STOCKHOLM SCHOOL OF ECONOMICS
Department of Accounting
Bachelor Thesis in Accounting & Financial Management
Spring 2016

SIGNAL OR NOISE – THE PREDICTIVE VALUE OF SOCIAL MEDIA FEATURES ON MARKET RETURNS

a study on the relationship between social media and the Swedish stock market

Eric De Geer (23016) & Benjamin Warberg (22908)

ABSTRACT: This study aims to investigate the relationship between social media data and the main market returns; stock price return, trading volume and volatility for the Swedish stock market. Using a data set of stock-specific social media mentions, classified by state of the art machine-learning algorithms, for the period of 2014-2016, we construct proxies for investor sentiment, attention and agreement. We then perform a time series analysis on daily financial and social media data, testing against an adjusted version of the OMX30 index and a sample of ten individual firms. Our results show a significant positive relationship between lagged changes in investor sentiment and stock price returns, implying that data from social media contains valuable information for investors. However, we find a negative relationship between the volumes of social media mentions, a proxy for attention, and trading volume. Our study contributes to the existing literature on the relationship between social media data and financial markets, as well as extends the relevance of previous research to the Swedish stock market.

TUTOR: Florian Eugster

KEYWORDS: investor sentiment, investor attention, social media, behavioural finance, market returns

ACKNOWLEDGEMENTS: We would like to thank our tutor Florian Eugster for his guidance and support. Furthermore, we would like to extend our most sincere gratitude to Modular Streams for giving us access to their data, which without this thesis would not have been possible.

DATE: May 16th, 2016

TABLE OF CONTENTS

1. INTRODUCTION

The use of social media keeps on growing and change the way we interact. Now, news stories travel across the world in a matter of seconds and with an unprecedented access to the Internet, sharing information and opinions has never been easier. The growing usage of social media and the Internet at large has led to an accumulation of user data. Coupled with increased computational power this vast amount of information lends itself to an array of research and experiments. Never before has there been such access to data documenting collective behaviour or collective thought. Given proper analysis, this type of data gives rise to possibilities for new insights within virtually every field. One such field, that has become an increasingly popular research area during the last years, is the relationship between social media and financial markets.

According to traditional financial theory, investors are *rational* and acting according to self-interest, what is commonly called *homo economicus.* Theories like the efficient market theory rest on this assumption, and state that asset prices reflect all available information. According to the efficient market hypothesis, further price movements are simply the result of unseen news or events, since only new information will move stock prices. However, theories from behavioural finance gives another perspective, studying the investor behaviour deriving from psychological principles of decision making, to explain why people make buy or sell decision in the market (Shefrin, 2001). A large amount of research exists examining the irrational behaviour of market participants (Barberis, 2007) and some researchers mean that standard economic models have neglected the influence of emotions on individual decision-making (Sanfey et al. 2003).

Traditional theory states that prices of assets move towards the fundamental value by rational arbitrators (Friedman, 1953) whereas behavioural finance talks about irrational participants, noise traders (Black, 1986) driving the prices away from fundamentals. These schools of thought offer opposing views on the factors affecting markets and why. Many financial economists now claim that stock prices and other financial indicators are at least partially predictable. While many believe that behavioural theories have some truth, it has been difficult to measure investor emotions and sentiments, or attention, until now. The vast amount of data from online sources has made it possible to perform large-scale studies examining what investors say, how they feel, what assets are grabbing attention, and how this is tied to market returns.

While several studies on the subject between social media data and financial markets has seen a spur during the last years, most studies are focused on the North American market. The Swedish market is relatively unexplored, only one study has previously been conducted on the relationship between social media data and the Swedish financial market (Lundell and Shuangyu, 2015). To the best of our knowledge this study is the first to examine the relationship between social media data and its connection to the main market returns: price return, trading volume and volatility. By performing a time-series analysis, using multiple linear regression models, we aim to examine the relationship between social media and the Swedish stock market.

Previous research performed on North American data and markets

Previous research analysing internet-sourced data and financial markets has focused on web news, search engines, or social media. Tetlock (2007) finds that the number of mentions of a company in financial news is a significant predictor of trading volume, additional research also finds a significant relation between the *sentiment* of news and stock returns (Allen et al. 2014 and Borovka, 2015). Antweiler and Frank (2002) find that sentiment from financial forums has significant predictive power between message posting and trading volume. These results have been supported further by Antweiler and Frank (2004) showing that heavily discussed stocks that are also particularly heavily traded. Sprenger et al. (2014), also confirm an association between tweet sentiment and stock returns, message volume and trading volume as well as disagreement and volatility. Studies focusing on search engine data mainly uses Google Trends have been focused on the relation between attention, measured by specific searches, and trading volume (Preis et al. 2010). Just recently, Molnár et al. (2016) found that high Google search volume has a leading effect on negative returns. However, the most popular area of research is conducted on social media data, and specifically micro blogs like Twitter. Most of the recent studies concerning social media data and its relationship with financial markets concern *sentiments* from the data. Ranco, Aleksovski, Caldarelli, Grčar and Mozetič (2015) have showed significant correlation between Twitter sentiment and CAR for the DJIA and Bollen et al. (2015), find that high twitter bullishness lead to subsequent positive stock price returns. Just recently Azar and Lo (2016) show that the content of tweets can be used to predict future returns as well as showing that a tweet-based asset-allocation strategy outperforms most benchmarks.

This research area is in its nativity for the Swedish stock market. Lundell and Shuangyu (2015) conducted the first study on Swedish data examining the relationship between social media sentiment and abnormal returns. They find a significant predictive power of sentiments on abnormal returns using an event study, built on a unique data set on Swedish firms. Their results extend the generalizability of previous results on US markets. However, to the best of our knowledge, no prior study has performed a time-series analysis on social media data and examined the relationship between social media features and the main market returns (price return, trading volume and volatility). This study contributes to the previous research in the field by providing insights into the Swedish market and its relationship with social media data. By examining price return, trading and volatility, and their connection to investor sentiment, attention and disagreement, we contribute to the current debate on traditional financial theory vs. behavioural finance and complement the broader literature in trying to extent the generalizability of previous research to the Swedish market. Our study is somewhat inspired by Antweiler and Frank (2004) and Sprenger et al. (2014) but we complement their studies by performing a time-series analysis, using similar models, and thus increase the granularity of analysis.

1.1 Purpose of study

The overall purpose of this study is to examine the relationship between social media features and market returns for the Swedish stock market. A way to use social media data is to measure collective emotions by classifying sentiment on stock-specific content from social media, as well as the frequency of a firm being mentioned. If social media sentiment and attention captures investor sentiment and attention, and there are behavioural aspects in investors' choices, these proxies might have a predictive value over market returns. Hence, our study aims to answer the following research question for the Swedish market:

Are stock-specific social media features associated with market returns on the Swedish market?

1.2 Research boundaries

The scope of the study is to investigate the relationship between social media features and market returns for Swedish public firms for the period 2014 to 2016. The sample period has been limited to the two year period from March 2014 to March 2016 due the social media data[1] provided by Modular Streams. More specifically we examine the predictive power of social media features on market returns for an adjusted version of OMX30, containing 22 of the 29 OMX30 firms (Atlas Copco is included twice with A and B shares), as well for 10 individual firms separated by industry and chosen based on social media data density over the sample period. Our social media features are sentiment, volume of mentions and agreement among mentions. The market return features are stock price return, trading volume and volatility. Our statistical tests are time series analysis

[1] The selection was based on density of data points, for more explanation see section 4.1.

consisting of a multivariate OLS-model with adjustments for internal data structures (see section 5.2). However we will not use ARMA or ARIMA-models in our time series adjustments as these methods are outside the scope for a bachelor thesis[2]

1.3 Outline

Chapter 2 presents the theoretical frameworks as well as previous related research on the relationship between social media and financial markets. From the theory logic and reasoning in Chapter 2 the general hypotheses are presented in Chapter 3. Chapter 4 covers the data collection, selection and formatting as well as a describing our sample. In Chapter 5 the research method and statistical tests are presented. Chapter 6 includes our findings/results and chapter 7 includes a discussion on our results as well as discussion on robustness tests. Lastly, Chapter 8 presents our conclusions drawn from the study as well as suggestions for future research in the field. References and appendix are presented in Chapter 9 and 10 respectively.

2. THEORETICAL FRAMEWORK AND PREVIOUS RESEARCH

In this chapter we introduce relevant theoretical frameworks and previous research that **motivates our study** on the relationship between investor sentiment and Swedish financial markets.

2.1 Introduction to previous research

Generally, previous research analysing internet-sourced data and financial markets can be divided into three segments:

- Web news
- Search engine queries
- Social media

Most previous research regarding *web news* and financial markets has been centred on either the stock price reaction to news or the number of mention of a company in financial news and trading volume (Tetlock, 2007 and Alanyali et al. 2013). Some of the research also focuses on the relation between the actual sentiment of news and return predictability (Allen et al. 2014 and Borovka, 2015).

---

[2] For more information of these models please see Edlund, P. 1989, "Preliminary estimation of transfer function weights: a two-step regression approach".

Within search engine queries Google is the most popular choice of research, more specifically Google Trends. Most of the research using Google Trends examines the relation between attention and trading volume (Preis et al. 2010) as well as google trends to forecast stock returns and abnormal returns. Molnár et al. 2016 showed that high google search volume shows a leading effect on negative returns. Möllborg and Berglin (2015) also find a significant predictive power of google trends on abnormal returns as well as trading volume.

Since around 2009, when the microblog Twitter had its beginning, research relating to the connection between social media data and financial markets has really catapulted. Some studies focus on the volume of mentions for a company and financial markets, – Mao, Wei, Wang, Liu, 2011 for example tested data from twitter against S&P 500 and found significant relationship between trading volume and volume of mentions (used as a proxy for attention). However more recently, most of the new studies concerning social-media data and its relationship to financial markets is focus on sentiments from the data. Bollen et al. 2011 spurred the research in this area when their study showed a predictive model, using *public sentiment from twitter,* of the next-day direction of the DJIA with an 87 per cent accuracy. Although there has been studies preceding Bollen et al. 2011, which have shown the connection between social media sentiment and stock returns (Zhang, Fuehres and Gloor, 2009, A. Yi, 2009). Sentiment data from a range of social media has also been shown to be predictive over abnormal returns (Lundell and Shuangyu, 2015). More recently one studies have shown that tweet sentiment help explain time-series variations of security returns beyond the variation explained by Fama-French's Five-Factor Model (Liew and Budavari, 2016). Azar and Lo, 2016, also show that the content of tweets can be used to predict future returns as well as showing that a tweet-based asset-allocation strategy outperforms most benchmarks.

So mostly, previous research seems to show that there is some valuable information in internet-sourced data and its relation to market returns. To be able to understand this as well as any results our study may yield, it is important to examine the underlying theoretical frameworks governing market returns such as asset prices, trading volume etc. Moreover, in contrast to most other studies (except Lundell and Shuangyu, 2015) our data, while sourced from social media, is not limited to just content by individuals or news but contain both web news and content from individuals. Accordingly, in the following sections we will cover the main theoretical frameworks for what affect assets as well how both news and individuals might play into this.

2.2 Efficient market hypothesis and behavioural finance

Our research area demands mentioning the two, often competing, financial theory frameworks. *Firstly*, on the one hand we have the orthodox *homo economicus* view, according to which all investors are *rational* and acting according to self-interest. Extensions of this view include the *capital asset pricing model (CAPM)* and the *efficient market hypothesis,* or EHM (Fama, 1965), suggesting asset prices fully reflect all available information and that further price movement are the result of unseen news or events. More specifically the EHM can be broken down into three versions:

- ***Weak-form efficiency***: Assumes efficient market, reflecting all market information. Assumes independent rate of returns, no dependencies with past or future rates of return.
- ***Semi-strong-form efficiency***: Assumes efficient market reflecting all publicly available information. Implicitly assumes no investor can benefit over the market trading on new information.
- ***Strong-form efficiency***: Assumes efficient market reflecting all information, both public and private. Accordingly, no investor can outperform the market even given new information.

So according to EHM, only new information will move stock prices. The weak form of EHM is also in line with the *random walk hypothesis* (Regnault, 1863 and Bachelier, 1900), stating that changes in stock prices are independent to each other and have the same distribution. That is, that prices move on random and are not influenced by past events (no momentum). However, there are some empirical evidence contradicting the random walk hypothesis, Lo and MacKinlay (2000) has shown that there are trends in the stock market and that the market as such is somewhat predictable.

Not only in academia but also within the financial industry, EHM has dominated during the major part of the last century and constitutes the key hypotheses behind stock market predictions. According to this classical view on asset prices, our study will not yield any significant results, at least for stock price returns, however any result contradicting the EHM might be explained by its short-term anomalies[3]

[3] Fama (1998) show some empirical evidence consistent with the EMH that anomalies are change results, that apparent overreaction to information is as common as underreaction. He also finds that anomalies, consistent with the EMH, can be due to methodology and that most long-term return anomalies tend to disappear with

However, and *secondly*, theories of behavioural finance (BF) gives us another perspective. BF is the study of investor market behaviour deriving from psychological principles of decision making, to explain why people make buy or sell decisions in the market (Shefrin, 2001). Building on the forefathers of BF (Kahneman & Tversky, 1974), a vast amount of research exists examining the irrational behaviour of market participants resulting from biases, heuristics and psychological variables (Barberis, 2007). Some researchers mean that standard economic models of decision making have typically minimised or ignored the influence of emotions on individual decision making (Sanfey et al. 2003). Some research in BF, such as Arkes, Herren and Isen (1988) or Nygren, Isen, Taylor, and Dulin (1996), regards market sentiment and its influence on asset pricing and volatility. Sentiment is here defined as the collective mode and attitudes of investors (Shiller, 2005).

"Behavioural economics studies reveal that negative sentiment driven by bad mood and anxiety affects investment decisions and may hence affect asset pricing"

- Kaplanski and Levy, 2010

Typically, changes in mood caused by either news or political events can affect asset prices in two ways. They can produce swings of optimism or pessimism in market participants, leading to biases in the expectations of future cash flows and over or underweighting of the related risks of future cash flows, which may materialise in investors (Boyle, Hagan, O'Connor and Whitwell, 2004).

2.3 Sentiment and Asset Prices

Clearly, there are opposing views between the traditional and behavioural school of thought as to what moves prices. According to traditional financial theory, the prices of assets are moving towards their fundamentals by rational arbitrators trading in the market on rational beliefs (Friedman, 1953). However, the behavioural theories proclaim that there are some market participants whose beliefs when trading are irrational. These irrational participants are called noise traders, a notion first developed by Black, 1986. The theory of noise traders beholds that markets

---

reasonable changes in technique. For more information on anomalies please see Latif, Madiha, et al. "Market efficiency, market anomalies, causes, evidences, and some behavioral aspects of market anomalies." *Research Journal of Finance and Accounting* 2.9/10 (2011): 1-14.

consist of some informational inefficiencies, limiting the possibilities for arbitrage. Due to this, investors do not make investments based solely on fundamentals, and can thus drive prices away from the true, or fundamental, value of assets by trading on their beliefs of sentiment (Black, 1986, Schleifer and Summers, 1990, De Long, 1990). However, early studies by Fama (1965) maintain that irrational market participants do not affect the prices of assets, but instead the rational investors do. Some mean that individual investors are the least informed market participants and because of this pays a penalty for active trading, being an underdog in relation to the more informed investors (Hirshleifer and Teoh, 2003, Barber and Odean, 2000). This is also implicitly supported by other theories suggesting that informed investors with a limited trading capital have incentives to spread imprecise, but informative, trading advice and stock tips (Van Bommel, 2003). This in term will likely drive uninformed investors beliefs of sentiments, especially if similar information is repeated, as bounded rationality, the fact that people fail to account for the repetitiveness in the information they receive (e.g. The more we here the same thing, the more we will believe it is true), leaves must of us vulnerable to persuasion (DeMarzo et al, 2003).

2.3.1 News as a source of investor sentiment

If there are irrational investors, or noise traders, how do they form their beliefs of sentiment? Intuitively, one form a subjective opinion based mostly on all the information available. Whether that information comes from stock tips (as in the reasoning above), news or other information channels, we cannot know. One information channel traditionally viewed as being a main factor affecting investors' beliefs is news. There are generally two main theories on how news typically affect investors' behaviour in the markets: the *information theory* and the *salience theory* (Solomon et al, 2012). The information theory simply states that news or media reduce the cost of information leading to investors improving their decision-making. The salience theory states that media steers attention of investors towards particular assets so that the more an asset is covered in news channels the more the demand for their stock will increase, this is also in line with the theory on bounded rationality (see above).

Empirical Evidence

A lot of research has been made on the relationship between investor sentiment and asset prices. There are several studies that have found that sentiment can predict asset prices – both using sentiment from social media, like twitter, and from traditional news (using it as a proxy for investor sentiment). By using sentiment form traditional news, Tetlock (2007) find that high media pessimism puts a downward pressure on market prices. Lemmon and Portniaguina (2006) find that for small stocks, sentiment has a predictive value of the return. Also Sprenger et al. (2014), looking at the content of microblogs like twitter, find an association between tweet sentiment and stock returns. One paper found significant correlation between twitter sentiment and CAR for the DJIA (Ranco, Aleksovski, Caldarelli, Grčar and Mozetič, 2015). By measuring online bullishness, Bollen et al. (2015), find that high twitter bullishness lead to subsequent positive stock price returns. Some studies also echo these results in broader terms, by measuring public sentiment, not linked to financial markets and find significant correlation between public sentiment and movements on the DJIA and S&P 500 (Bollen J, Mao H and Zeng X, 2011 & Zhang, Fuehres and Gloor, 2009). Souza and Ante (2016) also find evidence suggesting a non-linear causal relationship between social media and market prices.

2.4 Attention and Disagreement

How do individual investor choose which financial assets to invest or trade with? According to Odean, 1999, investors tend to choose assets that are in their available memory and have recently caught their *attention*. This is also in line with the behavioural bias called the availability bias (Kahneman and Twersky, 1973). Which assets specifically investor choose will depend on their personal beliefs, or *sentiment.* In our study we will use the total number of messages as a proxy for attention, where both the news will be included (professional source of mention) as well as individual preferences, or sentiments. Furthermore, people tend to misjudge and/or underestimate statistical, abstract and base-rate information (Kahneman and Twersky, 1973) which implicitly leads to that the attention-level of for example a specific financial asset might not be equal to, or correspond, to its fundamental value. Coupled with the aforementioned *saliency theory* and *bounded rationality*, news and other repetitive information will steer attention towards specific assets and increase its attractiveness. *One can logically and intuitively arrive at the notion then that increased attention might lead to an increase in trading volume*. This is also supported further by Cao et al. 2003, stating that conversation (or information exchange) will lead to increased market activity from noise traders, trading on other investors' signals or outlooks.

It is important to note, however, that the same does not necessarily apply to institutional investors. For institutional investors, attention is not a scarce resource and they devote more time in their search process than do individual investors (Odean and Barber, 2006). Thus, financial markets with a relatively high amount of institutional investors might show weaker results between attention and market returns. As of 2007, 18 per cent of the total value on the Stockholm exchange was owned by institutional investors (Fristedt and Sundqvist, 2007). Seeing as the large majority of previous research in related fields have been on the North American market, where the proportion of institutional ownership was around 67 per cent of the total market value as of 2010[4], we might be able to attain better results on the test adhering to attention and market returns.

Returning to investors' beliefs about assets, there are some notable theories about the implications of the *differences* of these beliefs. The traditional idea in financial theory states that an increase in *disagreement* will cause a rise in trading volume because trading occurs when market participants have differing beliefs about the prices of assets (Harris and Raviv, 1993). According Das et al. 2005, extensive debate will also follow from disagreement about market information and Danthine and Moresi, 1993, means that more information should reduce volatility. Although, both theory and empirical data have wide support on that disagreement *is* actually connected to increased volatility.

Empirical evidence

Antweiler and Frank (2001) show that sentiment from financial forums has significant predictive power between message posting and trading volume. Furthermore, their results were also supported by Antweiler and Frank (2004) showing that stocks that are heavily discussed are also particularly heavily traded. Sprenger et al. (2014) find similar results, confirming an association between tweet sentiment and stock returns, message volume and trading volume as well as disagreement and volatility.

[4] http://www.sec.gov/News/Speech/Detail/Speech/1365171515808

3. TEST LOGIC AND GENERAL HYPOTHESES

The theoretical background and empirical evidence from related research above provides the theory-driven hypotheses we will examine under our research question.

Research question:

Are stock-specific social media features associated with market returns on the Swedish market?

General hypotheses:

1. *SENTIMENT*
    a. Aggregated stock-specific sentiment from social media mentions is associated with stock price returns
2. *ATTENTION*
    a. Increased volume of mentions in social media is associated with an increase in trading volume
3. *AGREEMENT*
    a. Increased disagreement among mentions is associated with higher volatility

4. DATA COLLECTION

Here we describe our general outline of the collection of the data as well as methods used for formatting and processing the data. While most previous studies focus on North American or global markets, our study examines the Swedish stock market. To test our hypotheses, we use an adjusted version of the OMX30 (see section 4.1). We also test our hypothesis on a sample of ten individual firms, chosen by the density of data during our sample period (4.1).

The data can be divided into two areas: *social media data* and *financial data.* Our social media data includes proxies for investor sentiment. The financial data includes unadjusted (raw) opening and closing prices, unadjusted trading volume and unadjusted intraday high and low prices (except for our own index). All data has a daily frequency and is collected for the period 2014-03-30 to 2016-04-01.

4.1 Social media data

Our social media data has been donated by the company Modular Streams, a firm specialized in data analytics. By using a textual web mining approach, this data is obtained from a number of social media platforms (Twitter, Google+, YouTube, RSS feeds) and news sites. Their software scrapes these sources for *stock-specific mention* (such as name + TICKER symbol) of any firm in a list of 488 firms they follow, of which the lion share is Swedish. The textual content of every mention is then classified in terms of sentiment using a sophisticated machine-learning algorithm based on a supported vector machine. The software supports 49 languages and has a classification accuracy of around 90 per cent. The initial data set contains 162 278 data points. Each observation contains some additional information such as *time stamp, company name, sentiment classification, information type (question, opinion), user (individual or professional) and source (Twitter, YouTube etc.)*.

Adjustments in social media data

The raw data contains mentions collected from January 2002 until end of March 2016. However up until the beginning of 2014 the data is very scarce, so for the purpose of having denser data we have excluded any data before 2014-03-30, giving us a social media sample period from 2014-03-30 to 2016-03-30. This adjustment leaves us with 160 582 data points for all the firms. Moreover, the collection software unfortunately has a cap of 1000 mentions, trailing a collection of 1000 when this is exceeded, which leaves our maximum number of observations per firm at 1000 over the period. To mitigate this effect somewhat, we merged our data set with an identical data set from

Lundell and Shuangyu (2015), which consist of the same data but with the last collection date being 2015-07-30. By merging the data sets, we maximise the number of data points in our period and are left with 232 074 data points with the maximum number of mentions being 2000 per firm. Moreover, we adjusted the time stamp for closing of markets and weekends. Social media mentions made from 17:15 and onwards have been allocated to t+1 due to the markets being closed. Mentions from Friday until Monday have been allocated to Friday, as to "affect" the market on Monday.

Index creation and individual firms

We want to test our hypotheses against the OMX30 as well as against individual firms. However our social media data set only contains 22 out of the 30 (Atlas Copco is included twice with A and B shares) - firms on OMX30, why we *create our own index*[5] (hereinafter labelled OMX22) from these firms – representing OMX30. The seven missing companies are four banks (Nordea, Handelsbanken, SEB, Swedbank), two manufacturing companies (Skanska, Volvo) and a telecom company (Nokia). For testing our hypotheses against individual firms we choose the top ten firms, in terms of total mentions during our sample period, from different industries (see 4.3 for more on sample).

4.2 Financial Data

Using Thomson Reuters DataStream, for the period 2014-03-30 to 2016-04-01, we then obtain a time series of unadjusted opening and closing prices, unadjusted daily trading volume for the firms in our *own index* OMX22, as well as for the ten individual firms. For our ten individual firms we also obtain unadjusted intraday high and low prices. We use an adjusted version OMX30 in our research to ensure that the index constituents have enough public attention and that the companies included are also part of our sentiment data.

4.3 Sample

Here we present summary statistics over our samples, both for OMX22 and for our ten individual firms. Social media data summaries, as well as financial data summaries are presented below.

[5] For more information on the index creation, see *Appendix 1.*

Table 1

Summary statistics – Index level sample social media data

This table shows the summary of our social media mentions for our sample of 22 firms, constituting our index OMX22. This sample include 22 stocks for the period of March 2014 to March 2016.

Firm	Positive mentions	Neutral mentions	Negative mentions	Total
ABB	1153	54	390	1597
Alfa Laval	609	44	296	949
Assa Abloy	490	38	138	666
AstraZeneca	273	62	220	555
Atlas Copco	811	42	258	1111
Boliden	1390	71	539	2000
Electrolux	1002	70	552	1624
Ericsson	940	74	986	2000
Fingerprint Cards	1038	112	846	1996
Getinge	519	105	444	1068
HM	634	37	325	996
Investor	843	91	348	1282
Kinnevik	1634	61	305	2000
Lundin Petroleum	926	100	872	1898
Sandvik	821	105	1074	2000
SCA	1352	56	592	2000
Securitas	457	47	243	747
SKF	849	59	1067	1975
SSAB	898	99	1003	2000
Swedish Match	400	104	238	742
Tele2	1129	61	441	1631
TeliaSonera	803	31	1166	2000
Total	18971	1523	12343	32837

Table 2

Summary statistics – Index level sample financial data

This table shows the summary of the financial data for our sample of 22 firms, constituting our index OMX22. This sample include 22 stocks for the period of March 2014 to March 2016. P^C is the unadjusted opening price and TV being the trading volume in number of shares in thousands.

Variable	Obs	Mean	Std.Dev.	Min	Max
P^C	524	113.181	8.216	98.101	133.316
TV	524	4048.71	370.118	3305.25	4927.07

Table 3

Summary statistics – Firm level sample social media data

This table shows the summary of our social media mentions for our sample of ten individual firms, constituting our firm level sample. This sample include ten stocks for the period of March 2014 to March 2016.

Firm	Negative mentions	Neutral mentions	Positive mentions	Total
Anoto Group	533	143	820	1496
Arcam	268	134	938	1340
Betsson	424	105	835	1364
Boliden	539	71	1390	2000
Elekta	786	91	925	1802
Ericsson	986	74	940	2000
MTG	345	54	741	1140
NCC	141	26	1497	1664
Sandvik	1074	105	821	2000
SAS	1091	43	865	1999
Total	6187	846	9772	16805

Table 4
Summary statistics – Firm level sample financial data

This table shows the summary of the financial data for our sample of ten individual firms. This sample include ten stocks for the period of March 2014 to March 2016. P^C is the unadjusted opening price and TV being the trading volume in number of shares in thousands. P^H is the unadjusted intraday high price, and P^L is the unadjusted intraday low price.

Variable	Obs	Mean	Std. Dev.	Min	Max
P^C	5,247	127.776	90.7278	.21	379.2
TV	5,007	4076.169	9628.028	13.8	202781.7
P^H	5,007	129.36	91.63071	.23	380.3
P^L	5,007	125.9546	89.51381	.21	374.4

5 METHOD

For our baseline model we draw upon the research by Antweiler and Frank (2004) and Sprenger et al. (2014). We are interested in the predictive value of our social media data on Swedish market returns and will test the social media features sentiment, volume of mentions and agreement against the financial market features stock prices return, trading volume and volatility.

Using our financial and social media data we construct variables acting as proxies for investor sentiment and attentions. The baseline model is similar for all our hypothesis and is presented below:

$$Y_{ti} = \alpha_i + \left(\sum_{i=1}^{5} \beta_i\right) \times R_{t-1} + \left(\sum_{i=1}^{5} \gamma_i\right) \times S_{t-1} + \left(\sum_{i=1}^{5} \delta_i\right) \times MV_{t-1} + \left(\sum_{i=1}^{5} \theta_i\right) \times A_{t-1} + \varepsilon_{ti}$$

$Y = Dependent\ variable$

$R = Price\ returns$

$S = aggregated\ sentiment$

$MV = volume\ of\ mentions$

$A = agreement\ among\ mentions$

$t = time\ variable, t\ measured\ in\ days$

$i = firm\ and\ index$

We use a multiple linear regression for three different dependent variables on three independent variables and use lagging market returns as a control variable for momentum.

Below will follow a short description of our independent (5.1) and dependent variables (5.3). However, since we are looking at time series data rather than cross-sectional data, there are a few issues that need to be accounted for, such as internal structures of the daily data points (autocorrelation, trend, seasonal variation, etc.). This will be covered in sections 5.3.

5.1 Independent variables

From our social media data we construct three independent variables, sentiment, mentions volume, and agreement, in line with previous research by Antweiler and Frank (2004) and Bollen et al. (2015). For our own index OMX22 we only use the social media data relevant for the 22 firms and for our individual companies we only use data relevant for each company.

Sentiment

The sentiment variable is constructed as a ratio of aggregated positive (S_t^+) and negative (S_t^-) sentiments on day t, a variable used successfully in previous research by Antweiler and Frank (2004), Sprenger et al. (2014) and Bollen et al. (2015). We use the logarithmic value to adjust for large variances in our sample. We create ten firm specific sentiment indices and one for the OMX22 index. For OMX22, the sentiment index was constructed by weighting all 22 firms' individual sentiment score by their daily market capitalisation share of the index, and then aggregating these values into a new index[6].

$$SENTIMENT_t = S_t = \ln\left(\frac{1 + \|S_t^+\|}{1 + \|S_t^-\|}\right)$$

$SENTIMENT_t = aggregated\ sentiment\ score\ for\ day\ t$

$S_t^+ =$ total number of positive mentions on day t

$S_t^- =$ total number of negative mentions on day t

A positive value of SENTIMENT implies that $S_t^+ > S_t^-$, negative value implies $S_t^+ < S_t^-$ and a value of zero implies $S_t^+ = S_t^-$. For large differences between the positive and negative sentiments the magnitude of the SENTIMENT will increase.

Mentions volume

As a proxy for investor attention, we construct a variable in line with Antweiler and Frank (2004) and Sprenger et al. (2014). The proxy is simply the total number of mentions on day t. Again, for

[6] For more information on the index construction, see Appendix 1.

our index, this proxy has been aggregated by weighing the attention for each firm each day using market capitalisation share of the index to be able to capture any significant relationship with index trading volume. For our sample of individual firms, no such aggregation has been made, here the total volume of mentions, unadjusted, are used as the proxy for investor attention. To adjust for large variances in our data we also use the logarithmic values for this variable.

$$MENTIONS\ VOLUME_t = MV_t = \ln\left(1 + (S_t^+ + S_t^-)\right)$$

Mentions volume can only be positive or zero and here the magnitude of the total number positive and negative sentiments determines the magnitude of the variable.

Agreement

The construction of our variable agreement, measuring agreement among social media mentions, have been construction in line with the equation proposed by Tetlock (2007) and further used by Antweiler and Frank (2004) as well as Sprenger et al. (2014). The variable is a ratio between the difference in number of negative and positive sentiments, and acts as a proxy for investor agreement (or disagreement) on day t.

$$AGREEMENT_t = A_t = 1 - \sqrt{1 - \left(\frac{S_t^+ - S_t^-}{S_t^+ + S_t^-}\right)^2}$$

The variable ranges from zero to one, where numbers close to zero implies a large amount of disagreement and a values equal or close to one implies little to no variation in the number of positive (negative) sentiments in one day. Here we do not use logarithmic values since the variation range in the variable is limited.

5.2 Time series analysis and data adjustments

When conducting a time series analysis, we must first adjust the data for any trends and seasonal effects that might affect how our financial data changes. By identifying the underlying trends, we can adjust our data to become more stationary, if we do not adjust for any trends or seasonal effects in our data, the estimated coefficients would capture this effect and would cause a violation of the assumption about zero conditional mean. Normally, when using cross-sectional data, the assumption about stationarity is not needed for the ordinary multiple linear regression model, but since we are performing a time series analysis the assumption about stationarity needs to be considered.

5.3 Dependent variables

As we will test our hypotheses against stock price returns, trading volume and volatility, these will be our dependent variables, yielding three versions of or baseline model. Below we cover adjustments in order to perform a proper time-series analysis.

Price movements

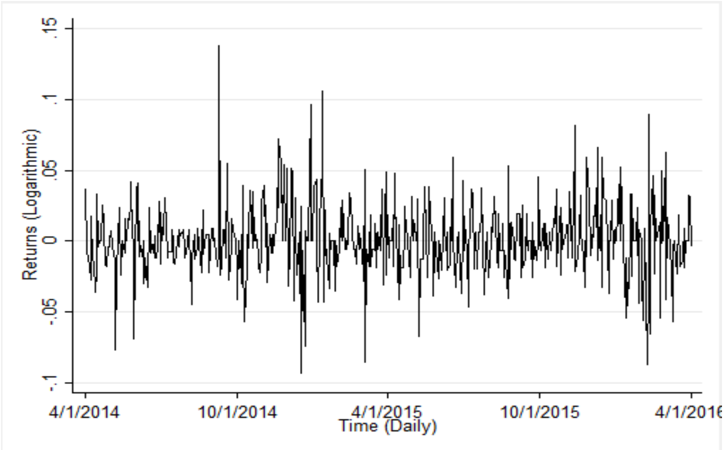
Our financial data for our two samples are collected over a time series of two consecutive years. The first step in our analysis will be to adjust the data into a functional form suitable for testing our hypothesis. An example of how closing price can move over time can be seen below in the graph. To account for any large variances in the closing prices we use the logarithmic values in our model. In the graph we can see a clear increase in closing price as time progress, indicating an upwards trend in closing price for this time series. The price development for each individual firm does not exhibit the same trend over time but the same principle applies for these firms as well.

Graph 1
Movement in price: unadjusted closing price

This graph shows the daily unadjusted closing price for SAS during the period of March 2014 to March 2016. The price development for each individual firm does not exhibit the same trend over time but this graph illustrates an example of upwards trending price movements.

In order for us to develop a model applicable for both our index and individual firms, the model needs to be generally applicable. A common approach in time series analysis is the use of a Seasonal ARIMA process to determine the right number of seasonal adjustments and lagging variables to include in the model process. By looking at historical values it is then possible to establish trends and seasonal factors affecting the time series variable. Although this process would yield a better fit for predicting and forecasting each individual company's closing price it would be one unique model for each company making the interpretation and generalisation of our results less applicable. Instead we chose to use the method of first differencing to adjust the data. Below we can see the results of such an adjustment leaving us with a stationary trend line as displayed in the graph.

Graph 2
Unadjusted closing price de-trended by first differencing



This graph shows the logarithmic daily unadjusted closing price for SAS during the period of March 2014 to March 2016. The de-trended prices do are not identical, but this graph illustrates the overall process of de-trending price movements.

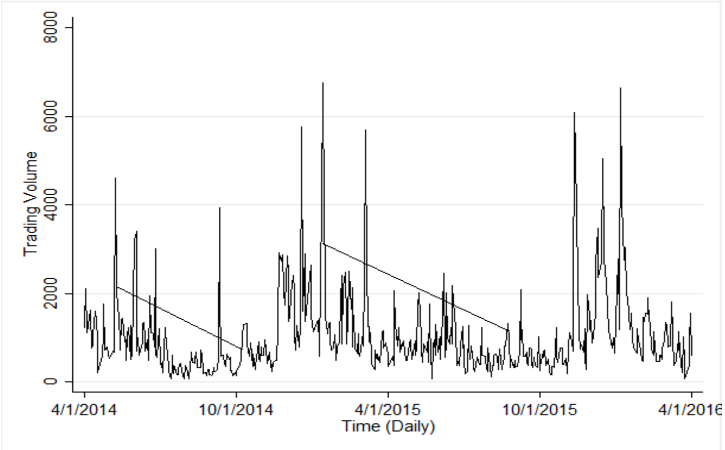
The first differencing of our closing prices is the same as price returns and we end up with the following formula for our first dependent variable.

$$RETURN_t = R_t = \ln(P_t^C) - \ln\,(P_{t-1}^C)$$

Trading volume

Our second hypothesis regards trading volume. Plotting the trading volume for one of our ten firms below, we can see a pattern displaying some seasonal trends as well as an upward trend over time.

Graph 3
Movements of logarithmic trading volume



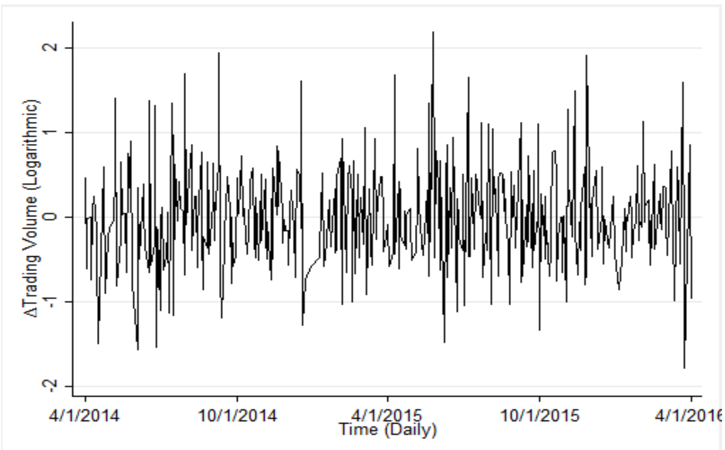
This graph shows the changes in logarithmic trading volume at a daily granularity for SAS during the period of March 2014 to March 2016. Trends in changes in logarithmic trading volume are not identical for each individual firm but this graph illustrates an example of trending changes in logarithmic trading volume.

We use the natural logarithm of trading volume to adjust for any extreme variance in trading volume and in order to adjust the variable to become stationary we use first differencing. The variable used in our regression model can be seen below.

$$\Delta TV_t = ln(TV_t) - ln\,(TV_{t-1})$$

The variable can thus be interpreted as the day-to-day change in trading volume rather than absolute values. When plotting our new variable ΔTV, the spike pattern seems to still exhibit some seasonal trend but we regard it as sufficient to use in our model.

Graph 4
Changes in logarithmic trading volume de-trended by first differencing



This graph shows the de-trended changes in logarithmic trading volume at a daily granularity for SAS during the period of March 2014 to March 2016. The de-trended changes in logarithmic trading volumes are not identical for all firms in our sample, but this graph illustrates the overall look of de-trending trading volumes.

By using a seasonal ARIMA process, the right functional form could be constructed, however as mentioned above the patterns in follows individual trends for each company making it hard to find a general model.

Volatility

The volatility on day t is defined in line with Sprenger et al. (2014), using the same measure as proposed by Parkinson (1980), and is based on intraday unadjusted high and low prices. Since we were not able to recreate intraday high and low prices for our own index the volatility measurement is only hypothesis tested on company basis.

$$VOLATILITY_t^{PARK} = V_t = \frac{(\ln(H_t - \ln(L_t)))^2}{4\ln(2)}$$

$H_t = intraday\ unadjusted\ high\ price\ on\ day\ t$

$L_t = intraday\ unadjusted\ high\ price\ on\ day\ t$

An example of the movements of volatility for one of our firms can be seen in the graph below and displays an upwards going trend over our time series. Volatility displays a similar pattern as seen in Graph 1 for the closing price trend and the same method is used here for de-trending the variable.

Graph 5
Movements of volatility

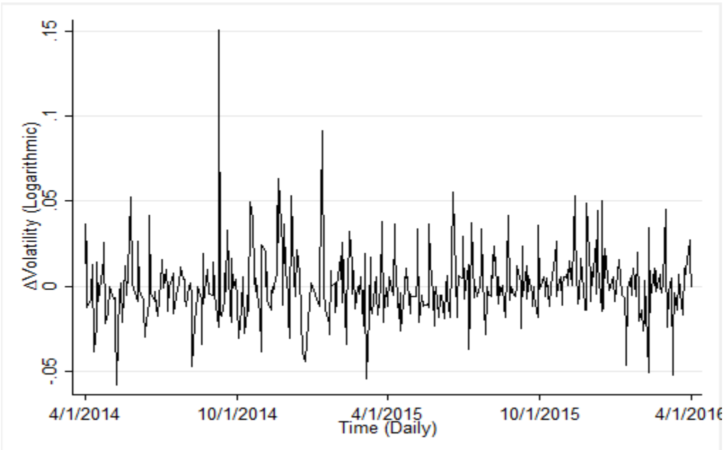
This graph shows the daily volatility, as constructed by Parkinson (1980), for SAS during the period of March 2014 to March 2016. Trends in volatility are not identical for each individual firm but this graph illustrates an example of the look of a trending volatility series.

After calculating the volatility measurement for each company we apply the same method as previously to make the variable stable.

$$\Delta VOLATILITY_t = \ln\left(VOLATILITY_t\right) - \ln\left(VOLATILITY_{t-1}\right)$$

The dependent variable volatility can now be interpreted as the day-to-day *changes* in volatility for each company. Plotting the changes in volatility we end up with a pattern similar to the two previous dependent variables

Graph 6
Volatility de-trended by first differencing



This graph shows the de-trended volatility at a daily granularity for SAS during the period of March 2014 to March 2016. The de-trended graphs are not identical for all firms in our sample, but this graph illustrates the overall look of de-trended volatility.

5.4 Model I and hypotheses

Too investigate the predictive power of social media features on the Swedish stock market we use one baseline model with different dependent variables. Each hypothesis have its own regression model. Below this Model I:

Returns - Aggregated stock-specific sentiment from social media mentions is associated with stock price returns

$$R_t = \alpha + \left(\sum_{i=1}^{5} \beta_i\right) \times R_{t-1} + \left(\sum_{i=1}^{5} \gamma_i\right) \times S_{t-1} + \left(\sum_{i=1}^{5} \delta_i\right) \times MV_{t-1} + \left(\sum_{i=1}^{5} \theta_i\right) \times A_{t-1} + \varepsilon_t$$

Trading volume - Higher volume of mentions in social media is associated with an increase in trading volume

$$\Delta TV_t = \alpha + \left(\sum_{i=1}^{5} \beta_i\right) \times R_{t-1} + \left(\sum_{i=1}^{5} \gamma_i\right) \times S_{t-1} + \left(\sum_{i=1}^{5} \delta_i\right) \times MV_{t-1} + \left(\sum_{i=1}^{5} \theta_i\right) \times A_{t-1} + \varepsilon_t$$

Volatility - Disagreement among mentions is associated with higher volatility

$$\Delta VOL_t^{PARK} = \alpha + \left(\sum_{i=1}^{5} \beta_i\right) \times R_{t-1} + \left(\sum_{i=1}^{5} \gamma_i\right) \times S_{t-1} + \left(\sum_{i=1}^{5} \delta_i\right) \times MV_{t-1} + \left(\sum_{i=1}^{5} \theta_i\right) \times A_{t-1} + \varepsilon_t$$

$R = stock\ price\ returns$

$\Delta TV = changes\ in\ trading\ volume$

$\Delta VOL = changes\ in\ volatility$

$A = agreement$

$\alpha = constant$

$\varepsilon = error\ term$

$t = time\ variable, t, measured\ in\ days$

$i = company\ and\ index$

The variables of interest are γ, δ, and θ. We use the following null hypothesis for our hypothesis testing and H_0 is rejected if our three variables of interest are significant on 10%.

Regression 1	Regression 2	Regression 3
$H_0: \gamma \leq 0$	$H_0: \delta \leq 0$	$H_0: \theta \geq 0$
$H_1: \gamma \neq 0$	$H_1: \delta > 0$	$H_1: \theta < 0$

Model development

Our first model outlines the relationship between the daily levels of our independent variables and changes in the dependent variable. The issue for our first model is that the assumption for stationarity might not hold for our independent variables and we thus adjust for any non-stationarity in these by using first differencing. Below is a presentation of the construction of the independent variables used in Model II.

$$\Delta SENTIMENT_t = SENTIMENT_t - SENTIMENT_{t-1}$$
$$\Delta MENTIONS\ VOLUME_t = MENTIONS\ VOLUME_t - MENTIONS\ VOLUME_{t-1}$$
$$\Delta AGREEMENT_t = AGREEMENT_t - AGREEMENT_{t-1}$$

Model II investigate the relationship between *changes* in *returns, trading volume and volatility* against *changes* in *sentiment, mentions volume and agreement*. We do not include lagged values of returns since we want to see the explanatory effect of social media data on our dependent variables.

5.5 Model II and hypotheses

The new baseline model for testing our hypothesis have the same dependent variables but the independent variables are now interpreted as changes in their values rather than fixed values.

Returns – Changes in aggregated stock-specific sentiment from social media mentions is associated with stock price returns

$$R_t = \alpha + \left(\sum_{i=1}^{5} \gamma_i\right) \times \Delta S_{t-i} + \left(\sum_{i=1}^{5} \delta_i\right) \times \Delta MV_{t-i} + \left(\sum_{i=1}^{5} \theta_i\right) \times \Delta A_{t-i} + \varepsilon_t$$

Trading volume - Increased volume of mentions in social media is associated with an increase in trading volume

$$\Delta TV_t = \alpha + \left(\sum_{i=1}^{5} \gamma_i\right) \times \Delta S_{t-i} + \left(\sum_{i=1}^{5} \delta_i\right) \times \Delta MV_{t-i} + \left(\sum_{i=1}^{5} \theta_i\right) \times \Delta A_{t-i} + \varepsilon_t$$

Volatility - Increased disagreement among mentions is associated with higher volatility

$$\Delta VOL_t^{PARK} = \alpha + \left(\sum_{i=1}^{5} \gamma_i\right) \times \Delta S_{t-1} + \left(\sum_{i=1}^{5} \delta_i\right) \times \Delta MV_{t-1} + \left(\sum_{i=1}^{5} \theta_i\right) \times \Delta A_{t-1} + \varepsilon_t$$

We are still interested in the same coefficients (γ, δ, θ) as in Model I but their interpretation changes. We use the following null hypothesis for our hypothesis testing and H_0 is rejected if our three variables of interest are significant on 10%.

Regression 1	Regression 2	Regression 3
$H_0: \gamma \leq 0$	$H_0: \delta \leq 0$	$H_0: \theta \leq 0$
$H_1: \gamma > 0$	$H_1: \delta > 0$	$H_1: \theta > 0$

6. RESULTS

The findings of our research are presented below. All results are presented with robust standard errors for the coefficients. The models contain five lags for each independent variable and lagged returns is used as a control variable for momentum. We include one section for each model, in section 6.1 the results from Model I is presented and in section 6.2 the results from Model II is shown.

The research question we aim to answer is:

Are stock-specific social media features associated with market returns?

General hypotheses[7] (see section 3 for test logic)**:**

1. *SENTIMENT*
   a. Aggregated stock-specific sentiment from social media mentions is associated with stock price returns
2. *ATTENTION*
   a. Increased volume of mentions in social media is associated with an increase in trading volume
3. *AGREEMENT*
   a. Increased disagreement among mentions is associated with higher volatility

6.1 Model 1 Regression

The results from our OLS-regression model I are presented below. There are three version of model I, testing *RETURNS*, ΔTV (trading volume) and $\Delta VOLATILITY$ respectively. The results for the tests on an index level will be presented first followed by the results of our test on our firm level sample. However, an index level test with the dependent variable $\Delta VOLATILITY$ is not made due to the lack of data on volatility for the constructed OMX22.

[7] Hypotheses H1A and H2A will be tested on both the index level and firm level, whereas hypotheses H3A will only be tested on our sample with ten individual firms due to the fact that no volatility variable could be constructed for our index OMX22.

Own index- OMX22

Table A in Appendix 1 shows the results of our Model I test against the OMX22. The model tests the lagged relationship between the social media features and price returns and changes in trading volume for the index.

Looking at the table, we can see that there are two instances when we have significant results for our independent variables, on *VMENTIONS_t-3* and *AGREEMENT_*t-3. Since we find no significant results in our sample on the sentiment variable, the null hypothesis for H1A cannot be rejected for our index sample. In contrast to previous studies such as Bollen et al. (2015), we find no predictive power in the level of *SENTIMENT* against returns using Model I, for our index OMX22 hence the EMH seems to hold true here.

The significant results for the *VMENTIONS* show however that returns of our index is affected by the *VMENTIONS_*t-3. The coefficient is -0.00682 on a 5% level of significance but since the value of the coefficient is so low the economic impact of *VMENTIONS* on returns is very small. The agreement variable shows significance on a 10% level with the coefficient -0.0124, which indicates a lower return for the index if there is relatively low *agreement* among mentions. Higher *AGREEMENT* three days prior to day *t* will lead to less returns.

The second test is for Δ*TV* against our three independent variables. Here we find significance for both the one and two days lag of *VMENTIONS*. For the first lag the coefficient is negative and significant on 10% level which implies that Δ*TV* the following day will decrease with 0.00248% if the *VMENTIONS* one-day prior increase by one percent. The coefficient for the second day lag is positive and significant on a 10% level and implies a one percentage increase in *VMENTIONS* two days prior will lead to a 0.00390% in Δ*TV*. The coefficient is very small but yet significant and we can thus reject the null hypothesis in H2.

The R^2 in the returns model is very small 4.9% and thus our model has little explanation to the variance seen in returns. For trading volume, the R^2 is 86.8% which makes it efficient at explaining the variation seen in the dependent variable.

Firm-level tests

Below will follow presentations of the results for Model I on our ten selected companies in our firm level sample. In the appendix, table B shows the results for the dependent variable returns, Table C shows the results for the dependent variable Δ*TV* and Table D shows results for the dependent variable Δ*VOLATILITY*.

Return as the dependent variable (Appendix 2 – Table A)

For our sample of ten firms we find a significance relationship between *SENTIMENT* and returns for *four* firms. However, the coefficients are not uniform in terms of direction for the companies and there is no distinguishable pattern for the time lag. The significant coefficients' economic impact on returns are also very small and thus we cannot reject the null hypothesis for H1 on the firm level.

The mean R^2 in our Model I for returns as dependent variable is 3.69% and our model is not very good at explaining the variation in returns for our ten companies.

VMENTIONS is only a significant predictor of returns for two out of ten firms, and here as well the beta value is very small, however significant. As such we cannot reject the null hypothesis for H2A on a firm level.

Trading volume as the dependent variable (Appendix 3 – Table B)

Looking at table B in the appendix, we can see that for *SENTIMENT* we find significant relationship with returns for six firms. The coefficients for these firms have different signs and the values for some companies are quite high, indicating that they have a significant value for the companies. However, we can see no clear pattern in the distribution of the significant lagged variables and no general conclusion can be made on the predictive power of *SENTIMENT* on Δ*TV* for our firm sample.

The independent variable *VMENTIONS* has a strong relationship for one days lag in nine out of ten firms. The coefficients are significant on a 10% level for nine firms and seven of them are significant on a 5% level. The coefficients all have a negative sign indicating that a higher VMENTIONS causes a negative change in trading volume. The null hypothesis H2 still cannot be rejected however, since we observe an opposite relationship to our hypothesis. For our firm sample, a one per cent increase in *VMENTIONS_t-1* is associated with a decrease in trading volume on day

t. Furthermore, the relationship between *AGREEMENT* among mentions and Δ*TV* seem to differ, showing some significant relationships between increased *AGREEMENT* and Δ*TV* as well as increased *disagreement* (where the coefficients are negative) and Δ*TV*, hence no clear conclusion can be drawn on their relationship.

The mean R^2 value for our ten companies is 7.49% indicating that our model does not explain a lot of the variation in changes in trading volume. Accordingly, we cannot reject the null hypothesis for H3.

Volatility as the dependent variable (Appendix 4- Table C)

From the table C in the appendix we can see that we find some significant results between *SENTIMENT* and Δ*VOLATILITY*, however the coefficients are very small and the economic importance of these values is low. The relationship between *VMENTIONS* and Δ*VOLATILITY* is not showing any clear patterns. For five of our firms there is a significant value for at least one of the lags on *VMENTIONS* but the results across our firms is not uniform. Thus we cannot reject the null hypothesis H2. Moreover, we can see that we find significance between *AGREEMENT* among mentions and Δ*VOLATILITY* in four firms. Looking at the table, the relationship is not general and we see that the coefficients show values in both positive and negative directions. For the coefficients with a positive (negative) significant value, higher *AGREEMENT* among mentions will lead to higher (lower) changes in volatility.

The mean R^2 for the model is 16.24% which is higher than the R^2 for previous dependent variables. In conclusion we can however not see any clear pattern in the predictive value of agreement, or disagreement, and volatility and hence cannot reject the null hypothesis for H3.

6.2 Model II Regressions

The results from regression using Model II are presented below. Table E in the appendix contains the results for OMX22 and tables F-H contain the results for our ten firms with each dependent variable assigned a separate table. These results are rather different in their interpretation compared to Model I. We now investigate whether *changes* in sentiments, message volume and agreement have a predictive power on any of our dependent variables. Compared to Model I which explores the predictive power of an absolute value on a change in the dependent variable this models captures the movements (or *delta)* of the social media features. All of our variables can still be interpreted as elasticities (being logarithmic functions) except for agreement.

We test our hypotheses H1A-H3A on our ten firms but only hypotheses H1A and H2A on OMX22, since the lack of volatility data still prevails.

Own index- OMX22 (Appendix 6 – Table E)

Table E in the appendix shows the results from Model II on the daily returns as dependent variable as well as changes in trading volume. Looking at the table, can see that one of our independent variables shows significant results in relation to returns. It is the third day lag on Δ*SENTIMENT* that shows a positive coefficient of 0.00924 on a 5% significance level. Although the coefficient is significant the economic impact on returns is still small, for a 1% change in returns *SENTIMENT* would have to change by close to 110% in one day. The general pattern, although insignificant, is that *SENTIMENT* have positive coefficients for all lags and *VMENTIONS* and *AGREEMENT* seems to follow the same pattern of positive coefficients for the first two days followed by a reversal effect in the two following lags.

In the second regression the dependent variable is Δ*TV*. Here none of our variables show statistical significance and we cannot draw any conclusions regarding the impact of our independent variables on trading volume for our index level.

The R^2 for both our regressions are very small (2-3%) indicating that the models used is not good for explaining and estimating daily returns, this might be because we look at an index were more noise can be present in the data compared to individual stocks.

In summary, we can reject the null hypothesis for H1 on our tests against OMX22 since changes sentiment show a significant, lagged, relationship with returns. However, we cannot to reject the null hypothesis for H2A, for all lags and variables.

Firm-level tests

The results using Model II on our firm level sample are presented below. Table F in the appendix show the results for the dependent variable return, Table G in the appendix shows the results for the dependent variable changes in trading volume and Table H, also in the appendix, shows results for the dependent variable changes in volatility.

Return as the dependent variable (Appendix 7 – Table F)

Using our second set of models we start by analysing the relationship between returns and *changes* in the independent variables. For our sample of ten companies we only find a significant relationship in one case between returns and Δ*SENTIMENT*. We can further see that the coefficients are very small and there seems to be little to zero economic impact on stock price returns for our sample. When looking at our second independent variable, Δ*VMENTIONS* we find significant results in two of our companies. The coefficients are significant on a 5% level and all show positive values. However, no general conclusion can be made since only two of our companies display this relationship. Δ*AGREEMENT* has significant values, on returns, for two firms but due to no clear pattern and differing sign for the coefficients, no clear conclusion can be made.

The R^2 for the model is quite low for all companies indicating that our model does not explain much of the variation in returns on a firm level. The average R^2 for our sample is 3.29% which is higher than for the index model.

Trading volume as the dependent variable (Appendix 8 – Table G)

The changes in our independent variables seem to have some impact when it comes to changes in trading volume. Nine out of ten firms display significant results for Δ*VMENTIONS* for at least one of the lagging variables. Eight of these companies do it for the first lag and there is a clear pattern for our firm sample. The coefficients are all *negative* which implies an increase (decrease) in mentions volume will cause negative (positive) change in trading volume the following day. The results are significant on a 5% level for six of the companies and on a 1% level for three companies. The other lags for Δ*VMENTIONS*, although not displaying the same statistical significance, have the same direction as the first lag in almost every case, but the coefficients' values are decreasing over time. The coefficient value for the first lag in our sample is 0.103 implying a 0.1% increase in Δ*TV* volume is associated with a 1% increase in Δ*VMENTIONS* on the previous day.

The other independent variables, Δ*SENTIMENT* and Δ*AGREEMENT*, do not display such a clear pattern as Δ*VMENTIONS*. For the ΔSENTIMENT we observe significant results on the first two lags in four companies. Δ*AGREEMENT* do have a few significant results scattered over the various lags but it is hard to determine a broader effect for our sample. Based on these results, we cannot reject the null hypothesis for H2.

The mean R^2 for the sample is quite low, 6.06%, but it is higher than the model II looking a return as dependent variable, indicating that this model is doing a better job in explaining the variance of ΔTV than it is at predicting returns.

Volatility as the dependent variable (Appendix 9 – Table H)

Looking at Table H in the appendix, the results show scattered significance for some values of the independent variables. The Δ*SENTIMENT* have significant values for three of the firms in our sample. The significant coefficients for these firms do not show a consistent pattern however. Δ*VMENTIONS a*re significant for four out of ten firms, but the coefficients are very small. There seems to be no relationship between Δ*VMENTIONS* and Δ*VOLATILITY* in our sample. Δ*AGREEMENT* show significance for two firms, but the directions of the coefficients are positive for one company and negative for the other. The coefficients' values do not indicate any substantial economic explanation for Δ*AGREEMENT* and Δ*VOLATILITY.* We thus cannot reject the null hypothesis for H3 for our firm level sample.

The mean R^2 for our sample is 3.84 %, which is somewhat higher than for the model II with return as dependent variable, but it is still very small and we thus conclude that there are probably other factors better explaining the variance of the volatility than we have included in our model.

7. DISCUSSION

We have used a similar baseline model to existing research on the subject but in contrast to Tetlock (2007), Antweiler and Frank (2004) and Sprenger et al. (2014), we have not performed an event study looking at abnormal returns, but instead we have investigated time-series data with a multivariate OLS-model adjusted for internal structures in the data. So while our results can be compared to aforementioned research, we cannot completely benchmark our study against them. Most other studies find some predictive relationship between social media data and market features, whether by analysing abnormal returns (Tetlock, 2007) or regressing time-series data (Bollen et al. 2015). The main results of our statistical tests show significant predictive value for lagged changes in sentiment over price returns at the index level and a negative relationship between mentions volume and trading volume at both index and firm level, however we find no pattern on the significant values for agreement (or changes in agreement) among mentions and volatility (or changes in volatility) on firm or index level. At the firm level we find no significant relationship between changes in sentiment and price returns, the difference in these results on changes in sentiment and price returns is probably due to firm effects for our individual firms, whereas our larger sample at the index level show the more general relationship. Thus we conclude that *for our sample period over the Swedish market, lagged changes in sentiments on social media has significant predictive value over price returns*. However, with a relatively small economic implication. Moreover, most of the previous studies have, as aforementioned, performed tests on abnormal returns or for longer time periods by performing event studies whereas we are looking at a daily granularity using time-series analysis. Our significant coefficients for changes in sentiment are all positive in our model II, however they are relatively low compared to previous event studies. Such studies isolate event windows with a particularly high frequency in social media mentions and thus observe higher significant coefficients in their regression models. We observe the same relationship but with a lesser economic importance.

Our result on sentiment and asset prices on our tests against our adjusted OMX30 index, OMX22, is in line with results from previous research and to some extent we extend the generalisability of these results to the Swedish market. Moreover, our results are in line with behavioural finance theories on irrational noise traders, trading on held beliefs of sentiment and driving prices away from fundamental asset values (Black, 1986, Schleifer and Summers, 1990, De Long, 1990) spurred by word of mouth or *bounded rationality* (DeMarzo et al, 2003). This stands in contrast to theories by Fame (1965), claiming that the rational investors affect the asset prices, not noise traders.

Our results on the relationship between volume of mentions and trading volume indicate that a rise in mentions will lead to a decrease in trading volume, both for absolute values (model I) and delta values for the independent variable (model II). These results are opposite to our hypothesis H2, that increased mentions would be associated with an increase in trading volume. These results also stand in contrast to research like Cao et al. (2003) and salience theory (Solomon et al, 2012), stating that media steers attention of investors towards particular assets so that the more an asset is mentions the more the demand for the asset will increase – thus increasing trading volume. It is possible however, that in some cases, returns are driving mentions and sentiment and not the other way around. Such a relationship would help explain why we observe a negative relationship between mention volume and trading volume for both our models. If returns are driving mentions, the decrease in trading volume we observe could be a reversal effect from previously high returns, driving mentions volume. We tested this reversed relationship on our sample of ten firms, but only found a significant relationship between returns and mentions volume for one firm (see appendix

10 – Table I) and thus this offers no explanatory value and we can draw no conclusion on their being a revers relationship.

Do our proxies measure investor sentiment and attention?

In some respect our unique social media data set differentiates our study even further from previous research, aside from the sample geography. Previous research on social media has typically focused on either twitter data (Bollen et al. 2015, Ranco, Aleksovski, Caldarelli, Grčar and Mozetič, 2015) or Google Trends data (Preis et al. 2010, Molnár et al. 2016). Our study, in contrast, has social media data from a range of platform such as Twitter, Google+, news sites, RSS feeds, Instagram etc. Investor sentiment can never be directly measured, but in previous research using sentiment classified Twitter content as a proxy for investor sentiment has shown significant results and seems to be a solid proxy, the same holds true for stock-specific firm searches on Google Trends but as a proxy for attention, the question is whether our scattered data sources are as solid. Having our data collected from a range of social media platform might capture a lot of noise that studies using only one specific platform (like Twitter) are not affected by. Moreover, data sources such as news sites or RSS feeds are very hard to analyse and we have no information on what type of properties these sites or RSS feeds have, in terms of following, readers, legitimacy etc. It is possible that thousands or tens of thousands of people have read some mentions in our data whether for some, only a handful of people may have read their content. Obviously this is an issue we cannot tackle due to lack of following, or readership data from all of the social media platforms. But ideally, we would weigh the mentions by importance in terms of the number of people having read, or possibly being directly reached by the content. As of now, all mentions are treated equally, in terms of impact as well as the degree of sentiment expressed, since we have no intensity range for our sentiments but instead use binary classifications. The issue with readership, or following, for some of the data sources likely affect our attention and sentiment proxies and probably does not capture investor sentiment or attention as effectively as the use of only Twitter data would. So, there is a lot of potential noise likely affecting our proxies and social media features in turn affecting our results and can probably explain, at least partly, some reasons as to why our results are differing in some regards from similar studies on US markets.

Difference in social media use

Generally social media in Sweden is probably not used as widely as in US markets, both overall and as a source of information. Hence, the word of mouth effect on social media might not be present in the same regard in Sweden as in North America and any noise traders might not react to or become influenced by information, or opinions on social media as quickly as for US markets. This slowness would also help explain why we find the significance of changes in sentiment over price return on day t-3, showing a lagged relationship, which is an occurrence not observed in previous research.

7.1 Sensitivity analysis

In Model II the lagged control variable return is omitted to only capture the explanatory power of the social media variables. By including it we can see if Model is affected by an omitted variable bias when returns are excluded. For the first regression using Model II with returns as dependent variable, the average R^2 increases from 3.29% to 4.7%. Omitting the lagged returns as a control variable has a negative effect for the model as such but we are able to single out the effect of the social media data. There are no changes in significance for the independent variables and the coefficients show little to no difference between the two models.

The second regression yields a mean R^2 of 7% compared to a mean of 6.06% for Model II excluding lagged returns. The coefficients are not affected and are significant on the same level as for the baseline Model II – omitting the lagging returns. This implies that the use of lagging variables of return in the second models has little to zero impact in explaining the variation in changes in trading volume.

For our third regression model, with changes in volatility as the dependent variable, there is a large difference when including the lagged returns. The average sample R^2 increases from 3.48% to 16.09%, further there is a significant change for some of the coefficients as lagged returns are included. For the third regression model the use of these lagged returns have a significant impact on changes in volatility and explains a lot of the variation seen.

Stationarity

Stationarity implies that our times series' mean and variance are constant over time When adjusting our variables to become stationary we use the first differencing. While this method works fairly well for us we have not fully adjusted for trends or seasonal effects that might better explain the variations in our variables. The assumption about stationarity is made in order to identify the true effect of the independent variables. If a trend or seasonal effect is still present after the adjustments have been made it will be captured in either the coefficients of our independent variable or in the estimated standard errors depending on what factor are violated in the stationarity assumption.

7.2 Sample bias

For our social media data, we have a sample bias since the data from Modular Streams has a maximum level of mentions gathered on 2000 per company. This implies that the social media data for companies with many mentions is collected on a shorter time period than for companies with less dense sentiment mentions. The use of social media data as a predictor of market returns has been extremely dense in previous research (e.g. Ranco, Aleksovski, Caldarelli, Grčar and Mozetič, 2015). When there is a limited amount of mentions for each company, any effect on the financial markets become harder to identify. A firm with much social media attention will only have data on a small part of the time series and have the same level of mentions as firms with less dense frequency of mentions. Because the raw data was collected by an external source we cannot adjust for this bias without imposing strong assumptions for the data. Further the classification process used has a hit ratio of 90%. The accuracy will not be of any problem as long as the wrongly classified mentions are not caused by systematic errors in the classification process. We can then assume that the wrongly classified sentiment has an expected mean equal to zero.

7.3 Robustness tests

Below follow the results and discussion on our robustness tests.

Autocorrelation (Appendix 11 – Table J)

If a variable in a model is dependent on previous values of itself, there is a presence of autocorrelation in the data. We test for autocorrelation in our model using the Durbin Watson test. The results are different depending on which dependent variable we are looking at. For changes in volume we can only see one occasion of autocorrelation being positive. Changes in returns indicate that there are several occasion of autocorrelation both in our first and second model. The first model where we include lagged changes in returns do not show any autocorrelation for our ten companies but it does however displays some autocorrelation for the OMX22. The second model where changes in returns are excluded, autocorrelation is positive for six of our firms and the OMX22. For the third dependent variable, changes in volatility, both Model I and II display signs of autocorrelation. To adjust for autocorrelation in our regressions we have used the first differencing of all of our dependent variables, we also included lagged changes in returns for our model I. The autocorrelation could be adjusted for further by using a Seasonal ARIMA (p, q) model, but finding the correct number of p lags and the moving average term q for each lag is out of the scope for this thesis.

Heteroscedasticity (Appendix 12 – Table K)

The variance of the standard errors is assumed to be constant for all independent variables. When this assumption does not hold our coefficients' values might still be accurate but the predicted standard errors will not be valid. Any effect of heteroscedasticity will then be present in our hypothesis testing and t-statistics, thus making it possible that we reject H0 when it should have been accepted. In our regression models we therefore test for heteroscedasticity using a Breusch-Pagan test. For our models there is sometimes heteroscedasticity present and in other cases it is not. We therefor to correct this issue by using robust standard errors for all our regressions.

Multicollinearity (Appendix 13 – Table L)

Multicollinearity between the independent variables can occur when two or more of the independent variables in the model are highly correlated with one another. This will not violate any of the OLS assumptions i-iv, but there will be a difficulty in interpretation of the coefficients. If two independent variables have a high correlation their unique effect on the dependent variable will be difficult to distinguish. The effect of multicollinearity in a regression model will be higher variance. We use a variance inflation factor (VIF) test to establish whether or not our coefficients might be affected by any correlation between the independent variables. The test looks at how much of the variation in the independent variable *i* is explained by the other independent variables and the R^2 from the regression on the independent variable on all other independent variable is then used in the below formula:

$$VIF = \frac{1}{1 - R_i^2}$$

The VIF value used as a cut off point for establishing multicollinearity is 10, implying that at least 90% of the variation in the independent variable i can be explained by the other independent variables.

For all our regression the VIF values are below 10 with one exception. One of the firms shows a high multicollinearity for some of their independent variables.

8. CONCLUSION

This study aims to examine the relationship between social media and the Swedish stock market. More precisely we measure the relationship between social media features sentiment, mentions volume and agreement among mentions against the main market returns stock price return, trade volume and volatility. By conducting a time-series analysis on daily price and social media data, we construct two OLS-regressions models that are used to test our hypotheses for our sample period of Swedish firms between March 2014 and March 2016. We build two regression models, where model II differs from model I tests by including the delta values of the independent variables instead of absolute values.

Our main result is that, on an index level, using model II, lagged *changes* in sentiment have a significant predictive power over stock price returns for the Swedish market. These results confirm both theory of behavioural impact of irrational investors, or noise traders, on asset prices as well as previous research on the subject outside of the Swedish context. We also find a significant *negative* relationship between the volume of mentions and trading volume, in contrast to previous findings by Sprenger et al. (2014) and others. However, we find no significant relationship between the agreement (or disagreement) among mentions on social media platforms and volatility.

We believe our results, although preliminary, can be useful for investors looking for an edge in the Swedish stock market, since sentiment on social media mentions contain valuable information that should not be neglected in the investment process, especially for traders with a short investment window. The value of social media data for financial markets has been observed before on larger markets, Azar and Lo (2016) show that a tweet-based asset-allocation strategy outperforms most benchmarks, and this study extends that conclusion to the Swedish stock market, however with a lagged effect. The implications of our preliminary results for market participants, are that social media content may contain valuable information for predicting market returns. In conclusion, for our index, OMX22, the answer to our research question is ***yes***.

8.1 Future research

Even if our results show promise, they offer no insights into the extent to which a causal relationship between social media data and stock price returns exists. This would be an interesting area for future research, where the causative relationship is examined more thoroughly. For the scope of this study we have not consider whether the content on social media is a reflection of traditional news or if the content contains any new information, this as well would be an interesting question to examine. Moreover, as shown by Souza and Aste (2016) there is some evidence suggesting that there is in fact a causal relationship between social media and asset prices. Although, they only find this being observed when performing non-linear (asymmetrical) statistical methods, and conclude that for their sample, the social media significance on stock's returns is *purely non-linear.* Similar methods could be applied in the Swedish context to examine the casual relationship between social media and asset prices.

8.2 Validity, reliability, and generalizability

We considered the validity of our study to be high as we have constructed models and proxies in line with common methods used in previous research. However, investor sentiment and attention cannot be directly observed, hence we have created proxies in order to examine our research question. There are some weaknesses for using proxies like this, as we cannot be sure that they catch what they intend to catch, which should be kept in mind when looking at the results we get. It is also worth to note that the sentiment classification in our social media data is 90 %, high among classifiers but with 10 % error over time. These potential classification errors could affect our results, although probably in minor or neglectable ways.

The reliability of our study can be said to be relatively high. The financial data was collected from a legitimate source, Thomson Reuters DataStream. In order to mitigate any effect on our results from extreme values, and verify the data, we transformed all of our variables (except AGREEMENT) into logarithmic functions. Our samples consist of 22 firms, forming an index, as well as ten individual firms. It could be said that the sample of individual firms suffer from some sample bias, however our significant results were mainly found in tests on our larger sample. Moreover, both the research methodology as well as baseline regression model (model I) relies on earlier research on the subject, as such the replicability of the study is concluded to be good.

Our study is the first on the Swedish market examining the relationship between social media data and stock price return, trading volume and volatility. Although our preliminary results on H1A are in line with previous research outside the Swedish context, however with a lagged relationship, it is not unproblematic to claim that our results are generalizable. Our study only covers a time period of two years, and a total of 29 firms, which is a relatively small sample of the total firms on the Swedish market, thus it is not possible generalise our results. Also, there is the question of causality, which need to be examined further.

REFERENCES

Alanyali, M., Moat, H.S. & Preis, T. 2013, "Quantifying the relationship between financial news and the stock market", *Scientific reports,* vol. 3.

Allen, D.E., McAleer, M. & Singh, A.K. 2015, "Daily Market News Sentiment and Stock Prices", .

Antweiler, W. & Frank, M.Z. 2004, "Does talk matter? Evidence from a broad cross section of stocks", *University of British Columbia Working Paper,* .

Frank, M.Z. & Antweiler, W. 2002, "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards", *The Information Content of Internet Stock Message Boards (August 21, 2001).AFA,* .

Arkes, H.R., Herren, L.T. & Isen, A.M. 1988, "The role of potential loss in the influence of affect on risk-taking behavior", *Organizational behavior and human decision processes,* vol. 42, no. 2, pp. 181-193.

Azar, P. & Lo, A.W. 2016, "The Wisdom of Twitter Crowds: Predicting Stock Market Reactions to FOMC Meetings via Twitter Feeds", *Available at SSRN 2756815,* .

Bachelier, L. 1900, "Théorie de la spéculation", .

Barber, B.M. & Odean, T. 2000, "Trading is hazardous to your wealth: The common stock investment performance of individual investors", *The journal of Finance,* vol. 55, no. 2, pp. 773-806.

Barber, B.M. & Odean, T. 1999, "The courage of misguided convictions", *Financial Analysts Journal,* vol. 55, no. 6, pp. 41-55.

Barber, B.M., Odean, T. & Zhu, N. 2006, "Do noise traders move markets?", *EFA 2006 Zurich meetings paper*.

Barberis, N. & Huang, M. 2007, "Stocks as lotteries: The implications of probability weighting for security prices", .

Berglin, H, Möllborg, K. 2015, "Predicting abnormal stock returns and trading volumes using Google Search", .

Bijl, L., Kringhaug, G., Molnár, P. & Sandvik, E. 2016, "Google searches and stock returns", *International Review of Financial Analysis,* vol. 45, pp. 150-156.

Black, F. 1986, "Noise", *The journal of finance,* vol. 41, no. 3, pp. 528-543.

Bollen, J., Mao, H. & Zeng, X. 2011, "Twitter mood predicts the stock market", *Journal of Computational Science,* vol. 2, no. 1, pp. 1-8.

Bommel, J.V. 2003, "Rumors", *The Journal of Finance,* vol. 58, no. 4, pp. 1499-1520.

Boyle, G., Hagan, A., O'connor, R.S. & Whitwell, N. 2004, "Emotion, fear and superstition in the New Zealand stockmarket", *New Zealand Economic Papers,* vol. 38, no. 1, pp. 65-85.

Cao, H.H., Coval, J.D. & Hirshleifer, D. 2002, "Sidelined investors, trading-generated news, and security returns", *Review of Financial Studies,* vol. 15, no. 2, pp. 615-648.

Danthine, J. & Moresi, S. 1993, "Volatility, information and noise trading", *European Economic Review,* vol. 37, no. 5, pp. 961-982.

Das, S., Martínez-Jerez, A. & Tufano, P. 2005, "eInformation: A clinical study of investor discussion and sentiment", *Financial Management,* vol. 34, no. 3, pp. 103-137.

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

DeMarzo, P.M., Zwiebel, J. & Vayanos, D. 2001, "Persuasion bias, social influence, and uni-dimensional opinions", *Social Influence, and Uni-Dimensional Opinions (November 2001).MIT Sloan Working Paper,* , no. 4339-01.

Edlund, P. 1989, "Preliminary estimation of transfer function weights: a two-step regression approach", .

Frank, M.Z. & Antweiler, W. 2002, "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards", *The Information Content of Internet Stock Message Boards (August 21, 2001).AFA,* .

Friedman, M. 1953, "The methodology of positive economics", .

Harris, M. & Raviv, A. 1993, "Differences of opinion make a horse race", *Review of Financial Studies,* vol. 6, no. 3, pp. 473-506.

Hirshleifer, D. & Teoh, S.H. 2003, "Limited attention, information disclosure, and financial reporting", *Journal of Accounting and Economics,* vol. 36, no. 1, pp. 337-386.

Kaplanski, G. & Levy, H. 2010, "Sentiment and stock prices: The case of aviation disasters", *Journal of Financial Economics,* vol. 95, no. 2, pp. 174-201.

Lemmon, M. & Portniaguina, E. 2006, "Consumer confidence and asset prices: Some empirical evidence", *Review of Financial Studies,* vol. 19, no. 4, pp. 1499-1529.

Liew, J.K. & Budavári, T. 2016, "Identification of a Social Media Equity Factor Derived Directly from Tweet Sentiments", *Available at SSRN 2711825,* .

Lo, A.W. & MacKinlay, A.C. 2002, "A non-random walk down Wall Street", .

Lo, A.W. & MacKinlay, A.C. 1988, "Stock market prices do not follow random walks: Evidence from a simple specification test", *Review of Financial Studies,* vol. 1, no. 1, pp. 41-66.

Lundell, E, Shuangyu, P. 2015, "Investor sentiment and stock returns", .

Madiha, L, Arshad, S, Fatima, M, Farooq, S. 2011, " Market efficiency, market anomalies, causes, evidences, and some behavioral aspects of market anomalies ", *Research Journal of Finance and Accounting* 2.9/10 (2011): 1-14

Mao, H., Counts, S. & Bollen, J. 2014, "Quantifying the effects of online bullishness on international financial markets", *ECB Workshop on Using Big Data for Forecasting and Statistics, Frankfurt, Germany*.

Mao, Y., Wei, W., Wang, B. & Liu, B. 2012, "Correlating S&P 500 stocks with Twitter data", *Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research*ACM, , pp. 69.

Nygren, T.E., Isen, A.M., Taylor, P.J. & Dulin, J. 1996, "The influence of positive affect on the decision rule in risk situations: Focus on outcome (and especially avoidance of loss) rather than probability", *Organizational behavior and human decision processes,* vol. 66, no. 1, pp. 59-72.

Preis, T., Reith, D. & Stanley, H.E. 2010, "Complex dynamics of our economic life on different scales: insights from search engine query data", *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences,* vol. 368, no. 1933, pp. 5707-5719.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M. & Mozetič, I. 2015, "The effects of Twitter sentiment on stock price returns", *PloS one,* vol. 10, no. 9, pp. e0138441.

Regnault, J. 1863, "Calcul des chances et philosophie de la bourse", .

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E. & Cohen, J.D. 2003, "The neural basis of economic decision-making in the ultimatum game", *Science,* vol. 300, no. 5626, pp. 1755-1758.

Shefrin, H. 2001, "Behavioral corporate finance", *Journal of Applied Corporate Finance,* vol. 14, no. 3.

Shiller, R.J. 2005, "Behavioral economics and institutional innovation", .

Solomon, D.H. 2012, "Selective publicity and stock prices", *The Journal of Finance,* vol. 67, no. 2, pp. 599-638.

Souza, T.T. & Aste, T. 2016, "A nonlinear impact: evidences of causal effects of social media on market prices", *arXiv preprint arXiv:1601.04535,* .

Sprenger, T.O., Tumasjan, A., Sandner, P.G. & Welpe, I.M. 2014, "Tweets and trades: The information content of stock microblogs", *European Financial Management,* vol. 20, no. 5, pp. 926-957.

Tamerus, S., Van de Looij-Jansen, P., Baerveldt, C. & De Waart, F. 2014, "De samenhang tussen de kwaliteit van de wijk en psychosociale problemen bij jongeren in Rotterdam", *Tijdschrift voor Gezondheidswetenschappen,* vol. 92, pp. 163.

Tetlock, P.C. 2007, "Giving content to investor sentiment: The role of media in the stock market", *The Journal of Finance,* vol. 62, no. 3, pp. 1139-1168.

Tversky, A. & Kahneman, D. 1974, "Judgment under uncertainty: Heuristics and biases", *Science,* vol. 185, no. 4157, pp. 1124-1131.

Tversky, A. & Kahneman, D. 1973, "Availability: A heuristic for judging frequency and probability", *Cognitive psychology,* vol. 5, no. 2, pp. 207-232.

Zhang, X., Fuehres, H. & Gloor, P.A. 2011, "Predicting stock market indicators through twitter "I hope it is not as bad as I fear"", *Procedia-Social and Behavioral Sciences,* vol. 26, pp. 55-62.

APPENDIX

Appendix 1 - Index creation

For our 22 firms on the OMX30 we create a separate stock index. The firms are denoted as $f_1, \dots, f_n$ for $n = 23$ firms. We define the index for day t as follows:

$$OMX23_t = \sum_{f=1}^{23} P_{ft}^{c} \times weight_f$$

Where P_t^c the unadjusted closing price on day t, and weight is the importance of each firm. We define the important for each company in terms of market cap as:

$$weight_f = \frac{MarketCap_f}{\sum_{f=1}^{23} MarketCap_f}$$

For the index trade volume on day t we define:

$$TV_t^{OMX23} = \sum_{f=1}^{23} TV_{ft} \times weight_f$$

Appendix 2 - Table A

Index level regression Model I – OMX22

VARIABLES	Return	ΔTrading Volume
RETURN_t-1	-0.0925*	0.984***
	(0.0557)	(0.0227)
RETURN_t-2	-0.0378	0.0328
	(0.0539)	(0.0207)
RETURN_t-3	0.129**	0.0206
	(0.0571)	(0.0196)
RETURN_t-4	-0.0158	0.0477*
	(0.0501)	(0.0277)
RETURN_t-5	-0.0288	0.0135
	(0.0527)	(0.0332)
SENTIMENT_t-1	0.00343	0.000480
	(0.00343)	(0.00132)
SENTIMENT_t-2	0.00147	0.000776
	(0.00326)	(0.00108)
SENTIMENT_t-3	0.00301	-0.000422
	(0.00337)	(0.00162)
SENTIMENT_t-4	-0.00275	0.000429
	(0.00355)	(0.00143)
SENTIMENT_t-5	-0.00443	0.000294
	(0.00312)	(0.00149)
VMENTIONS_t-1	6.67e-06	**-0.00248***
	(0.00327)	(0.00137)
VMENTIONS_t-2	-0.00151	**0.00390***
	(0.00311)	(0.00200)
VMENTIONS_t-3	**-0.00682****	-0.00229
	(0.00336)	(0.00214)
VMENTIONS_t-4	0.00207	-0.00157
	(0.00341)	(0.00139)
VMENTIONS_t-5	0.00269	0.00228
	(0.00300)	(0.00180)
AGREEMENT_t-1	0.00407	-0.00161
	(0.00783)	(0.00307)
AGREEMENT_t-2	-0.000803	0.00538
	(0.00698)	(0.00392)
AGREEMENT_t-3	**-0.0124***	-0.000870
	(0.00709)	(0.00379)
AGREEMENT_t-4	-0.00377	-0.00354
	(0.00703)	(0.00298)
AGREEMENT_t-5	0.00438	0.00268
	(0.00689)	(0.00347)
Constant	0.00952	-0.00188
	(0.0162)	(0.00809)

Observations	518	415
R-squared	0.049	0.868

Results presented above are attained by an OLS-regression using our model I. This table shows the lagged relationship, on an index level, between the social media features (SENTIMENT, VMENTIONS, AND AGREEMENT) and firms' returns and changes in trading volume for the time period March 2014 to March 2016. Return is the difference between the logarithmic value of the unadjusted closing price and unadjusted opening price. ΔTrading Volume *is the first difference of the trading volume. All independent variables are lagged for a period of five days. SENTIMENT is the aggregated sentiment score. VMENTIONS is the volume of mentions in social media. AGREEMENT is the agreement among mentions, constructed in line with Antweiler and Frank (2004) as well as Sprenger et al. (2014). RETURN is used as a control variable to capture any momentum effect. Due to the variables being logarithmic functions, the coefficients should be interpreted as elasticities except for AGREEMENT, which is not logarithmic. We use robust standard errors, shown in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level.*

Appendix 3 - Table B

Firm level regressions Model I – Return as the dependant variable

VARIABLES	(1) Return	(2) Return	(3) Return	(4) Return	(5) Return	(6) Return	(7) Return	(8) Return	(9) Return	(10) Return
RETURN_t-1	-0.0853	0.0124	-0.00721	-0.0556	0.0512	-0.0791*	-0.0189	0.0205	-0.0832	-0.0245
	(0.0533)	(0.0501)	(0.0101)	(0.0565)	(0.0550)	(0.0438)	(0.0529)	(0.0509)	(0.0528)	(0.0507)
RETURN_t-2	0.0281	-0.0865	5.72e-06	-0.0583	-0.0309	-0.0473	-0.0731	-0.0425	-0.0547	-0.00703
	(0.0346)	(0.0631)	(0.0277)	(0.0526)	(0.0450)	(0.0468)	(0.0483)	(0.0447)	(0.0594)	(0.0510)
RETURN_t-3	-8.90e-05	0.0139	-0.0313**	0.0716	0.0750*	0.0670	0.0219	0.0801*	0.116**	0.0455
	(0.0348)	(0.0412)	(0.0159)	(0.0529)	(0.0446)	(0.0523)	(0.0523)	(0.0431)	(0.0550)	(0.0455)
RETURN_t-4	0.0449	-0.0395	-0.0232	-0.0793	-0.0498	-0.0443	-0.0738*	0.0121	-0.0318	-0.0217
	(0.0418)	(0.0460)	(0.0177)	(0.0553)	(0.0546)	(0.0441)	(0.0402)	(0.0486)	(0.0513)	(0.0458)
RETURN_t-5	0.0829**	-0.0371	-0.0102	0.0128	0.00418	-0.0428	-0.0578	-0.0576	-0.0658	-0.00258
	(0.0398)	(0.0483)	(0.0259)	(0.0641)	(0.0476)	(0.0549)	(0.0468)	(0.0463)	(0.0504)	(0.0496)
SENTIMENT_t-1	0.00352	0.00355	-0.00189	-0.000218	0.00106	**0.00233***	0.00151	-1.27e-05	0.000536	0.00263
	(0.00587)	(0.00276)	(0.00253)	(0.00147)	(0.00167)	(0.00136)	(0.00136)	(0.00257)	(0.00107)	(0.00228)
SENTIMENT_t-2	0.00442	0.00459	-0.000141	-3.87e-05	-0.00145	0.000138	-0.000110	-0.000511	0.000624	-0.000789
	(0.00349)	(0.00284)	(0.00382)	(0.00189)	(0.00137)	(0.00143)	(0.00163)	(0.00198)	(0.00109)	(0.00223)
SENTIMENT_t-3	-7.47e-05	**0.00469***	-0.000202	-0.000195	-0.00188	0.000516	-0.000108	-0.00208	-0.000407	6.62e-05
	(0.00364)	(0.00242)	(0.00167)	(0.00175)	(0.00151)	(0.00130)	(0.00129)	(0.00175)	(0.00116)	(0.00194)
SENTIMENT_t-4	**-0.0109****	-0.00547	0.00175	**0.00268***	0.000929	0.00102	-0.00156	1.69e-05	0.00156	-0.00293
	(0.00453)	(0.00425)	(0.00168)	(0.00161)	(0.00244)	(0.00130)	(0.00156)	(0.00212)	(0.00114)	(0.00232)
SENTIMENT_t-5	-0.00239	0.00220	-0.00290	-0.000714	-0.00166	0.00109	-0.00119	-0.00278	-0.000262	0.00263
	(0.00358)	(0.00246)	(0.00242)	(0.00136)	(0.00144)	(0.00156)	(0.00136)	(0.00198)	(0.00100)	(0.00217)
MVOLUME_t-1	-0.00145	**-0.00669*****	-0.00141	0.00261	0.000879	-0.00164	-0.00174	0.000830	0.000597	0.00192
	(0.00477)	(0.00256)	(0.00192)	(0.00178)	(0.00176)	(0.00142)	(0.00169)	(0.00267)	(0.00150)	(0.00241)
MVOLUME_t-2	-0.000604	**-0.00560***	-0.00832	-0.000919	0.000915	2.15e-05	0.000638	0.00106	-0.000576	-0.00206
	(0.00362)	(0.00296)	(0.00678)	(0.00221)	(0.00175)	(0.00166)	(0.00179)	(0.00196)	(0.00136)	(0.00218)
MVOLUME_t-3	-0.00197	-0.00163	0.00332	-0.00338*	-0.000795	-0.000295	-0.000217	0.00123	0.00130	0.000747
	(0.00341)	(0.00257)	(0.00266)	(0.00180)	(0.00166)	(0.00147)	(0.00163)	(0.00173)	(0.00147)	(0.00230)
MVOLUME_t-4	0.00324	0.00409	0.000558	-2.11e-05	-0.00199	-0.00226	0.00206	0.000870	0.000774	0.000831
	(0.00433)	(0.00463)	(0.00262)	(0.00167)	(0.00210)	(0.00150)	(0.00167)	(0.00203)	(0.00133)	(0.00219)
MVOLUME_t-5	0.00142	-0.00190	-0.00167	-0.000877	0.00261	0.00106	0.000389	0.00150	0.000146	0.00112
	(0.00319)	(0.00280)	(0.00158)	(0.00157)	(0.00185)	(0.00165)	(0.00145)	(0.00202)	(0.00121)	(0.00227)
AGREEMENT_t-1	0.00946	-0.00780	0.000389	0.00104	-0.00218	-0.00387	**-0.00753***	0.00294	-0.00271	**0.00972****
	(0.00889)	(0.00601)	(0.00401)	(0.00350)	(0.00456)	(0.00288)	(0.00410)	(0.00525)	(0.00298)	(0.00485)
AGREEMENT_t-2	-0.00745	**-0.0188*****	-0.0124	-0.00160	0.00344	-0.00355	0.00257	-0.00109	0.00220	-0.00530
	(0.00867)	(0.00636)	(0.0106)	(0.00445)	(0.00392)	(0.00334)	(0.00479)	(0.00398)	(0.00272)	(0.00504)
AGREEMENT_t-3	-0.0103	-0.00731	-0.000722	-0.00455	-0.000354	**0.00671****	-0.00104	0.00473	0.00196	0.00321
	(0.0106)	(0.00604)	(0.00346)	(0.00368)	(0.00375)	(0.00303)	(0.00416)	(0.00380)	(0.00287)	(0.00443)
AGREEMENT_t-4	0.00672	0.00722	-0.000890	-0.00153	-0.00344	0.00172	0.00594	0.00412	-0.00153	0.00199
	(0.0133)	(0.00839)	(0.00351)	(0.00359)	(0.00447)	(0.00308)	(0.00396)	(0.00424)	(0.00298)	(0.00499)
AGREEMENT_t-5	**0.0138***	-0.00289	**-0.0107****	-0.00491	0.00489	**-0.00739***	4.77e-05	0.000538	-0.00342	0.00105
	(0.00799)	(0.00594)	(0.00488)	(0.00328)	(0.00460)	(0.00408)	(0.00372)	(0.00407)	(0.00276)	(0.00498)
Constant	-0.0108	0.0310*	0.0254	0.0113	-0.00342	0.00690	-0.000921	-0.0109	0.000103	-0.00893
	(0.0113)	(0.0182)	(0.0165)	(0.0101)	(0.00860)	(0.00665)	(0.00904)	(0.0102)	(0.00577)	(0.00801)
Observations	518	518	518	519	519	519	519	519	519	519
R-squared	0.040	0.056	0.022	0.039	0.026	0.066	0.032	0.032	0.057	0.025

*Results presented above are attained by an OLS-regression using our model I. This table shows the lagged relationship, on an index level, between the social media features (SENTIMENT, VMENTIONS, AND AGREEMENT) and firms' returns for the time period March 2014 to March 2016. Return is the difference between the logarithmic value of the unadjusted closing price and unadjusted opening price. All independent variables are lagged for a period of five days. SENTIMENT is the aggregated sentiment score. VMENTIONS is the volume of mentions in social media. AGREEMENT is the agreement among mentions, constructed in line with Antweiler and Frank (2004) as well as Sprenger et al. (2014). RETURN is used as a control variable to capture any momentum effect. Due to the variables being logarithmic functions, the coefficients should be interpreted as elasticities except for AGREEMENT, which is not logarithmic. We use robust standard errors, shown in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level.*

Appendix 4 - Table C

Firm level regressions Model I – Changes in Trading Volume as the dependant variable

VARIABLES	(1) ΔTrading Volume	(2) ΔTrading Volume	(3) ΔTrading Volume	(4) ΔTrading Volume	(5) ΔTrading Volume	(6) ΔTrading Volume	(7) ΔTrading Volume	(8) ΔTrading Volume	(9) ΔTrading Volume	(10) ΔTrading Volume
RETURNS_t-1	0.311	0.0789	-0.453	0.189	-0.952	3.063**	-1.443	-0.854	1.822	0.348
	(0.594)	(1.058)	(1.752)	(0.766)	(1.316)	(1.382)	(1.402)	(1.320)	(1.190)	(1.182)
RETURNS_t-2	-1.009	-1.214	-0.517***	0.00505	0.00928	0.397	0.265	0.247	-0.567	-0.248
	(0.777)	(0.920)	(0.190)	(0.707)	(1.252)	(1.183)	(1.236)	(1.418)	(1.212)	(1.379)
RETURNS_t-3	-0.521	-0.198	-0.303*	-0.384	1.582	-0.188	0.117	1.319	-1.218	-1.180
	(0.669)	(0.915)	(0.171)	(0.682)	(1.297)	(1.386)	(1.343)	(1.305)	(1.191)	(1.108)
RETURNS_t-4	-0.121	-0.571	-0.277	-0.186	-0.553	0.311	-0.552	0.874	-0.0110	-1.027
	(0.648)	(0.917)	(0.186)	(0.673)	(1.193)	(1.242)	(1.319)	(1.524)	(1.097)	(1.101)
RETURNS_t-5	1.016	0.247	0.369*	0.704	1.404	2.706**	0.914	3.953***	0.0680	-0.439
	(0.814)	(0.803)	(0.207)	(0.549)	(1.030)	(1.298)	(1.191)	(1.466)	(1.001)	(1.000)
SENTIMENT_t-1	-0.182**	-0.0149	0.00882	0.0618**	0.0174	0.0218	-0.0469	0.209***	0.0333	-0.0224
	(0.0806)	(0.0735)	(0.0547)	(0.0281)	(0.0465)	(0.0461)	(0.0408)	(0.0548)	(0.0264)	(0.0457)
SENTIMENT_t-2	-0.0813	-0.0338	-0.00121	0.0203	-0.0635	0.0714	-0.0190	0.0270	-0.0613**	0.0145
	(0.0696)	(0.0679)	(0.0421)	(0.0328)	(0.0456)	(0.0544)	(0.0439)	(0.0596)	(0.0255)	(0.0671)
SENTIMENT_t-3	0.0419	0.0589	0.0121	0.00951	0.0148	-0.0486	-0.00307	0.155***	0.0300	0.0141
	(0.0728)	(0.0737)	(0.0392)	(0.0283)	(0.0454)	(0.0458)	(0.0447)	(0.0564)	(0.0256)	(0.0556)
SENTIMENT_t-4	0.0437	-0.108**	-0.0817*	0.00423	-0.0193	-0.0334	-0.0158	0.0304	-0.0231	-0.0612
	(0.0639)	(0.0500)	(0.0487)	(0.0240)	(0.0440)	(0.0383)	(0.0528)	(0.0790)	(0.0260)	(0.0626)
SENTIMENT_t-5	0.0109	0.0753	-0.0142	0.00957	0.0354	0.0157	-0.0283	-0.00641	-0.00455	0.0124
	(0.0601)	(0.0617)	(0.0513)	(0.0258)	(0.0396)	(0.0471)	(0.0399)	(0.0731)	(0.0254)	(0.0621)
VMENTIONS_t-1	-0.168**	-0.184**	-0.160***	-0.122***	-0.102**	-0.0803*	-0.00200	-0.296***	-0.0541*	-0.107**
	(0.0741)	(0.0798)	(0.0547)	(0.0314)	(0.0462)	(0.0421)	(0.0446)	(0.0599)	(0.0311)	(0.0541)
VMENTIONS_t-2	0.0142	0.0605	0.0266	0.0114	0.0521	0.0373	-0.0255	-0.0332	0.0306	0.0643
	(0.0636)	(0.0726)	(0.0476)	(0.0339)	(0.0449)	(0.0489)	(0.0498)	(0.0611)	(0.0306)	(0.0692)
VMENTIONS_t-3	0.0894	-0.0405	-0.0466	0.0135	-0.0205	-0.0572	-0.0467	-0.106*	-0.0326	0.0417
	(0.0698)	(0.0752)	(0.0440)	(0.0320)	(0.0486)	(0.0571)	(0.0495)	(0.0546)	(0.0316)	(0.0770)
VMENTIONS_t-4	-0.0265	-0.0501	0.0133	0.0280	0.0371	-0.0226	0.0363	-0.0458	0.0273	0.128*
	(0.0651)	(0.0616)	(0.0538)	(0.0306)	(0.0495)	(0.0536)	(0.0558)	(0.0744)	(0.0307)	(0.0711)
VMENTIONS_t-5	0.0466	-0.0261	-0.0403	-0.0317	-0.00503	0.110*	-0.0124	-0.00106	-0.0757**	-0.0976
	(0.0613)	(0.0667)	(0.0503)	(0.0308)	(0.0442)	(0.0567)	(0.0436)	(0.0710)	(0.0293)	(0.0753)
AGREEMENT_t-1	0.0241	0.0576	-0.103	-0.0488	0.0802	-0.0191	0.114	-0.464***	-0.0587	0.00625
	(0.138)	(0.152)	(0.108)	(0.0627)	(0.119)	(0.112)	(0.107)	(0.159)	(0.0612)	(0.112)
AGREEMENT_t-2	-0.192	0.0899	-0.0806	0.0174	0.203*	-0.0221	-0.0190	-0.0857	0.0551	-0.140
	(0.149)	(0.161)	(0.0987)	(0.0676)	(0.110)	(0.128)	(0.130)	(0.133)	(0.0618)	(0.133)
AGREEMENT_t-3	0.0849	-0.0858	0.000179	-0.0590	-0.0942	0.0885	0.0982	-0.163	-0.0483	0.171
	(0.164)	(0.169)	(0.0921)	(0.0669)	(0.105)	(0.107)	(0.135)	(0.0992)	(0.0654)	(0.141)
AGREEMENT_t-4	-0.273*	0.0228	-0.143	0.0646	0.0142	-0.430***	0.178	-0.0510	0.0616	0.179
	(0.153)	(0.137)	(0.114)	(0.0565)	(0.113)	(0.127)	(0.132)	(0.149)	(0.0627)	(0.160)
AGREEMENT_t-5	0.467***	-0.128	-0.108	0.00338	-0.0288	0.326**	-0.123	-0.00916	-0.112*	-0.0756
	(0.135)	(0.129)	(0.102)	(0.0608)	(0.113)	(0.127)	(0.128)	(0.159)	(0.0648)	(0.177)
Constant	-0.0304	0.230	0.501*	0.102	-0.0887	0.0702	-0.157	0.824**	0.206*	-0.134
	(0.157)	(0.385)	(0.273)	(0.135)	(0.269)	(0.215)	(0.335)	(0.336)	(0.115)	(0.282)
Observations	385	385	385	386	386	386	386	386	386	386
R-squared	0.099	0.100	0.057	0.064	0.066	0.121	0.056	0.080	0.056	0.050

Results presented above are attained by an OLS-regression using our model I. This table shows the lagged relationship, on an index level, between the social media features (SENTIMENT, VMENTIONS, AND AGREEMENT) and firms' changes in trading volume for the time period March 2014 to March 2016. ΔTrading Volume *is the first difference of the trading volume. All independent variables are lagged for a period of five days. SENTIMENT is the aggregated sentiment score. VMENTIONS is the volume of mentions in social media. AGREEMENT is the agreement among mentions, constructed in line with Antweiler and Frank (2004) as well as Sprenger et al. (2014). RETURN is used as a control variable to capture any momentum effect. Due to the variables being logarithmic functions, the coefficients should be interpreted as elasticities except for AGREEMENT, which is not logarithmic. We use robust standard errors, shown in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level.*

Appendix 5 - Table D

Firm level regressions Model I – Change in Volatility as the dependant variable

VARIABLES	(1) ΔVolatility	(2) ΔVolatility	(3) ΔVolatility	(4) ΔVolatility	(5) ΔVolatility	(6) ΔVolatility	(7) ΔVolatility	(8) ΔVolatility	(9) ΔVolatility	(10) ΔVolatilit y
RETURN_t-1	-1.825***	0.137***	0.123***	0.149***	0.192***	0.159***	0.159***	0.144***	0.156***	0.278***
	(0.536)	(0.0215)	(0.0283)	(0.0228)	(0.0467)	(0.0287)	(0.0254)	(0.0165)	(0.0245)	(0.0513)
RETURN_t-2	-0.524	-0.0450*	-0.00238	-0.0237	-0.0206	-0.0326	-0.0156	-0.0314*	-0.0357	0.0254
	(0.624)	(0.0268)	(0.0129)	(0.0193)	(0.0269)	(0.0222)	(0.0203)	(0.0171)	(0.0253)	(0.0389)
RETURN_t-3	-0.638	-0.00331	-0.00991	0.0187	0.0501	0.0243	-0.00614	0.0267*	0.0281	-0.00812
	(0.610)	(0.0171)	(0.00717)	(0.0199)	(0.0317)	(0.0214)	(0.0198)	(0.0161)	(0.0237)	(0.0380)
RETURN_t-4	0.628	-0.0278	-0.00527	-0.0235	-0.0200	-0.0193	-0.00687	0.0102	-0.0336	-0.0259
	(0.553)	(0.0187)	(0.00635)	(0.0211)	(0.0345)	(0.0230)	(0.0140)	(0.0187)	(0.0221)	(0.0423)
RETURN_t-5	0.0902	-0.0227	-0.00472	0.00729	0.0406*	-0.00471	-0.00954	0.00654	-0.0194	-0.00581
	(0.511)	(0.0188)	(0.0103)	(0.0200)	(0.0229)	(0.0254)	(0.0164)	(0.0146)	(0.0229)	(0.0331)
SENTIMENT_t-1	-0.116	-0.000164	-0.00131	1.35e-05	-2.37e-05	0.000565	-1.99e-06	0.000869	-0.000466	-0.00161
	(0.0761)	(0.00126)	(0.00119)	(0.000584)	(0.00118)	(0.000733)	(0.000522)	(0.000891)	(0.000561)	(0.00127)
SENTIMENT_t-2	-0.0438	0.00163	-0.000482	-1.32e-05	-0.00150	-3.69e-05	-0.000694	0.000201	0.000454	0.00144
	(0.0758)	(0.00110)	(0.00188)	(0.000635)	(0.00109)	(0.000588)	(0.000631)	(0.000615)	(0.000514)	(0.00183)
SENTIMENT_t-3	0.0513	0.00159	0.000136	-0.000216	-0.00134	6.50e-05	**0.00125****	-0.000623	-2.95e-05	**-0.00253***
	(0.0689)	(0.00111)	(0.000766)	(0.000567)	(0.000813)	(0.000783)	(0.000545)	(0.000748)	(0.000537)	(0.00151)
SENTIMENT_t-4	**0.105****	-0.00236	0.000556	**0.00106***	0.000847	0.000302	**-0.00134***	0.000456	0.000518	0.00111
	(0.0530)	(0.00195)	(0.000746)	(0.000642)	(0.00146)	(0.000625)	(0.000739)	(0.000719)	(0.000574)	(0.00201)
SENTIMENT_t-5	-0.0740	**0.00208****	-0.00186	-8.42e-05	**-0.00174***	0.000483	2.66e-05	-0.000946	-0.000394	0.00135
	(0.0583)	(0.00102)	(0.00141)	(0.000540)	(0.000888)	(0.000907)	(0.000483)	(0.000672)	(0.000477)	(0.00153)
VMENTIONS_t-1	-0.0609	-0.000929	**-0.00168****	0.000321	-0.000897	-0.00117	**-0.00130***	-0.000880	-0.000446	-0.00107
	(0.0588)	(0.00107)	(0.000745)	(0.000794)	(0.00119)	(0.000966)	(0.000716)	(0.000953)	(0.000742)	(0.00184)
VMENTIONS_t-2	-0.0418	-0.00143	-0.00426	-0.000304	0.00103	-0.000699	0.000720	-0.000104	0.000142	-0.00107
	(0.0434)	(0.00118)	(0.00328)	(0.000715)	(0.00111)	(0.000890)	(0.000684)	(0.000657)	(0.000634)	(0.00171)
VMENTIONS_t-3	0.112**	-0.000783	0.00200	-0.000533	0.00135	-5.42e-05	**-0.00151****	0.000324	0.000306	0.00131
	(0.0493)	(0.00112)	(0.00135)	(0.000590)	(0.000879)	(0.000761)	(0.000639)	(0.000759)	(0.000590)	(0.00155)
VMENTIONS_t-4	**-0.1000****	0.00176	0.000248	-0.000249	-0.00188	-0.00118	**0.00143***	-4.80e-05	0.000193	0.00146
	(0.0498)	(0.00217)	(0.00117)	(0.000615)	(0.00126)	(0.000966)	(0.000792)	(0.000665)	(0.000598)	(0.00200)
VMENTIONS_t-5	0.0276	**-0.00211***	-0.000403	-0.000467	**0.00204***	0.000493	-0.000677	0.000496	-0.000362	-0.000546
	(0.0488)	(0.00115)	(0.000734)	(0.000636)	(0.00114)	(0.000796)	(0.000497)	(0.000707)	(0.000541)	(0.00156)
AGREEMENT_t-1	-0.140	0.00184	-0.00222	3.15e-05	0.000844	-0.00247	-0.00175	-4.24e-05	-0.00132	0.00161
	(0.131)	(0.00254)	(0.00214)	(0.00132)	(0.00270)	(0.00178)	(0.00158)	(0.00174)	(0.00130)	(0.00320)
AGREEMENT_t-2	-0.0463	**-0.00591****	-0.00530	-0.000786	0.00295	-0.000750	0.00137	-0.00121	0.00153	-0.00372
	(0.129)	(0.00283)	(0.00504)	(0.00146)	(0.00250)	(0.00170)	(0.00178)	(0.00135)	(0.00131)	(0.00377)
AGREEMENT_t-3	0.198	0.00184	-0.000356	-0.00104	0.000406	0.00218	**-0.00349****	**0.00295***	-0.000177	0.00232
	(0.125)	(0.00260)	(0.00152)	(0.00131)	(0.00190)	(0.00164)	(0.00167)	(0.00178)	(0.00135)	(0.00350)
AGREEMENT_t-4	**-0.235***	0.000943	-0.00184	-0.00134	-0.00219	-0.00209	0.00363**	-0.00100	0.000129	0.00432
	(0.122)	(0.00341)	(0.00174)	(0.00129)	(0.00265)	(0.00183)	(0.00179)	(0.00143)	(0.00140)	(0.00430)
AGREEMENT_t-5	0.184	-0.00281	-0.00343	-0.000855	0.00173	-0.00325	-0.00191	0.000941	-0.00217	-0.00138
	(0.119)	(0.00213)	(0.00228)	(0.00137)	(0.00266)	(0.00215)	(0.00153)	(0.00137)	(0.00133)	(0.00423)
Constant	0.0898	0.00510	0.0137	0.00440	-0.00388	0.00671*	0.00291	-0.00124	0.00154	-0.00257
	(0.176)	(0.00753)	(0.00913)	(0.00339)	(0.00499)	(0.00364)	(0.00352)	(0.00356)	(0.00251)	(0.00583)
Observations	385	385	385	386	386	386	386	386	386	386
R-squared	0.124	0.169	0.042	0.169	0.147	0.176	0.229	0.242	0.177	0.149

Results presented above are attained by an OLS-regression using our model I. This table shows the lagged relationship, on an index level, between the social media features (SENTIMENT, VMENTIONS, AND AGREEMENT) and firms' changes in trading volume for the time period March 2014 to March 2016. ΔVolatility is the first difference of the volatility variable, constructed in line with Parkinson (1980). *All independent variables are lagged for a period of five days. SENTIMENT is the aggregated sentiment score. VMENTIONS is the volume of mentions in social media. AGREEMENT is the agreement among mentions, constructed in line with Antweiler and Frank (2004) as well as Sprenger et al. (2014). RETURN is used as a control variable to capture any momentum effect. Due to the variables being logarithmic functions, the coefficients should be interpreted as elasticities except for AGREEMENT, which is not logarithmic. We use robust standard errors, shown in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level.*

Appendix 6 - Table E

Index level regression Model II – OMX22

VARIABLES	Return	ΔTrading Volume
ΔSENTIMENT_t-1	0.00463	0.00189
	(0.00337)	(0.00353)
ΔSENTIMENT_t-2	0.00527	0.00236
	(0.00417)	(0.00436)
ΔSENTIMENT_t-3	**0.00924****	0.00351
	(0.00449)	(0.00470)
ΔSENTIMENT_t-4	0.00702	0.00745
	(0.00436)	(0.00457)
ΔSENTIMENT_t-5	0.000243	0.00473
	(0.00351)	(0.00368)
ΔVMENTIONS_t-1	0.000964	-0.000173
	(0.00347)	(0.00363)
ΔVMENTIONS_t-2	0.000446	0.00592
	(0.00435)	(0.00456)
ΔVMENTIONS_t-3	-0.00483	0.000290
	(0.00444)	(0.00464)
ΔVMENTIONS_t-4	-0.00143	-0.00686
	(0.00424)	(0.00444)
ΔVMENTIONS_t-5	0.00315	-0.00166
	(0.00348)	(0.00364)
ΔAGREEMENT_t-1	0.00645	-0.00190
	(0.00800)	(0.00837)
ΔAGREEMENT_t-2	0.00397	0.0137
	(0.0101)	(0.0105)
ΔAGREEMENT_t-3	-0.00782	0.00566
	(0.0108)	(0.0113)
ΔAGREEMENT_t-4	-0.00770	-0.00574
	(0.0101)	(0.0105)
ΔAGREEMENT_t-5	0.00176	6.89e-05
	(0.00812)	(0.00850)
Constant	0.000273	0.000343
	(0.000589)	(0.000617)
Observations	415	415
R-squared	0.026	0.029

Results presented above are attained by an OLS-regression using our model II. This table shows the lagged relationship, on an index level, between the social media features (ΔSENTIMENT, ΔVMENTIONS, AND ΔAGREEMENT) and firms' returns and changes in trading volume for the time period March 2014 to March 2016. Return is the difference between the logarithmic value of the unadjusted closing price and unadjusted opening price. ΔTrading Volume *is the first difference of the trading volume. All independent variables are lagged for a period of five days. ΔSENTIMENT is the change in the aggregated sentiment score. Δ VMENTIONS is the change in volume of mentions in social media. ΔAGREEMENT is the change in agreement among mentions, constructed in line with Antweiler and Frank (2004) as well as Sprenger et al. (2014). Due to the variables being logarithmic functions, the coefficients should be interpreted as elasticities except for ΔAGREEMENT, which is not logarithmic. We use robust standard errors, shown in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level.*

Appendix 7 – Table F

Firm level regressions Model II – Return as the dependant variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	Return	Return	Return	Return	Return	Return	Return	Return	Return	Return
ΔSENTIMENT_t-1	0.00290	0.000349	-0.00181	-0.00132	0.00221	0.000910	0.00213	0.000494	-0.000167	0.000714
	(0.00392)	(0.00308)	(0.00466)	(0.00183)	(0.00159)	(0.00137)	(0.00160)	(0.00236)	(0.00116)	(0.00199)
ΔSENTIMENT_t-2	**0.00869***	0.000693	-0.000606	0.000298	0.00176	-0.00180	0.00206	0.00241	0.000903	0.000917
	(0.00498)	(0.00372)	(0.00513)	(0.00207)	(0.00192)	(0.00174)	(0.00176)	(0.00290)	(0.00138)	(0.00235)
ΔSENTIMENT_t-3	**0.0116****	0.00303	-0.000292	-0.000312	0.00117	-0.000941	0.00246	0.00163	0.000149	-0.00107
	(0.00530)	(0.00392)	(0.00537)	(0.00215)	(0.00211)	(0.00189)	(0.00191)	(0.00315)	(0.00146)	(0.00255)
ΔSENTIMENT_t-4	0.00160	-0.00458	0.00325	0.00224	0.00253	-0.000288	0.000645	0.00447	0.00183	-0.00141
	(0.00524)	(0.00375)	(0.00530)	(0.00205)	(0.00189)	(0.00179)	(0.00180)	(0.00298)	(0.00140)	(0.00256)
ΔSENTIMENT_t-5	0.00378	-0.00257	0.00240	0.000931	0.00257	-0.000939	-0.000775	0.00170	0.00162	0.000207
	(0.00414)	(0.00310)	(0.00455)	(0.00189)	(0.00167)	(0.00141)	(0.00159)	(0.00239)	(0.00115)	(0.00211)
ΔVMENTIONS_t-1	-0.00161	-0.00300	0.000138	**0.00489****	-1.99e-05	-0.000662	-0.00252	0.000427	5.13e-05	0.000976
	(0.00404)	(0.00340)	(0.00508)	(0.00218)	(0.00182)	(0.00163)	(0.00172)	(0.00236)	(0.00134)	(0.00247)
ΔVMENTIONS_t-2	-0.00533	-0.00568	-0.00883	0.00215	0.00100	-0.000743	-0.00155	-0.000842	0.000565	-0.000852
	(0.00469)	(0.00409)	(0.00569)	(0.00253)	(0.00219)	(0.00180)	(0.00198)	(0.00290)	(0.00163)	(0.00301)
ΔVMENTIONS_t-3	-0.00521	-0.00367	-0.00257	-0.00121	0.000710	6.23e-05	-0.00250	-0.000804	**0.00371****	-0.00145
	(0.00484)	(0.00428)	(0.00579)	(0.00262)	(0.00236)	(0.00188)	(0.00220)	(0.00308)	(0.00164)	(0.00313)
ΔVMENTIONS_t-4	-0.00316	0.00329	9.08e-05	0.000265	-0.000442	-0.00219	-0.000951	-0.00222	**0.00280***	-0.00124
	(0.00461)	(0.00399)	(0.00563)	(0.00243)	(0.00222)	(0.00183)	(0.00201)	(0.00292)	(0.00155)	(0.00298)
ΔVMENTIONS_t-5	-0.00538	0.00242	-0.000828	-0.000268	0.00120	0.000498	-0.00149	-0.00124	**0.00403*****	-0.00178
	(0.00403)	(0.00337)	(0.00503)	(0.00216)	(0.00188)	(0.00160)	(0.00172)	(0.00236)	(0.00135)	(0.00246)
ΔAGREEMENT_t-1	-0.000123	-0.000618	0.00538	0.00448	-0.00265	-0.000924	**-0.01000****	-0.000803	-0.00139	0.00547
	(0.00883)	(0.00768)	(0.00970)	(0.00409)	(0.00440)	(0.00363)	(0.00420)	(0.00459)	(0.00287)	(0.00512)
ΔAGREEMENT_t-2	-0.00480	-0.0118	-0.00455	0.00290	-0.000849	-0.00379	-0.00732	-0.00653	0.00524	-0.00484
	(0.0115)	(0.00919)	(0.0118)	(0.00509)	(0.00532)	(0.00474)	(0.00522)	(0.00587)	(0.00354)	(0.00668)
ΔAGREEMENT_t-3	-0.00944	-0.0100	0.000602	-0.000758	-0.000860	0.00651	**-0.0103***	-0.00109	**0.00989*****	-0.00471
	(0.0126)	(0.00957)	(0.0125)	(0.00522)	(0.00549)	(0.00503)	(0.00571)	(0.00640)	(0.00363)	(0.00678)
ΔAGREEMENT_t-4	0.00604	0.00262	0.00344	0.000923	-0.00349	0.00627	-0.00413	-0.00257	**0.00727****	-0.00474
	(0.0113)	(0.00890)	(0.0120)	(0.00488)	(0.00530)	(0.00476)	(0.00522)	(0.00602)	(0.00347)	(0.00673)
ΔAGREEMENT_t-5	0.0126	0.00633	-0.00243	-0.00156	-0.000892	0.00311	-0.00378	-0.00276	**0.00767*****	-0.00708
	(0.00892)	(0.00772)	(0.00958)	(0.00414)	(0.00430)	(0.00355)	(0.00416)	(0.00469)	(0.00289)	(0.00530)
Constant	-0.00308	-0.000256	-0.00199	0.000550	-0.000814	-0.000197	-7.16e-05	0.000544	-0.000632	0.000580
	(0.00262)	(0.00193)	(0.00290)	(0.00133)	(0.00125)	(0.000850)	(0.000922)	(0.000742)	(0.000919)	(0.00127)
Observations	415	415	415	416	416	416	416	416	416	416
R-squared	0.037	0.038	0.020	0.031	0.017	0.048	0.030	0.032	0.059	0.017

*Results presented above are attained by an OLS-regression using our model II. This table shows the lagged relationship, on a firm level, between the social media features (ΔSENTIMENT, ΔVMENTIONS, AND ΔAGREEMENT) and firms' returns for the time period March 2014 to March 2016. Return is the difference between the logarithmic value of the unadjusted closing price and unadjusted opening price. All independent variables are lagged for a period of five days. ΔSENTIMENT is the change in the aggregated sentiment score. Δ VMENTIONS is the change in volume of mentions in social media. ΔAGREEMENT is the change in agreement among mentions, constructed in line with Antweiler and Frank (2004) as well as Sprenger et al. (2014). Due to the variables being logarithmic functions, the coefficients should be interpreted as elasticities except for ΔAGREEMENT, which is not logarithmic. We use robust standard errors, shown in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level.*

Appendix 8 - G

Firm level regressions Model II – Change in Trading Volume as the dependant variable

VARIABLES	(1) ΔTrading Volume	(2) ΔTrading Volume	(3) ΔTrading Volume	(4) ΔTrading Volume	(5) ΔTrading Volume	(6) ΔTrading Volume	(7) ΔTrading Volume	(8) ΔTrading Volume	(9) ΔTrading Volume	(10) ΔTrading Volume
ΔSENTIMENT_t-1	-0.166**	-0.0263	0.0243	0.0408	-0.00310	0.0162	-0.0352	0.145**	0.0332	-0.0243
	(0.0653)	(0.0578)	(0.0470)	(0.0273)	(0.0351)	(0.0375)	(0.0415)	(0.0708)	(0.0252)	(0.0494)
ΔSENTIMENT_t-2	-0.240***	-0.0740	0.0270	0.0390	-0.0722*	0.0919*	-0.0420	0.1000	-0.0246	0.0149
	(0.0852)	(0.0701)	(0.0521)	(0.0313)	(0.0427)	(0.0481)	(0.0456)	(0.0885)	(0.0300)	(0.0598)
ΔSENTIMENT_t-3	-0.213**	-0.0351	0.0485	0.0231	-0.0517	0.0420	-0.0253	0.169*	0.0120	0.0407
	(0.0911)	(0.0735)	(0.0548)	(0.0327)	(0.0457)	(0.0534)	(0.0495)	(0.0960)	(0.0318)	(0.0669)
ΔSENTIMENT_t-4	-0.188**	-0.170**	-0.0316	0.0103	-0.0699*	0.00435	-0.0255	0.120	-0.00435	-0.0285
	(0.0896)	(0.0699)	(0.0548)	(0.0302)	(0.0416)	(0.0506)	(0.0468)	(0.0906)	(0.0303)	(0.0640)
ΔSENTIMENT_t-5	-0.141**	-0.120**	-0.0335	0.00399	-0.0329	0.00896	-0.0293	0.0384	0.00600	-0.00538
	(0.0693)	(0.0577)	(0.0462)	(0.0278)	(0.0364)	(0.0390)	(0.0416)	(0.0723)	(0.0255)	(0.0525)
ΔVMENTIONS_t-1	-0.186***	-0.135**	-0.120**	-0.0942***	-0.0885**	-0.0806*	0.0198	-0.205***	-0.0247	-0.120*
	(0.0688)	(0.0631)	(0.0515)	(0.0324)	(0.0412)	(0.0456)	(0.0446)	(0.0709)	(0.0291)	(0.0620)
ΔVMENTIONS_t-2	-0.146*	-0.0282	-0.0609	-0.0703*	-0.0342	-0.0528	0.00510	-0.153*	0.00938	-0.0584
	(0.0802)	(0.0774)	(0.0578)	(0.0380)	(0.0496)	(0.0519)	(0.0514)	(0.0888)	(0.0357)	(0.0775)
ΔVMENTIONS_t-3	-0.0323	-0.00415	-0.0861	-0.0421	-0.0553	-0.110**	-0.0312	-0.169*	-0.0162	-0.0378
	(0.0826)	(0.0808)	(0.0600)	(0.0395)	(0.0516)	(0.0521)	(0.0569)	(0.0937)	(0.0360)	(0.0805)
ΔVMENTIONS_t-4	-0.0388	-0.0280	-0.0325	-0.00320	-0.00617	-0.143***	0.0159	-0.129	0.0215	0.0678
	(0.0793)	(0.0746)	(0.0584)	(0.0353)	(0.0485)	(0.0506)	(0.0521)	(0.0890)	(0.0339)	(0.0745)
ΔVMENTIONS_t-5	0.0106	0.0162	-0.0600	-0.0216	-0.0140	-0.0868*	0.0168	-0.0480	-0.0501*	-0.0623
	(0.0679)	(0.0625)	(0.0513)	(0.0316)	(0.0412)	(0.0454)	(0.0448)	(0.0716)	(0.0297)	(0.0615)
ΔAGREEMENT_t-1	0.0144	0.0745	-0.0443	-0.0458	0.0735	0.00447	0.0758	-0.352**	-0.0326	-0.0132
	(0.148)	(0.145)	(0.0997)	(0.0601)	(0.0959)	(0.103)	(0.110)	(0.138)	(0.0629)	(0.128)
ΔAGREEMENT_t-2	-0.204	0.167	-0.0736	-0.0302	0.283**	-0.0356	-0.000873	-0.291	0.0264	-0.148
	(0.194)	(0.172)	(0.121)	(0.0743)	(0.116)	(0.136)	(0.136)	(0.179)	(0.0775)	(0.171)
ΔAGREEMENT_t-3	-0.157	0.138	-0.0384	-0.0914	0.162	0.0438	0.0496	-0.305	0.0110	0.00158
	(0.211)	(0.177)	(0.129)	(0.0759)	(0.119)	(0.145)	(0.148)	(0.196)	(0.0804)	(0.174)
ΔAGREEMENT_t-4	-0.423**	0.189	-0.0982	-0.0355	0.177	-0.379***	0.169	-0.194	0.0834	0.170
	(0.193)	(0.166)	(0.124)	(0.0706)	(0.115)	(0.135)	(0.135)	(0.186)	(0.0772)	(0.167)
ΔAGREEMENT_t-5	0.0497	0.144	-0.157	-0.0338	0.156*	-0.0664	-0.0184	-0.0266	-0.000651	0.0902
	(0.151)	(0.146)	(0.0987)	(0.0605)	(0.0934)	(0.102)	(0.109)	(0.147)	(0.0641)	(0.132)
Constant	0.00190	-0.0103	-0.0326	0.0150	0.0208	0.0191	0.00456	-0.00963	0.0301	-0.00204
	(0.0438)	(0.0364)	(0.0299)	(0.0195)	(0.0278)	(0.0231)	(0.0242)	(0.0230)	(0.0203)	(0.0316)
Observations	385	385	385	386	386	386	386	386	386	386
R-squared	0.092	0.089	0.047	0.041	0.060	0.109	0.029	0.044	0.041	0.054

Results presented above are attained by an OLS-regression using our model II. This table shows the lagged relationship, on a firm level, between the social media features (ΔSENTIMENT, ΔVMENTIONS, AND ΔAGREEMENT) and firms' changes in trading volume for the time period March 2014 to March 2016. ΔTrading Volume is the first difference of the trading volume. All independent variables are lagged for a period of five days. ΔSENTIMENT is the change in the aggregated sentiment score. ΔVMENTIONS is the change in volume of mentions in social media. ΔAGREEMENT is the change in agreement among mentions, constructed in line with Antweiler and Frank (2004) as well as Sprenger et al. (2014). Due to the variables being logarithmic functions, the coefficients should be interpreted as elasticities except for ΔAGREEMENT, which is not logarithmic. We use robust standard errors, shown in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level.

Appendix 9 - Table H

Firm level regressions Model II – Volatility as the dependant variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	ΔVolatility	ΔVolatility	ΔVolatility	ΔVolatility	ΔVolatility	ΔVolatility	ΔVolatility	ΔVolatility	ΔVolatility	ΔVolatility
ΔSENTIMENT_t-1	-0.120**	0.000375	-0.000272	0.000820	0.00227***	0.000548	0.00114*	0.000302	-0.000307	-0.00188
	(0.0491)	(0.00117)	(0.00185)	(0.000690)	(0.000814)	(0.000641)	(0.000601)	(0.000819)	(0.000504)	(0.00167)
ΔSENTIMENT_t-2	-0.142**	0.000918	-0.000281	0.000386	0.00147	0.000471	0.000647	0.000792	0.000342	0.000737
	(0.0641)	(0.00141)	(0.00206)	(0.000793)	(0.000989)	(0.000822)	(0.000660)	(0.00102)	(0.000599)	(0.00202)
ΔSENTIMENT_t-3	-0.0928	0.00212	0.000635	0.000102	0.000547	4.37e-05	0.00185**	0.000564	0.000344	-0.00270
	(0.0685)	(0.00148)	(0.00216)	(0.000826)	(0.00106)	(0.000914)	(0.000718)	(0.00111)	(0.000635)	(0.00226)
ΔSENTIMENT_t-4	0.0373	-0.000936	0.00153	0.000671	0.00141	-0.000322	0.000673	0.000876	0.000673	-0.00198
	(0.0674)	(0.00141)	(0.00216)	(0.000764)	(0.000963)	(0.000865)	(0.000677)	(0.00105)	(0.000604)	(0.00216)
ΔSENTIMENT_t-5	-0.00470	-0.000548	0.000508	0.000785	0.00101	-0.000538	0.000641	0.000180	0.000584	-0.00180
	(0.0521)	(0.00116)	(0.00182)	(0.000702)	(0.000844)	(0.000667)	(0.000603)	(0.000836)	(0.000509)	(0.00177)
ΔVMENTIONS_t-1	-0.0474	0.000437	-0.000638	-0.000409	**-0.00193****	-0.000836	**-0.00172*****	-0.000229	0.000190	-0.000679
	(0.0517)	(0.00127)	(0.00203)	(0.000818)	(0.000954)	(0.000781)	(0.000646)	(0.000820)	(0.000581)	(0.00209)
ΔVMENTIONS_t-2	-0.0900	-0.00140	**-0.00420***	4.54e-05	-0.000672	-0.00104	-0.000899	-0.000470	0.000179	-0.00196
	(0.0603)	(0.00156)	(0.00228)	(0.000962)	(0.00115)	(0.000887)	(0.000744)	(0.00103)	(0.000712)	(0.00262)
ΔVMENTIONS_t-3	0.0229	-0.00161	-0.00156	-4.97e-06	0.000433	-0.000362	**-0.00242*****	-0.000497	0.000701	-0.00239
	(0.0621)	(0.00163)	(0.00237)	(0.001000)	(0.00120)	(0.000891)	(0.000824)	(0.00108)	(0.000719)	(0.00272)
ΔVMENTIONS_t-4	-0.0826	0.000721	-0.000299	-0.000137	-0.00167	-0.00104	-0.00123	-0.000549	**0.00116***	-0.00150
	(0.0596)	(0.00151)	(0.00230)	(0.000892)	(0.00112)	(0.000866)	(0.000754)	(0.00103)	(0.000678)	(0.00251)
ΔVMENTIONS_t-5	-0.0645	-0.000108	-0.000107	0.000185	-0.000493	-0.000311	**-0.00166****	-0.000257	**0.00101***	-0.00214
	(0.0510)	(0.00126)	(0.00203)	(0.000799)	(0.000955)	(0.000777)	(0.000648)	(0.000828)	(0.000593)	(0.00208)
ΔAGREEMENT_t-1	-0.164	0.00314	5.91e-05	-0.000365	-0.00128	-0.00144	-0.00248	0.00149	-0.000277	-0.000172
	(0.111)	(0.00293)	(0.00394)	(0.00152)	(0.00222)	(0.00176)	(0.00160)	(0.00159)	(0.00126)	(0.00432)
ΔAGREEMENT_t-2	-0.215	-0.00237	-0.00234	0.000101	0.000814	-0.00106	-0.00207	-0.000465	0.00107	-0.00328
	(0.146)	(0.00347)	(0.00478)	(0.00188)	(0.00269)	(0.00233)	(0.00198)	(0.00207)	(0.00155)	(0.00577)
ΔAGREEMENT_t-3	-0.0335	-0.000845	-0.000318	-0.000133	0.000851	0.00164	**-0.00551****	0.00117	**0.00277***	-0.00360
	(0.159)	(0.00358)	(0.00510)	(0.00192)	(0.00277)	(0.00248)	(0.00214)	(0.00227)	(0.00161)	(0.00588)
ΔAGREEMENT_t-4	**-0.272***	0.000587	0.000851	-0.000779	-0.00229	0.00240	-0.00178	8.01e-05	**0.00374****	-0.000105
	(0.145)	(0.00334)	(0.00488)	(0.00179)	(0.00267)	(0.00230)	(0.00195)	(0.00215)	(0.00154)	(0.00563)
ΔAGREEMENT_t-1	-0.104	0.000912	0.000346	-0.000446	-0.00171	0.00163	**-0.00297***	0.00111	**0.00263****	-0.00102
	(0.113)	(0.00294)	(0.00389)	(0.00153)	(0.00217)	(0.00175)	(0.00158)	(0.00170)	(0.00128)	(0.00445)
Constant	-0.000213	-0.000369	-0.000907	0.000247	-0.000258	5.23e-05	-7.02e-05	0.000181	-3.57e-05	0.000164
	(0.0330)	(0.000735)	(0.00118)	(0.000494)	(0.000644)	(0.000395)	(0.000351)	(0.000266)	(0.000406)	(0.00106)
Observations	385	385	385	386	386	386	386	386	386	386
R-squared	0.088	0.043	0.024	0.011	0.048	0.033	0.058	0.027	0.033	0.019

Results presented above are attained by an OLS-regression using our model II. This table shows the lagged relationship, on a firm level, between the social media features (ΔSENTIMENT, ΔVMENTIONS, AND ΔAGREEMENT) and firms' changes in volatility for the time period March 2014 to March 2016. ΔVolatility is the first difference of the volatility variable, constructed in line with Parkinson (1980). *All independent variables are lagged for a period of five days. ΔSENTIMENT is the change in the aggregated sentiment score. ΔVMENTIONS is the change in volume of mentions in social media. ΔAGREEMENT is the change in agreement among mentions, constructed in line with Antweiler and Frank (2004) as well as Sprenger et al. (2014). Due to the variables being logarithmic functions, the coefficients should be interpreted as elasticities except for ΔAGREEMENT, which is not logarithmic. We use robust standard errors, shown in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level.*

Appendix 10 - Table I

Sensitivity analysis using Model II: SANDVIK – Volume of mentions as the dependent variable

VARIABLES	(1) VMENTIONS
RETURN_t-1	**4.822***
	(2.590)
RETURN_t-2	-1.017
	(2.580)
RETURN_t-3	1.098
	(2.627)
RETURN_t-4	0.400
	(2.386)
RETURN_t-5	-0.830
	(2.523)
ΔTV_t-1	**0.314*****
	(0.113)
ΔTV_t-2	**0.442****
	(0.178)
ΔTV_t-3	**0.399****
	(0.201)
ΔTV_t-4	0.190
	(0.185)
ΔTV_t-5	0.102
	(0.120)
Constant	1.052***
	(0.0479)
Observations	417
R-squared	0.032

Results presented above are attained by an OLS-regression using model II. This table shows the lagged relationship, on a firm level, for Sandvik, between RETURN, TRADING VOLUME and VMENTIONS for the time period March 2014 to March 2016. VMENTIONS *is the volume of mentions on social media. RETURN is the difference between the logarithmic value of the unadjusted closing price and unadjusted opening price.* ΔTV is the change in trading volume and is added a control variable. *All independent variables are lagged for a period of five days.We use robust standard errors, shown in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level.*

Appendix 11 - Table J

Durbin-Watson test for autocorrelation

	Index	Anoto Group	Arcam	Betsson	Boliden	Elekta	Ericsson	MTG	NCC	Sandvik	SAS
Regression 1	**1.525547**	1.581987	1.646919	**1.106045**	1.686014	1.790268	1.602792	1.6	1.638303	1.646871	1.707484
Regression 2	1.80014	***1.956119***	1.942214	1.883887	1.923791	1.827191	2.005041	1.796881	1.761376	1.76296	1.769735
Regression 3	-	***2.059259***	**1.525637**	**1.072753**	1.691341	1.704981	1.6035	1.897561	1.862837	1.755657	1.94455
Regression 4	**1.554471**	1.714594	**1.32264**	**1.077931**	1.725718	**1.552331**	**1.51**	**1.524983**	**1.504934**	1.612586	1.780265
Regression 5	1.698984	1.925192	1.93804	1.886259	1.916162	1.850114	2.008167	1.796467	1.739576	1.755131	1.776461
Regression 6	-	***2.040346***	**1.236868**	**1.054625**	**1.297151**	**1.358377**	**1.340506**	**1.49783**	**1.386215**	**1.4224235**	**1.464078**

The table above shows the results obtained from Durbin Watson testing for autocorrelation in our regressions from both models. Regression 1-3 is the regression in the first model and 4-6 shows the results obtained for model II. The third and sixth regression have no value for our index due to the lack of volatility variable. The results in bold shows values indicating a positive autocorrelation for the model and results in italics indicates negative autocorrelation.

Appendix 12 - Table K

Breusch-Pagan test for heteroskedasticity

	Index	Anoto Group	Arcam	Betsson	Boliden	Elekta	Ericsson	MTG	NCC	Sandvik	SAS
Regression 1	0.5424	**0**	**0**	**0**	0.6928	**0.0554**	**0**	0.1297	0.3719	**0.0033**	0.5052
Regression 2	**0**	**0.002**	**0.0017**	0.8015	0.3127	0.5153	0.2394	0.4744	**0.0059**	0.1912	**0.0481**
Regression 3	-	0.3875	**0**	**0**	**0.043**	**0**	**0.0156**	**0.0233**	**0.0454**	**0.0091**	**0.0681**
Regression 4	**0.039**	**0**	**0**	**0**	0.0541	**0.0012**	**0.0011**	**0.0046**	0.4379	0.6607	**0.084**
Regression 5	**0.0141**	**0.042**	**0.0757**	0.5212	0.467	0.2884	0.3342	0.8128	**0.0566**	**0.06632**	0.1895
Regression 6	-	**0.0008**	**0**	**0**	0.6789	**0**	**0**	0.4471	**0.0023**	**0**	0.5235

Shows the p-values obtained from our Breusch-Pagan test determining the presence of heteroscedasticity. The null hypothesis is that the regression is homoscedastic and values in bold shows when there is heteroscedasticity present and the null hypothesis can be rejected.

Appendix 13 - Table L

VIF-test for multicollinearity

	Index		Anoto Group		Arcam		Betsson	
	Max	Mean	Max	Mean	Max	Mean	Max	Mean
Regression 1	4.13	2.29	2.08	1.44	2.67	1.89	2.28	1.62
Regression 2	4.44	2.33	2.18	1.47	2.75	1.89	2.46	1.65
Regression 3	-	-	2.18	1.47	2.75	1.89	2.46	1.65
Regression 4	6	3.82	3.49	2.44	6.17	4.16	3.42	4.8
Regression 5	6	3.82	3.56	2.49	6.02	4.07	4.8	3.32
Regression 6	-	-	3.56	2.49	6.02	4.07	4.8	3.32

	Boliden		Elekta		Ericsson		MTG	
	Max	Mean	Max	Mean	Max	Mean	Max	Mean
Regression 1	2.86	1.82	2.55	1.73	5.61	1.73	2.81	1.87
Regression 2	3.07	1.88	2.69	1.75	5.92	1.75	3.4	1.94
Regression 3	3.07	1.88	2.69	1.75	5.92	1.75	3.4	1.94
Regression 4	4.11	2.87	4.08	2.83	4.68	2.83	6.44	4.18
Regression 5	4.04	2.92	4.03	2.84	5	2.84	6.53	4.17
Regression 6	4.04	2.92	4.03	2.84	5	2.84	6.53	4.17

	NCC		Sandvik		SAS	
	Max	Mean	Max	Mean	Max	Mean
Regression 1	**10.2**	5.76	2.44	1.67	5.6	2.76
Regression 2	**12.74**	5.85	2.55	1.7	6.39	2.9
Regression 3	**12.74**	5.85	2.55	1.7	6.39	2.9
Regression 4	**17.74**	**11.03**	4.53	2.8	4.61	2.94
Regression 5	**16.96**	**10.67**	4.44	2.74	4.73	3.04
Regression 6	**16.96**	**10.67**	4.44	2.74	4.73	3.04

Above is presented the max and mean VIF values for all regressions. The VIF test is used to investigate whether or not there is multicollinearity present in our regressions. Values in bold indicates that at least two of the independent variables displays multicollinearity.