Investigating the Impact of Media Sentiment and Investor Attention on Financial Markets

Michael Siering

Goethe University Frankfurt, Grüneburgplatz 1,
60323 Frankfurt, Germany
siering@wiwi.uni-frankfurt.de

Abstract. Media sentiment has been shown to be related to stock returns. However, one prerequisite for this influence has not been taken into account yet: the question of whether investors actually pay attention to news and the related financial instruments. Within this study, we close this research gap by examining the interplay between media sentiment and investor attention. Thereby, we find that the positive impact of media sentiment on returns is increased when investor attention is high. Furthermore, we evaluate whether these variables can be used to forecast future market movements. Although our results reveal that the obtained forecasting accuracy cannot be achieved by chance, we conclude that further information has to be included in the forecasting model to obtain satisfying results.

Keywords: Media Sentiment, Investor Attention, Behavioral Finance.

1 Introduction

Within recent years, the impact of media sentiment on financial markets has been of great interest. Recent studies have found that sentiment expressed in traditional mainstream media like newspapers or social media such as blogs and message boards is related to stock returns [1, 2]. In addition, the question of whether investors are aware of financial news has also been investigated in order to explain anomalies like the slow incorporation of new information into stock prices on Fridays [3] or the processing of news related to economically linked firms [4]. Furthermore, the impact of an increased number of irrelevant news on the processing of relevant ones has been examined as well [5].

However, to the best of our knowledge, previous research has neglected the interplay between investor attention and media sentiment as well as the following influence on financial markets. Nevertheless, it seems intuitive to take the interaction between both variables into account: the prerequisite for investors being influenced by media sentiment is that investors actually notice the news articles published within the media. Consequently, it can be assumed that a higher level of investor attention leads to an increased media sentiment impact on financial markets.

Furthermore, the instantaneous influence of media sentiment on financial markets still remains underexplored: recent studies investigate the impact of media sentiment

F.A. Rabhi and P. Gomber (Eds.): FinanceCom 2012, LNBIP 135, pp. 3–19, 2013.

with at least a one day delay [2, 6]. In contrast, we show that many news articles are published during the day. Since new information contained within these news articles is usually processed within short periods of time [7], it also seems reasonable to investigate the instantaneous impact of media sentiment. Finally, recent studies also focus on forecasting future market movements by means of machine learning techniques in order to cover non-linear relationships [8]. In this context, the integration of media sentiment and investor attention has also not been analyzed yet.

In this paper, we contribute to the literature in several ways. At first, we investigate the instantaneous impact of media sentiment and investor attention on stock returns. Second, we examine the interplay of media sentiment as well as investor attention and its impact on the following stock market reactions. Third, we forecast future returns taking into account those variables.

Based on an analysis of the media sentiment related to the Dow Jones Industrial Average (DJIA), we confirm the impact of media sentiment on DJIA returns and enhance the previous understanding related to the fact that media sentiment already has an influence on financial markets at the same day when the corresponding news are published. Additionally, we find that positive media sentiment has an increased influence on DJIA returns when investor attention is high. Consequently, when investors are actually interested in a stock and are aware of the related news articles, the media sentiment impact on financial markets rises. Related to forecasting future DJIA returns by means of machine learning techniques, we conclude that the accuracy that can be obtained is higher compared to results that are achieved by chance. However, additional information has to be included within the forecasting model to achieve satisfying results.

The remainder of this paper is structured as follows. At first, we present related work concerning the influence of media sentiment and investor attention on financial markets. Thereafter, we outline the data used within our study, derive our sentiment measure and describe the proxy for quantifying investor attention. Next, we consider the joint impact of media sentiment and investor attention on DJIA returns. Thereafter, we evaluate whether future market movements can be forecasted by taking into account these variables. Finally, we discuss the results and conclude this paper.

2 Related Work

Within previous research, a large number of studies have already considered the impact of media sentiment and investor attention on financial markets separately. In this context, the following section provides an overview about the corresponding research streams. Furthermore, the main results are presented in order to provide a proper grounding for our study.

2.1 Media Sentiment and Financial Markets

Generally, investors decide to trade because of new fundamental information like dividend announcements or management decisions [9]. Besides, they may also rely on expectations that do not follow rational rules [10]. For instance, these expectations can be based on the advice of "financial gurus" [11] or simply on the sentiment prevailing in the media that causes them to be overconfident in making the right decisions [12]. In this context, sentiment expressed in the media covers opinions, expectations or beliefs of market participants towards certain companies or towards certain financial instruments [13]. If many investors take media sentiment into account, have similar (irrational) expectations and follow each other, this can influence stock prices [14]. Recent research has provided evidence for these assumptions. It has been shown that sentiment expressed in media has an impact on investors' decision-making activities and thus affects several financial variables. Consequently, investors act according to their expectations and buy or sell the respective financial instrument.

Different studies investigate the impact of sentiment expressed in traditional media like newspapers. Within this context, [6] analyzes a daily Wall Street Journal column and finds that high media pessimism leads to a decline in market prices. Additionally, an abnormal high or low level of pessimism is supposed to be related to high trading volumes. A similar study is conducted in [2]. On a daily basis, the authors analyze the news stories published in the Wall Street Journal as well as in the Dow Jones News Service and confirm that stock prices react to media sentiment. In contrast, [15] evaluates the sentiment expressed in corporate disclosures (i.e. 10-K company reports) and finds that, compared to general approaches, domain-specific sentiment measures are more appropriate for sentiment detection within this document type. Furthermore, the authors confirm the relation of sentiment and several financial variables.

Compared to these studies, another stream of research focuses on the impact of sentiment expressed in social media. A seminal study that investigates the impact of sentiment on a stock level is presented in [1], which collects and analyzes messages posted on two finance message boards. The authors find that trading volume increases when a disagreement in sentiment among the messages prevails. Additionally, they observe that the number of messages posted during a day can help to predict the stock returns during the following day. [16] follows a similar approach and investigates messages which are published on stock message boards. However, it focuses on an index rather than stock level. Thereby, the sentiment is determined for every message whereas these messages are then used to calculate an overall sentiment index. [16] finds that this sentiment index has explanatory power for the level of the corresponding stock index. In contrast to this result, it only provides weak evidence that the sentiment concerning an individual stock can forecast daily stock price movements. Apart from these results, recent studies have also found a link between the sentiment prevailing in microblogging services like twitter and financial markets [17].

The studies presented above provide evidence that sentiment expressed in news articles or message board postings is related to stock returns. However, these studies mainly focus on the long-term effect of media sentiment on financial variables. Instead of taking into account instantaneous effects, stock returns are related to the previous days' media sentiment. Nevertheless, due to the fact that new information is often processed within minutes rather than days [7] it seems possible that the related sentiment has a instantaneous effect as well.

2.2 Investor Attention and Financial Markets

In recent years, several financial market anomalies have been investigated theoretically and empirically, such as underreaction and overreaction to financial news [5], the influence of weekdays on investors' reactions [3] as well as the impact of advertisements on investors' decisions [18]. Many of these anomalies have been attributed to the level of investor attention, i.e. the question whether investors are aware of the current market situation or not. Thus, there is a large number of studies investigating which instruments receive attention, how corporate advertising impacts the level of investor attention and how investors pay attention to news published by firms or by the media in general.

Concerning the question of which financial instruments are of interest to different groups of investors, [19] examines how individual and institutional investors react to "attention-grabbing" stocks. Thereby, the authors find that individual investors especially pay attention to stocks which are discussed within the media, exhibit high abnormal trading volumes and high returns.

Next to the question of which financial instruments gain attention in general, another stream of research investigates how stock recommendations published within the media influence investor attention. Within this context, it has been found that trading volumes increase after a stock has been discussed on television [20]. Additionally, it has been figured out that a firm's advertising expenses lead to an increased number of individual investors buying a stock [18]. In this case, a spillover effect of advertisements can be measured: although a firm advertises its product and intends to increase the product related attention, there is also a higher interest related to its stocks. These results are confirmed in [21]. Furthermore, [21] finds that advertisements lead to an increase in stock returns in the contemporary year, but they also note that stock returns decrease in the following year. Considering the company size, it is found that this effect is larger for small firms [21].

Next to stock recommendations and advertisements, investor attention is also influenced by ordinary news published within the media. In this context, [4] analyzes the market reactions on news of economically linked firms. In this context, news are incorporated slower when they are not directly related to the firm under investigation but deal with an economically linked firm. This effect is attributed to a small degree of investor attention. Additionally, a study by [5] investigates whether the amount of news articles published within the same period of time has an impact on market reactions. [5] finds that an increased amount of unimportant news decreases the investors' reactions to relevant news. Thereby, price and volume reactions are lower and the post-announcement adjustment to the news is stronger. Thus, as a result, an increased number of news articles published in the same period is said to reduce investor attention towards specific news items. Considering investor attention on different days of the week, [3] finds that the response to earnings announcements is slower on Fridays compared to the remaining days of the week. This effect is attributed to a lower level of investor attention on Fridays, whereas investors are said to be more distracted because of the following weekend. Similar results concerning limited investor attention on Fridays are also reported in [22]. Additionally, [23] provides evidence that the level of investor attention is related to stock returns.

Based on these studies, it can be noted that the level of investor attention has an influence on the market reactions following the publication of financial news. In this context, investors who are aware of the news articles published are also confronted with the corresponding sentiment. As a consequence, it is more likely that their expectations and trading decisions are influenced by media sentiment. Thus, a joint effect of both variables on financial markets can be expected. However, previous studies have not focused on this specific relation of media sentiment and investor attention.

3 Research Methodology

Before measuring the impact of media sentiment and investor attention on DJIA returns, we give an overview on the research methodology applied within this study. Therefore, as described in the next paragraphs, an unsupervised dictionary-based sentiment analysis approach has been used. Furthermore, the Google Search Volume Index (SVI) has been taken into account as a measure for investor attention. Finally, several financial news articles have been acquired from Dow Jones Newswires in order to be able to determine the corresponding media sentiment index.

3.1 Measuring Media Sentiment

In general, sentiment analysis encompasses the investigation of documents like news articles, message board postings or product reviews in order to determine their tone concerning a certain topic [24, 25]. There are two broad strategies to perform sentiment analysis: it can be distinguished between supervised and unsupervised approaches [26]. Supervised approaches require a dataset composed of documents that are manually labeled according to the respective sentiment. After several pre-processing steps, this dataset is used to train machine learning classifiers like naive bayes or support vector machines. During the training phase, the classifiers search for patterns within the documents. These patterns can thereafter be used to determine the sentiment of further documents or sentences. In contrast, unsupervised approaches rely on external knowledge such as predefined dictionaries providing lists of words that are connected with a positive or negative sentiment. These word lists are usually created manually with a couple of precoded terms and are applied to determine a sentiment measure [27].

Within our study, we decide to follow an unsupervised dictionary-based approach which determines the sentiment taking into account a dictionary containing sentiment bearing words [27]. This is appropriate because dictionary-based approaches have proven to be very promising within the financial domain [2, 6, 15]. In contrast, applying a supervised machine learning-based approach would require a manually labeled dataset for training whereas manual labeling would be time-consuming and error-prone.

For unsupervised approaches, different dictionaries are available that contain positive and negative expressions. Within this study, we make use of the Harvard-IV-4 dictionary. This dictionary has often been applied in the financial context [2, 6]. Since we analyze general financial news articles rather than specific corporate disclosures, we make use of this dictionary instead of using the specific dictionary proposed in [15] which was suggested for the analysis of corporate disclosures.

To calculate a daily sentiment index, we first determine the sentiment of each document. Accordingly, we obtain the occurrences of positive and negative words by comparing each news article with the positive and negative word lists. To take negations into account, we follow [15] and reverse the interpretation of a word if it is preceded by a negation so that positive words are counted as negative and vice versa. Thereafter, we adapt a document-level sentiment polarity measure which determines the direction of the sentiment (i.e. ranging from negative to positive) as well as its strength [2, 28]:

$$sent_{doc} = \frac{pos - neg}{pos + neg} \tag{1}$$

The measure defined in equation 1 takes into account the number of positive words *pos* as well as the number of negative words *neg,* calculated as described above. If a document contains neither positive nor negative words, $sent_{doc}$ is defined as zero. In line with [2], this measure assumes that all positive and negative words are equally important, i.e. no weights are assigned to certain words.

As a next step, we determine a daily sentiment index by aggregating the document-level sentiment on a daily basis. Therefore, we calculate the average of $sent_{doc}$. In the following, the resulting daily sentiment index *sent* that takes into account the sentiment related to the DJIA is used to investigate the research questions of our study.

3.2 Measuring Investor Attention

Within previous studies, different approaches for measuring investor attention have been proposed. In general, we can distinguish between indirect and direct measures of investor attention. On the one hand, a large body of literature deals with indirect measures of investor attention. Exemplary proxies used are unusual trading volumes or returns as well as the number of news articles published per day [4, 5, 19, 21]. In this case, it is assumed that large trading volumes or extreme returns indicate that investors are extremely aware of a stock and respond more timely to new information, i.e. they trade this stock. As follows, these measures can be denoted as ex post measures of investor attention. In contrast, the number of news articles per day can be seen as an ex ante measure of investor attention: an increased amount of news articles is assumed to lead to an increased amount of investor attention related to the corresponding financial instruments [19]. However, an increase in these indirect measures only expresses the results of investors buying or selling a stock (ex post measures) or a general increase in media attention (ex ante measures). In contrast, these indirect measures do not indicate whether investors are interested in a financial instrument or whether the news articles in the media are actually noticed by them at all [19].

In consequence, [23] represents a seminal paper about the direct measurement of investor attention, i.e. the measurement of investor attention without relying on trade-based proxies or the number of news articles published. Instead, they propose to take into account the amount of web searches related to the company under investigation, assuming that investors being interested in a financial instrument also search for related information. In this case, [23] makes use of Google's search volume index. Thereby, the authors find that this measure is correlated with indirect proxies for investor attention but that it encompasses investor attention in a more timely way. Furthermore, they note that this measure is especially suited to cover retail investor attention [23]. Since studies from other domains have already proven the applicability of the amount of search queries to forecast housing sales, car sales or the outbreak of influenza (e.g. [29]), we also decide to use SVI as a direct measure of investor attention.

The SVI can be obtained via Google Trends for different search terms and for different time horizons (beginning from January 2004). However, SVI is only displayed for search terms that received attention exceeding a certain (unknown) threshold. Thus, identifying the correct search term to cover investor attention can be seen as a crucial step. Within our study, we decide to take the SVI related to the search term "DJIA" into account to represent investor attention related to the DJIA. An alternative would have been to download the SVI for each constituent of the DJIA separately. However, in this case, several problems would arise: As already noted by [23], some company names are ambiguous (e.g. searching for "Kraft" which can also represent the German word for power). Using ticker symbols instead could be an alternative, however, there are also some pitfalls in this case. At first, SVI is not available for every ticker symbol and second, some ticker symbols are ambiguous, too. For example, searching for "T" as ticker symbol for AT&T also leads to results related to T-Mobile, "HD" for Home Depot could also be interpreted as a search for the technical abbreviation "high definition" (as in HD-DVD) and the same applies to "BA" (Bank of America) which can also be an abbreviation for British Airways. In these cases, the SVI would not cover the corresponding ticker symbol and would be inappropriate to measure investor attention. Thus, we decide to use the SVI for "DJIA" as a proxy for DJIA investor attention. In contrast to [23], which makes use of the weekly SVI, we take into account the daily SVI in order to measure instantaneous effects.

3.3 Dataset Acquisition

Within this study, we consider three data sources. First, we acquire financial news articles in order to determine the media sentiment index. Therefore, we make use of news articles published by Dow Jones Newswires (DJNS). Second, we download the SVI related to the DJIA from Google Trends. Third, we acquire the corresponding DJIA closing prices and trading volumes from Yahoo! Finance.

The news articles by DJNS are accessed via the application programming interface provided by Interactive Data. Thereby, we search for all news articles that are tagged by DJNS to deal with the constituents of the DJIA. We see DJNS as a representative source for financial news since DJNS is a major news provider that publishes financial news throughout the day and whose news are accessed by a large audience [6]. As revealed by a manual review of the news articles at hand, the assigned labels are too broad: news articles are already tagged to be related to a certain company when they mostly deal with its competitors. Thus, we only include those news articles into our analysis that contain the corresponding search term within the headline. Due to licence terms, we were able to request all news articles from 2011/01/01 until 2012/02/29, so that 292 trading days could be analyzed. In total, the news article dataset obtained for this study consists of 13,696 news articles. Thereby, the dataset covers different news categories. First, 6,454 regular financial news articles are included. Second, 7,176 news articles are included that explicitly deal with corporate disclosures and press releases. Finally, there are 66 news articles included in the dataset that contain analyst opinions. Thus, our news article dataset covers the full spectrum of articles that is usually available within a regular financial newspaper.

As already discussed above, some ticker symbols and company names are ambiguous. As a result, the SVI cannot be acquired with an adequate accuracy for each ticker symbol separately. Thus, we decided to acquire the daily SVI for the search term “DJIA” from Google Trends to measure investor attention. In this context, the SVI can be downloaded relative to the beginning of the corresponding month or relative to the beginning of the year 2004. Since the first option does not allow us to compare the SVI across several months, we have chosen to download the SVI for our sample period relative to the search volume in 2004. Finally, we acquire the DJIA prices and trading volumes for the sample period from Yahoo! Finance.

4 Empirical Results

The following section provides the descriptive results and the results of our explanatory analysis. Thereby, we first focus on the daily and hourly distributions of the news articles published as well as the SVI time series in general. Furthermore, we provide the regression estimates related to our analysis.

4.1 Descriptive Results

At first, we consider the daily number of news articles related to the DJIA and its constituents published by Dow Jones Newswires. Taking into account the daily distribution as indicated in Fig. 1, it can be noted first that the number of news articles published per day is not constant over time. Instead, a much smaller amount of news articles is published on weekends as compared to the rest of the week.

Second, it can be observed that the number of news articles published from Monday to Thursday is relatively constant, except from a peak on Tuesday. It is notable that the number of news articles on Friday is smaller than during the remaining trade days. This result may be attributed to the fact that the general number of financial news issued by firms is smaller on Fridays, as already reported in [3], [30] as well as [31]. Thereby, these related studies find that on Fridays, a lower fraction of earnings announcements is published. As follows, the number of news articles published dealing with these events is smaller, too. Another explanation could be that next to investor inattention, also journalists are distracted on Fridays because of the following weekend.

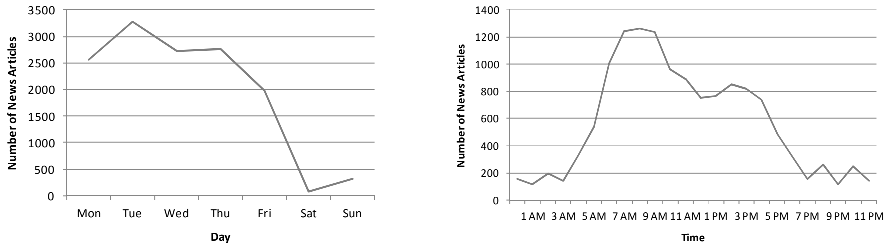

Fig. 1. Total Number of News Articles Published per Day / per Hour

Next, considering the time of day when news articles are published, it can be noted that a large peak can be found in the morning at the start of the trading hours (all times reported in Eastern Standard Time) and a small peak can be found at the end of the trading hours. Since the news articles published by Dow Jones Newswires are delivered electronically, a lot of information (and related sentiment) is released during the day after traditional newspapers have been printed in the morning.

Furthermore, we consider the amount of Google searches for the Keyword “DJIA”. In this case, we find abnormal high search volumes in August 2011 (see Fig. 2). These high levels occur simultaneously with a decline of the DJIA caused by weak economic perspectives. Thus, we control for this abnormal movement within our further analyses. Additionally, it is shown that SVI is low on weekends. Thus, investors are distracted on Saturdays and Sundays [3].

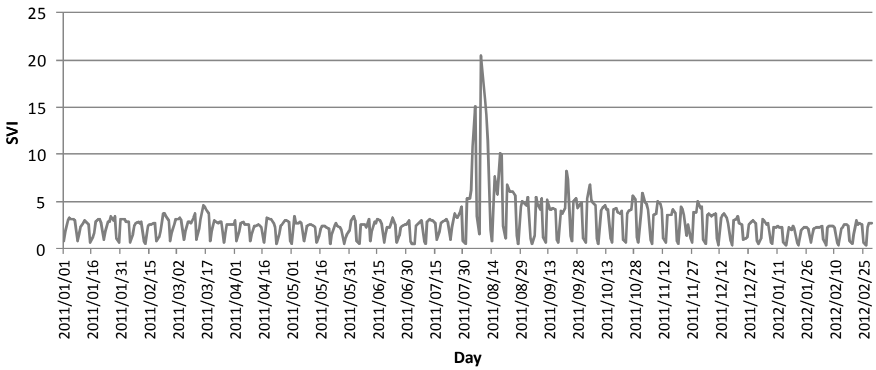

Fig. 2. Daily raw SVI Time Series

Finally, Table 1 presents the means, standard deviations and correlations of the main variables of interest. Since stock returns are measured as a percentage change and are only available for trading days, we do not include the raw measures of dow_t, $sent_t$ and SVI_t. Instead, we calculate the percentage change of sentiment and investor attention on trading days, report these values in the following table and include them in our further analyses. Non-trading days are excluded. As Table 1 shows, we find small positive correlations between the variables, except for SVI. In this case, SVI is negatively correlated with DJIA returns. Furthermore, we find a small negative correlation between SVI and sent. However, this correlation is not statistically significant ($p = 0.84$). Finally, the average value of the sentiment index is above zero which leads to the conclusion that in general, the news articles that are the basis for the sentiment index have a slightly positive sentiment.

Table 1. Means, Standard Deviations and Correlations

	Mean	Std. Dev.	dow	sent	sent x SVI	SVI
dow	0.00046	0.00073	1.0000			
sent	0.04286	0.02078	0.0754	1.0000		
sent x SVI	-0.00017	0.00290	0.0657	0.0066	1.0000	
SVI	0.01162	0.00935	-0.3205***	-0.0118	0.1380**	1.0000

* / ** / *** = significant at a 10% / 5% / 1% level of significance

4.2 Impact of Media Sentiment and Investor Attention on Stock Returns

To investigate the impact of media sentiment and investor attention on DJIA returns, we regress the DJIA returns (dow_t) on our daily sentiment measure ($sent_t$), investor attention (SVI_t) as well as the moderating effect taking into account both variables ($sent_t\, x\, SVI_t$):

$$dow_t = \beta_1 + \beta_2 \mathrm{sent}_t + \beta_3(sent_t \times SVI_t) + \beta_4 SVI_t + \beta_5 Controls_t + \varepsilon_t \tag{2}$$

Within Equation 2, ε_t denotes the error term. Additionally, $Controls_t$ stands for several control variables that are also included to ensure that the results are not biased because of further effects possibly influencing stock returns. To control for day patterns of stock returns and for the January effect that can cause abnormal stock returns, we include dummy variables for the different trading days as well as for January [6]. Furthermore, to account for the developments in August 2011, we also include a dummy variable for this month. Additionally, we incorporate variables for past volatility[1], previous trading volume[2] as well as previous DJIA returns up to five lags [6]. Within the regression, we use heteroscedasticity- and autocorrelation-consistent standard errors [33]. The results of the regression are denoted in Table 2. In order to test for multicollinearity, the variance inflation factor was calculated for each independent variable. Thereby, no multicollinearity was detected since the highest score of 2.17 is below common thresholds of 4 and 10 [34].

---

[1] Thereby, the approach proposed in [6] is followed: to account for past volatility, the daily returns of the DJIA are demeaned to obtain a residual, this residual is squared and the past 60-day moving average is subtracted.

[2] Specifically, the detrended logvolume is used as proposed in [32].

At first, the results confirm the impact of media sentiment on stock prices. As indicated by a positive coefficient for the sentiment measure, we can note that an increase in media sentiment leads to an increase in the corresponding DJIA return. Thereby, the coefficient is significant at a 5% level of significance.

Considering the joint impact of investor attention and media sentiment on DJIA returns, we also find a positive relationship which is significant at a 10% level of significance. In this context, the positive effect of media sentiment on DJIA returns is increased when investor attention is high. Although an increased SVI does not imply that investors actually read the news articles published via DJNS, it can be noted that an augmented interest in the corresponding topic (i.e. the DJIA) prevails. Since the news articles at hand are published on several websites as well, it is more likely that the news articles are actually read by investors' searching for information via Google. As follows, more investors are confronted with a certain level of media sentiment and consequently, their trading decisions are influenced.

Interestingly, the sole impact of investor attention on DJIA returns is negative, whereas the coefficient is significant at a 1% level of significance. At first sight, this result contradicts previous research. In this context, [23] finds a positive relationship of investor attention (measured by the amount of Google searches) and stock returns. However, [23] shows that, when controlling for market capitalization, the positive price pressure is only present among the smaller half of their stock sample. Furthermore, in their study, an interaction effect of market capitalization and SVI has a negative impact on returns [23]. Since the constituents of the DJIA have a high market capitalization and are analyzed on an aggregated level, these results do not necessarily contradict previous studies. Considering the control variables, the results remain robust when including a dummy variable for August 2011. Thereby, the day-of-week dummy variables as well as the dummy variable for January have no significant influence, whereas few of the lagged control variables for previous returns, volatility and trading volumes have a significant influence (not reported in Table 2 due to space constraints).

Table 2. Impact of Media Sentiment and Investor Attention on DJIA Returns

	Coefficient	Standard Error
const	0.0007622	(0.0017192)
sent	0.0033198**	(0.0015118)
sent x SVI	0.0323227*	(0.0187595)
SVI	-0.0284493***	(0.0068986)

* / ** / *** = significant at a 10% / 5% / 1% level of significance. Controls included.

5 Predicting Bidirectional Market Movements Based on Media Sentiment and Investor Attention

In order to examine the economic impact of the relation between media sentiment and investor attention and to cover possible non-linear relationships, we focus on a bidirectional forecast of market movements based on machine learning techniques. Therefore, we present the general setup of our approach to evaluate the predictive value of the variables under consideration.

5.1 General Setup

In the following, we investigate whether the influence of media sentiment and investor attention on DJIA returns can be taken into account to forecast future market movements. Thereby, we focus on predicting DJIA returns by means of machine learning techniques. In this case, machine learning techniques are advantageous because of two main reasons. At first, the evaluation becomes more reliable since evaluation methodologies like 10-fold cross validation can be used [35]. In this respect, 10-fold cross validation offers the possibility to use an increased number of data items for evaluating the trained model. Second, machine learning classifiers like Support Vector Machine (SVM) are also suitable to cover non-linear relationships within the data which may improve forecasting results [36].

For predicting DJIA returns, we make use of supervised learning and train a machine learning classifier with labeled (historical) training data in order to find patterns within the data that can serve for future predictions related to new, (unlabeled) datasets. Thus, every observation of the training dataset (i.e. each trading day) is labeled according to the corresponding daily DJIA return. Thereby, we assign two labels: the first label is assigned according to the instantaneous DJIA return (T_0), the second label is related to the one day ahead return (T_{0+1}). We follow previous studies and focus on two classes [37]: the class *negative* is assigned if the corresponding DJIA return is lower than zero, otherwise, the class *positive* is assigned. In total, for T_0 and T_1 forecasts, 161 observations are labeled as *positive* and 131 observations are assigned to the class *negative*.

Within this study, we train a Support Vector Machine (SVM) classifier since SVMs have proven to be a good choice for financial forecasting [38]. Thereby, the same input variables are used that were already defined in section 4.2, i.e. media sentiment, investor attention, the interaction term as well as the control variables.

5.2 Evaluation

To evaluate the proposed machine learning setup, we make use of k-fold stratified cross validation (k=10) [35]. In this case, the whole dataset is split into 10 subsets with equal class distributions, whereas nine subsets are used for classifier training and one subset is used for classifier testing. In total, this procedure is repeated ten times so that each subset is used 9 times for classifier training and once for classifier testing. At the end of each iteration, a contingency table is created that contains the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Finally, a global contingency table is created by summing up the different contingency tables (micro-averaging) [39]. Based on this global contingency table, different performance metrics are calculated. Thereby, we focus on *Accuracy*, indicating the percentage of cases classified correctly as well as *Precision*, *Recall* and the F_1-measure [40, 41]. These metrics are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3) \qquad Recall = \frac{TP}{TP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5) \qquad F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (6)$$

The results of our evaluation for the instantaneous as well as the T_{0+1} forecasts are depicted in Table 3. Thereby, the classification results of SVM classifiers depend on the choice of a kernel function as well as on the choice of several parameters. In order to choose a proper configuration, we followed the procedure proposed in [42]: a SVM classifier using a radial basis function kernel has been used and the corresponding parameters have been selected via grid search[3].

Table 2. Forecasting DJIA Market Movements Using Media Sentiment and Investor Attention

		Class: positive			Class: negative		
Forecast	Accuracy	Precision	Recall	F_1	Precision	Recall	F_1
T_0	58.55	58.40	86.34	69.67	59.26	24.43	34.60
T_{0+1}	58.56	59.71	76.40	67.03	55.81	36.64	44.24

All values are given as percentages.

In relation to the predictability of DJIA market movements based on media sentiment and investor attention, it can be observed that the obtained results are better than results being achieved just by chance. This is evidenced by the fact that the precision scores are above 50% in all cases. Additionally, instantaneous market movements as well as the returns of the following day can be predicted with similar accuracy of 58.55% and 58.56% respectively. Considering the class recall, we find that the recall for the class *positive* is higher than the recall for the class *negative*, which can be attributed to the class distribution within our sample: 55% of all observations are labeled as positive and, as follows, the SVM is trained respectively.

However, taking the economic value of these results into account, an accuracy of below 60% cannot be considered as promising. Thus, using only these structured variables as input data for a decision model can hardly be seen as a source for significant profit. On the contrary, many cases are classified incorrectly. This may be attributed to the noisy nature of financial markets and to the fact that the decision model does not take into account the textual information published within the news articles under investigation. As a consequence, media sentiment and investor attention should not be used solely to forecast market movements. Instead, they should be incorporated in existing forecasting models to improve forecasting results.

---

[3] We followed [42] and evaluated the proposed values for C, a penalty parameter and γ, a parameter of the radial basis function. For $T_{o+1,}$ $C = 512$ and $\gamma = 2^{-15}$ lead to the best results. In the case of T_0, $C= 32$ and $\gamma = 2^{-15}$ were selected.

6 Discussion

Based on an empirical analysis of the sentiment expressed within 13,696 financial news articles, we find that higher investor attention increases the impact of media sentiment on DJIA returns. Thus, when investors actually pay attention to a financial instrument and search for related information, the impact of media sentiment on these financial variables is higher. Furthermore, this effect is already measured at the same day rather than with a delay of several days. As follows, media sentiment is processed at least within the trading day. If the variables under investigation are used to forecast DJIA returns by means of a machine learning approach, it can be observed that the results obtained are higher than results being achieved just by chance. However, there are still many cases which are classified incorrectly. Consequently, further input variables have to be incorporated within the decision model in order to improve forecasting results. For example, textual inputs or technical indicators may also be considered to incorporate the information published as well as current market trends [43, 44].

Within our study, media sentiment and investor attention are measured on a daily basis. As a consequence, our study does not cover intraday effects of media sentiment and investor attention on stock returns. However, new information is often processed and reflected within stock prices within short periods of time [7], which could also lead to the effect that the related sentiment is processed accordingly. To take this into account, the intraday stock price impact of media sentiment may be measured by considering financial news articles and relating them to the corresponding tick-by-tick trading data. However, since SVI is only available on a daily basis, we are aware of the limitation that actually, intraday effects of both variables cannot be measured adequately. Thus, these effects are not covered and should be investigated as soon as an intraday SVI is available.

Additionally, this study focuses on an index perspective, i.e. media sentiment and investor attention related to the DJIA are covered. This has the advantage that a sufficient amount of financial news articles is published each day so that a corresponding sentiment index can be determined. However, due to the focus on an index level, specific stock characteristics are not taken into account. In this context, a focus on single companies may also result in several methodological issues. In this case, investor attention measured by company names may also depend on other factors apart from a general interest in the financial instrument. For instance, investors may also be interested in a firm's products which could not necessarily be related to the firm's stock returns. Furthermore, the SVI used within this study covers the attention of investors searching for information within Google. Thus, the index does not cover search queries of users using different search engines or directly accessing web sites like stock message boards. However, this approach of measuring investor attention has proven to be related to other measures of investor attention [23]. Additionally, the attention paid related to stock message boards can hardly be measured since corresponding statistics tracking user activity are not publicly available.

From a methodological perspective, an unsupervised sentiment analysis approach has been used which has been found to be appropriate in previous studies [2, 6, 15].

However, sophisticated language constructs such as irony are not covered by such an approach relying on dictionaries containing positive and negative terms. Although it may be debatable whether irony is present within newspaper articles at all, future research may focus on a more detailed analysis of media sentiment.

7 Conclusion

In recent years, the impact of media sentiment on financial variables like stock prices has been of great interest. However, one crucial prerequisite of relating media sentiment to financial variables has not been taken into account: current studies do not consider whether the news articles expressing sentiment are actually noticed by investors. As a consequence, we examine the interplay between media sentiment and investor attention in order to investigate the joint impact of both variables on Dow Jones Industrial Average returns. Thereby, we find a instantaneous impact of media sentiment on DJIA returns and a moderating role of investor attention. However, if these variables are used as input variables for a machine learning approach for bidirectional market forecasts, it can be found that further input variables should be included to obtain satisfactory results.

From a theoretical perspective, we contribute to the literature related to Behavioral Finance dealing with the media sentiment impact on financial markets as well as to the literature on investor attention. Thereby, we enhance the understanding that media sentiment has a instantaneous impact on stock returns by introducing the moderating effect of investor attention and media sentiment. From a practical perspective, we take this relationship into account in order to provide a model to forecast bidirectional market movements of the DJIA. This model may rather be used solely but could be combined with existing forecasting models in order to evaluate whether forecasting results can be improved.

Within further research, the interplay between media sentiment and investor attention as well as its impact on financial variables should also be examined at a stock level in order to investigate stock-specific effects. Thereby, it has to be taken care of stocks with low media coverage which hampers the determination of an appropriate sentiment index. Furthermore, since the effect of investor attention among small-capitalized stocks has been found to be higher [23], less frequently traded stocks should also be incorporated in a related analysis. Finally, contemporary research reports that traditional news media and social media are interconnected. In this case, topics that are discussed within newspapers are also talked about within blogs [45]. Thus, the discussions within social media could also be analyzed in order to develop a more fine-grained indicator for measuring investor attention at a topic level.

Acknowledgements. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) within the context of the Project FIRST, Large scale information extraction and integration infrastructure for supporting financial decision making, under grant agreement n. 257928.

References

1. Antweiler, W., Frank, M.Z.: Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. Journal of Finance 59, 1259–1294 (2004)
2. Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S.: More Than Words: Quantifying Language to Measure Firms' Fundamentals. Journal of Finance 63, 1437–1467 (2008)
3. Dellavigna, S., Pollet, J.M.: Investor Inattention and Friday Earnings Announcements. Journal of Finance 64, 709–749 (2009)
4. Cohen, L., Frazzini, A.: Economic Links and Predictable Returns. Journal of Finance 63, 1977–2011 (2008)
5. Hirshleifer, D., Lim, S.S., Teoh, S.H.: Driven to Distraction: Extraneous Events and Underreaction to Earnings News. Journal of Finance 64, 2289–2325 (2009)
6. Tetlock, P.C.: Giving Content to Investor Sentiment: The Role of Media in the Stock Market. Journal of Finance 62, 1139–1168 (2007)
7. Muntermann, J., Guettler, A.: Intraday stock price effects of ad hoc disclosures: the German case. Journal of International Financial Markets, Institutions and Money 17, 1–24 (2007)
8. Bollen, J., Huina, M.: Twitter Mood as a Stock Market Predictor. Computers and Operations Research 44, 91–94 (2011)
9. Shleifer, A., Summers, L.H.: The Noise Trader Approach to Finance. Journal of Economic Perspectives 4, 19–33 (1990)
10. Black, F.: Noise. Journal of Finance 41, 529–543 (1986)
11. Shleifer, A.: Inefficient markets. An introduction to behavioral finance. Oxford Univ. Press, Oxford (2000)
12. de Bondt, W.F.M.: A portrait of the individual investor. European Economic Review 42, 831–844 (1998)
13. Brown, G.W., Cliff, M.T.: Investor sentiment and the near-term stock market. Journal of Empirical Finance 11, 1–27 (2004)
14. Shiller, R.J.: Stock Prices and Social Dynamics. Brookings Papers on Economic Activity 1984, 457–510 (1984)
15. Loughran, T., McDonald, B.: When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. Journal of Finance 66, 35–65 (2011)
16. Das, S.R., Chen, M.Y.: Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. Management Science 53, 1375–1388 (2007)
17. Bollen, J., Mao, H., Zeng, X.-J.: Twitter mood predicts the stock market. Journal of Computational Science 2, 1–8 (2011)
18. Lou, D.: Attracting investor attention through advertising. Discussion Paper 644. Financial Markets Group, London School of Economics and Political Science, London (2009)
19. Barber, B.M., Odean, T.: All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. Review of Financial Studies 21, 785–818 (2008)
20. Busse, J.A., Green, T.C.: Market efficiency in real time. Journal of Financial Economics 65, 415–437 (2002)
21. Chemmanur, T.J., Yan, A.: Advertising, Attention, and Stock Returns. Working Paper (2009), `http://ssrn.com/abstract=1340605`
22. Louis, H., Sun, A.: Investor Inattention and the Market Reaction to Merger Announcements. Management Science 56, 1781–1793 (2010)
23. Da, Z., Engelberg, J., Gao, P.: In Search of Attention. Journal of Finance 66, 1461–1499 (2011)

24. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Philadelphia, Pennsylvania, pp. 79–86 (2002)
25. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval 2, 1–135 (2008)
26. Zhou, L., Chaovalit, P.: Ontology-supported polarity mining. Journal of the American Society for Information Science and Technology 59, 98–110 (2008)
27. Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment in Twitter events. Journal of the American Society for Information Science and Technology 62, 406–418 (2011)
28. Zhang, W., Skiena, S.: Trading Strategies To Exploit Blog and News Sentiment. In: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, Washington, DC (2010)
29. Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L.: Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014 (2009)
30. Penman, S.H.: The distribution of earnings news over time and seasonalities in aggregate stock returns. Journal of Financial Economics 18, 199–228 (1987)
31. Patell, J.M., Wolfson, M.A.: The intraday speed of adjustment of stock prices to earnings and dividend announcements. Journal of Financial Economics 13, 223–252 (1984)
32. Campbell, J.Y., Grossman, S.J., Wang, J.: Trading Volume and Serial Correlation in Stock Returns. Quarterly Journal of Economics 108, 905–939 (1993)
33. Newey, W.K., West, K.D.: A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55, 703–708 (1987)
34. O'Brien, R.M.: A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & Quantity 41, 673–690 (2007)
35. Mitchell, T.: Machine learning. McGraw-Hill, London (1997)
36. Dhar, V., Stein, R.: Intelligent decision support methods. The science of knowledge work. Prentice Hall, Upper Saddle River (1997)
37. Groth, S.S., Muntermann, J.: Supporting Investment Management Processes with Machine Learning Techniques. In: Proceedings of the 9th Internationale Tagung Wirtschaftsinformatik, Vienna, vol. 2, pp. 275–284 (2009)
38. Tay, F.E.H., Cao, L.: Application of support vector machines in financial time series forecasting. Omega 29, 309–317 (2001)
39. Chau, M., Chen, H.: A machine learning approach to web page filtering using content and structure analysis. Decision Support Systems 44, 482–494 (2008)
40. Hotho, A., Nürnberger, A., Paaß, G.: A Brief Survey of Text Mining. GLDV Journal for Computational Linguistics 20, 19–62 (2005)
41. van Rijsbergen, C.J.: Information retrieval. Butterworths, London (1979)
42. Hsu, C.W., Chang, C.C., Lin, C.J.: A practical guide to support vector classification. National Taiwan University (2003), `http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf`
43. Mittermayer, M.-A.: Forecasting Intraday Stock Price Trends with Text Mining Techniques. In: Proceedings of the 37th Hawaii International Conference on System Sciences, Big Island, Hawaii (2004)
44. Lam, M.: Neural network techniques for financial performance prediction: integrating fundamental and technical analysis. Data mining for financial decision making. Decision Support Systems 37, 567–581 (2004)
45. Leskovec, J., Backstrom, L., Kleinberg, J.: Meme-tracking and the Dynamics of the News Cycle. In: Proceedings of the 15th International Conference on Knowledge Discovery and Data Mining, Paris (2009)