The predictive characteristic of the social sentiment on the stock market: Twitter and the stock trend

Moreno Lo Giudice
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT

The study analysed the relationship between the stock return and the investors' sentiment. One month of daily return for four stocks have been computed, as well as the sentiment data for the same period. The sentiment data has been collected through Twitter, by recording every tweet posted on the period containing a cashtag with the ticker of the companies. Two measures have been implemented to categorise the sentiment. The data gathered shows a large increase of the usage of Twitter by investors, compared to findings of previous studies. The hypotheses have been tested using the vector autoregression model (VAR) and the causal relationship between the two variables is further evaluated with the Granger causality test. The findings reject the possibility of one direction of the causal relationship, thus the investors' sentiment fails to affect the return of the stocks, contrasting the findings of previous studies. However the results show a significant causal relationship in the opposite direction, thereby the stock returns seem to affect significantly the investors' sentiment. Furthermore, evidence from the VAR model suggests that some stocks do not follow the random walk of the returns, but there is a correlation between the stock returns in a short period. In addition, investors tend to believe in certain pattern of the stock returns, thus they themselves do not believe in the random walk. In conclusion, the sentiment data cannot be exploited to build a trading strategy aimed at forecasting the stock price, but more research on the topic has to be performed to further explain the relationship.

Supervisors:

Xiaohong Huang, Rezaul Kabir, Henry van Beusichem, Peter-Jan Engelen, Samy A.G. Essa, George Iatridis.

Keywords

Twitter, social sentiment, cashtag, stock price, forecasting, behavioural finance, sentiment analysis, efficient market

1. INTRODUCTION

The idea that the stock market is efficient holds to some extent in the real world. The fundamental theory of an efficient market advocates that the value of a security is based on future cash flows and the opportunity cost of capital. The average market value of a security on a long period is fairly close to its intrinsic value. Any shift in price is due to the disclosure of new information, and the shift is correlated to that information. However, most of the time investors tend to over- and under-react to new information, often generating abnormal return (Barberis, Shleifer, & Vishny, 1998). The abnormal return is defined as the difference between the actual stock return and the expected stock return; such difference can be negative or positive, reflecting an overvaluation or an undervaluation of the security. Market bubbles are evidence of such overvaluation and investors' overconfidence. The dot-com bubble, which saw the high peak in the year 2000, is one example. The investors felt extremely confident towards any internet stock; Yahoo! rose by 1400% in four years, companies increased their stock price just by adding *e-* as a prefix or *.com* at the end of the name. Needless to say that when the bubbles burst, many companies go bankrupt or in state of financial distress. Moreover, the world crisis of 2007 also may have the roots in a real estate bubble, when the ease of obtaining mortgages and the (wrong) assumption that the house prices would rise further, ended up in a world-wide economic collapse.

Many questions thus arise, one above all concerns the cause to such wide market fluctuations. It is argued that the steep declines in the stock market are corrections, whereby the market was overvalued, and the decrease brings it back to its fundamental value, or at least close to it. Therefore, in the long run the market may be in fact efficient. However, in the short term, the market tends to follow the irrationality of the investors. After all, the rule of supply and demand is the basis of the economy and the free market; indeed the financial markets follow that rule. If the investors believe that a sector, or a company, will definitely perform well, the high demand will bring the price to high levels, higher than the fundamental value. Such feelings and beliefs of the investors can be synthesised in two words: *social sentiment*.

Several studies investigated the relationship between the social sentiment and the stock trend, and most of them found a significant relationship (Bollen, Mao, & Zeng, 2011; Joseph, Babajide Wintoki, & Zhang, 2011; Zhang, Fuehres, & Gloor, 2012). Therefore, the overall market trend is correlated to the social sentiment and thus to the investors' expectations. The market price has been described as "an equilibrium price of expectations of investors about the value of the company" (Engelen & van Essen, 2011), therefore the correlation between the social sentiment and the stock trend might be expected. However, whether companies analysed individually present the same correlation is not yet explored thoroughly by the literature. If the market index is related to the social sentiment, and the market index is related to the companies listed on the same stock exchange, it is logic to assume that the same correlation exists between the company and the sentiment toward that company.

This study will try to enlighten the relationship, taking into account the current research on the broad correlation and narrow it to individual companies. Hence, the formulated research question is:

"To what extent the investors' sentiment affects companies' stock price?"

To further specify, the research is aimed at analysing the possible relationship between the investors' sentiment toward a company and the underlying security price.

The study will add value to the existing literature also from a practical point of view, providing insights to both investors and companies; if the correlation in fact exists, it will be possible to exploit it for innumerous applications like forecasting the market trend. Moreover, the study challenges the efficient market hypothesis, checking in the specific to what extent it holds on common stocks.

The paper is structured by first examining the literature, in order to learn from previous research and give insights to the reader on the topic. The methods follow in the next section, as well as the collection of the data and the measurement. Two measurements are implemented, to reduce the bias brought about by a fallacious categorization of the data. Both measurements are expected to yield similar results. After the exploration of the theory and the presentation of the methods, the core of the study shows the results with the subsequent analysis and discussion of the findings. In conclusion a brief summary outlines the most significant results, commenting the limitations of the study and including suggestions for future research.

2. LITERATURE REVIEW

The starting point of this study lies on the literature and on previous researches on the topic. Initially, studies presenting evidence of the efficient market are explored, in order to check to what extent the efficient market hypothesis holds in practice. Secondly, explanation of the factors which might cause the deviation of the financial market from the fundamental value are investigated and challenged. Furthermore, the social sentiment is considered to be one of the factors that influence the market prices, thus the notion is explored among the existent literature. The social sentiment is analysed both as a general mood which might affect the investing decisions, as well as a specific feeling and expectation toward a company, which in turn may drive the price of the underlying security. The literature indeed provides

insightful suggestion and evidence that must be taken into consideration.

2.1 Evidence of the efficient market hypothesis

The variety of antecedent papers that investigated the discrepancy between the efficient market hypothesis and the actual market value of assets foster the spread of studies towards factors that produce those discrepancies. The efficient market hypothesis states that the share price always incorporates all relevant information, and that the stocks trade at their fundamental value. According to the theory, there are no under- or over- valued stocks, therefore it is impossible to outperform the market.

However, the stock market does not always follow the theory, in facts Barberis et al. (1998) investigated underreaction and overreaction of stock prices. The authors advocated that stocks tend to incorporate new information in a period of up to 12 months, which means that stock prices do not reflect the fundamental value in the short term. Furthermore, companies that release good news in a consistent pattern present overvaluation in the stock price even in long term horizons, up to 5 years (Barberis et al., 1998). Barberis et al. (1998) introduce psychological factors as explanation for the price discrepancies, like the concept of conservativism, which is associated with "the slow updating of models in the face of new evidence", consistent with underreactions when good news arrive (Barberis et al., 1998). The authors suggest that investors do not believe in the random walk of companies' earnings; the authors describe the investors as Bayesian and thus a positive earnings surprise will be followed by another one. The empirical findings of Bernad & Thomas (1990) show that earnings actually present a slight correlation between themselves in a short term period such as one to three quarters horizons, therefore the random walk of the earnings is not completely accurate.

Bertone, Paeglis & Ravi (2015) looked into the deviation of stock prices from the intrinsic value, finding a significant gap between the two values. Moreover, the authors continue, the gap is especially large for short time intervals. The study covers a long period, from 1998 to 2010, and it shows that the deviations decrease overtime; the authors argue that such decreases of deviations from the intrinsic value are due to the increase of the speed of information availability. Therefore the authors state the markets are actually becoming more efficient (Bertone, Paeglis, & Ravi, 2015).

2.2 The general mood as a factor for market deviations

It is evident that the financial market is not as efficient as hypothesised by the theories; the prices do deviate from the intrinsic value of the asset. Furthermore the deviations seem to be larger in the short term. There is not a coherent explanation of the reason of such discrepancies, however the literature point the cause to the investors' beliefs and expectations.

The social sentiment reflects those thoughts and expectations of the investors. In facts, recently, the interest in studying the

social sentiment and the effects of non-rational investors increased, implying that the actions of non-rational investors are of key importance in the financial market (Wang, 2001). Before going deeper into the topic, it is important to make a distinction. The investor's sentiment, or social sentiment (both used interchangeably), concerns directly the stock market, while the mood is considered as a general state of mind. The twos are not strictly related, an investor may be in a bad mood for personal reasons and at the same time very positive regarding a stock. Thus, the relationship between the general mood and the stock price movements may not be that evident. However, studies found evidence affirming that relationship, thus the mood of the investors has been found to affect significantly the stock market (Edmans, GarcÍA, & Norli, 2007). Edmans et al. (2007) investigated sport events and how these influence the people's mood, which in turn is reflected on the stock market. The authors collected data of 30 years of sport events; they correlated each event with the day following the match. The findings show that international sporting events affect significantly the stock market; furthermore, in countries where football is the predominant sport, the loss of the national team negatively affect by a large extent the stock market on the day following the defeat (Edmans et al., 2007).

The study by Palomino, Renneboorg, & Zhang (2009) also presents similar findings. The authors investigated the British football and found that the market strongly reacts to match results. They analysed the listed British football teams, reporting that a win triggers a positive average abnormal return of 53 basis point on the next day, and 88 basis point in the next three days. Furthermore, the authors argue that the market is faster at processing good news than processing bad news (Palomino, Renneboog, & Zhang, 2009).

2.3 Social sentiment and stock market

People are thus affected by their mood, and therefore investors are biased in their decision by their beliefs and by their sentiment. Consciously or unconsciously our sentiment affects the decisions we make, and it is difficult to completely nullify such bias.

The sentiment is defined as "the expectations of market participants" (Brown & Cliff, 2004). It may take two forms, either positive or negative. The positive is the so called bullish sentiment, the negative one is the bearish sentiment; the former reflects the investors' expectations of an above average return, the latter the opposite outcome (Brown & Cliff, 2004). A third classification often adopted is a neutral sentiment.

Brown & Cliff (2004) gathered several surveys to asses and measure the investors' sentiment, which is correlated to the market returns. The results present indeed a correlation, but not the causality. The authors implement the Granger-causality test, which fails to reject the null hypothesis (Brown & Cliff, 2004). Therefore the sentiment possess a rather limited ability to predict the market, both in a short term and in a long term period. Brown & Cliff (2004) hypothesised that sentiment would affect more individual investors rather than institutional ones; however the authors found the strongest relationship is actually with institutional investors and large stocks.

One drawback of the study by Brown & Cliff (2004) is the method in which the data is gathered, the sample surveyed may

not be representative of all the investors, given the fact that it is mostly aimed at experienced investors operating at the stock exchanges.

Twitter can help in overcoming the issue of gathering data. In facts Twitter is a social network that urges people to share publically what is happening in a determined moment. It allows users to tweet a state of mind and their perception of what is around them. The tweets can accommodate a maximum of 140 characters, therefore users tend to be direct and straight to the point. Bollen et al. (2011) found a correlation between the tweets collected randomly in a period of time and the general mood. The authors used a period with public holidays in order to validate the hypothesis that the tweets can in fact be a proxy for the mood. Furthermore, in the same study, the authors look into the correlation of the public mood and the Dow Jones industrial average; more precisely the causality of the mood towards the index trend. The results show that the mood has some predictive power (Bollen et al., 2011).

Another study by Joseph et al. (2011) looked into web searches for a given ticker and how these may predict abnormal returns of the stock. The authors state that analysing the web searches may be time consuming and therefore require high transaction costs; they suggest that implementing the same method on Twitter may lower the costs significantly (Joseph et al., 2011). Zhang et al. (2012) analysed the social sentiment and the relationships between the three major American indices, two commodity prices, and dollar exchange rate. The authors argue that there is a causal relationship between social and market data, the only variable not presenting significant relationship is the one concerning the Dollar exchange rate. (Zhang et al., 2012).

The sentiment thus presents significative relationships with several other variables constituted by market data. It is also been noted that the relationship is often of causal nature. Oh & Sheng (2011) used Stocktwits.com[1] to perform the sentiment analysis. The authors collected more than 70000 postings, they classified the posts according to the related stock. They found that 70% of all postings concern just 10 companies, Apple leading the top discussed. The authors advocated that the microblogging posts have strong predictive value. The authors continue stating that the microblogging posts are to be considered as a discussion between investors, rather than mere noise created by irrational investors in order to speculate (Oh & Sheng, 2011).

Explanations and reasons of such predictive power might be several. Primarily, if the data from which the sentiment is extrapolated is considered as a discussion between investors, it is likely that those investors will tend to decide considering that discussion. Hence the sentiment reflected contains information about future actions of the investors.

Another argument is that investors, and thus their sentiment, assimilate new information faster than the market itself, therefore the postulation of a possible predicting power does not necessarily go against the efficient market hypothesis (Hengelbrock, Theissen, & Westheide, 2010). The investors

first have to digest the relevant information, which will be reflected on the market according to their actions. Therefore, if the tweets of the investors are a discussion aimed at digesting new information, the sentiment analysed via those tweets is assumed to have a positive relationship with the stock price.

In contrast, the sentiment may reflect a state of the market, whereby a large number of positive (bullish) sentiment might actually reflect an overvalued market, with stock prices above their fundamental value; thus a correction may be foreseeable. In this last context, the sentiment is expected to have a negative relationship with the stock price (Hengelbrock et al., 2010).

The study is centred on the effect the social sentiment induce on the trading operations undertaken by the investors. The hypothesis to be tested is formalised as:

"the social sentiment influences the stock price".

To further specify, the sentiment, which is created before the opening of the market, influences that day of trading. Neither positive nor negative relationship is assumed, furthermore the two variables, Sentiment and Stock price, are both tested as dependent variables, since the hypothesis that the stock price influences the sentiment must be taken in consideration. Further explanations follow in the next section.

3. DATA AND METHODOLOGY

This section outlines the methods and the models used to test the data. First the methodology is presented, with the model implemented for the study. The model is the vector autoregression, the hypotheses are further tested with the Granger causality test. Subsequently, the next section describes the collection and measurement of the data. The data consists of tweets gathered from the social platform Twitter.

3.1 Methods

The hypothesis that the tweets might cause the stock trend needs to be tested statistically. The hypothesised relationship is a causal one, whereby the tweets data cause the stock price to increase and/or decrease, according to the sentiment the data reflects. The null hypothesis is thus formulated:

H_0: the sentiment does not cause any price movement in the stock trend.

The model used is the vector autoregression model (VAR) in order to include the time variable. A linear regression model

---

[1] Stocktwits.com is a microblogging platform like Twitter. Stockwits focuses is business in postings about stocks and the stock market, thus the users are mostly investors. It has not

been used for this study because it is not yet very popular and the sample may not be representative enough.

would not be appropriate, because the observations are assumed to be independent between each other, and thus it does not implicate a time series. According to the efficient market hypothesis the returns are not correlated, thus the market data satisfies the independence condition, but the sentiment observations may not be independent. Although results from the linear regression may, or may not, support the hypothesis, the observations have to be considered with the time series and dependent on each other.

To tackle the issue of the dependency of the observations, which means that Sentiment at day T may be dependent to the Sentiment at day T-1, the vector autoregression (VAR) model is undertaken. The model implies that the variable is affected by the same variable taken one, or more, periods before. The number of lags chosen for the autoregression is found to be two. The reason is that the lagged 2 variable shows a statistical significance stronger than further variables, according to both Akaike Information Criterion (AIC) and Schwarz' Bayesian Information Critetion.(SBIC) criteria. The variables for VAR model are thus three for each of the original variables Sentiment and Return.

The Granger test is widely used in econometrics, it is used to investigate the causal relationship between two variables. The test states that a variable X Grange-causes a variable Y if the latter can be predicted by using the history of both variables (Granger, 1969). Therefore applied to this study, the sentiment variable predicts the variable return by using the lagged variables of both. It is well suited for the hypothesis of this study because it includes a time series variable, which means that the two variables to be tested are linked together by the third variable that is the time. Furthermore both variables are assumed as dependent, thereby the test is aimed at investigating the causal relationship in both directions.

The model is formulated:

(1) $Y = a_0 + a_1 Y_{lag1} + a_2 Y_{lag2} + b_1 X_{lag1} + b_2 X_{lag2}$

(2) $X = c_0 + c_1 X_{lag1} + c_2 X_{lag2} + d_1 Y_{lag1} + d_2 Y_{lag2}$

Where Y is the variable return, X is the sentiment variable, $Y_{lag(n)}$ and $X_{lag(n)}$ are the lagged n variables of the original X and Y. The coefficients of each variables are denominated *a*, *b*, *c*, and *d*, numbered according to the correspondent lagged variables. In the first equation the variable X Granger-causes the variable Y, the second equation delineates the Granger-causality in the opposite direction.

The results of the test have three possible outcomes:
- no statistical significance, whereby the null hypothesis that the sentiment does not influence the stock price cannot be rejected;
- statistical significance of a causal relationship between the sentiment and the stock price, whereby the sentiment influences the stock price;
- statistical significance of a causal relationship between the stock price and the sentiment, whereby

the daily return influences the sentiment of the investors.

The reasoning for the last point is that a stock performing well, thus increasing its value day by day, creates positive feelings, and therefore a positive sentiment on its investors.

The statistical significance of each coefficient is checked with the VAR output, where the values are calculated.

However, the significance of the coefficient does not mean that there in facts is a causal relationship. The Granger test is aimed at evaluating that relationship. The Granger causality test assesses the hypothesis that the conditional probability of Y given the lagged variables of X and Y is not equal to the conditional probability of Y given only the lagged variable(s) of Y:

$$E(Y|Y_{t-k}, X_{t-k}) \neq E(Y|Y_{t-k})$$

More specifically, the Granger-causality test evaluates the significance of the difference between the two variances; thereby assessing whether the difference between the variance of estimating the variable Y using Y_{t-k} and X_{t-k} compared to the variance of estimating the same variable with solely Y_{t-k} is significant. Therefore if the granger causality is significant, the model can be used to forecast the next period, with the coefficients *a, b, c,* and *d* of the VAR model.

For this study the statistical significance level considered is an alpha equal to, or less than, 10% ($\alpha \leq 0.1$), unless specified otherwise.

3.2 Collection and sentiment measurements

The data required to measure the sentiment is collected from Twitter. However historic data cannot be obtained free of charge since Twitter Inc. does not provide access to its database. Historic data may just be accessed by paying large amount of money. Gnip is the company, owned by Twitter Inc., which offers the service of providing tweets data, but requests start at 1000$ for 40 days of tweets[2]. To bypass the issue, the tweets have been recorded from one day onwards, due to the fact that tweets are shared publicly.

Therefore the data has been collected for the study through the Twitter API[3]. Using a script for Google sheet the tweets have been recorded directly to the spreadsheet. The triggers of the script has been set to record all the tweets every five minutes given a determined query. The query is aimed at recording only the relevant tweets for the study. Four datasets have been created corresponding at the four different companies investigated.

In contrast to previous studies, which analysed the overall mood through keywords (Bollen et al., 2011; Joseph et al., 2011; Zhang et al., 2012), this research focuses the interest in a few selected stocks. Instead of looking for selected keywords among all tweets, the data is gathered using cashtag[4] containing the ticker of the company. The use of cashtags rather than hashtags serves to reduce noise among the data, since the cashtag contains information specific for stocks. The

---

[2] A representative of Gnip has been contacted via email for the purpose, although the precise estimation for this study is still unknown, requests to obtain data from Gnip start at 1000$

[3] Application Programming Interface
[4] The cashtag compared to the hashtag differs in that the prefix is the dollar symbol $, instead of the hash sign # (e.g. $TWTR)

number of cashtags on Twitter is growing significantly, more than 65% increase in the same month one year apart (Hentschel & Alonso, 2014). The same study offers an overview of the most tweeted companies in the year 2013, as well as the sector the most tweeted stocks belong to; the technology sector is by far the most tweeted one, almost twice as much as the follower (Hentschel & Alonso, 2014).

The query thus is aimed at the most tweeted stocks, as to have enough data. All the tweets containing the following cashtags have been recorded: $AAPL, $FB, $GOOGL, and $MSFT. The companies are listed on the American stock exchanges.

The market data has been collected via Orbis, containing daily price series for the period of analysis, with the stock closing price of each day.

The number of tweets collected is in line with the results reported by Hentschel & Alonso (2014). The most tweeted cashtag remains $AAPL; the tweets for the cashtag $FB increased significantly, from ≈12000 in 2013 to ≈56000 in 2015. The data gathered for cashtags $GOOG and $MSFT increased in number, but $MSFT grew by a larger extent. The overall increase in tweets was expected as the use of Twitter becomes more popular. Although it is not completely clear the reason behind the massive growth of tweets for $FB. The reason might be that when Hentschel & Alonso (2014) collected the data in April 2013, Facebook was still fresh from the IPO, held in May 2012. Hence the Facebook stock was not traded and discussed as it is three years after the initial public offering. Microsoft Corporation, on the other hand, saw a revolution in the year 2014, as the historic CEO stepped down; perhaps the company came to the centre of the attention by the investors as the market evaluation of the new CEO performance is still undergoing.

The data constituted by text strings has to be converted in quantitative values corresponding to the sentiment. Prior studies have divided the social sentiment in different groups, therefore adding different variants to a positive or negative mood, such as *calm*, *alert*, *happy*, etc., and found causative correlation to some, but not all groups (Bollen et al., 2011). However, there is not a logical expectations on the reactions a *kind* or *alert* mood would produce that might alter the stock price. For this reason, this study classifies the sentiment as *positive*, *neutral*, or *negative*. Any new information available to the market is incorporated with the sentiment data; after an earnings release many tweets follow, reflecting a positive or negative sentiment.

Pak & Paroubek (2010) with their study concerning Twitter data for the sentiment analysis support the validity of the source for several reasons: the number of posts is large enough and it grows increasingly; many different people make use of it, thus it is a source of peoples' opinions, and the users are significantly heterogeneous (Pak & Paroubek, 2010).

The data collected through Twitter presents a high degree of noise, such as spam posts and words which do not possess any informative value. Hence, the first step is to clean and filter the database. All the links to websites are removed, thus every

word containing a URL link such as *http://* are excluded from the data. Furthermore tagged username presenting the tag prefix @ are also deleted from the data. The same filtering method is used by Pak & Paroubek (2010). In doing so, many spam posts are cleaned out from the database, since those tweets containing solely a URL do not fulfil the scope of this study.

The tweets, being of a maximum of 140 characters, are usually only a phrase. Therefore the informative part of the phrase can be extrapolated by just a few key words. The most accurate are found to be the bigrams, as the unigrams do not express the sentiment accurately, and the trigrams might cut out valid information from the database (Pak & Paroubek, 2010). Although the method described by Pak & Paroubek (2010) is deemed to be the most accurate, it is also extremely time consuming. Other methods involve the use of software designed for the task.

The developers of CX Data Science provided the Excel add-in Simply Sentiment[5] to analyse the tweets. The model classifies the sentiment with values from -5 to 5, reflecting a negative and a positive sentiment respectively. Values close to 0 are regarded as neutral. Furthermore, a second measurement has been implemented as well. A macro[6] for Excel scanning the textual data looking for keywords that possess informative characteristics for the sentiment. The two measurements differ in the fact that the first method contextualises the words, while the macro simply looks for keywords. Therefore, the first measurement is considered more accurate than the second one. Both measurements assign the same values to the sentiment variable, from -5 to 5. Those are the values used for the statistical procedure; the notions of negative, neutral, and positive sentiments are solely used for discussion sake.

In measurement 2 however, the value 0 has been scrapped out, because the method gives the value 0 in two cases, either the text in facts reflects a neutral sentiment, or when the software does not find any keywords. Hence, the 0 values must be taken out from the tests, otherwise the results may be biased by a large number of false neutrals.

The statistical testing is performed on both sentiment variables separately. Thereby the variable Sentiment1 is created using the results of Simply Sentiment; the variable Sentiment2 uses the results of the macro. It is assumed that the results of the statistical tests are not too far apart, since both measurements reflect the sentiment of the same text strings. The study is focused on a day to day basis, therefore the sentiment for each day has to be calculated. Thus, the average daily sentiment has been computed, for each day the tweets taken into consideration are the ones tweeted from one closing market day to the next one. Therefore to calculate the sentiment of day T, the tweets computed are those presenting the time stamp between T-1 16:00 to T 16:00[7].

The variables used for the hypothesis testing are four; two for the sentiment measurements, one for the price return, and one

---

[5] www.cxdatascience.com

[6] The macro in computer science is a sequence of commands to perform a procedure, like a rudimental software. In this case

the macro is designed in Excel to look for specific words, assign a value, and calculate the net score of the text string.

[7] The time stamp of the tweets has been synchronised with the American stock exchanges (GMT +5)

time variable. Thus, they are indicated with the following nomenclatures:

- Daily return = the return calculated with opening and closing price
- DailySentiment1 = the sentiment variable measured with the first measurement described previously
- DailySentiment2 = the sentiment variable measured with the second measurement
- Date = the time series variable

3.3 Sample

The sample consists of tweets and market data for a period of 30 trading days, collected between the months of March and May 2015. The total number of observations valid for the test is 28, due to the fact that the VAR model uses lag variables up to two days, therefore such computation for the first two observations is not feasible.

The sentiment data gathered with the first measurement and the second measurement is presented in Table 1. For the four companies Simply Sentiment recognised a total of 49619 tweets, which contain sentiment characteristics. The total numbers of the tweets recognised to have informative value for the second measurement is 75740. The number is almost twice as much as the total of measurement 1. The majority of tweets appears to reflect positive sentiment. Similarly, Zhang et al (2012) in the study on the predictive characteristic of Twitter, found close results. The authors categorize the tweets in *Hope*, *Fear*, and *Worry*, they found that *Hope* tweets were almost six times as much as the others (Zhang et al., 2012). The *Hope* tweets denominated by Zhang et al. (2012) are comparable to the positive tweets used in this study, since the data mining process is very similar. It appears the people tend to express an opinion when excited, or, as argued by Zhang et al. (2012), people tend to express negative opinion using positive words. According to the last argument, the data would be biased by a large number of false positive. For this reason this study implements two different measurements; with the measurement 1 the words are contextualised, hence the false positive bias is minimised.

Furthermore, Table 1 presents descriptive statistics of the sample, with average, median, and standard deviation values both for the sentiment and market data. The mean figured in Panel D of Table 1 is mostly positive, with the exception of the GOOGL case. The average of Panel C is also positive for every value of the different cases.

Table 1

Panel A	Positive	Neutral	Negative	Total
MSFT	3597	816	1391	5804
AAPL	14028	2959	6171	23158
GOOGL	4895	667	2119	7681
FB	7308	1832	3836	12976

Panel B	Positive	Neutral	Negative	Total
MSFT	7471	-	2479	9950
AAPL	26190	-	9183	35373
GOOGL	7313	-	2715	10028
FB	14744	-	5645	20389

Panel C	Mean		Median		St. Deviation	
Meas.	1st	2nd	1st	2nd	1st	2nd
MSFT	1.32	0.97	3.26	2	3.18	1.7
AAPL	1.15	1.08	3.26	1.5	3.3	1.79
GOOGL	1.17	1.3	3.26	2	3.36	2.14
FB	0.98	0.85	3.12	1	3.31	1.75

Panel D	Mean	Median	St. Deviation
MSFT	.0019	-.0004	.0233
AAPL	.0011	-.0017	.0147
GOOGL	-.0009	.0002	.0125
FB	.0002	-.0006	.013

Panel A shows the number of tweets with informative value categorised by the measurement 1;
Panel B presents the valid tweets categorised through the measurement 2;
Panel C and D outline descriptive statistics of the sentiment data and market data respectively, in Panel C Meas. stands for measurement, the numbers 1 and 2 indicate the two different measurements

4. ANALYSIS & DISCUSSION

After having described the data and the models used for the research, this section highlights the results of the statistical testing. The VAR model is considered part of the Granger test, since the results from the VAR are used to evaluate the Granger-causal relationship. There are two measurements for the investors' sentiment and they are discussed together case by case. The first part consists on the findings of tests, thus the coefficients are explained and analysed by case. The final part presents the discussion of the results, with the implications of the statistical findings.

4.1 Findings

To investigate the statistical significance of a possible causal relation between the variables Daily sentiment and Daily return, the vector autoregression (VAR) model and the Granger causality test have been performed. The Granger test evaluates the possibility of a causal relationship in both directions, therefore the two null hypotheses tested are: (1) *the daily sentiment does not cause the daily return* and (2) *the daily return does not cause the daily sentiment*.

The coefficients for the VAR model are presented in Table 2. Panel A and C represents the causal relationship of the first equation mentioned earlier between the variables DailySentiment (X) and the DailyReturn (Y). Therefore, according to the model previously formulated, the coefficients are those related to the lagged variables of the sentiment (X_{Lag1}, X_{Lag2}) and to the lagged variables of the return (Y_{Lag1}, Y_{Lag2}). The coefficients are thus denominated coherently with the model with the letter a and b. The panels B and D show the coefficients for the second equation and thus the inverse relationship, Y→X, whereby the return causes the sentiment. The coefficients are denominated according to the model d and c. The statistical significance of the coefficients is reported with the star symbol (*), as already mentioned the significance is at the 0.1 level. However, some of the coefficients is significant even at smaller levels, which is reported on the tables with two and three stars for .05 and .01 levels respectively.

Nevertheless, the significance of the individual coefficients does not imply the Granger causal relationship. To test that, the Granger causality test needs to be performed. Of course, if all the coefficients are not significant the Granger causality is likely rejected. The results of the VAR model are first presented in details case by case.

The AAPL case in the direction X → Y shows a significant correlation of the return coefficients b_1 and b_2, indicating a significant relation of the stock price returns one and two days before. Both coefficients in either measurements are negative and significative at the 5% level. The sentiment coefficients a_1 and a_2 are found to be not significative for both measurements; thus the sentiment shows little power in affecting the stock price return. In the opposite direction, Y → X, the sentiment coefficients d_1 and d_2 appear rather significant. Only the coefficient d_2 of the first measurement is not significant. All four coefficients show positive relationship with the sentiment variable. Overall it appears that the variables are correlated to their own lagged variables, as shown by the coefficients. Therefore the variables are better explained by using solely the same variable taken in the previous periods. In facts, the Granger causality is rejected for the AAPL case, in either directions (see Table 3).

The FB case presents no significance of any coefficients in affecting the stock return. Neither the lag return variables, nor the lag sentiment variables are significantly related to the stock price return. Therefore, the Granger causality X → Y is not confirmed for this case; thus the sentiment does not affect the return. However, in the opposite direction the coefficients appear significative. On the first measurement the return coefficient c_2 is negative and significant in affecting the sentiment, on the other hand c_1 is positive but not significant. With second measurement the same coefficients are both significant; similarly to the first measurement the polarity of the coefficients is positive for c_1 and negative for c_2. Furthermore, the d_2 coefficient for the second measurement appear significative too, thus implying the auto-correlation of the sentiment variable within different periods. The Granger causality of the return in affecting the sentiment, Y→X, is confirmed for the both measurements.

The GOOGL case does not present significance of any coefficient. The polarity of the coefficients is largely coherent between the measurements, indicating a reliability of the sentiment categorisation. The Granger causality thus is not confirmed for GOOGL, both in different directions and with different measurements.

Finally, the MSFT case presents little statistical significance in the direction X→Y, the coefficients a and b are not significant with both measurements. Thus the Granger causality of the sentiment in affecting the return is rejected. In the opposite direction the coefficient d_2 for the first measurement is significant at the 5% level, with negative sign. Thus the sentiment is auto-correlated with its lag 2 variable. On the second sentiment measurement one of the return coefficients appear significative at the 10% level. The sign is positive, indicating that the investors' sentiment is positively related to the stock return of two days before. The Granger causality Y→X is rejected for the first measurement, but is confirmed with the second one at the 5% level. The MSFT case is the only one presenting such discrepancy between the two sentiment measurements.

Overall, the results show a statistical significance mostly in rejecting the second null hypothesis, hence affirming that the daily return affects the daily sentiment. Both FB and MSFT reveal a significant causal relationship of the return in affecting the sentiment. The GOOGL and AAPL cases do not display any statistical significance, in either directions of the Granger-causality relationship.

Table 2 – VAR coefficients
Measurement 1

	X → Y			
Panel A	Sentiment coefficients		Return Coefficients	
	a_1	a_2	b_1	b_2
AAPL	.0005 (0.18)	.0028 (0.83)	-.5826*** (-3.66)	-.29** (-2.48)
FB	.0071 (1.38)	.0031 (0.60)	-.052 (-0.26)	-.0391 (-0.20)
GOOGL	-.0032 (-1.43)	-.0022 (-0.95)	-.2021 (-1.11)	.1076 (0.60)
MSFT	-.0034 (-0.58)	-.0054 (-0.94)	-.0478 (-0.26)	.216 (1.31)

	Y → X			
Panel B	Sentiment coefficients		Return Coefficients	
	d_1	d_2	c_1	c_2
AAPL	.4909** (2.35)	.3438 (1.50)	.12 (0.01)	-7.367 (-0.91)
FB	-.0187 (-0.10)	.016 (0.08)	11.791 (1.59)	-15.929** (-2.20)
GOOGL	.149 (0.80)	-.105 (-0.55)	3.126 (0.21)	-14.908 (-1.01)
MSFT	.1171 (0.67)	-.3581** (-2.13)	-1.948 (-0.36)	-2.9057 (-0.59)

Measurement 2

	X → Y			
Panel C	Sentiment coefficients		Return Coefficients	
	a_1	a_2	b_1	b_2
AAPL	-.0006 (-0.08)	.0058 (0.71)	-.5689*** (-3.56)	-.2746** (-2.27)
FB	-.0058 (-0.38)	.0181 (1.29)	.1101 (0.58)	.0056 (0.02)
GOOGL	-.0043 (-1.16)	-.0041 (-1.08)	-.1664 (-0.93)	.1129 (0.63)
MSFT	-.0196 (-1.19)	.0134 (0.97)	.0785 (0.41)	.167 (0.92)

	Y → X			
Panel D	Sentiment coefficients		Return Coefficients	
	d_1	d_2	c_1	c_2
AAPL	.5331*** (2.68)	.4185* (1.94)	4.299 (1.02)	.2073 (0.07)
FB	-.1919 (-1.06)	.383** (2.31)	8.3487*** (3.70)	-4.733* (-1.75)
GOOGL	.0561 (0.30)	-.0558 (-0.29)	-2.991 (-0.33)	-3.474 (-0.38)
MSFT	-.0468 (-0.26)	-.0744 (-0.48)	.9734 (0.46)	4.06* (2.02)

Panel A and Panel C list the coefficients of equation 1 of the VAR model, where the DailyReturn is the dependent variable, thus DailySentiment affects DailyReturn; Panel B and D list the coefficients for the second equation, where DailyReturn affects DailySentiment.

The subscript numbers represent the lag (e.g. a_1 = Lag_1); in brackets the t-statistics value.

4.2 Discussion

The findings highlight the statistical significance in some coefficients, but not in others. Furthermore, the Granger causality in the expected direction is not confirmed, but it is found in the opposite direction. In any case, the individual coefficients, where significant, need to be further discussed and analysed as they might provide insightful information.

First the expected relationship is discussed, where the sentiment was hypothesised to affect the stock return. The VAR coefficients show weak statistical significance in affecting the return. What is surprising is the fact that the lagged variables of the stock return of Apple seem to affect significantly the present return. To recall the VAR model presented previously, these coefficients are denominated b_1 and b_2. Moreover they appear to be negative, therefore the stock return of yesterday affects negatively the present stock return. Thus, such evidence suggests that the efficient market hypothesis does not hold firmly in practice. Similar findings have been presented by other studies (Bertone et al., 2015). Furthermore Bertone et al. (2015) stated that the deviations from the theory are pronounced in the short period. Likewise, this study analysed short periods, thus the findings confirm what previously discovered by Bertone et al. (2015).

The polarity of the coefficients, which is negative, suggests that the Apple stock price experiences some sort of mini-corrections. The investors think that the stock price is overvalued and they expect the stock price to decrease, up to the next two days of trading; the same is expected in the opposite condition, when investors feel the stock to be undervalued. The coefficients are stronger for one day lag than for two days lag variable.

In the four cases analysed, the coefficients a present no statistical significance in affecting the dependent variable. Furthermore, the Granger causality test suggests that the sentiment has little ability of affecting the return of the stocks. Hence, this study confirms the findings that the stock return is hardly forecasted by using investors' sentiment (Brown & Cliff, 2004).

In the opposite direction, the relationship appear stronger and more significative. Although the results are quite contrasting. With the first measurement (Table 2, Panel B) the return coefficient c_1 is mostly positive but no significant. On the other hand, the coefficient for two lags c_2 is negative for all the cases, but significant just for Facebook. In contrast the second measurement (Table 2, Panel D) shows many statistical significant coefficients.

Overall the results are not coherent, for some companies the return coefficients appear negatively related to the sentiment, for some others the opposite. The sentiment coefficients d_1 and d_2 for the lagged variables Lag 1 and Lag 2 respectively, present a rather strong relationship to the sentiment variable. The finding was expected, in fact the VAR model has been implemented due to that expectation. The Granger causality is confirmed for two cases, namely FB and MSFT. Therefore, with a significance level of 5%, it is safe to state that the stock return Granger-causes the investors' sentiment for the companies Facebook and Microsoft.

The findings are thus contrasting those found by other papers (Joseph et al., 2011; Oh & Sheng, 2011; Zhang et al., 2012), whereby the authors found the Granger causality in opposite direction and advocating the sentiment to have predictive power. Furthermore, it is also been argued that the tweets are to be considered a discussion between investors, and therefore the business activities follow rather than leads that discussion (Nofsinger, 2005; Oh & Sheng, 2011). In that context the sentiment was expected to have predictive power, but the evidence suggests the contrary.

It appears that investors communicate on Twitter their opinion following an operation. Hence they express a negative sentiment following a trading operation that did not performed as planned. Furthermore, investors expect the stock price to increase following two days of negative performance, as shown by negative coefficients. Hence the study confirms that the investors do not believe on the random walk of the stock return. The same suggestion was postulated by Barberis et al. in 1998. However, it is worth to mention that the findings of this study assert a lack of the random walk for the stock return of Apple in the short term; thus the investors trading the stock are following a winning strategy. According to second measurement, the results show that the yesterday stock return tends to affect positively the today sentiment, while the two days before stock return affects the today sentiment negatively. Therefore confirming the lack of faith by the investors in the random walk of the stock returns.

Table 3 – Granger causality

Panel A	X$\rightarrow$Y	Y$\rightarrow$X
AAPL	.0.606 (1.00)	0.648 (0.88)
FB	0.363 (2.03)	0.018[**] (7.99)
GOOGL	0.196 (3.26)	0.549 (1.20)
MSFT	0.523 (1.3)	0.29 (2.47)
Panel B	X$\rightarrow$Y	Y$\rightarrow$X
AAPL	0.703 (0.71)	0.587 (1.07)
FB	0.265 (2.66)	0.000[***] (17.61)
GOOGL	0.268 (2.63)	0.899 (0.21)
MSFT	0.381 (1.93)	0.011[**] (9.09)

X is the variable DailySentiment; Y is the DailyReturn;
Panels A and B represents the two different measurements respectively; the values reported are the probability of the χ^2 distribution, in brackets the χ^2 value.

5. CONCLUSION

The study analysed and explored the relationship between the investors' sentiment and the stock price. The first part outlined the findings from previous papers; the literature point out the existence of such relationship, and in some cases a causal relationship. The causality implies a forecasting ability of the investor sentiment in projecting the future returns. However, few studies centred the research on whether such causality is in fact real with individual stocks, hence the possibility of exploiting it with a tailored trading strategy. Thus this study was aimed at fulfilling that gap in the literature.

The social sentiment has been measured through the raw data collected from the social platform Twitter. The stocks analysed are the most tweeted ones, in order to gather a sample of data large enough to validate the study. The findings propose contrasting results with previous literature. No significant causal relationship has been found of the investors' sentiment in affecting the stock return. However the causal relationship appear in the opposite direction.

In answering the research question formulated at the beginning of the paper, the investors' sentiment do not affect the stock return by a great extent. The findings point to the picture that investors uses the micro blogging posts to communicate opinions on the stocks, following the trading operations. Therefore the opinions are biased by the outcome of those actions. It seems that investors want to express their feelings in order to influence other investors' opinions on that stock. Therefore the conclusion of the study postulates that a trading strategy built to forecast the stock price analysing the investors' sentiment will not be feasible, because the sentiment is shaped by the market trend, and not the other way around. However the study has some intrinsic limitations, which may foster the spread for future research on the topic.

5.1 Limitations and future research

The study, being extremely experimental, is affected by several limitations. The measurement of the data, which influence the results by a great extent, is indeed a compelling issue. A third party software has been used, which is highly reliable. However the lexicon implemented for the text analysis does not include financial jargon. The second measurement has been designed for the study, therefore it includes specific financial related words, but the drawback is the lack of the ability to contextualise the words. Nevertheless, both measurements yielded similar results, therefore the categorisation of the data is assumed to be reliable, although the results are not exactly alike. The lack of an extensive and complete method to perform such text analysis, concerning financial related text strings, is a limitation to any study focused on similar topics.

The study focused on a small number of stocks, but as Twitter becomes more popular, similar studies might be extended to a larger group of stocks, considering also securities listed on stock exchanges different than just those from the United States.

As mentioned in previous sections, Twitter does not provide free access to historical data, therefore the data need to be

recorded day by day. The method requires long time of preparation, depending on the hypothesis to be tested and thus the amount of data needed. For this study one month of data has been collected due to tight time frame, but larger sample of data might yield more significant results.

Future studies may diverge the same procedure to different stocks and compare the findings. As previous studies found the Granger causality between the sentiment and the indices, the same relationship should be found for transitivity with individual stocks listed on the same indices. The findings of this study did not find evidence confirming it, but it is expected that the transitivity holds to some extent, either in a larger sample of stocks, or in longer period of data. Moreover, if such Granger causality is further confirmed by other studies and thus the sentiment possess forecasting ability, the sentiment data would be highly valuable. An interesting point of research is measuring the value of the sentiment data, hence how much a tweet is worth.

All the tweets analysed within this study have been treated equally, because no in depth study have been performed on the number of followers. Arguably, that number can make a third variable which can be exploited further. A tweet made by an investor whose account has many followers might influence in a greater extent comparing to a tweet from an account with fewer followers (Rui & Whinston, 2012).

Furthermore, the relationship assumed is linear, thus the tests are aimed at finding such linear relationship. However it may be possible that the nature of the relationship is not linear. A non-linear relationship may be suggested by the hypothesis that the sentiment reflects the action of the investors and thus presenting a positive relationship with the stock return. But at some point the large number of positive sentiments might reflect an overvaluation of the stock, thus the relationship might become negative. Therefore such non-linear relationship might be investigated further.

This study added to the literature insights on the investors' sentiment, and therefore improved the knowledge the topic recently highly investigated, which forms theories on the large field of behavioural finance.

6. AKNOWLEDGEMENTS

7. REFERENCES

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics, 49*(3), 307-343.

Bernard, V. L., & Thomas, J. K. (1990). Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics, 13*(4), 305-340.

Bertone, S., Paeglis, I., & Ravi, R. (2015). (How) has the market become more efficient? *Journal of Banking & Finance, 54*, 72-86.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1-8.

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance, 11*(1), 1-27.

Edmans, A., GarcÍA, D., & Norli, Ø. (2007). Sports Sentiment and Stock Returns. *The Journal of Finance, 62*(4), 1967-1998.

Engelen, P.-J., & van Essen, M. (2011). Reputational Penalties in Financial Markets: An Ethical Mechanism? *Responsible investment in times of turmoil* (pp. 55-74): Springer.

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

Hengelbrock, J., Theissen, E., & Westheide, C. (2010). Market response to investor sentiment. *Available at SSRN 1343798*.

Hentschel, M., & Alonso, O. (2014). Follow the money: A study of cashtags on Twitter. *First Monday, 19*(8).

Joseph, K., Babajide Wintoki, M., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting, 27*(4), 1116-1127.

Nofsinger, J. R. (2005). Social mood and financial economics. *The Journal of Behavioral Finance, 6*(3), 144-160.

Oh, C., & Sheng, O. (2011). *Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement*. Paper presented at the ICIS 2011, Shanghai.

Pak, A., & Paroubek, P. (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. Paper presented at the LREC.

Palomino, F., Renneboog, L., & Zhang, C. (2009). Information salience, investor sentiment, and stock returns: The case of British soccer betting. *Journal of Corporate Finance, 15*(3), 368-387.

Rui, H., & Whinston, A. (2012). Designing a social-broadcasting-based business intelligence system. *ACM Trans. Manage. Inf. Syst., 2*(4), 1-19.

Wang, F. A. (2001). Overconfidence, investor sentiment, and evolution. *Journal of Financial Intermediation, 10*(2), 138-170.

Zhang, X., Fuehres, H., & Gloor, P. A. (2012). Predicting asset value through twitter buzz *Advances in Collective Intelligence 2011* (pp. 23-34): Springer.