAI and II values. Therefore, they may permit a satisfactory anticipation of these sentiment indicators or an acceptable alternative whenever they are unavailable.

We obtained results suggesting that mining social media may provide valuable actionable intelligence to investors. For instance, intelligent systems may support investors in their decisions by providing instant access to social media analytics such as customized sentiment indicators or predictions.

In future, we intend to identify different types of microblog users and communities and assess their contribution to the forecasting of specific stocks. For instance, social influence analysis has been applied for diverse purposes (e.g., advertisement [13, 33]) but has been scarcely explored for the stock market topic. However, the identification of influent social media users may allow the creation of sentiment indicators of informed users and anticipate the overall sentiment of investors. Moreover, there are diverse communities of users focused on specific stocks. The analysis of their information content for the prediction of those stocks variables may permit the identification of relevant communities. Additionally, most studies apply only one source of Web data so the complementarity value of different data sources remains unclear. These sources have distinct characteristics that can be complementary and enable better predictions. For instance, blogs have more complete opinionated content, microblogging contents have greater objectivity, interactivity and posting frequencies and Google searches represent a superior number of users. The dynamic combination of diverse Web data sources may result in more informative financial indicators.

## Acknowledgments

This work was supported by FCT - Fundação para a Ciência e Tecnologia within the Project Scope UID/CEC/00319/2013.

## References

- [1] Yakov Amihud and Haim Mendelson. Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17(2):223–249, 1986.

- [2] Werner Antweiler and Murray Z. Frank. Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, 59(3):1259–1294, 2004.
- [3] Nelson Areal and Stephen J Taylor. The realized volatility of ftse-100 futures prices. *Journal of Futures Markets*, 22(7):627–648, 2002.
- [4] J Scott Armstrong. Evaluating forecasting methods. In *Principles of forecasting*, pages 443–472. Springer, 2001.
- [5] J Scott Armstrong and Fred Collopy. Error measures for generalizing about forecasting methods: Empirical comparisons. *International journal of forecasting*, 8(1):69–80, 1992.
- [6] Malcolm Baker and Jeffrey Wurgler. Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4):1645–1680, 2006.
- [7] Malcolm Baker and Jeffrey Wurgler. Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, 21(2):129–151, 2007.
- [8] Malcolm Baker, Jeffrey Wurgler, and Yu Yuan. Global, local, and contagious investor sentiment. *Journal of Financial Economics*, 104(2):272–287, May 2012.
- [9] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.
- [10] Gregory W. Brown and Michael T. Cliff. Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, 11(1):1–27, 2004.
- [11] Gregory W. Brown and Michael T. Cliff. *Investor Sentiment and Asset Valuation*, 2005.
- [12] CL Philip Chen and Chun-Yang Zhang. Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, 275:314–347, 2014.
- [13] Min Chen, Shiwen Mao, and Yunhao Liu. Big data: a survey. *Mobile Networks and Applications*, 19(2):171–209, 2014.
- [14] Ray Chen and Marius Lazer. Sentiment analysis of twitter feeds for the prediction of stock market movement. *Stanford Education*, 25, 2013.
- [15] Tsan-Ming Choi, Hing Kai Chan, and Xiaohang Yue. Recent development in big data analytics for business operations and risk management. 2016.

- [16] San-Lin Chung, Chi-Hsiou Hung, and Chung-Ying Yeh. When does investor sentiment predict stock returns? *Journal of Empirical Finance*, 19(2):217–240, 2012.
- [17] Pilar Corredor, Elena Ferrer, and Rafael Santamaria. Investor sentiment effect in stock markets: Stock characteristics or country-specific factors? *International Review of Economics & Finance*, 27:572–591, 2013.
- [18] P. Cortez. Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool. In P. Perner, editor, *Advances in Data Mining – Applications and Theoretical Aspects, 10th Industrial Conference on Data Mining*, pages 572–583, Berlin, Germany, July 2010. LNAI 6171, Springer.
- [19] Paulo Cortez and Mark J. Embrechts. Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, 225:1–17, March 2013.
- [20] Sven F Crone, Michele Hibon, and Konstantinos Nikolopoulos. Advances in forecasting with neural networks? empirical evidence from the nn3 competition on time series prediction. *International Journal of Forecasting*, 27(3):635–660, 2011.
- [21] S R Das and M Y Chen. Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science*, 53(9):1375–1388, 2007.
- [22] Sanjiv Das, Asís Martínez-Jerez, and Peter Tufano. einformation: A clinical study of investor discussion and sentiment. *Financial Management*, 34(3):103–137, 2005.
- [23] Gene D’Avolio. The market for borrowing stock. *Journal of Financial Economics*, 66(2-3):271–306, 2002.
- [24] Shangkun Deng, Takashi Mitsubuchi, Kei Shioda, Tatsuro Shimada, and Akito Sakurai. Combining technical analysis with sentiment analysis for stock price prediction. In *Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on*, pages 800–807. IEEE, 2011.
- [25] Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 2012.
- [26] Eugene F Fama. Market efficiency, long-term returns, and behavioral finance. *Journal of financial economics*, 49(3):283–306, 1998.

- [27] Weiguo Fan and Michael D Gordon. The power of social media analytics. *Communications of the ACM*, 57(6):74–81, 2014.
- [28] Kenneth L. Fisher and Meir Statman. Investor Sentiment and Stock Returns. *Financial Analysts Journal*, 56(2):16–23, 2000.
- [29] Diego Garcia. Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300, 2013.
- [30] Tomer Geva and Jacob Zahavi. Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news. *Decision support systems*, 57:212–223, 2014.
- [31] Ken Goldberg, Theresa Roeder, Dhruv Gupta, and Chris Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2):133–151, 2001.
- [32] Axel Groß-Klußmann and Nikolaus Hautsch. When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, 18(2):321–340, 2011.
- [33] Pritam Gundecha and Huan Liu. Mining social media: a brief introduction. *Tutorials in Operations Research*, 1(4):1–17, 2012.
- [34] Jonghyun Han and Hyunju Lee. Characterizing the interests of social media users: Refinement of a topic model for incorporating heterogeneous media. *Information Sciences*, 358:112–128, 2016.
- [35] Trevor Hastie, Robert Tibshirani, Jerome Friedman, and James Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.
- [36] Jördis Hengelbrock, Erik Theissen, and Christian Westheide. Market response to investor sentiment. *Journal of Business Finance & Accounting*, 40(7-8):901–917, 2013.
- [37] Chienwei Ho and Chi-Hsiou Hung. Investor sentiment as conditioning information in asset pricing. *Journal of Banking & Finance*, 33(5):892–903, May 2009.
- [38] Paul Hribar and John McInnis. Investor sentiment and analysts’ earnings forecast errors. *Management Science*, 58(2):293–307, 2012.

- [39] Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.
- [40] Alexander Kurov. Investor sentiment, trading behavior and informational efficiency in index futures markets. *Financial Review*, 43(1):107–127, 2008.
- [41] Alexander Kurov. Investor sentiment and the stock market’s reaction to monetary policy. *Journal of Banking & Finance*, 34(1):139–149, January 2010.
- [42] Doug Laney. 3d data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6:70, 2001.
- [43] Charles Lee, Andrei Shleifer, and Richard Thaler. Investor sentiment and the closed-end fund puzzle. *The Journal of Finance*, 46(1):75–109, 1991.
- [44] Wayne Y. Lee, Christine X. Jiang, and Daniel C. Indro. Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking and Finance*, 26(12):2277–2299, 2002.
- [45] Michael Lemmon and Evgenia Portniaguina. Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies*, 19(4):1499–1529, 2006.
- [46] Qing Li, TieJun Wang, Ping Li, Ling Liu, Qixu Gong, and Yuanzhu Chen. The effect of news and public mood on stock movements. *Information Sciences*, 278:826–840, 2014.
- [47] J. Bradford De Long, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4):703, 1990.
- [48] Masoud Makrehchi, Shalin Shah, and Wenhui Liao. Stock prediction using event-based sentiment analysis. In *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on*, volume 1, pages 337–342. IEEE, 2013.
- [49] Huina Mao, Scott Counts, and Johan Bollen. Predicting financial markets: Comparing survey, news, twitter and search engine data. *arXiv preprint arXiv:1112.1051*, 2011.
- [50] Andrew McAfee, Erik Brynjolfsson, Thomas H Davenport, DJ Patil, and Dominic Barton. Big data. *The management revolution. Harvard Bus Rev*, 90(10):61–67, 2012.
- [51] Robert Neal and Simon M Wheatley. Do Measures of Investor Sentiment Predict Returns? *Journal of Financial and Quantitative Analysis*, 33(04):523–547, 1998.

- [52] Chong Oh and Olivia R Liu Sheng. Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement. In *ICIS 2011 Proceedings*, Shanghai, China, 2011. AIS.
- [53] Nuno Oliveira, Paulo Cortez, and Nelson Areal. On the predictability of stock market behavior using stocktwits sentiment and posting volume. In *Progress in Artificial Intelligence*, volume 8154 of *Lecture Notes in Computer Science*, pages 355–365. Springer Berlin Heidelberg, 2013.
- [54] Nuno Oliveira, Paulo Cortez, and Nelson Areal. Stock market sentiment lexicon acquisition using microblogging data and statistical measures. *Decision Support Systems*, 85:62–73, 2016.
- [55] David Omand, Jamie Bartlett, and Carl Miller. Introducing social media intelligence (socmint). *Intelligence and National Security*, 27(6):801–823, 2012.
- [56] Giovanni Petris et al. An r package for dynamic linear models. *Journal of Statistical Software*, 36(12):1–16, 2010.
- [57] José C Pinheiro and Douglas M Bates. Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6(3):289–296, 1996.
- [58] Lily Qiu and Ivo Welch. Investor Sentiment Measures. *Seminar*, No. 10794:1–39, 2006.
- [59] Philip Russom et al. Big data analytics. *TDWI Best Practices Report, Fourth Quarter*, pages 1–35, 2011.
- [60] Sanjiv Sabherwal, Salil K Sarkar, and Ying Zhang. Do internet stock message boards influence trading? evidence from heavily discussed stocks with no fundamental news. *Journal of Business Finance & Accounting*, 38(9-10):1209–1237, 2011.
- [61] Maik Schmeling. Institutional and individual sentiment: Smart money and noise trader risk? *International Journal of Forecasting*, 23(1):127–145, 2007.
- [62] Maik Schmeling. Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance*, 16(3):394–408, June 2009.
- [63] Robert P. Schumaker, Yulei Zhang, Chun-Neng Huang, and Hsinchun Chen. Evaluating sentiment in financial news articles. *Decision Support Systems*, 53(3):458–464, June 2012.

- [64] Catherine T Shalen. Volume, volatility, and the dispersion of beliefs. *Review of Financial Studies*, 6(2):405–434, 1993.
- [65] Robert J Shiller. From efficient markets theory to behavioral finance. *The Journal of Economic Perspectives*, 17(1):83–104, 2003.
- [66] Andrei Shleifer and Robert W Vishny. The limits of arbitrage. *The Journal of Finance*, 52(1):35–55, 1997.
- [67] Jasmina Smailović, Miha Grčar, Nada Lavrač, and Martin Žnidaršič. Predictive sentiment analysis of tweets: A stock market application. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*, pages 77–88. Springer, 2013.
- [68] Jasmina Smailović, Miha Grčar, Nada Lavrač, and Martin Žnidaršič. Stream-based active learning for sentiment analysis in the financial domain. *Information Sciences*, 285:181–203, 2014.
- [69] Michael E. Solt and Meir Statman. How Useful is the Sentiment Index? *Financial Analysts Journal*, 44(5):45–55, 1988.
- [70] Timm O. Sprenger, Andranik Tumasjan, Philipp G. Sandner, and Isabell M. Welpe. Tweets and trades: the information content of stock microblogs. *European Financial Management*, 20(5):926–957, 2014.
- [71] Robert F. Stambaugh, Jianfeng Yu, and Yu Yuan. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2):288–302, May 2012.
- [72] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 807–816. ACM, 2009.
- [73] Leonard J Tashman. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4):437–450, 2000.
- [74] Paul C Tetlock. Giving Content to Investor Sentiment : The Role of Media in the Stock Market. *Journal of Finance*, 62(3):1139–1168, 2007.
- [75] Allan Timmermann. Elusive return predictability. *International Journal of Forecasting*, 24(1):1–18, 2008.

- [76] Kristina Toutanova, Dan Klein, Christopher D Manning, and Yoram Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology NAACL 03*, 1(June):173–180, 2003.
- [77] Robert Tumarkin and Robert F Whitelaw. News or noise? internet postings and stock prices. *Financial Analysts Journal*, 57(3):41–51, 2001.
- [78] Rahul Verma and Gökçe Soydemir. The impact of individual and institutional investor sentiment on the market price of risk. *The Quarterly Review of Economics and Finance*, 49(3):1129–1145, 2009.
- [79] Rahul Verma and Priti Verma. Are survey forecasts of individual and institutional investor sentiments rational? *International Review of Financial Analysis*, 17(5):1139–1155, 2008.
- [80] Jeffrey Wurgler and Ekaterina Zhuravskaya. Does Arbitrage Flatten Demand Curves for Stocks? *The Journal of Business*, 75(4):583–608, 2002.
- [81] Peter D Wysocki. Cheap talk on the web: The determinants of postings on stock message boards. *University of Michigan Business School Working Paper*, (98025), 1998.
- [82] Jianfeng Yu and Yu Yuan. Investor sentiment and the mean-variance relation. *Journal of Financial Economics*, 100(2):367–381, 2011.
- [83] Yang Yu, Wenjing Duan, and Qing Cao. The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, 55(4):919–926, November 2013.
- [84] Xue Zhang, Hauke Fuehres, and Peter A Gloor. Predicting stock market indicators through twitter “i hope it is not as bad as i fear”. *Procedia-Social and Behavioral Sciences*, 26:55–62, 2011.
- [85] Xiaolian Zheng and Ben M Chen. Identification of stock market forces in the system adaptation framework. *Information Sciences*, 265:105–122, 2014.
- [86] Mohamed Zouaoui, Genevieve Nouyrigat, and Francisca Beer. How does investor sentiment affect stock market crises? evidence from panel data. *Financial Review*, 46(4):723–747, 2011.