HØGSKOLEN I OSLO
OG AKERSHUS

Chris-André Einarsen

Media's effect on the Norwegian share

A study on trading volume and intraday price spread of the Norwegian Air Shuttle ASA share

Master thesis in Business and Administration
Oslo Business School
2017

Masteroppgave i økonomi og administrasjon
Handelshøyskolen ved HiOA
2017

Abstract

In this paper, I study the relationship between media intensity and trading volume and intraday price spread for the share of the Norwegian airline Norwegian Air Shuttle ASA from January 2010 to June 2016. The company Norwegian was chosen as it is a company that is heavily covered by the Norwegian press, and the investor profile is a good mixture of institutional and non-institutional investors, suitable for my study.

I use an exploratory method to arrive at the final models. My data consists of non-stationary time series and I look at various ways of making them stationary, as well as various combinations of explanatory variables. A large portion of the study is devoted to model testing and finding the news related variables that best explain the variation in trading volume and intraday price spread.

I perform correlation analysis, as well as linear regressions, and find that the investors are indeed attention-driven, i.e. they trade when the company is mentioned in the news. Rational investors would allocate higher importance to the financial press than the tabloids, and indications of this is found in my study. However, the novelty of my study is that I look at the ratios between financial- and non-financial press articles, and my models show indications that the investors in Norwegian have difficulties screening out the relevant articles when the financial press is drowning in a high general media intensity.

My conclusions are robust to the assumptions behind linear regression on time series data, but must however be viewed with some caution due to lack of details in the news data, and the potential issue of reversed causation.

Foreword

This master thesis was written as the final part of my Master of Finance at Oslo Business School, part of Oslo and Akershus University College of Applied Sciences (HiOA). The topic of the thesis was chosen based on a curiosity that was sparked in me when reading the article "How news affects the trading behaviour of different categories of investors in a financial market" by Lillo and al. (2015). My research did not turn out exactly the same as theirs, but it inspired me to look at the Norwegian media's role on the trading of the shares of a company which I have personal interest in, namely the airline Norwegian Air Shuttle ASA.

I have previously worked for the airline, and I had a feeling that it was a lot in the news. I was aware that this could very well be a completely subjective feeling, but my research later confirmed that the company is indeed covered by the Norwegian media to a much greater extent than other Norwegian companies of similar size. There might be something about the airline industry that triggers people's interest, and this makes it the perfect candidate for a study such as this one.

The research has been exciting and a great challenge to my econometrics skills. It is the first time I have collected and prepared a large data set from scratch, and used a statistical analysis software for my own personal research. The experience has been very stimulating and educational, and given me the drive to continue my interest for research in the future.

The last semester before this one, I was on exchange to INSEEC Business School in Paris, and during the whole process of writing this thesis I have remained in Paris. I would therefore like to give a special thank you to my mentor, Einar Belsom, who has been willing to give me guidance and advice over Skype and e-mail without ever meeting in person. I also take this opportunity to thank Jari Kätsyri for initial discussions regarding the sentiment of news; Andreas Kammerlander for fruitful discussions and feedback; Thomas Stenborg at Atekst; Linn Furuvald Næss at Oslo Børs; and everyone behind the Titlon database, in particular Espen Sirnes, who helped me find some missing data.

Paris, 4th of May 2017.

Index

1. Introduction

It is safe to say that the market hypothesis is widely accepted, and that the discussion nowadays revolves more around to which *degree* the market is efficient. And what is market efficacy? Berry and Howe (1994, 1) puts it perfectly: "A fundamental tenet of market efficiency is that investors react to new information as it arrives, resulting in price changes that reflect investors' expectations of risk and return.". An interesting question is how do investors acquire this information, and how exactly does it affect the behaviour of the assets? I will contribute to this by studying how the Norwegian media affects trading volume and intraday price spread of the Norwegian airline Norwegian Air Shuttle ASA.

In a recent study by Lillo et al. (2014), they assess different investor groups' trading behaviour in response to news articles and -sentiment (among other factors). This is the main article that initially inspired this paper, although my research focuses more on the asset behaviour, rather than the individual investor's behaviour.

Although the group of research on the investor reaction is slimmer than that on the asset itself, I still believe that my research will complement the asset research group. The investor specific research is interesting, but demands an immensely detailed data frame, which I was unable to gain access to. But the total volume traded, without details on how the volume is divided among different categories of investors, still indirectly reflects investor behaviour. For whatever reason they trade, trading volume is a reflection of the ensemble of investors' decisions to either buy or sell. The second variable, intraday price spread, is a more direct measure on asset behaviour, as it is related to the price movement during one trading day.

In the existing literature on asset behaviour in relation to news, the focus has primarily been on returns, but some have also looked at volatility- and liquidity measures. I landed on the variables trading volume and intraday price spread since the literature on returns is so comprehensive already.

As summarised by Jonathan M. Karpoff (1986), there are two ways informational events affect trading volume: either the investors interpret the new information differently, or they interpret the new information identically, but had divergent expectations prior to the release. Researchers have found empirical evidence for both of these explanations (see, among others, Kim and Verrecchia (1991), Harris and Raviv (1993), Paul C. Tetlock (2010) and Birz and Lott Jr (2011)), so either way, there is an established relationship between news releases and trading volume.

It is intuitive to think that the sentiment of the news is a very important part of this subject. A wide range of research show that negative stimuli elicit greater affective, cognitive, and behavioural influences than equally intense positive stimuli (see, among others, Kahneman and Tversky (1984), Shelley E. Taylor (1991), and Rozin and Royzman (2001)). More recently, Katsyri et al. (2016) find that negative tweets draw longer viewing time and more attention to themselves than positive tweets, and Yechiam and Hochman (2013) conclude that losses elicit more on-task attention than equivalent gains. In the study of Lillo et al. (2014), however, they find little importance of the sentiment. Other researchers have found sentiment to be of relevance, but even when considering the statistically significant findings on sentiment, researchers have difficulties identifying profitable sentiment-based trading strategies (see for example Tetlock (2007) and Groß-Klußmann and Hautsch (2011)). Finally, including sentiment in the research requires a finely tuned qualitative method to determine the sentiment, and so far, there is no widespread agreement on a good method to do this. For example, a news article could contain a large number of negative words, but if the market was expecting even more negative news, this article could still be associated with a positive market response (Engelberg, Reed, and Ringgenberg 2012). The sum of all this is why I have decided not to consider sentiment in my paper.

The next decision was the source of news. Tauchen and Pitts (1983) treat "news arrival" as an unobserved mixing variable, and Engle and Ng (1993) define "news" as unexpected increases or decreases in returns. Several researchers argue, however, that these methods are inadequate and that news wires provide better insight (for example Kalev et al. (2004) and Joulin et al. (2008)). I also personally find it more interesting to look at the actual media's role in asset behaviour. A long series of studies indicate that the media has the power to influence the opinion of many stakeholders, even to such an extent that it is recognised by the companies themselves, which have been proven to change their actions when faced with massive media attention (see for example McCombs and Shaw (1972), David L. Deephouse (2000) and Zyglidopoulos et al. (2012)). In my study, I therefore use the number of articles in the Norwegian press as the news source.

My hypothesis is that a high intensity of news will be positively correlated with a high level of trading. When new information arrives to the market, the investors change their views of earning potential (be it based on rational or sentimental evaluations) and thus trading is induced due to the different conclusions reached by different investors. Increased trading volume and different opinions about whether to sell or buy, create price changes. But some news stories might change the prices of the share without it being driven directly from volume. This is why it is natural to assume that the two variables volume and intraday price spread are correlated, but reflect different sides of investor behaviour related to news.

My variable "intraday price spread" is a measure which expresses the difference between the highest and lowest price, divided by the average (formal definition on page 20), which is identical to how Lillo et al. (2014) defines volatility in their study. However, I will use the term *intraday price spread* as opposed to *volatility* to avoid confusion with the variance of returns, which is more commonly associated with the term *volatility*.

In terms of the financial data, I will look at one company only: Norwegian Air Shuttle ASA (commonly known as Norwegian). This company's shares were chosen because it is a company that virtually everyone in Norway knows about; it is the 9th largest airline in Europe with 29.3 million passengers carried in 2016, right behind Scandinavian Airlines (SAS) with their 29.4 million passengers; it is the 31st largest company in Norway with 22.491 million NOK in revenue (2016) (and the biggest in the transport industry where SAS is in 4th place); and it is considered one of the most liquid shares on the Oslo Stock Exchange, having been part of the OBX-list the whole time period covered by this study.[1] From a media attention point of view, the company is also an interesting candidate: in the period covered by this paper, Norwegian is mentioned in 100 000 articles, compared to the companies above and below Norwegian on the list of largest companies (DNV GL and Tine Gruppen), which are mentioned in 18 000 and 13 000 articles, respectively. Its biggest competitor, SAS, is mentioned in practically the same number of articles as Norwegian, but articles where Norwegian is mentioned and SAS is not mentioned, is 68 000, compared to 50 000 articles where SAS is mentioned and not Norwegian. This indicates that Norwegian is a company constantly in the public eye, and widely covered by the media. The company's investor profile is also suitable for this study as the company has a 70/30 ratio of institutional to non-institutional investors, and 75% of the shares are held by Norwegian investors (Bloomberg 2017).

According to researchers, there is a divide between "rational traders" and "sentiment traders", and the way in which they acquire information, interpret it, and act on it (see for example Barber and Odean (2008), Engelberg, Reed, and Ringgenberg (2012) and Lillo et al. (2014)). Individual investors are considered to be more sentiment based and display attention-driven buying behaviour, whereas the opposite is true for institutional investors (Barber and Odean 2008). In addition, studies have shown that different news categories display different level of relevance, and only the "relevant" news had statistically significant effects on asset behaviour (Groß-Klußmann and Hautsch 2011). That is why I find it interesting to divide the news sources into financial press and tabloid press, to see if there is a difference here. The hypothesis is that rational investors should not care about the number of articles per se, but rather the content, and as an extension of this, they should care more about the financial

[1] https://en.wikipedia.org/wiki/List_of_largest_airlines_in_Europe (accessed: 11.04.2017), http://www.kapital500.no/index.php?option=com_kapital500&view=companydetail&cid[]=31&Itemid=101 (accessed: 11.04.2017), https://www.oslobors.no/markedsaktivitet/#/details/OBX.OSE/overview (accessed: 11.04.2017).

press than the tabloid press. But are the investors in Norwegian rational or sentiment-based? I just mentioned that Norwegian is subject to high media attention compared to other comparable companies, and my data shows that VG and Dagbladet alone had from the same to double the number of articles as the whole financial press together in one third of the weeks covered by this study. In addition to being tabloid, these two news sources were the two most viewed online in Norway in 2016, and they hold a second and fourth place, respectively, in the TV, radio and print category (Newman et al. 2016). Is this tabloid intensity affecting investors, or are they able to screen out the relevant news from the tabloid?

The objective of this paper is to contribute to the understanding of the market dynamics in response to news. Tauchen and Pitts (1983) illustrate that there is an equilibrium between a share's price and all investors' individual reservation prices for said share, and that trading volume is a product of investors changing their reservation prices. These reservation prices should be based on the investors' expectations of risk and return (Berry and Howe 1994), which, for a rational investor, should be based on analysis of the fundamental financial information of the company and its environment (see e.g. (Cornell 2013)). There might also be other reasons to trade, for example individual liquidity or speculative desires (Karpoff 1986), but the majority should come from the investors' views on the company. Norway is a small country, and the role of the tabloid media is big, with VG and Dagbladet being the preferred source of news for many. In order to better understand the dynamic between media and asset behaviour in Norwegian market, it is therefore necessary to take the whole media picture into account.

My aim is to develop two parsimonious linear regression models that describe the relationships between the Norwegian media and asset behaviour in terms of trading volume and intraday price spread, respectively. It should be parsimonious in that it contains weekly data, and few and simple variables. At the same time, it should add some novelty to the existing literature.

Current research has found significant correlation and relationships, although with varying magnitude and robustness, between trading volume and number of mentions of the company in financial news wires, such as the Dow Jones news wire, Reuters' North American Securities News Service, the Securities Industry Research Centre of Asia-Pacific (SIRCA), and the print version of Financial Times (see for example Mitchell and Mulherin (1994), Berry and Howe (1994), Kalev et al. (2004), Tetlock (2010), Groß-Klußmann and Hautsch (2011) and Alanyali, Moat, and Preis (2013)). The relationship between volume and financial- and "relevant" articles is thus established, and the novelty of my research is that I look at the total amount of firm-specific news, not only that which is provided by the financial news sources, and various ratios between the financial and non-financial news sources. In the research mentioned above, there has been used different techniques and different variables, but to my knowledge, no one has studied any news sources that are not in some way considered “financial”.

The predefined aim is, as mentioned, to develop two models: one on volume and one on intraday price spread. My approach is exploratory in that I employ different methods to make the time series stationary, and I test different model specifications to figure out which media related variables are best at explaining trading volume and intraday price spread. The data in my study consists of non-stationary times series on daily share price and trading volume, and weekly numbers of news articles. I look at the total number of articles, the total number of articles in the financial press and the total number of articles in the tabloid press; I look at ratios such as financial to total, tabloids to total and tabloids to financial; all the while keeping in mind that the combination of variables must make substantive and logical sense. I also include control variables which is known to have an effect on the dependent variables in question. I end up using first difference on all variables, and get models on trading volume where the total number of articles and the ratio of financial press to total have clear and statistically significant explanatory power, and the ratio of tabloids to financial press has a less clear relationship. On intraday price spread it seems like the effects are not direct, but indirect through spill-over from the effects on trading volume.

The next section contains a more detailed literature review on the field of media and asset-/investor behaviour; section 3 provides detailed definitions and explanation of the data frame of this study; section 4 contains analysis and empirical results, alongside methodology specifications; section 5 presents diagnostics of the regressions; and section 6 wraps up the paper with a conclusion.

2. Literature Review

One of the earliest papers on how investors access information through news stories is written by Thompson, Olsen, and Dietrich (1987). They look at a vast amount of news stories, categorize them and study them in depth. They discover that security returns associated with the release of firm-specific news items appear to differ systematically, either in mean or variance (or both), from returns on days without firm-specific news. Many have since looked at the relationship between information flow and various aspects of asset behaviour.

A great deal of this research revolves around stock prices and returns, and they have in common that they often find drift and lagging response. This supports theories that investors are slow to react to news, and that they overreact to price shocks, causing excess trading and volatility which leads to quick reversal; see for example Wesley S. Chan's (2003) study on stock price movements related to media attention to similar price movements unrelated to media attention. Chan (2003) finds that stocks with news exhibit momentum (up to 12 months for bad news, less for good news), while stocks without news have little to no drift. Sprenger et al. (2014) show that the stock market impact of news events differs substantially across different categories (e.g. *Technical Trading*, *Financial Issues*, *Legal Issues*, etc.). They distinguish between good and bad news and demonstrate that the returns prior to good news events are more pronounced than for bad news events. This suggests that leakage of information is more widespread before good news than before bad news, meaning that bad news come more surprisingly than good news. In addition, they show that the different categories of news display different levels of surprise to the marked (e.g. news related to M&As and earnings surprise more than those related to product development and joint ventures).

Many researchers have looked at different types of news individually: Bernard and Thomas (1990), amongst others, study surprises in earnings announcements; Michaely, Thaler, and Womack (1995) study news regarding dividend initiations and omissions; stock splits are covered by Ikenberry and Ramnath (2002); and Michaely and Womack (1999) examine responses to changes in analyst recommendations. Ederington and Lee (1993) study the impact of scheduled macroeconomic news on interest rate and foreign exchange futures; Birz and Lott Jr (2011) examined macroeconomic news concerning GDP growth, unemployment, retail sales, and durable goods, and find that the news about GDP growth and unemployment significantly affect stock returns.

Joulin et al. (2008), on the other hand, argue that the frequency and amplitude of price jumps cannot be explained by idiosyncratic – or even market wide – news. They find, however, that the price jumps that were in fact accompanied by news display lower volatility levels than those not accompanied by news.

In contrast to Chan (2003), Clara Vega (2006) finds that sometimes investors underreact to public announcements and sometimes public announcements make prices more informative. She looks at how private and public information received by agents prior to earnings announcements affected the post-earnings announcement drift, and her key finding is that public announcements that generate underreaction are associated with the arrival rate of noise traders, while public announcements that make markets more efficient are associated with the arrival rate of informed traders. Her empirical results show that the more information (private or public) investors have about the true value of an asset, and the more they agree and trade on this information, the smaller the abnormal return drift. This is also consistent with "the stylized fact that small firms, on average, realize greater post-announcement drift than large firms, which tend to be more transparent." (Vega 2006, 131). Tetlock's (2010) findings also suggest that each news story conveys more information for small companies than for big ones, and consequently that news resolves asymmetric information to a greater extent for these firms.

Tetlock (2010) uses an extensive news variable which includes every article on publicly traded American companies in the Dow Jones news archive from 1979 to 2007, and proposes a model (on stock price) in which public news stories eliminate an information asymmetry between informed and uninformed investors. He finds four patterns in postnews returns and trading volume that are consistent with this theory, and some evidence which is inconsistent with alternative theories wherein traders interpret news differently for rational or behavioural reasons, such as those of Harris and Raviv (1993) and Kim and Verrecchia (1991), among others. For example, Tetlock (2010) shows that public news is connected to substantially lower reversals of stock returns, higher volume-induced momentum in returns, temporarily higher correlations between absolute returns and volume, and lower price impacts. In Harris and Raviv's (1993) model, they disregard the private information and assume that investors' different positions are due to different interpretation of public news announcements. Kim and Verrecchia's (1991, 302) view is that investors are "diversely informed and *differ in the precision* of their private prior information; they therefore respond differently to the announcement" (italics in the original).

Engelberg, Reed, and Ringgenberg (2012) address the fundamental question "how do informed traders become informed?". In large, their results suggest that "informed traders" (here defined as short-sellers) gain their information advantage not from the ability to create or anticipate news, but merely from superior processing of public news compared to "uninformed traders". These results are thus in contrast to those of Tetlock (2010), and more in line with Kim and Verrecchia (1991). Additionally, Engelberg, Reed, and Ringgenberg's (2012) study shows that although short-selling's predictive power on future returns is twice as strong on news days, they cannot find evidence of profitable short-selling strategies around news events, because of the associated increase in the bid-ask spread.

Tauchen and Pitts (1983) study the relationship between the variability of the daily price change and the daily volume of trading on the speculative markets. They construct a model where each trader's desired position in a given stock depends on the stock's market price and the trader's personal reservation price, and trading volume is determined by all the traders' reservation prices in that the average of the reservation prices clears the marked. Following this logic, they depict a dynamic where a piece of news arrives to the market and changes the traders' reservation prices, resulting in a change in the market price by the average of the increments to the traders' reservation prices. The associated volume of trading would be half the sum of the absolute values of the changes in the traders' positions. This is linked to the Mixture of Distribution Hypothesis (MDH), which argues that the rate of information arrival influence both volatility (variance of returns) and volume simultaneously (Clark 1973, Epps and Epps 1976, Tauchen and Pitts 1983, among others). In the study of Tauchen and Pitts (1983), "news arrival" is not based on actual news data, but is treated as an unobserved mixing variable.

Engle and Ng (1993) study how news is incorporated into volatility measures, and test various ARCH-type models to decide which one is best equipped to handle the asymmetric effect of news on volatility. They find that the best model is the one proposed by Glosten, Jagannathan, and Runkle (1993). Similar to Tauchen and Pitts (1983), Engle and Ng (1993) do not look at actual news articles; their definition of "news" is unexpected increases or decreases in the return (compared to the expected return given yesterday's available information). Following the arguments put forward by Joulin et al. (2008), though, this does not constitute a satisfying proxy for news, as far from all jumps could be explained by news-flow. Also Kalev et al. (2004) argue that volume is a noisy and inadequate measure of information flow. Later research uses data on actual news articles and endorses that augmenting GARCH- and GARCH-Jumps models with news intensity improve their explanatory power compared to "pure" GARCH- and GARCH-Jump models respectively (Sidorov, Date, and Balash 2013, Sidorov et al. 2014).

Mitchell and Mulherin (1994) try to link volume and volatility to observed measures of information, when they study the relation between the number of news announcements reported daily by Dow Jones & Company and aggregate measures of market activity (across NYSE, AMEX, and OTC), including trading volume, the absolute value of market returns, and the sum of the absolute value of firm-specific returns. They find statistically significant and positive correlation between news and volume and news and absolute values of market returns. They also include variables such as the size of the

New York Times headlines, and days with major macroeconomic announcements, and find above-average market returns on days with large headlines. Trading volume is not affected by this, however, only the market returns. Macroeconomic news, on the other hand, has no effect on neither returns nor volume, which is inconsistent with Ederington and Lee (1993) who find that periodic macroeconomic news induce higher trading volumes. Although the links Mitchell and Mulherin (1994) find are robust to various tests and alterations to their model, significance is varying and sometimes weak. According to themselves this might be due to them using aggregate levels of complex data, and a more detailed study could give different results.

Also Berry and Howe (1994) look at aggregate data when they test the relationship between public information and measures of trading volume and volatility. They use NYSE as a whole (volume) and the absolute value of the S&P500 (volatility), as well as a broad-based, comprehensive measure for public information which includes all news stories from Reuters' North American Securities News Service: firm-specific and industry information, macroeconomic, political, and international stories relevant to American financial markets. Their intraday results suggest a positive, moderate relationship with trading volume, but an insignificant relationship with price volatility.

Kalev et al. (2004) study information flow, in terms of firm-specific announcements provided in the Securities Industry Research Centre of Asia-Pacific's (SIRCA) news service, and its relation to volatility and volume on the Australian Stock Exchange (ASX). Using a sample of index and company returns in Australia, their analysis reveals a positive and significant impact of the selected news variable on the conditional variance of stock returns and on the de-trended trading volume, however the latter being more inconsistent and not very strong – they meet thus some of the same difficulties as reported by Mitchell and Mulherin (1994). In a later study, Kalev and Duong (2011) measure the same information proxy against the volatility of the S&P/ASX 200 Index and the SPI 200 Futures. Also here they find a positive impact on volatility, and that the level of volatility persistence is significantly reduced in both equity and futures markets after controlling for the effect of news arrivals on volatility.

Alanyali, Moat, and Preis (2013) find that the number of times a company's name is mentioned in the print issue of the *Financial Times* is significantly correlated with both trading volume and absolute return (ignoring direction) of the said company. They find no evidence for a similar relationship when direction of the return is taken into account. Yu, Mitra, and Yu (2013) study how quantified news sentiment data can be used to predict return, volatility and liquidity (here: bid-ask spread), and their study shows that inclusion of news impact data does in fact improve the prediction of volatility and liquidity, but not of return.

Tetlock (2007) studies the relationship between sentiment of the *Wall Street Journal*'s "Abreast of the Market" column and the U.S. stock market. His findings are that high values of media pessimism induce downward pressure on market prices, and unusually high or low values of pessimism lead to temporarily high market trading volume. However, the negative returns following negative sentiment are reversed over the next few days of market activity, and pessimism weakly predicts increases in market volatility. In sum, he could not say that pessimism represented negative fundamental information not yet incorporated into prices. It is unclear whether a sentiment-based trading strategy would be profitable after accounting for transaction costs, due to the high-frequency arbitrage such a strategy would require. He suggests that these limitations may prevent markets from responding efficiently to the information embedded in media content. The following year, Tetlock, Saar-Tsechansky, and Macskassy (2008) suggest that linguistic media content captures otherwise hard-to-quantify aspects of firms' fundamentals, which investors quickly incorporate into stock prices. In certain aspects, these two papers are hence contradicting each other, where the former "reject[s] the hypothesis that media content contains new information about fundamental asset values" (Tetlock 2007, 1140).

Groß-Klußmann and Hautsch (2011) use an automated news analytics tool by the Reuters Company to analyse news about companies traded at the London Stock Exchange (LSE). The tool automatically disentangles relevant news items from irrelevant ones, identifies the sentiment (by using linguistic pattern recognition algorithms) and the novelty of the news items, as well as excluding earnings announcements. They find that only cumulative trading volumes and volatility are directly influenced by news releases, whereas reactions in bid–ask spreads and trade sizes are only indirectly influenced (through dynamic spill-overs from volatility and volume effects). Secondly, they find that significant market response is only observed for news items which the machine identified as "relevant". What the machine identifies as relevant might be related to the categories identifies by Sprenger et al. (2014), but such a comparison does not exist. In regard to the sentiment of the news items, they find a certain predictability for future price trends, but are unsuccessful in identifying profitable sentiment-based trading strategies.

Attention is a scarce cognitive resource (Kahneman 1973), and this affects the way investors act on the stock market. Barber and Odean (2008) discuss the jungle of offers in the stock market, and hypothesise that many investors only consider purchasing shares after they first somehow have caught their attention. They point out that this is different for sellers, as they only have a limited amount of shares to choose from (their own portfolio), whereas buyers have thousands (the whole market). This is true, however, only for the individual investors, as institutional ones generally have bigger portfolios, they short-sell to a greater extent, and time is less of a scarce resource for them. Since there for every buyer must be a seller, their derivative anticipation is that individual investors will exhibit higher buying tendency than institutional ones on high-attention days – and vice versa. They use three proxies for attention: news stories on the Dow Jones News Service, unusual trading volume, and extreme previous-day returns. They study trading records from brokerage firms, and find that individual investors do indeed display attention-driven buying behaviour: they are net buyers on high-volume days, following both extremely negative and extremely positive one-day returns, and when stocks are in the news. The opposite is true for the institutional investors.

In a later study on different investor categories' trading behaviour relative to endogenous (price return and volatility) and exogenous (number of news articles and their sentiment) factors, Lillo et al. (2014) find that all the different categories of investors are sensitive to the number of articles, although at varying levels. For example, governmental and non-profit investors were the only categories (out of 6) where news had higher impact than volatility, but these were also the two categories where the model had the lowest explanatory power overall. For households and companies (where the model had higher explanatory power) exhibited a sensitivity to news that was lot lower than to volatility. Their study argues thus that news is significantly correlated with the decision to trade, but that volatility has a higher explanatory power. Note that in their study, volatility is defined as the difference between the highest and lowest price, divided by the average, not as the variance of returns. Finally, they find little importance of the sentiment of the news, but they admit that their sentiment indicator might not be strong enough, as its correlation with returns is very low. A different explanation proposed, is that investors interpret news differently, in line with other researchers already mentioned (e.g. Harris and Raviv (1993)).

The fact that they find little importance of the sentiment, is an interesting contest to the intuitive, and to research coming from other fields than finance and economics. A wide range of research show that negative stimuli elicit greater affective, cognitive, and behavioural influences than equally intense positive stimuli, see, among others, Kahneman and Tversky (1984), Taylor (1991), and Rozin and Royzman (2001). Katsyri et al. (2016) study people's allocation of cognitive resources, and they conclude that negative tweets draw longer viewing time and more attention to themselves than positive tweets. It is therefore natural to think that you cannot do any news research without taking sentiment into account, but fact is that several researchers have found interesting findings which excludes sentiment (for example Lillo et al. (2014)). Also, as mentioned already, several researchers have difficulties creating profitable trading strategies based on sentiment, even when considering the sentiment findings that are indeed significant.

Other researchers use Google Search Volume Index (GSVI) as a measure of attention, and look at its effect on stock prices, see for example Da, Engelberg, and Gao (2011). Search engine data has, in other fields, also proven to have predictive power, see for example Ginsberg et al. (2009), and Choi and Varian (2012). Da, Engelberg, and Gao (2011) find that SVI is correlated with, but different from, existing proxies of investor attention, and that it likely measures the attention of retail investors in particular. Moreover, their results show that an increase in SVI predicts higher stock prices in the next 2 weeks, and an eventual price reversal within the year.

Ranco et al. (2015) study Twitter volume and sentiment about the 30 stock companies forming the Dow Jones Industrial Average, and find a significant dependence between the Twitter sentiment and abnormal returns during the peaks of Twitter volume – both when they do, and when they do not, correspond with expected peaks (such as around earnings announcements). They find that the sentiment polarity of the tweets implies the direction of the cumulative abnormal returns (CAR), and the length of the period of impact (longer for negative than for positive). The polarity variable is, however, not useful for predicting the price returns, as only three companies passed the Granger causality test.

Souza et al. (2015) investigate the relationship between Twitter sentiment and returns and volatility for five retail brands listed on the U.S. stock market. They find that Twitter sentiment is relevant for prediction of the next day's excess return, and exhibits stronger Granger-causality than traditional newswires (Dow Jones Newswire, the *Wall Street Journal* and *Barron's*). On the volatility side, the traditional news wires show stronger prediction power and Granger-causality. In both cases, positive Twitter sentiment has stronger relevance than negative, which in some cases shows no relevance at all. In total, this shows that Twitter is a good and reliable complement to the traditional news proxies, but it does not in any way discredit the traditional news wires.

3. Data

This chapter describes the time frame for this paper, the variables used in the study, and the sources from which the data stems.

3.1. Time period

The period covered by this research paper is from Monday 28.12.2009 to Sunday 03.07.2016. This is a period of 340 full weeks, or 2 380 days. In this period, the Oslo Stock Exchange had 1 633 open trading days and Norwegian's shares were traded on every one of these open trading days.

I chose this periode because I wanted to avoid the turmoil of the financial crisis of 2008, and as we can see in figure 1, it seems like the NAS stock had completely recovered from the crisis by the end of 2009. The period ends Sunday 03.07.2016 because by the time this study started, the financial data available in TITLON was up until Friday 01.07.2016.

All dates presented in this paper are in the following format: [day].[month].[year].

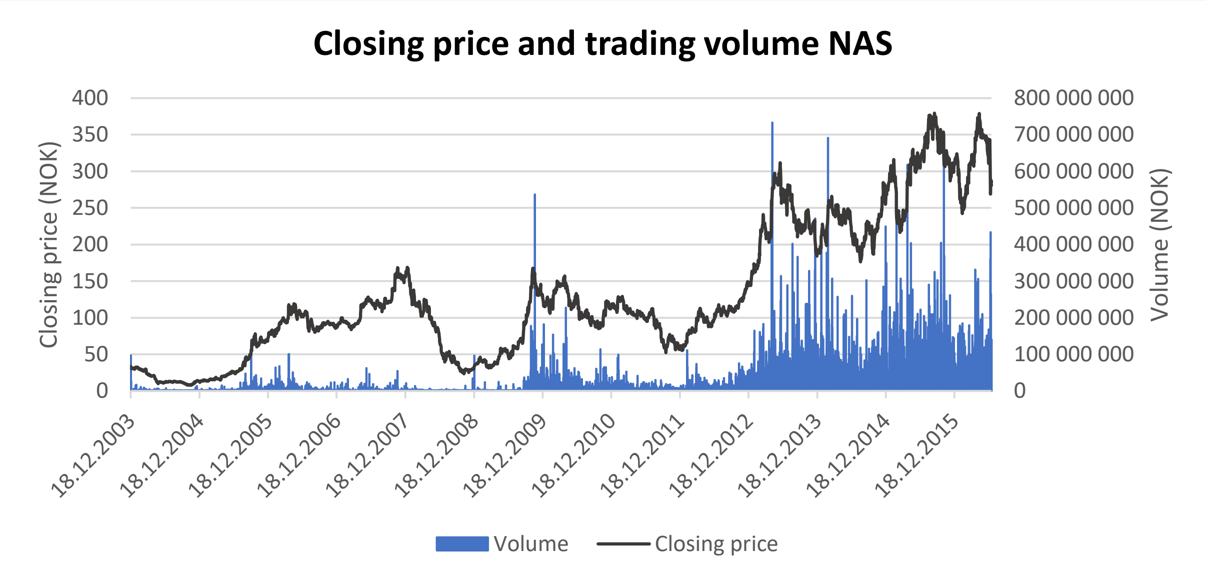


Figure 1: Closing price and trading volume NAS, from its IPO 18.12.2003 up until 01.07.2016.

3.2. Financial data

The financial asset in question is the common stock of the Norwegian airline Norwegian Air Shuttle ASA (commonly known as Norwegian), traded on the Oslo Stock Exchange (OSE) with the ticker NAS. The choice of this company is discussed in the introduction.

The trading volume and price data of the share is collected from TITLON, a financial database for Norwegian universities and university colleges, containing daily data from Oslo Stock Exchange. The trading volume for the Oslo Stock Exchange All Share Index (OSEAX) was provided to me directly from the Oslo Stock Exchange.

DESCRIPTIVE SUMMARY OF THE FINANCIAL DATA			
	Price NAS (daily)	**Volume NAS** (weekly)	**Volume OSEAX** (weekly)
Mean value	189.7	325 195 503	20 412 858 504
Minimum value	52.3	13 405 708	3 079 337 643
Maximum value	379.5	1 285 962 285	45 897 092 095

Table 1:Descriptive summary of the financial data.

3.3. News data

The data regarding the number of news articles is collected from ATEKST, a comprehensive database of all articles from 317 Norwegian newspapers and magazines in paper format, 1618 in web format, 7 in the format of web-TV and 1 radio channel.

Finding the exact number of articles mentioning the company Norwegian was a challenge. The company's full name is Norwegian Air Shuttle ASA, but it is very well known in Norway by simply its brand name, Norwegian. A number of articles will therefore mention the company by its brand name, without using the full nume even once. But the word *Norwegian* also appears in a number of different company names, names of ships etc. Ergo, searching for *Norwegian Air Shuttle ASA* will exclude relevant articles, and searching for *Norwegian* will include irrelevant ones. My solution to maximise the number of relevant articles and minimising the number of irrelevant ones, was the shearch *Norwegian + fly**. This search picks up all articles which includes both the word *Norwegian* and any word starting with *fly*, for example *flyselskap* (airline), or simply the word *fly* (airplane).

This search gave me a total of 100 077 articles in the period covered by this paper. The sources were categorised (by myself) as either financial, tabloid or other. For details on what counts as financial and what counts as tabloid, see table 3 "Explanation of the variables" on the next page.

DESCRIPTIVE SUMMARY OF THE NEWS DATA			
	All articles	**Financial**	**Tabloid**
Total	100 077	6 666	4 699
Mean, weekly	294.3	19.6	13.8
Minimum in one week	61	2	2
Maximum in one week	2 726	258	137

Table 2: Descriptive summary of the news data.

3.4. Descriptive statistics

This section provides a descriptive summary of all the variables used in the process of this study.

EXPLANATION OF THE VARIABLES	
price	Closing price (NOK) each trading day.
volume	Sum of the traded volume of each open trading day in a normal calendar week, Monday to Sunday. Traded volume defined as "Official volume" + "Unofficial volume", as defined by TITLON, measured in NOK.
oseax	Sum of the traded volume of each open trading day in a normal calendar week, Monday to Sunday, of the Oslo Stock Exchange All Share Index (OSEAX), measured in NOK. The OSEAX contains all listed shares on the Oslo Stock Exchange (OSE), adjusted for share events, total number of shares outstanding, and dividends. Norwegian Air Shuttle ASA has been included in the index the whole period covered by this paper.
spread	Arithmetic average of the intraday price spread (NOK) of each open trading day in a normal calendar week, Monday to Sunday. Intraday price spread defined as the following: $2\frac{\text{Price}_{\text{highest(t)}} - \text{Price}_{\text{lowest(t)}}}{\text{Price}_{\text{highest(t)}} + \text{Price}_{\text{lowest(t)}}}$
oseax.spread	Intraday price spread, defined as above, for the Oslo Stock Exchange All Share Index (OSEAX).
articles	Total number of articles in the period covered by this paper.
finpress	Number of articles in the period covered by this paper, found in sources defined as "financial press". This definition covers the following sources: Dagens Næringsliv, E24, Dine Penger Dn.no, E24 AKSJELIVE, Dine Penger Pluss Dn.no Pluss, NA24, Oslo Børs DN Investor, Stocklink, Business Q4 DN Tekno, Næringsavisen, Gründer Økonomisk Rapport
tabloids	Number of articles in the period covered by this paper, found in sources defined as "tabloid press". This definition covers the following sources: VG Nett, Dagbladet VG (paper version), Dagbladet (paper version) VG Pluss, Dagbladet Pluss VG Nyhetsdøgnet, Dagbladet Web-TV VG TV
finpress.tot	$\frac{\text{finpress}}{\text{articles}} = \frac{\text{Number of articles in financial press}}{\text{Total number of articles}}$
tabloids.finpress	$\frac{\text{tabloids}}{\text{finpress}} = \frac{\text{Number of articles in tabloid press}}{\text{Number of articles in financial press}}$
Δ or **[ ].diff**	The variables marked with Δ or [variable].diff have been differenced once, in the following manner: $[\text{variable}]_t - [\text{variable}]_{t-1}$.

Table 3:Explanation of the variables.

Table 4 below provide a descriptive summary of all the variables.

DESCRIPTIVE STATISTICS: SUMMARY OF THE VARIABLES					
Variable	**N**	**Minimum value**	**Mean value**	**Maximum value**	**Std. deviation**
price	1633	52.3	189.7	379.5	89.98
volume	340	13 405 708	325 195 503	1 285 962 285	287 383 359
oseax	340	3 079 337 643	20 412 858 504	45 897 092 095	7 864 145 749
spread	340	0.0003	0.0354	0.0818	0.0125
oseax.spread	340	0.0055	0.0140	0.0530	0.0066
articles	340	61	294.3	2 726	258.0
finpress	340	2	19.6	258	20.1
tabloids	340	2	13.8	137	11.3
finpress.tot	340	0.0089	0.0693	0.1842	0.0299
tabloids.finpress	340	0.1111	0.9761	5.0000	0.7389

Table 4: Descriptive statistics: summary of the variables.

3.4.1. Remarks on the ratios

The mean value of tabloids.finpress is almost 1, meaning that, on average, there was about the same number of articles in VG and Dagbladet combined as there was in all financial press combined. On the most tabloid intense week, the tabloids had 5 times as many articles as the financial press. On the least tabloid intense week, the tabloid press had 11.11% of the number of articles in the financial press – meaning that there were about 9 times as many articles in the financial press than in the tabloids that week.

In 211 weeks, there were more articles in the financial press than in the tabloids, in 107 weeks the tabloids had more articles, and in 22 weeks they had the same amount. Table 5 below shows a more detailed description of the ratios distribution.

Tabloids > Finpress: 107		Tabloids=Finpress:	Tabloids < Finpress: 211	
$\mathbf{T/F} \in \langle \mathbf{1, 2} \rangle$	75	22	$\mathbf{T/F} \in \langle \mathbf{1,0, 0,8} \rangle$	31
$\mathbf{T/F} \in [\mathbf{2, 3} \rangle$	21		$\mathbf{T/F} \in [\mathbf{0,8, 0,6} \rangle$	58
$\mathbf{T/F} \in [\mathbf{3, 4} \rangle$	7		$\mathbf{T/F} \in [\mathbf{0,6, 0,4} \rangle$	66
$\mathbf{T/F} \in [\mathbf{4, 5} \rangle$	2		$\mathbf{T/F} \in [\mathbf{0,4, 0,2} \rangle$	50
$\mathbf{T/F} \in [\mathbf{5, \infty} \rangle$	2		$\mathbf{T/F} \in [\mathbf{0,2, 0,0}]$	6

Table 5: Distribution of the tabloids.finpress ratio.

This detailed table show us that, even though the tabloids had less articles than the financial press in the majority of the weeks, the most common occurrence was that the tabloid press had from the same to double the number of articles as the financial press: in 102 out of 340 weeks, the ratio is $\in [1, 2]$. The median value is 0.75, which means that in half of the weeks, the tabloids had 75% or more of the number of articles as the financial press.

3.5. Plot of the variables

The next two figures show relevant plots of all the variables used in the linear regressions. It starts with the dependent variables volume and intraday price spread, and continues with the independent variables, presented both as level and differenced.

PLOTS OF THE DEPENDENT VARIABLES USED IN THE LINEAR REGRESSIONS

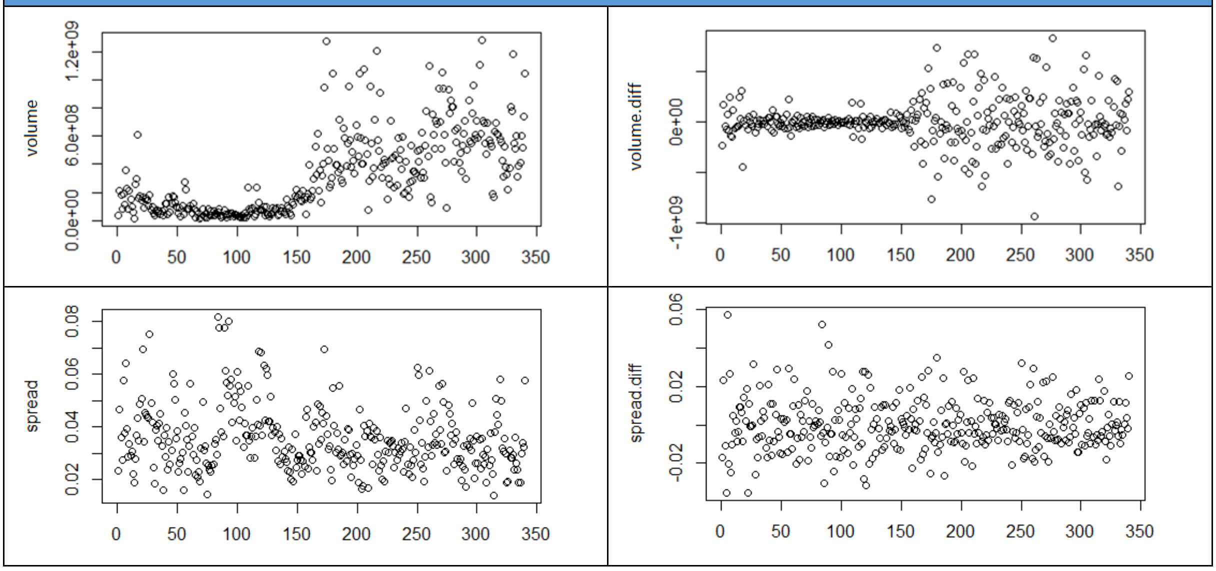


Figure 2: Plots of the dependent variables used in the linear regressions. Note that [variable].diff = Δ[variable].

PLOTS OF THE INDEPENDENT VARIABLES USED IN THE LINEAR REGRESSIONS

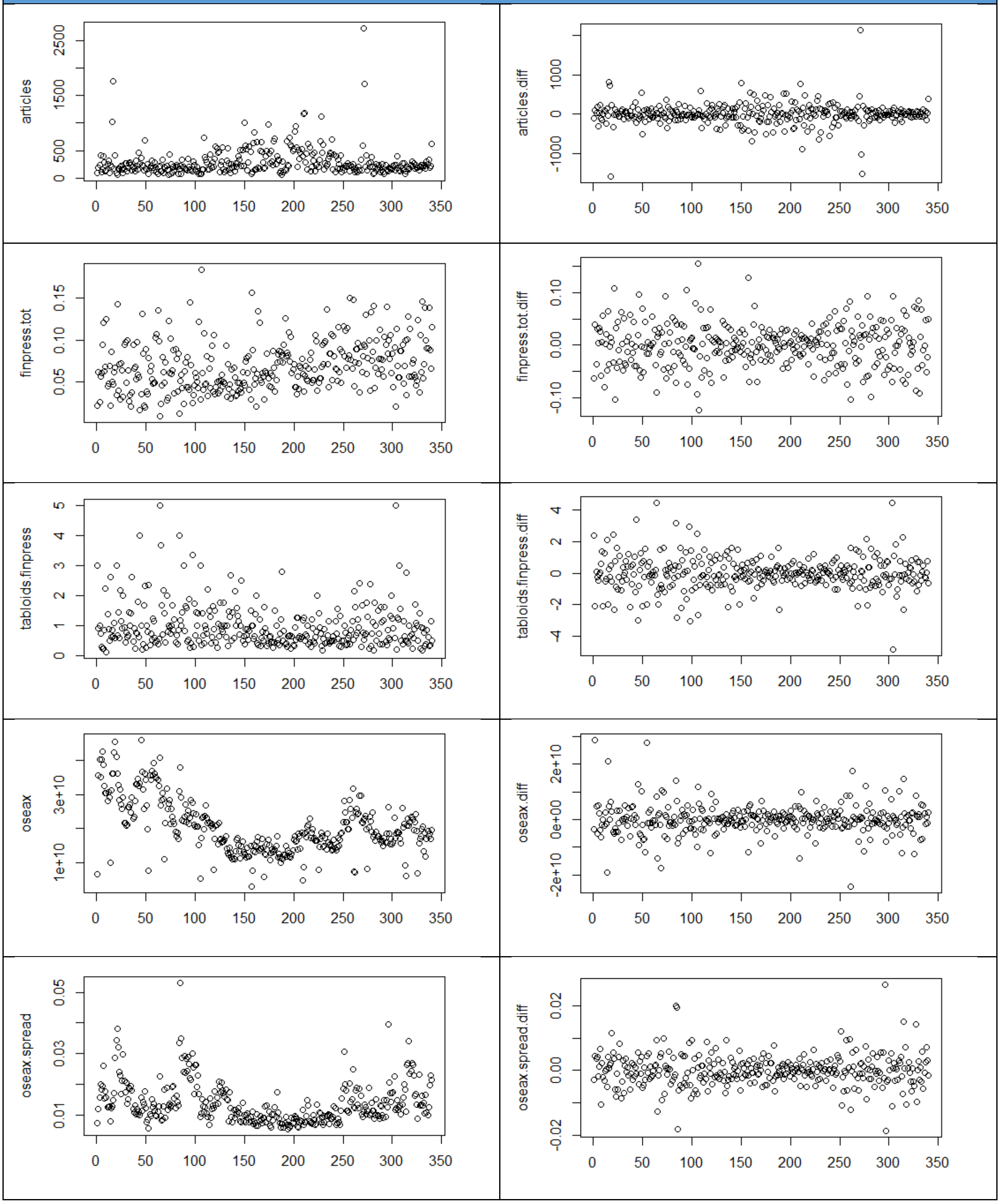


Figure 3: Plots of the independent variables used in the linear regressions. Note that [variable].diff = Δ[variable].

4. Findings and discussion

In this section, I will present the results from the linear regressions performed in my study, and discuss their implications. But first, let us remind ourselves that we are looking at contemporaneous variables and strictly speaking, it can therefore formally not be attached causality to my findings. Still, they provide insight into the dynamics that exists between news intensity and asset behaviour.

4.1. Statistical significance

In any comments on the statistically significance, if nothing else is specified, "statistically significant" means at a confidence level of 95% or higher. If indicated with symbols, they are to be interpreted as follows:

	***	**	*	.	(no symbol)
Confidence level	99,9%	99%	95%	90%	< 90%
Significance level	0,1%	1%	5%	10%	> 10%

4.2. Alignment of data

In my regressions, I use weekly contemporaneous data, meaning that all financial data and all news data are from the same calendar week, from Monday to Sunday. This means that financial data from open trading days in one week is regressed against news data from that whole week. For example: trading volume from Monday to Friday is regressed against news data from Monday to Sunday the same week.

The use of weekly data, as opposed to daily, is chosen for two reasons related to simplicity: (1) restrictions in ATEKST hinders collection of detailed daily news data over such a long period of time; and (2) a time series with daily news data would be much longer than one with daily financial data, as the stock exchange has closed days and (online) newspapers do not. Both of these issues would demand a great deal of manual adaptation of the data, which I consider would not yield adequate payoff. It is an accepted practice in the field of finance to assume five open trading days per week, and therefore use a +/- five days method in constructing weekly data from daily data. As the stock exchange in several weeks per year has less than five trading days, due to bank holidays etc., this process introduces some imprecision, but is nonetheless widely accepted. In my alignment, the financial news has been manually summed up per actual calendar week, meaning that bank holidays have been accounted for. Due to this, I believe my method contains no more imprecision than that of the +/- five method.

In addition, weekly data is also chosen in an attempt to catch the clusters of volume around clusters of news, in an uncomplicated manner. Several researchers have shown a delay in investor behaviour in reaction to news, even after the price has been adjusted. For example, Karpoff (1986) explains that (1) some investors are late in the information queue and adjust their positions disregarding that their information is "old", (2) demand is not always cleared immediately, and (3) "mistaken decisions" that later are corrected. Bradford Cornell (2013) explains that only a minority of the biggest stock price movements can be tied to actual change in the fundamental economic information, and that it seems like investors modify their long-run views on cash flows and discount rates, only to change their minds the next day.

To supplement my regressions, all of them have also been performed with financial data from Monday to Friday in $week_t$ and news data from Monday to Sunday in $week_{t-1}$. These models systematically have lower R^2, lower adjusted R^2, lower F statistics and less significant coefficients. Both for volume and for intraday price spread, all but the OSEAX coefficient become insignificant in the shifted models. The OSEAX coefficients remain significant and show that volume and price spread for Norwegian in $week_t$ can be explained by the previous week's OSEAX volume and -price spread. However, when both a contemporaneous and a lagged volume- and price spread element is included in the regression at the same time, only the contemporaneous one is significant.

4.3. Differencing of variables

All the variables (dependent and independent) in all the regressions are differenced once to induce stationarity and thereby making them more suited for a linear regression. As we can see visually in the plots in figure 2 and 3 on page 22 and 23, the variables all have very uneven spreads. The first difference transformation helps to even this out, and thereby avoid spurious results driven by similar trends. In addition, testing the regression models with level data gives a lot lower R^2, and it gives unfavourable results in specification tests (such as the RESET test). Mitchell and Mulherin (1994) use a 20 days-moving average model to difference "to avoid the loss of information around clustering of high levels of news (…) and to eliminate dependence on day-of-the-week patterns in news and market activity." (Mitchell and Mulherin 1994, 933). As my data is already at a weekly level, my first differenced data hopefully inhibits similar advantages (to a certain extent), although it expresses "jumps" from one average to the next, and not moving average.

In addition to the visual evaluation, I also performed KPSS- and ADF tests on all the variables. The interpretations of the tests are summarised in table 6 below.

UNIT ROOT TESTS OF THE VARIABLES				
	LEVEL DATA		DIFFERENCED DATA	
Variable	**KPSS**	**ADF**	**KPSS**	**ADF**
volume	Unit root	Stationary	Stationary	Stationary
spread	Unit root	Unit root	Stationary	Stationary
articles	Stationary	Stationary	Stationary	Stationary
finpress.tot	Unit root	Unit root	Stationary	Stationary
tabloids.finpress	Unit root	Stationary	Stationary	Stationary
oseax	Unit root	Unit root	Stationary	Stationary
oseax.spread	Unit root	Unit root	Stationary	Stationary

Table 6: Unit root test of the variables. For KPSS level and Dickey-Fuller stats of the tests, see appendix A.

4.4. Correlation analysis

This paper seeks to develop two main models based on ordinary least square (OLS) linear regressions: one with trading volume as dependent variable, and one with intraday price spread as dependent variable. But first, let us have a look at the correlations between the variables. The next two tables show the correlation matrix of all the variables, both as level and first difference. All variables are at a weekly level as described above, price is the weekly average.

LEVEL DATA										
	price	volume	spread	articles	finpress	tabloids	finpress.tot	tabloids.fin	oseax	oseax.spre
price	1.00									
volume	0.77	1.00								
spread	-0.27	0.05	1.00							
articles	0.03	0.31	0.08	1.00						
finpress	0.14	0.42	0.11	0.88	1.00					
tabloids	0.07	0.26	0.04	0.82	0.81	1.00				
finpress.tot	0.30	0.32	0.03	-0.10	0.26	-0.03	1.00			
tabloids.finpress	-0.16	-0.27	-0.02	-0.23	-0.36	0.06	-0.54	1.00		
oseax	-0.34	-0.14	0.30	-0.06	-0.06	-0.07	-0.04	0.04	1.00	
oseax.spread	-0.19	-0.12	0.56	-0.20	-0.17	-0.15	-0.05	0.12	0.46	1.00

Table 7: Correlation matrix of the variables, weekly level data.

DIFFERENCED DATA										
	Δprice	Δvolume	Δspread	Δarticles	Δfinpress	Δtabloids	Δfinpress.to	Δtabloids.fi	Δoseax	Δoseax.spr
Δprice	1.00									
Δvolume	0.06	1.00								
Δspread	-0.26	0.47	1.00							
Δarticles	-0.11	0.37	0.17	1.00						
Δfinpress	-0.13	0.47	0.24	0.85	1.00					
Δtabloids	-0.14	0.16	0.07	0.73	0.66	1.00				
Δfinpress.tot	0.01	0.23	0.13	-0.11	0.30	-0.13	1.00			
Δtabloids.finpress	-0.02	-0.30	-0.14	-0.17	-0.36	0.20	-0.60	1.00		
Δoseax	-0.03	0.35	0.32	0.10	0.10	0.01	0.05	-0.17	1.00	
Δoseax.spread	-0.26	0.13	0.40	0.05	0.02	-0.03	-0.02	-0.04	0.45	1.00

Table 8: Correlation matrix of the variables, weekly differenced data.

For Δvolume, all correlations are statistically significant, except for the one between Δvolume and Δprice. For Δspread, all correlations are statistically significant, except for the ones between Δspread and ∆price, and ∆spread and ∆tabloids. None of the insignificant variable-pairs are regressed against each other in my analysis.

We see that the ∆volume has a correlation with ∆articles of 0.37 (p-value = $1.273e^{-12}$), and an even higher with ∆finpress (0.47 (p-value = $2.2e^{-16}$)). This is not surprising, as financial press articles are expected to carry more substantive information. I am interested in the ratio of financial articles to total, and this correlation (∆volume and ∆finpress.tot) is 0.23 (p-value = $2.368e^{-05}$). The correlation with Δtabloids.finpress is -0.30 (p-value = $2.072e^{-08}$), which is strong, but negative. This indicates that when the tabloid intensity increases compared to financial press intensity, the trading volume of the Norwegian share decreases.

We also notice that the correlation with the OSEAX volume (Δoseax) is high (0.35 (p-value = $2.677e^{-11}$). It is reasonable to believe that movements in the OSEAX index would be associated with the trading volume of the Norwegian share, for a number of reasons. First of all, Norwegian is part of that very index. But more generally, as Norwegian is one of the biggest companies in Norway, and has one of the most liquid shares, it is reasonable to assume that Norwegian's share will move somewhat in line with the general economic development in Norway.

∆spread has a correlation with ∆articles of 0.17 (p-value = 0.001264), and also an even higher correlation with ∆finpress (0.24 (p-value = $6.637e^{-06}$)). I am interested in the ratio of tabloids to total, and this correlation (∆volume and ∆finpress.tot) is 0.13 (p-value = 0.01734). The correlation with Δtabloids.finpress is -0.14 (p-value = 0.009485), which is noteworthy and also negative. As above, this indicates that when the tabloid intensity increases compared to financial press intensity, the intraday price spread of the Norwegian share gets narrower.

We also notice that the correlation with the OSEAX volume (Δoseax) and -spread (Δoseax.spread) is high (0.32 (p-value $2.376e^{-09}$) and 0.40 (p-value = $3.443e^{-14}$) respectively). This for the same reasons as above: Norwegian is part of the index, but it is also reasonable to assume that Norwegian will, to a certain extent, follow the general economic development in the country.

We finally acknowledge that the correlation between ∆volume and ∆spread is 0.47 (p-value = $2.2e^{-16}$). This is reasonable for the reasons already mentioned in the introduction and literature review: it has been is established that greater volume of trading is correlated with greater movements in the price of a company's stock. This correlation is even stronger around news days (Tetlock 2010), so it is clear that this relationship is relevant to my analysis.

The correlation coefficient between Δvolume and Δarticles (0.37) is almost identical to that discovered by Mitchell and Mulherin (1994) between their broadest category of news stories and total trading volume across the NYSE, AMEX, and OTC. This correlation indicates that a mere intensity increase in the media, no matter the news source or content, increases the trading volume of the company's shares. Groß-Klußmann and Hautsch (2011) find that trading on updated news is much more pronounced than on the initial news, and, as a sidebar, Tetlock (2010) finds that the link between news and reduced return reversal is stronger for stories that consist of many newswire messages. Alanyali, Moat, and Preis (2013) find Spearman's Rank correlations from -0.14 to 0.43, with a mean of 0.10, between daily mentions in the print issues of the *Financial Times* of each of the 31 companies listed in the Dow Jones Industrial Average and the transaction volumes of the companies' stocks. This is also in line with my findings, but the fact that they find lower correlations than both Mitchell and Mulherin (1994) and myself could suggest that a print issue of a financial newspaper on its own is too weak of a news source. In any case, there is a correlation between trading volume and my measure of any firm-specific article which makes it look like the investor's exhibit attention-driven trading behaviour, but since my data shows an even stronger correlation between the financial press articles and volume (than between total articles and volume), I have stronger evidence that it is "relevant" information in the financial press that leads to updated beliefs of the fundamentals, and thereby trading behaviour to update one's positions.

The fact that the financial press variable exhibits stronger correlation than the tabloids one is also consistent with the "relevant" theory above. Mitchell and Mulherin (1994, 939) find that "stronger information content gives a larger correlation coefficient, while weaker content gives a smaller correlation coefficient.", and in Groß-Klußmann and Hautsch's (2011) study only "relevant" news displays significant responses.

When it comes to the correlations for Δspread, I have less comparable numbers in existing literature, as the definition of the variable is unlike most other research. But we see that although noteworthy and with the same sign as for Δvolume, Δspread has lower correlations with the variables of interest. The exception is the correlation of 0.40 between Δspread and Δoseax.spread, which is stronger than the analogue correlation of 0.35 between Δvolume and Δoseax. We gather from this that the news variables have less magnitude for the intraday price spread than for the trading volume, and the general economic development on the Oslo Stock exchange has a higher magnitude for the intraday price spread than for the trading volume.

4.5. Initial simple linear regressions

Next, let us see if simple linear regressions give any indications that the news proxies and control variables chosen carry any information of interest. The results, with OLS standard errors in parentheses, are presented in table 9 below.

INITIAL SIMPLE LINEAR REGRESSIONS			
DEPENDENT VARIABLE: ΔVOLUME			
Independent variable	**Coefficient**	**p-value**	**Adjusted R^2**
Δarticles	300 649 (40 762)	$1.27e^{-12}$ ***	0.1361
Δfinpress.tot	1 233 000 000 (287 600 000)	$2.37e^{-05}$ ***	0.04876
Δtabloids.finpress	-63 276 379 (11 017 682)	$2.07e^{-08}$ ***	0.08621
Δoseax	0.01440 (0.002.089)	$2.68e^{-11}$ ***	0.1207
Δspread	7 692 000 000 (778 300 000)	$< 2e^{-16}$ ***	0.2219
DEPENDENT VARIABLE: ΔSPREAD			
Independent variable	**Coefficient**	**p-value**	**Adjusted R^2**
Δarticles	0.000008655 (0.000002662)	0.00126 **	0.02746
Δfinpress.tot	0.04310 (0.01803)	0.0173 *	0.01373
Δtabloids.finpress	-0.001835 (0.0007034)	0.00949 **	0.01684
Δoseax.spread	1.091 (0.1378)	$3.44e^{-14}$ ***	0.1541
Δvolume	0.00000000002914 (0.000000000002949)	$< 2e^{-16}$ ***	0.2219

*Table 9: Initial simple linear regressions. OLS standard errors presented in parentheses. Statistically insignificant intercepts omitted. Statistical significance: *** p<0.001, ** p<0.01, * p<0.05, . p<0.1.*

These results, although maybe futile on their own, show an image worth further investigation. First of all, we have a statically significant coefficient for Δarticles which evidently explain 13.6% of the trading volume of Norwegian. Both ratios on financial press and tabloid press also produce significant coefficients, and seem to explain 4.9% and 8.6% of the trading volume, respectively. The coefficients' size and signs are also as expected: more articles means more trading volume, with the financial articles having a higher coefficient than that of "any article", and a higher relationship between tabloids and finpress means less trading volume. The latter can be interpreted as a sign that the investors are able to recognise that if high media intensity is driven by the tabloids and not the financial press, it should not lead to trading decisions. This indicates that the investors in Norwegian seem to be following the pattern of rational investors, rather than sentiment-based investors. But these variables are not on the same scale and therefore not straight forward to interpret. In the multiple regressions, I will look at them more in depth. Finally, we notice that Δspread and Δvolume has a relatively high explanatory power on each other, and that the sign of the coefficients and their statistically significance are comparable between the Δvolume regressions and the Δspread regressions.

As we will see, the final models developed in this study do indeed contain comparable coefficients to those presented above for Δvolume, but not for Δspread, and that the tabloids to financial press ratio is scrapped in most of the models.

Let us now continue to the multiple regressions and models developed by this study. The diagnostics of the following models will be discussed in detail in the next chapter, but to sum it up: all models on trading volume are robust to the assumptions on OLS regressions on time series, safe of some autocorrelation in the residuals. This has been accounted for by solely reporting Newey-West standard errors. For the intraday price spread model, we have the same situation, but in addition, I have to mention that a RESET test rejects linearity as appropriate shape. More details on this on page 48.

4.6. A model on trading volume

We start by looking at the dynamics between news intensity and trading volume of the shares of Norwegian. The dependent variable is the weakly NOK trading volume of the Norwegian share, for each week in the period covered. This variable is regressed against the following independent variables:

- the number of articles per week
- the ratio financial press to total articles per week
- the weakly NOK trading volume of the OSEAX index
- the average intraday price spread of the Norwegian share per week

For definitions and detailed descriptions of the variables, refer to chapter 3 “Data” on page 18.

The trading volume of the OSEAX index is included as it is reasonable to believe that movements in this variable would be associated with the trading volume of the Norwegian share, which was confirmed by the correlation analysis above. In the multiple regression model, exclusion of the variable leads to significant inflation of the other coefficients and their standard errors, suggesting that inclusion contributes to better isolating the effects of the news activity in the news related coefficients. A likelihood ratio test also strongly prefers the model where OSEAX is included.

The average intraday price spread, which is the dependent variable in the other main model, is included as a control variable because of the high correlation discovered and explained in the correlation analysis. As with the OSEAX index, the intraday price spread variable also reduces the other coefficients and their standard errors when included, and a likelihood ratio test strongly prefers the model where the variable is included.

The ratio tabloids to financial press (Δtabloids.finpress) was initially meant to be part of the analysis, but this solution was eventually scrapped because it systematically inflated standard errors and reduced statistical significance of the other coefficients, while failing to produce a statistically significant coefficient itself. Nor did it give noteworthy improvement of the adjusted R^2 (exclusion: 0.3846, inclusion: 0.3853), and exclusion gave higher F statistics than inclusion (53.96 vs. 43.50). Inclusion also gave fairly similar coefficients for all variables, compared to exclusion, except for Δfinpress.tot which was reduced by 20% in the inclusion model. This indicates that this coefficient might be slightly inflated in the exclusion model, but the argumentation above strongly support exclusion, and so does a likelihood ratio test with a p-value of 0.2338.

To interpret the role of news on trading volume, let us therefore consider the following model:

$$\Delta volume_t = \beta_0 + \beta_1 \Delta articles_t + \beta_2 \Delta finpress.tot_t + \beta_3 \Delta oseax_t + \beta_4 \Delta spread_t + \epsilon_t \qquad (1)$$

However, as we can see in figure 2 on page 22, the plot of the volume variable (even when differenced) shows a dramatic change of character about midway through the time period of this study. In a model based on equation (1) for the whole time period, standardised residual number 174 out of 340 is the first one to deviate more than two standard deviations from the mean. Since then, such deviation is a common occurrence.

Observation number 174 represent the week between 22.04.2013 and 28.04.2013, which is the week after some very important events for Norwegian: 16.04.2013 Norwegian announces plans to open a base at London Gatwick Airport; 18.04.2013 Norwegian publishes a much better 1Q report than expected by the market, sending the Norwegian share to an all-time high; and 19.04.2013 the Federal Aviation Administration (FAA) of the United States approves the new batteries for the Dreamliner airplanes (which Norwegian had a standing order on to launch their first long distance routes later the same year).[2] Evidently, this week mark a fundamental change in the behaviour of the Norwegian share.

[2] http://www.dailymail.co.uk/travel/article-2309903/Budget-airline-Norwegian-launches-12-new-routes-Gatwick-Airport.html (accessed 30.03.17)
http://e24.no/boers-og-finans/norwegian-taper-mindre-enn-ventet/20359118 (accessed: 30.03.2017),
http://e24.no/boers-og-finans/boerskommentar/norwegian-dundrer-til-med-boers-rekord/20359404 (accessed: 30.03.2017),
http://e24.no/boers-og-finans/naa-er-droemmeflyet-endelig-godkjent/20360104 (accessed: 30.03.2017)

I therefore decide to run the model for observations 1 to 173 and 174 to 340 separately, as a supplement to the full model.

In this process, it becomes evident that for the second half of the time period, the coefficient for Δtabloids.finpress actually does produce a statistically significant coefficient. I decided to exclude it in the initial model, but now it appears that inclusion/exclusion decision is not as clear.

When including the ratio of tabloid press to financial press, we consider the following model:

$$\Delta volume_t = \beta_0 + \beta_1 \Delta articles_t + \beta_2 \Delta finpress.tot_t + \beta_3 \Delta oseax_t + \beta_4 \Delta spread_t + \beta_5 \Delta tabloids.finpess_t + \epsilon_t \tag{2}$$

The second half is the only model in which the variable actually produces a statistically significant coefficient, suggesting that it might in fact carry some significant information; inclusion increases half of the standard errors, but reduces the other half; inclusion mostly reduces the statistically significance of the other coefficients, but as it is significant itself, this could constitute a valuable correction; inclusion lowers the F statistics from 51.28 to 43.24, but improves the R^2 from 0.5479 to 0.5599. This is a marginal improvement, but still stronger than for the full period and the first half model. Also here, inclusion gives fairly similar coefficients for all variables, compared to exclusion, except for Δfinpress.tot; but here the difference is even bigger, as the coefficient is nearly halved (-42%). Considering that it is inclusion of a statistically significant variable that causes this reduction, it is likely that exclusion leads to a significant overvaluation of the magnitude of the Δfinpress.tot variable. Finally, a likelihood ratio test for the second half shows a slight preference for inclusion with a p-value of 0.0185 (ergo significant at a 5% level).

In the first half, inclusion induces the same problems as in the full-period model, and a likelihood ratio test prefers exclusion with a p-value of 0.2926 (almost equal to that of the full period model). But, as in the second half model, inclusion reduces the Δfinpress.tot variable in the first half model quite a bit: by 39%.

The inconclusive discussion above is why I decide to report all six models on trading volume. Tables 10 (p. 32) and 11 (p. 33) show the models by presenting the values of the coefficients, Newey-West standard errors in parentheses, and their statistical significance, along with the models' F statistics and R^2 values. The models excluding and including the variable Δtabloids.finpress will hereafter be referred to as exclusion models and inclusion models, respectively.

Models with standardised coefficients are presented in appendix B and a descriptive summary of the relevant variables can be found in appendix D.

SUMMARY: EXCLUSION MODELS ON TRADING VOLUME				
Full period: t_1-t_{340}				
	Coefficient	**p-value**	**F statistics**	53.96
(intercept)	1 339 000 (3 779 300)	0.72333	**Multiple R^2**	0.3918
Δarticles	255 560 (45 605)	$4.39e^{-08}$ ***	**Adjusted R^2**	0.3846
Δfinpress.tot	1 135 300 000 (340 710 000)	0.00096 ***		
Δoseax	0.0083958 (0.0031428)	0.00792 **		
Δspread	5 305 700 000 (1 111 600 000)	$2.72e^{-06}$ ***		
First half: t_1-t_{173}				
	Coefficient	**p-value**	**F statistics**	29.93
(intercept)	3 034 500 (4 263 800)	0.46281	**Multiple R^2**	0.4161
Δarticles	170 960 (42 638)	$9.13e^{-05}$ ***	**Adjusted R^2**	0.4022
Δfinpress.tot	334 010 000 (145 710 000)	0.02313 *		
Δoseax	0.00092192 (0.0010541)	0.38302		
Δspread	2 986 900 000 (676 370 000)	$1.80e^{-05}$ ***		
Second half: t_{174}-t_{340}				
	Coefficient	**p-value**	**F statistics**	51.28
(intercept)	331 550 (8 529 100)	0.96904	**Multiple R^2**	0.5587
Δarticles	250 240 (90 740)	0.00649 **	**Adjusted R^2**	0.5479
Δfinpress.tot	1 761 200 000 (511 720 000)	0.00074 ***		
Δoseax	0.018956 (0.0037929)	$1.49e^{-06}$ ***		
Δspread	10 064 000 000 (2 085 200 000)	$3.19e^{-06}$ ***		

*Table 10: Summary of the exclusion models on trading-volume, Newey-West standard errors in parentheses. Statistical significance: *** p<0.001, ** p<0.01, * p<0.05, . p<0.1.*

SUMMARY: INCLUSION MODELS ON TRADING VOLUME					
Full period: t_1-t_{340}					
	Coefficient	**p-value**		**F statistics**	43.50
(intercept)	1 377 500 (3 778 800)	0.71570		**Multiple R^2**	0.3944
				Adjusted R^2	0.3853
Δarticles	243 230 (47 409)	4.91e^{-07}	***		
Δfinpress.tot	912 820 000 (403 270 000)	0.02424	*		
Δoseax	0.0080656 (0.0032575)	0.01378	*		
Δspread	5 324 600 000 (1 118 900 000)	2.90e^{-06}	***		
Δtabloids.finpress	-14 087 000 (16 653 000)	0.39821			
First half: t_1-t_{173}					
	Coefficient	**p-value**		**F statistics**	24.17
(intercept)	3 041 000 (4 149 100)	0.46463		**Multiple R^2**	0.4198
				Adjusted R^2	0.4024
Δarticles	161 320 (46 003)	0.00058	***		
Δfinpress.tot	203 740 000 (183 080 000)	0.26740			
Δoseax	0.00068633 (0.0010086)	0.49716			
Δspread	3 010 100 000 (675 100 000)	1.51e^{-05}	***		
Δtabloids.finpress	-7 361 700 (6 666 000)	0.27102			
Second half: t_{174}-t_{340}					
	Coefficient	**p-value**		**F statistics**	43.24
(intercept)	550 790 (8 018 000)	0.94532		**Multiple R^2**	0.5732
				Adjusted R^2	0.5599
Δarticles	220 580 (85 333)	0.01063	*		
Δfinpress.tot	1 030 200 000 (613 100 000)	0.09484	.		
Δoseax	0.018382 (0.0039386)	6.39e^{-06}	***		
Δspread	9 910 500 000 (2 073 900 000)	3.95e^{-06}	***		
Δtabloids.finpress	-52 602 000 (22 892 000)	0.02286	*		

*Table 11: Summary of the inclusion models on trading-volume, Newey-West standard errors in parentheses. Statistical significance: *** p<0.001, ** p<0.01, * p<0.05, . p<0.1.*

4.6.1. Exclusion models

Full period model

For the full period, we have a model which explains 39.18% of the variation in trading volume (R^2), with an F statistics of 53.96. The intercept is insignificant and only considered a technical constant. All other coefficients are significant at a 99% confidence level or higher.

The model says that an increase of 1 article per week leads to an increase of 255 560 in trading volume, all other variables held constant. In an average week in terms of trading volume, the Norwegian stock has a volume of 325 195 503, and in an average week in terms of articles, Norwegian is mentioned in 294.3 articles (see table 4 on page 21, or appendix D). This means that an increase of 1 article represent a 0.34% increase from the mean, and the associated increase in trading volume is 0.079% from the mean. Next, looking at the ratio of financial press articles to total, an increase of 1 leads to an increase of 1 135 300 000 in trading volume, ceteris paribus. The mean of the ratio is 0.0693, so this increase of 1 represents a 1443% increase from the mean, and the associated increase in trading volume is 349% from the mean.

But the values of the articles variable span from 61 to 2 2726, and the values of the ratio variable only span from 0.0089 to 0.1842, so an increase of 1 is a bit farfetched to consider. Let us therefore rather consider a 10% increase in the variables. In week 1, given average values for all variables, the full period exclusion model gives a trading volume of 514 430 655. In such a week, increasing from 294.3 articles to 323.7 articles (+10%) leads to a 1.462% increase in trading volume, ceteris paribus. The corresponding increase in the ratio, from 0.0693 to 0.0762 (+10%), leads to a 1.529% increase in trading volume. This indicates that the magnitude of the Δfinpress.tot variable is slightly stronger than that of Δarticles. This, in turn, suggests that an increase in articles leads to an increase in trading volume, and even more so when the increase in news intensity is driven by the financial press rather than the non-financial press.

If we turn our attention to the standardised coefficients (appendix B), we can see that Δarticles has a coefficient of 0.31650 (NW-SE[3]: 0.05630) and Δfinpress.tot a coefficient of 0.20909 (NW-SE: 0.06257). This shows that when their respective tendencies to vary is considered, it is the total number of articles, not the ratio of financial press to total, that has the highest effect on the variation in trading volume.

Let us therefore consider the change in trading volume as response to a change of one standard deviation in the explanatory variables. The standard deviation of the articles variable is 258, and the associated coefficient is, as mentioned, 255 560. This means that an increase of one standard deviation

[3] Newey-West Standard Errors

of articles leads to an increase of 65 934 480 in trading volume. This represents a 20% increase from the mean trading volume of 325 195 503, which is quite substantial. For the financial press to total ratio the standard deviation is 0.0299 and the associated coefficient is, as mentioned, 1 135 300 000. An increase of one standard deviation leads thus to an increase of 33 945 470 in trading volume. This represents a 10% increase from the mean trading volume, which is also quite substantial, but nonetheless only half of what the analogous increase in articles leads to.

First half model

For the first half, we have a model which explains 41.61% of the variation in trading volume (R^2), with an F statistics of 29.93. The intercept is insignificant and only considered a technical constant. The other coefficients have varying significance, but all except the Δoseax variable are significant on a 95% confidence level or higher. For the first half, the Δoseax is not statistically significant.

Analogous interpretations as in the full period model, show us that increases of 1 in the variables articles and financial press to total ratio in the first half represent increases of 0.38% and 1647%, respectively, and are associated with increases in trading volume of 0.15% and 289%, respectively. We have the same problems with the scales: in the first half, they range from 63 to 1 756 and 0.0089 to 0.1842, respectively (see appendix D). We will therefore rather consider a 10% increase in the variables. In week 1, given average values for all variables, the first half exclusion model gives a trading volume of 203 703 865. In such a week, increasing from 265.7 articles to 292.3 articles (+10%) leads to a 2.230% increase in trading volume, ceteris paribus. The corresponding increase in the ratio, from 0.0607 to 0.0668 (+10%), leads to a 0.995% increase in trading volume. This indicates that for the first half, the magnitude of the Δarticles variable is much stronger than that of Δfinpress.tot. This, in turn, suggests that an increase in articles leads to an increase in trading volume, even when the increase in news intensity is driven by the non-financial press.

If we turn our attention to the standardised coefficients for the first half, we can see that Δarticles has a coefficient of 0.39835 (NW-SE: 0.09785) and Δfinpress.tot a coefficient of 0.13750 (NW-SE: 0.06092). This also support the notion that Δarticle has a much higher magnitude than Δfinpress.tot.

Let us consider the change in trading volume as response to a change of one standard deviation in the explanatory variables. The standard deviation of the articles variable is in the first half 234.6, and the associated coefficient is 170 960. This means that an increase of one standard deviation of articles leads to an increase of 40 107 216 in trading volume. This represents a 35% increase from the mean trading volume of 116 001 225, which is quite substantial. For the financial press to total ratio the standard deviation in the first half is 0.0316, and the associated coefficient is 334 010 000. An increase of one standard deviation leads thus to an increase of 10 554 716 in trading volume. This represents a

9% increase from the mean trading volume, which is also noteworthy, but just more than a fourth of what the analogous increase in articles leads to.

Second half model

For the second half, we have a model which explains 55.87% of the variation in trading volume (R^2), with an F statistics of 51.28. The intercept is insignificant and only considered a technical constant. All other coefficients are significant at a 99% confidence level or higher.

Analogous interpretations as in the full period model, show us that increases of 1 in the variables articles and financial press to total ratio in the second half represent increases of 0.31% and 1280%, respectively, and are associated with increases in trading volume of 0.046% and 325%, respectively. We have the same problems with the scales: in the second half, that they range from 61 to 2 726 and 0.0207 to 0.1508, respectively. We will therefore rather consider a 10% increase in the variables. In week 1, given average values for all variables, the second half exclusion model gives a trading volume of 878 520 780. In such a week, increasing from 323.9 articles to 356.3 articles (+10%) leads to a 0.923% increase in trading volume, ceteris paribus. The corresponding increase in the ratio, from 0.0781 to 0.0859 (+10%), leads to a 1.566% increase in trading volume. This indicates that for the second half, the magnitude of the Δfinpress.tot variable is much stronger than that of Δarticles. This, in turn, suggest that an increase in articles leads to an increase in trading volume, and even more so when the increase in news intensity is driven by the financial press rather than non-financial press.

If we turn our attention to the standardised coefficients for the second half, we can see that Δarticles has a coefficient of 0.25477 (NW-SE: 0.09460) and Δfinpress.tot a coefficient of 0.22703 (NW-SE: 0.06634). This shows that when their respective tendencies to vary is considered, the total number of articles has a slightly higher magnitude than the ratio of financial press to total in explaining the variation in trading volume.

Let us consider the change in trading volume as response to a change of one standard deviation in the explanatory variables. The standard deviation of the articles variable is in the second half 300.8, and the associated coefficient is 250 240. This means that an increase of one standard deviation of articles leads to an increase of 75 272 192 in trading volume. This represents a 15% increase from the mean trading volume of 541 753 799, which is quite substantial. For the financial press to total ratio the standard deviation in the second half is 0.0269, and the associated coefficient 1 761 200 000. An increase of one standard deviation leads thus to an increase of 47 376 280 in trading volume. This represents a 9% increase from the mean trading volume, which is also noteworthy, but just more than half of what the analogous increase in articles leads to.

These models show findings consistent with other research: a greater number of news articles correspond to a greater trading volume of the company's shares. This suggest that when the Norwegian media show interest in Norwegian, so do investors. This makes sense, but my findings cannot reveal whether a higher number of articles directly leads to a higher trading volume, or the opposite, or a combination of the two. It is not unlikely that it is a combination of the two, but it is not very common that increased trading volume per se creates headlines in the media. Sudden increases or decreases in returns, on the other hand, commonly leads to articles in the financial press. The variable Δspread is meant to control for changes in prices (if not returns directly), and therefore we should be able to trust the news related coefficients in the model. We notice that between the first and the second half models, the change in Δarticles leads to a higher percentage change in trading volume in the first half. This could relate back to Tetlock's (2010) findings that each news story conveys more information, and consequently serve a stronger purpose in reducing information symmetry, for smaller companies than for bigger ones.

The novelty in my models is that they consider the ratio of financial press articles to the total number of articles, and it follows that this is where my most interesting findings are. For the full period, trading volume is affected slightly more by a 10% increase in the ratio compared to the equivalent increase in the total number of articles. For the second half, this effect is much more pronounced, whereas it is the opposite in the first half. This on its own could indicate that when the company was smaller, investors were less skilled in screening the relevant articles out from the general media intensity. When the company grew, and became more international, it can be argued that this attracted more "rational" investors and less "sentiment based" investors.

However, when we take into account the variables' tendencies to vary, it is revealed that for all three models, the mere media intensity, regardless of the source, has a higher explanatory power on the variation in trading volume. For the full period, one standard deviation increase in Δarticles and Δfinpress.tot is associated with respectively 20% and 10% in increased trading volume. When we split up the model, we see that this difference is much larger in the first half than in the second half. This again also provide support that investors do indeed become *more* rational as the company grows.

Looking at the models in a more elevated view, we can conclude that the model fits better in the second half than in the first half, and that it fits the least for the full period. It seems like the company change too much character in the middle of this period for the model to be a good match for the whole period, and that it fits more for a bigger than a smaller company. In particular, it is highly irregular that the OSEAX volume is insignificant in the first half model.

4.6.2. Inclusion models

Full period model

For the full period, we have a model which explains 39.44% of the variation in trading volume (R^2), with an F statistics of 43.5. The intercept is insignificant and only considered a technical constant. The other coefficients have varying significance, but all except the Δtabloids.finpress variable are significant on a 95% confidence level or higher. The disputed Δtabloids.finpress variable has a coefficient which is clearly insignificant in this model, which means that we cannot confidently say that its effect on trading volume is different from zero. I choose to concentrate my comment to the statistically significant variables only.

Analogous interpretations as in the exclusion models show us that increases of 1 in the variables articles and financial press to total ratio in the full period inclusion model represent increases of 0.34% and 1443%, respectively, and are associated with increases in trading volume of 0.075% and 281%, respectively. We have the same problems with the scales: they range from 61 to 2 726 and 0.0089 to 0.1842, respectively. We will therefore rather consider a 10% increase in the variables. In week 1, given average values for all variables, the full period inclusion model gives a trading volume of 475 600 986. In such a week, increasing from 294.3 articles to 323.7 articles (+10%) leads to a 1.505% increase in trading volume, ceteris paribus. The corresponding increase in the ratio, from 0.0693 to 0.0762 (+10%) leads to a 1.330% increase in trading volume. This indicates that the magnitude of the Δarticles variable is slightly stronger than that of Δfinpress.tot. This, in turn, suggest that an increase in articles leads to an increase in trading volume, even when the increase in news intensity is driven by the non-financial press.

If we turn our attention to the standardised coefficients, we can see that Δarticles has a coefficient of 0.30123 (NW-SE: 0.05860) and Δfinpress.tot a coefficient of 0.16812 (NW-SE: 0.07507). This also support the notion that Δarticle has a much higher magnitude than Δfinpress.tot.

Let us consider the change in trading volume as response to a change of one standard deviation in the explanatory variables. The standard deviation of the articles variable is 258, and the associated coefficient is 234 230. This means that an increase of one standard deviation of articles leads to an increase of 62 753 340 in trading volume. This represents a 19% increase from the mean trading volume of 325 195 503, which is quite substantial. For the financial press to total ratio the standard deviation is 0.0299 and the associated coefficient is 912 820 000. An increase of one standard deviation leads thus to an increase of 27 293 318 in trading volume. This represents an 8% increase from the mean trading volume, which is also noteworthy, but less than half of what the analogous increase in articles leads to.

First half model

For the first half, we have a model which explains 41.98% of the variation in trading volume (R^2), with an F statistics of 24.17. The intercept is insignificant and only considered a technical constant. Of the other coefficients, only that of Δarticles and Δspread are statistically significant. The other coefficients are insignificant, including both the news related ratios. I will therefore only comment on the effects of the Δarticles variable.

Analogous interpretations as in the full period model, show us that an increase of 1 in the articles variable in the first half represents an increase of 0.38% and is associated with increased trading volume of 0.14%. An increase of 10% in the variable articles (from 265.7 to 292.3) in week 1, given average values on all variables, leads to a 2.373% increase in trading volume, ceteris paribus. The fact that for the first half, only the Δarticle variable is significant strongly supports that an increase in articles leads to an increase in trading volume, regardless of whether the increase in news intensity is driven by the financial or the non-financial press. The standard deviation of the articles variable in the first half is 258, and the associated coefficient is 161 320. This means that an increase of one standard deviation of articles leads to an increase of 41 620 560 in trading volume. This represents a 36% increase from the mean trading volume of 116 001 225, which is quite substantial.

Second half model

For the second half, we have a model which explains 57.32% of the variation in trading volume (R^2), with an F statistics of 43.24. The intercept is insignificant and only considered a technical constant. All other coefficients are significant at a 95% confidence level or higher, except for the Δfinpress.tot variable which is only significant at a 90% confidence level. In spite of the lesser significance, I still choose to include comments on the ratio of financial press articles to total.

Analogous interpretations as in the full period model, show us that increases of 1 in the variables articles, financial press to total ratio, and tabloids to financial press ratio in the second half represent increases of 0.31%, 1280%, and 120%, respectively, and are associated with the following respective changes in trading volume: increase of 0.041% and 190%, and decrease of 9.7%.

We have the same problems with the scales: in the second half, they range from 61 to 2 726, 0.0207 to 0.1508, and 0.1591 to 5, respectively. We will therefore rather consider a 10% increase in the variables. In week 1, given average values for all variables, the second half inclusion model gives a trading volume of 753 051 585. In such a week, increasing from 323.9 articles to 356.3 articles (+10%) leads to a 0.949% increase in trading volume, ceteris paribus. The corresponding increase in the Δfinpress.tot ratio, from 0.0781 to 0.0859 (+10%), leads to a 1.068% increase in trading volume. The corresponding increase in the Δtabloids.finpress ratio, from 0.8352 to 0.9187 (+10%) leads to a 0.583% decrease in trading volume. This indicates that for the second half, the magnitude of the Δfinpress.tot variable is slightly stronger than that of Δarticles, and the weakest is that of Δtabloids.finpress. This, in turn, suggest that an increase in articles leads to an increase in trading volume, and even more so when the increase in news intensity is driven by the financial press rather than non-financial press. This effect is marginal though, in an average week. There are, nonetheless, some indications that investors are able to filter out financial press articles from the rest, but that they get a little confused when the intensity in the tabloids is running high. A ceteris paribus increase in Δtabloids.finpress means that the number of articles do not change and the number of financial press articles do not change (to maintain the financial press to total constant), which means that articles only shift from "other" news sources to the tabloid VG and Dagbladet. In fact, this shift should not affect a rational investor, whereas it might lead to increased trading for a sentiment based investor. It is peculiar that it is negative whilst still statistically significant. Let us investigate it further.

If we turn our attention to the standardised coefficients for the second half, we can see that Δarticles has a coefficient of 0.22458 (NW-SE: 0.08259), Δfinpress.tot a coefficient of 0.13280 (NW-SE: 0.07607), and Δtabloids.finpress a coefficient of -0.15668 (NW-SE: 0.06774). This suggests that Δarticle has a much higher magnitude than Δfinpress.tot, and Δtabloids.finpress has a slightly higher magnitude than Δfinpress.tot (yet in the opposite direction).

Let us consider the change in trading volume as response to a change of one standard deviation in the explanatory variables. The standard deviation of the articles variable is in the second half 300.8, and the associated coefficient is 220 580. This means that an increase of one standard deviation of articles leads to an increase of 66 350 464 in trading volume. This represents a 12% increase from the mean trading volume of 541 753 799, which is quite substantial. For the financial press to total ratio the standard deviation in the second half is 0.0269, and the associated coefficient 1 030 200 000. An increase of one standard deviation leads thus to an increase of 27 712 380 in trading volume. This represents a 5% increase from the mean trading volume, which is not insignificant, but less than half of what the analogous increase in articles leads to. For the tabloids to financial press ratio the standard deviation in the second half is 0.6245, and the associated coefficient -52 602 000. An increase of one standard deviation leads thus to a decrease of 32 849 949 in trading volume. This represents a 6% decrease from the mean trading volume, which is about the same magnitude as the financial press to total ratio. This provide some further indications that the investors are sentiment-based and let themselves get confused by in increased intensity in the tabloids. This increased intensity in the tabloids does not lead to an increase in trading volume, but to a maybe misunderstood decrease in trading volume. As the number of financial articles remains unchanged, the relevant fundamental financial information should be unchanged as well.

As with the exclusion models, we can conclude that the inclusion models fit better in the second half than in the first half, and that they fit the least for the full period. In particular, it causes concern that neither the OSEAX volume, nor the financial press to total ratio is significant in the first half model.

Nevertheless, we notice that the significant percentage increases in trading volume associated with increases in the number of articles is comparable between the exclusion- and inclusion models, within each respective time period.

4.6.3. Exclusion or inclusion

I already briefly discussed some of the differences between exclusion and inclusion on page 32, but let us have another look at it. For all the models, the adjusted R^2 values are more or less identical, but inclusion gives a slight improvement for the second half model. The F statistics, on the other hand, are always higher for the exclusion models. For the full period model, all standard errors are smaller in the exclusion model compared to the inclusion model. For the for the first half, the standard errors for the two news variables are smaller, but bigger for the two control variables. One of these control variables (Δoseax) is insignificant, though, and the Δfinpress.tot variable goes from significant only at a 10% level in the exclusion model, to completely insignificant in the inclusion model. For the second half, the Δarticles and Δoseax variables have smaller standard errors in the exclusion model, and opposite for the Δfinpress.tot and Δspread variables. It is the only model where the Δtabloids.finpress coefficient is statistically significant, and the improvement of the adjusted R^2 is noticeable, although still only by one percentage point. It is also the only model where inclusion is preferred over exclusion in a likelihood ratio test. This being said, when the variables' tendency to vary is taken into account, the Δfinpress.tot and the Δtabloids.finpress have almost identical magnitude, but in opposite directions.

One can argue that for the first half, the exclusion model is the best fit, and for the second half, the inclusion model is the best fit. However, in a more elevated look at the whole picture, it is arguable that exclusion is preferable. The most important lesson from the inclusion process is maybe the revelation that the Δfinpress.tot coefficients might be too high in the exclusion models. If this is the case, then the conclusion might be different in the full period exclusion model regarding an increase of 10% in the variables. This increase showed only a marginal higher effect for the Δfinpress.tot variable compared to the Δarticles variable, so a smaller Δfinpress.tot coefficient might switch this around. The conflicting result between a 10% increase and one standard deviation increase in the second half exclusion model, might still remain, as the difference here was bigger (0.923% vs. 1.566% and 15% vs.

9%). If this is the case, then there is agreement between the exclusion- and inclusion models, within each respective time period, on the magnitude of each variable.

4.7. A model of intraday price spread

Let us continue with looking at the dynamics between news intensity and intraday price spread of the shares of Norwegian. The dependent variable is the average intraday price spread of the Norwegian share, for each week in the period covered. This variable is regressed against the following independent variables:

- the number of articles per week
- the ratio financial press to total articles per week
- the average intraday price spread of the OSEAX index per week
- the weakly trading volume of the Norwegian share

For definitions and detailed descriptions of the variables, refer to chapter 3 "Data" on page 18.

As with the volume models, I include comparable OSAEX data as it is reasonable to believe that there will be a relationship between the Norwegian share and the general economy, which should be controlled for, and a likelihood ratio test shows a strong preference for inclusion. Also here, the other dependent variable (now weekly trading volume) is included because of the established relationship between trading volume and price variation. It also decreases the other coefficients and their standard errors, and is preferred by a likelihood ratio test. And also here was the ratio tabloids to financial press was meant to be included, but was excluded due to the following reasons: (1) its coefficient is statistically insignificant and inclusion inflates all other coefficients' standard errors; (2) inclusion decreases the F statistics of the model, and even its adjusted R^2 value; and (3) exclusion is preferred by a likelihood ratio test.

To interpret the role of news on intraday price spread, let us therefore consider the following model:

$$spread_t = \beta_0 + \beta_1 \Delta articles_t + \beta_2 \Delta finpress.tot_t + \beta_3 \Delta oseax.spread_t + \beta_4 \Delta volume_t + \epsilon_t \tag{3}$$

Table 12 below shows the results of this model by presenting the values of the coefficients, Newey-West standard errors in parentheses, and their statistical significance, along with the model's F statistics and R^2 values.

The model's standardised coefficients are presented in appendix C and a descriptive summary of the important variables can be found in appendix D.

SUMMARY: MODEL ON INTRADAY PRICE SPREAD		
	Coefficient	**p-value**
(intercept)	-0.000 039 83 (0.00031600)	0.8998
Δarticles	-0.000 000 420 5 (0.000 002 758)	0.8789
Δfinpress.tot	-0.014 58 (0.021 123)	0.4905
Δoseax.spread	0.9417 (0.136 75)	$2.84e^{-11}$ ***
Δvolume	0.000 000 000 025 57 (0.000 000 000 003 664)	$1.62e^{-11}$ ***

F statistics	42.93
Multiple R^2	0.3389
Adjusted R^2	0.3310

*Table 12: Summary of the model on intraday price spread, Newey-West standard errors in parentheses. Statistical significance: *** p<0.001, ** p<0.01, * p<0.05, . p<0.1.*

Here we have a model which explains 33.89% of the variation in intraday price spread (R^2), with an F statistics of 42.93. The intercept is insignificant and only considered a technical constant. In addition to that, none of the news related coefficients are statistically significant, only the two control variables Δoseax.spread and Δvolume are.

For the sake of argument, let us still comment briefly on the two news variables. The standard deviation of articles is 258, and the associated coefficient is -0.0000004205. This means that an increase in one standard deviation of articles is associated with a decrease in spread of 0.3% from the mean of 0.0354 (see table 4 on page 21, or appendix D). For the financial press to total ratio the standard deviation is 0.0299, and the associated coefficient -0.01458. This means that an increase in one standard deviation of the ratio is associated with a decrease in spread of 1.2% from the mean. The standardised coefficient (see appendix C) of Δfinpress.tot is also bigger than that of Δarticles. It is interesting that this is the only model where the standardised coefficient for the ratio is bigger than that for total number of articles. It is also interesting that the coefficients are negative, indicating that a higher media intensity reduces the intraday price spread, and even more so when the increase in media intensity is driven by the financial press rather than the non-financial press. But, we cannot, with our model, say that these effects are actually significantly different from zero.

It is interesting to note that both news coefficients were statistically significant (and positive) when only Δoseax.spread was controlled for. When Δvolume was introduced, they lost their significance. And looking at the standardised coefficients, Δvolume is the variable with the highest magnitude in explaining the variation of the Norwegian share's intraday price spread. When trading volume increases with one standard deviation, the share's intraday price spread increases by 21%, which is quite substantial. This, however, only confirms the already established relationship between volume and price variation, and my model does not show whether this relationship is stronger on news days or not, which is what was found by Tetlock (2010). Though, with his findings in mind, it is not unreasonable to believe that this might be the same connection we see here.

In Berry and Howe (1994) they cannot identify a significant relationship between their information variable and the volatility or the intraday market returns. As stated by Kalev et al. (2004), though, most studies related to the Mixture of Distribution Hypothesis propose that the rate of information simultaneously influence both volume and volatility. They find support for this in their study, and later, Kalev and Duong (2011) find that the level of volatility persistence is significantly reduced when controlling for news arrival. Yu, Mitra, and Yu (2013) find that inclusion of news data improves prediction of volatility, and Groß-Klußmann and Hautsch (2011) find that volatility is directly influenced by news releases.

All researchers mentioned above use different methodology than I do, however, and they all define volatility as variance of returns. In the study by Lillo et al. (2014), volatility is defined like in my study, but their approach is different. Their conclusion is that volatility itself triggers decisions to trade, and for some categories of investors this factor has a higher impact on the decision to trade than volume. For other categories, changes in volume has higher impact on the decision to trade, than volatility. In any case, it is possible to image, based on the results I have found, that the media intensity affect trading volume (as my models on trading volume show), and this in turn affect the intraday price spread through spill-over effects. In Groß-Klußmann and Hautsch's (2011) study, the trading volume and volatility was influenced directly, as mentioned, but the bid-ask spread was affected indirectly through spill-over from trading volume- and volatility effects.

Although Berry and Howe (1994) did not find significant relations between their measure of public information and volatility or the intraday market returns, they do not infer that the news variable necessary is insignificant. They admit that it might be that their news variable is insufficient, and Groß-Klußmann and Hautsch (2011, 326) says that "the noisiness of less relevant news items is the major reason for the yet missing empirical evidence on statistically significant relationships between intraday news flow and high-frequency market activity." . It is possible to imagine that my news variable is too noisy and fails to adequately filter out relevant news stories from irrelevant ones. But if this is the case, then my findings offer support to the notion that investors too are unable to adequately filter out relevant news stories from irrelevant ones.

5. Regression diagnostics

As described by Wooldridge (2013), the three Gauss-Markov assumptions which need to be fulfilled for the OLS linear regressions on time series to give the best linear unbiased estimators are:

1. Linearity in the parameters
2. No perfect collinearity
3. Zero conditional mean

And two more are needed for the OLS standard errors to be reliable:

4. Homoscedasticity
5. No serial correlation

In addition, a normality assumption is added in order for the OLS standard errors, t statistics, and F statistics to be reliable. I will in this section test these assumptions and briefly discuss their implications.

The conclusion of the discussion below is that the assumptions regarding the unbiased OLS coefficients hold to a fair degree. The biggest problem in my data is the autocorrelation in the residuals, which is accounted for by reporting Newey-West standard errors.

5.1. Linearity in the parameters

To determine linearity, I evaluate the plots of the standardised residuals vs. fitted values, as well as perform a Ramsey Regression Equation Specification Error Test (RESET) on each model.

5.1.1. The models on trading volume

In table 13 below, the results from the RESET tests are presented, along with plots for each model.

The RESET tests suggest that we accept the null hypothesis that linearity is the appropriate shape for all six models. The conclusion is weaker for the first half, but it still holds with a significance level of 5%. The model was tested on data excluding observation 18, which represent the outlier fitted value = -281 630 028.4 (marked with red), and this gave a RESET-test with a p-value of 0.4797. The coefficients of the news related variables were reduced a little, the Δoseax coefficient was increased a little, and the Δspread coefficient remained practically unchanged. All standard errors were reduced, except for that of Δoseax, which increased. However, the statistically significance stayed within the same limits as before for each coefficient, and the adjusted R^2 for the model went down.

Therefore, I decided to *not* exclude the outlier in my analysis, as I cannot find any substantive reason to do this.

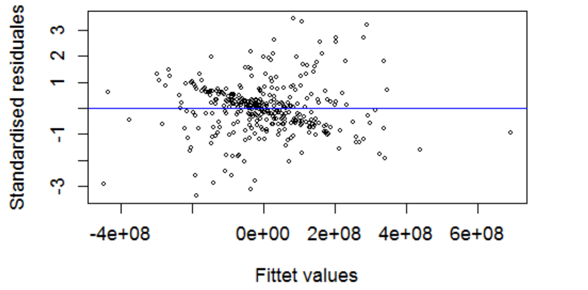
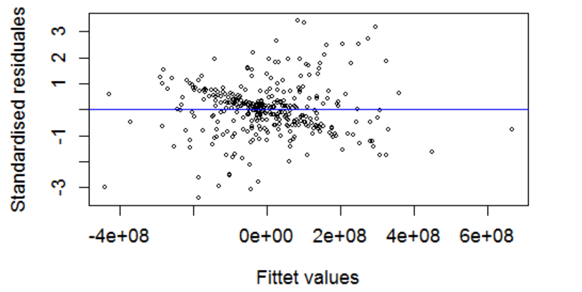
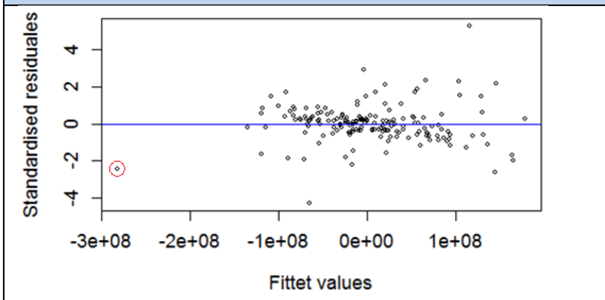
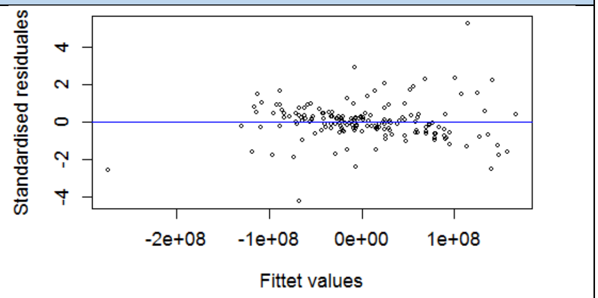
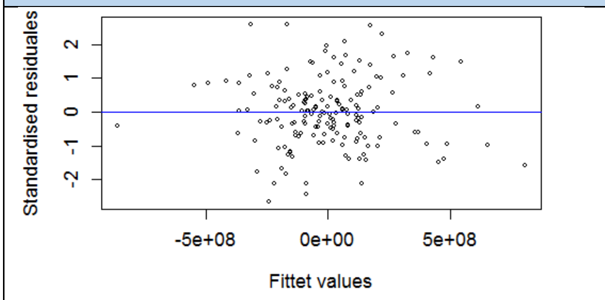
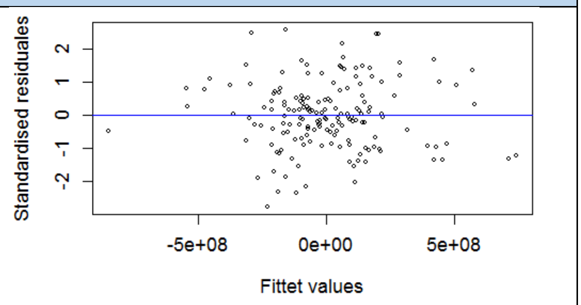
<table>
<tr><th colspan="2">RESET TESTS AND PLOTS OF STANDARDISED RESIDUALS VS. FITTED VALUES</th></tr>
<tr><th>EXCLUSION MODELS</th><th>INCLUSION MODELS</th></tr>
<tr><th colspan="2">Full period</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>RESET test
RESET-value: 0.87468
p-value: 0.418
Interpretation: linear shape is appropriate.</td><td>RESET test
RESET-value: 0.72117
p-value: 0.4869
Interpretation: linear shape is appropriate.</td></tr>
<tr><th colspan="2">First Half</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>RESET test
RESET-value: 2.4664
p-value: 0.08799
Interpretation: linear shape is appropriate.</td><td>RESET test
RESET-value: 2.8743
p-value: 0.05929
Interpretation: linear shape is appropriate.</td></tr>
<tr><th colspan="2">Second half</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>RESET test
RESET-value: 1.119
p-value: 0.3291
Interpretation: linear shape is appropriate.</td><td>RESET test
RESET-value: 0.70897
p-value: 0.4937
Interpretation: linear shape is appropriate.</td></tr>
</table>

Table 13: RESET tests and plots of standardised residuals vs. fitted values, models on trading volume.

5.1.2. The model on intraday price spread

The visual evaluation of the plot is a bit hard to draw conclusion from, but a RESET test for the model on intraday price spread suggests rejection of the null hypothesis that linearity is the appropriate shape. See the results in table 14 below.

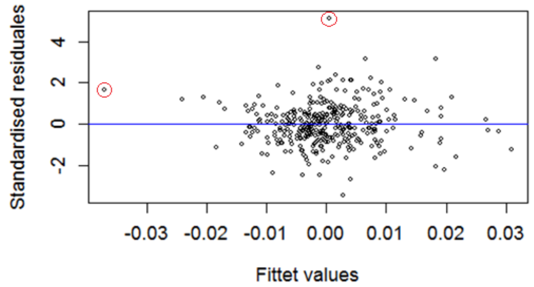
RESET TEST AND PLOT OF STANDARDISED RESIDUALS VS. FITTED VALUES	
RESET test RESET-value: 4.6025 p-value: 0.01067 Interpretation: linear shape is not appropriate.	

Table 14: RESET test and plot of standardised residuals vs. fitted values, model on intraday price spread.

In order to improve these results, I tried removing observation 261 and observation 5, individually and both at the same time. These observations respectively represent the two outliers fitted value = -0.037 and standardised residual = 5.13 (marked with red). This improved the plot noteworthy, but the p-value of the RESET test remained almost unchanged. The statistical significance of the model's coefficients remained unchanged as well.

It is interesting to note, that when Δvolume is not included in the model, the RESET test indicated that linear was the appropriate shape (p-value = 0.6763), and the coefficients for Δarticles and Δfinpress.tot were statistically significant (even with Newey-West standard errors). Although it fits the linear shape better, when not accounting for trading volume the model cannot be said to be complete.

I also tried to use logged variables instead of differenced variables, to see how this affected my results. This model gave a p-value of 0.1551, but an R^2 value of only 0.1441, and explains thus a lot less of the variation in volume, than the model based on differenced data (where R^2 = 0.4161). All variables are stationary as logged, except for the Δspread. However, the lagged model still does not produce statistically significant coefficients for the news related variables. This is why I stand by my discussion of the model in chapter 4 "Findings and discussion".

5.2. No perfect collinearity

To evaluate potential issued related collinearity I look at the correlation between each of the independent variables, as well as the VIF-values for each model.

5.2.1. The models on trading volume

There is little concern about collinearity in these models as correlation between the independent variables are (for the most part) low, and so are the VIF-values of the models. The correlation coefficients and VIF-values are presented in table 15 below.

CORRELATION MATRIX						VIF-VALUES		
Full period								
	Δarticles	Δfinpres	Δoseax	Δspread	Δtabloid		**EXCL**	**INCL**
Δarticles	1.00					Δarticles	1.05	1.14
Δfinpress.tot	-0.11	1.00				Δfinpres	1.04	1.70
Δoseax	0.10	0.05	1.00			Δoseax	1.12	1.14
Δspread	0.17	0.13	0.32	1.00		Δspread	1.16	1.16
Δtabloids.finpr	-0.17	-0.36	-0.17	-0.14	1.00	Δtabloid		1.74
First half								
	Δarticles	Δfinpres	Δoseax	Δspread	Δtabloid		**EXCL**	**INCL**
Δarticles	1.00					Δarticles	1.03	1.17
Δfinpress.tot	-0.14	1.00				Δfinpres	1.05	1.82
Δoseax	0.10	0.05	1.00			Δoseax	1.11	1.16
Δspread	0.07	0.14	0.29	1.00		Δspread	1.13	1.13
Δtabloids.finpr	0.19	-0.46	0.15	-0.11	1.00	Δtabloid		1.84
Second half								
	Δarticles	Δfinpres	Δoseax	Δspread	Δtabloid		**EXCL**	**INCL**
Δarticles	1.00					Δarticles	1.11	1.18
Δfinpress.tot	-0.08	1.00				Δfinpres	1.06	1.67
Δoseax	0.12	0.19	1.00			Δoseax	1.18	1.18
Δspread	0.30	0.12	0.36	1.00		Δspread	1.25	1.25
Δtabloids.finpr	-0.16	-0.60	-0.21	-0.19	1.00	Δtabloid		1.70

Table 15: Correlation matrices and VIF-values, models on trading volume. All correlation coefficients are statistically significant at a 5% level or stronger.

The correlation between Δfinpress.tot and Δtabloids.finpress in the second half-model is -0.6, which is quite high. However, the VIF-values remain very low, so this support the conclusion that collinearity is of little concern.

5.2.2. The model on intraday price spread

As with the volume models, the correlation between the independent variables, as well as the VIF-values of the model, are low, and thus the problem with collinearity is considered to be low. The correlation coefficients and VIF-values are presented in table 16 below.

CORRELATION MATRIX					VIF-VALUES	
	Δarticles	Δfinpress.tot	Δoseax.spread	Δvolume		VIF
Δarticles	1.00				Δarticles	1.216
Δfinpress.tot	-0.11	1.00			Δfinpress.tot	1.108
Δoseax.spread	0.05	-0.02	1.00		Δoseax.spread	1.021
Δvolume	0.37	0.23	0.13	1.00	Δvolume	1.291

Table 16: Correlation matrix and VIF-values, model on intraday price spread. All correlation coefficients are statistically significant at a 5% level or stronger.

5.3. Zero conditional mean

In order for the error term at time t to be uncorrelated with each explanatory variable in every time period, I must control for all variables that are related to the endogenous and any explanatory variable at the same time. Otherwise, it might be the case that my findings are due to some omitted variable causing both news and trading volume/intraday price spread to move together. Below follows a discussion on what I have done to avoid such issues.

5.3.1. The models on trading volume

I have tried to avoid dependencies that are due to day-of-the-week patterns in trading by using differences in weekly data in my analysis, I control for the established relationship between volume and price change by including the spread variable in the volume models (and vice versa), and I try to control for the general market movement by including the OSEAX variable.

Compared to the initial simple linear regression (see page 29), the models where we have included the control variables seem to show more reliable results. This is especially true for the model on intraday price spread, but also the models on trading volume show that including the chosen control variables avoided the error of accepting the initially apparent relationships. For example, the troublesome variable Δtabloids.finpress showed a highly significant coefficient in the simple regression, with a higher R^2 than that of Δfinpress.tot.

The fact that some of the results my models display are comparable to that of other research is also an indication that the quality of my control variables is in line with that of other research.

5.3.2. The model on intraday price spread

Same argumentation as for the models on trading volume above, in particular the comparison to the initial linear regressions, which showed significant coefficients for all three news related variables.

As noted alongside the linearity tests, including Δvolume as a control variable, made the coefficients for Δarticles and Δfinpress.tot go from statistically significant to insignificant. This shows that including the variable was a good decision to rule out a conclusion that otherwise would have been made based on spurious relationships.

An issue that goes for both models, is that of endogeneity in the news and revered causation. As Tetlock (2007, 1) puts it: “it is unclear whether the financial news media induces, amplifies, or simply reflects investors’ interpretations of stock market performance.”. Does news articles lead to change in trading volume and intraday price spread, or is it the other way around? Or a combination of the two?

Mitchell and Mulherin (1994) also acknowledge this endogeneity problem, and try to account for it by randomly surveying the content of five days in each year of their study. They find that such articles represent less than one percent of the headlines, and therefore conclude that this sort of endogeneity bias is not a serious concern in their sample. Birz and Lott Jr (2011) address the same problem in their study, and find evidence that such reverse effects are unlikely. For example, if S&P500 movements cause a stronger increase in news articles, than the other way around, then the number of news articles should be positively correlated to the lagged movement of the S&P500. Such a relation is not found in their sample. Like me, Lillo et al. (2014) deal with contemporaneous data, and thus do not attach causation to their findings. But they do perform a VAR analysis to check whether the news flux is what affects the trading, or the other way around. Their conclusion is that the flux of news of the previous day does affect trading activity, and evidence of the opposite is not observed.

As I mentioned in my model comments, I do believe that it is likely that there is a reciprocal relationship between news articles and volume, to a certain extent. But it is not very common that increased trading volume per se creates news articles, at least not to the same extent as sudden increases or decreases in returns does. And if my sample is similar to that of the researchers mentioned above, I can assume that the endogeneity problem is neglectable.

5.4. Homoscedasticity

To evaluate the homoscedasticity in the residuals I look at plots of the models’ standardised residuals, as well as performing a studentised Breusch-Pagan (BP) test on each model.

5.4.1. The models on trading volume

In table 17 below, I have presented the plots of the models’ standardised residuals, along with the results from the studentised Breusch-Pagan (BP) test on each model.

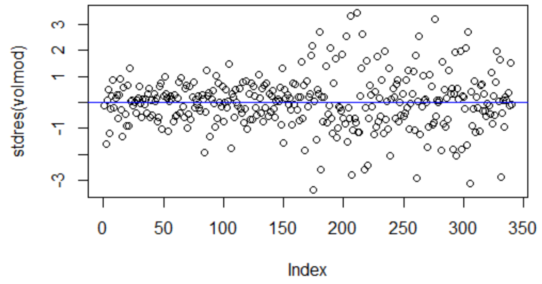
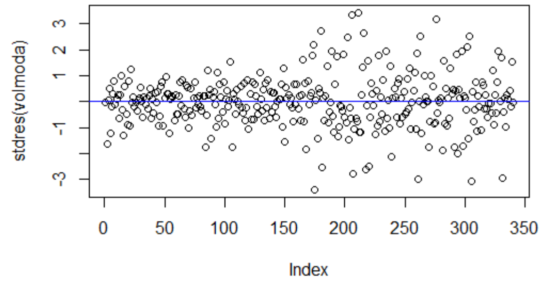
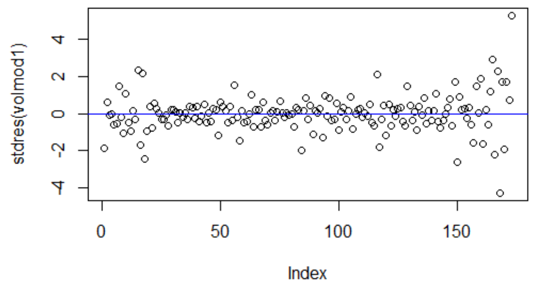
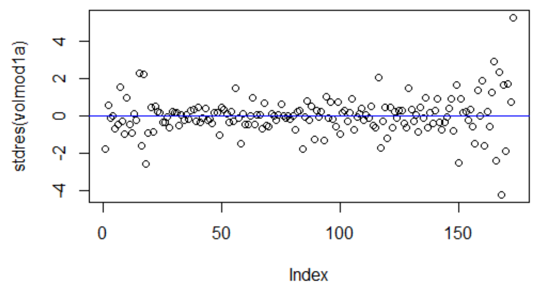
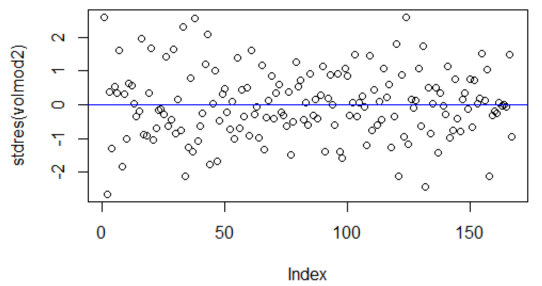
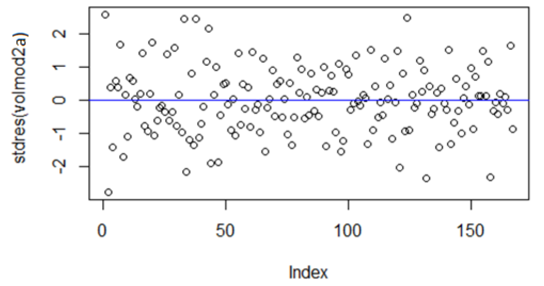
BREUSCH-PAGAN TESTS AND PLOTS OF STANDARDISED RESIDUALS	
EXCLUSION MODELS	**INCLUSION MODELS**
Full period	
	
Studentised Breusch-Pagan test: BP = 2.4517 Df = 4 p-value = 0.6533 Interpretation: homoscedasticity.	**Studentised Breusch-Pagan test:** BP = 4.0765 Df = 5 p-value = 0.5385 Interpretation: homoscedasticity.
First Half	
	
Studentised Breusch-Pagan test: BP = 3.872 Df = 4 p-value = 0.4237 Interpretation: homoscedasticity.	**Studentised Breusch-Pagan test:** BP = 3.9726 Df = 5 p-value = 0.5534 Interpretation: homoscedasticity.
Second half	
	
Studentised Breusch-Pagan test: BP = 3.3633 Df = 4 p-value = 0.499 Interpretation: homoscedasticity.	**Studentised Breusch-Pagan test:** BP = 3.6687 Df = 4 p-value = 0.499 Interpretation: homoscedasticity.

Table 17: Breusch-Pagan tests and plots of standardised residuals, models on trading volume.

We see from the plots that for the full period, the variance is quite different in the first and in the second half. This is, as already mentioned, precisely why I decided to split up this model. For the first half model, there seems to be a greater variance in the residuals in the beginning and in the end of the period, whereas the second half-model seems to be the most equal one. Nonetheless, they all have insignificant values on the BP tests, and we therefore accept the null hypotesis of homoscedasticity in all six models on trading volume. The increase in variance at the end of the second half, might indicate that the second "half" could have started slightly earlier. But, as argued before (see page 31), the point of division was decided based on the variance of the residuals of the full period-model.

5.4.2. The model on intraday price spread

In table 18 below, I have presented the plot of the model's standardised residuals, as well as the result from the studentised Breusch-Pagan (BP) test on the model.

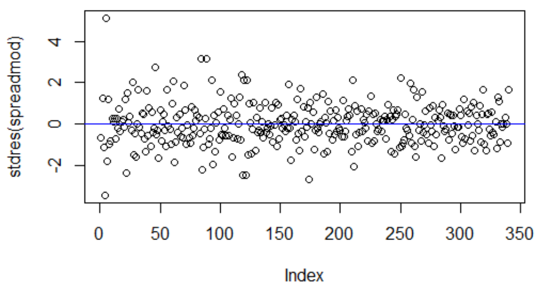
BREUSCH-PAGAN TEST AND PLOT OF STANDARDISED RESIDUALS	
	**Studentised Breusch-Pagan test:** BP = 2.8544 Df = 4 p-value = 0.5825 Interpretation: homoscedasticity.

Table 18: Breusch-Pagan test and plot of standardised residuals, model on intraday price spread.

We see that there is one positive and one negative outlier in the beginning of the time period, but overall it looks pretty equally distributed. The BP test gives insignificant values and we therefore accept the null hypotesis of homoscedasticity in model on intraday price spread.

5.5. No serial correlation

To determine serial correlation in the residuals, I evaluate ACF- and PACF plots of the standardised residuals for each model.

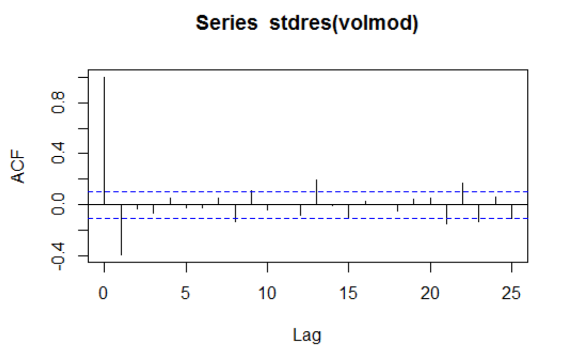
5.5.1. The models on trading volume

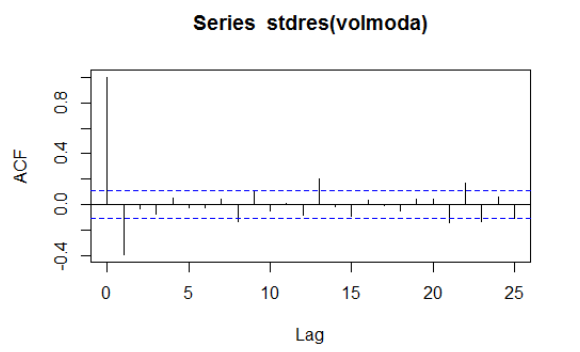
The ACF- and PACF plots of the standardised residuals for all six models on trading volume are found in table 19 below (over two pages).

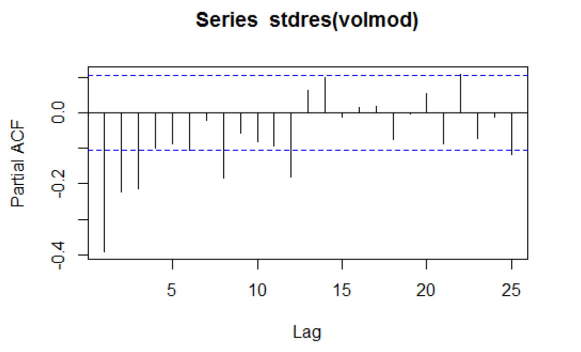
ACF- AND PACF PLOTS OF STANDARDISED RESIDUALS

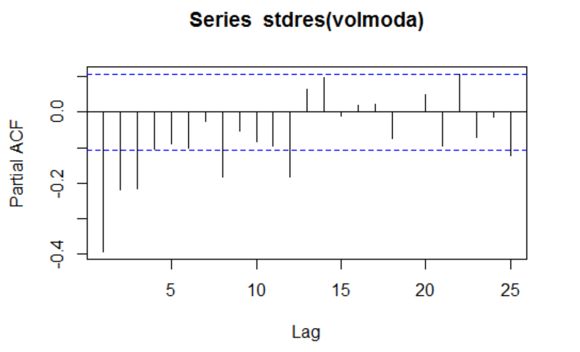
EXCLUSION MODELS	INCLUSION MODELS

Full period

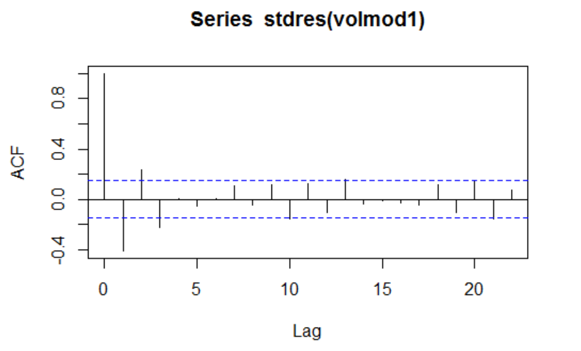


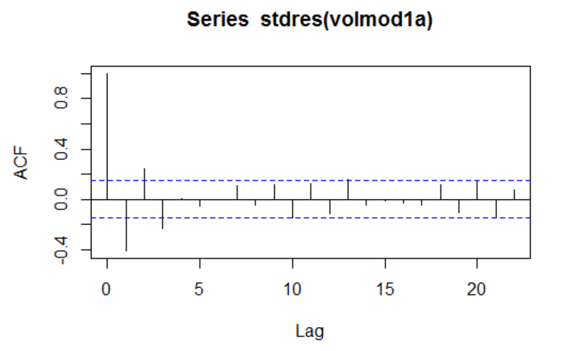


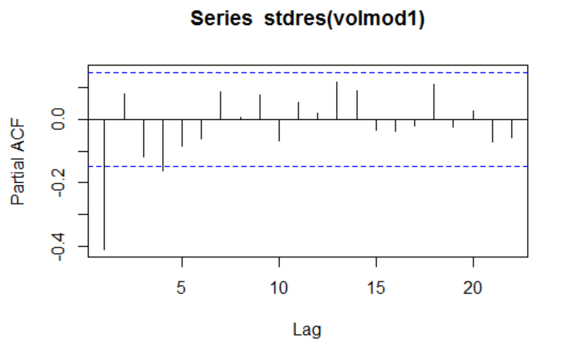


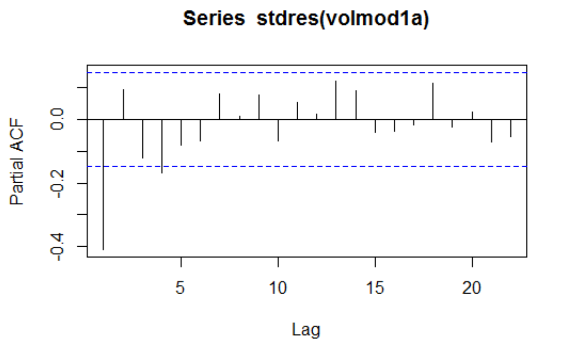


First Half









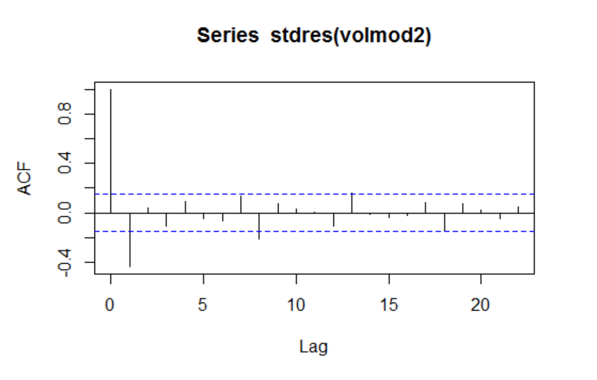
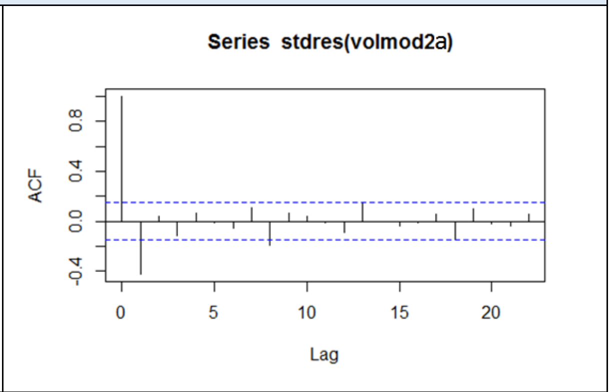
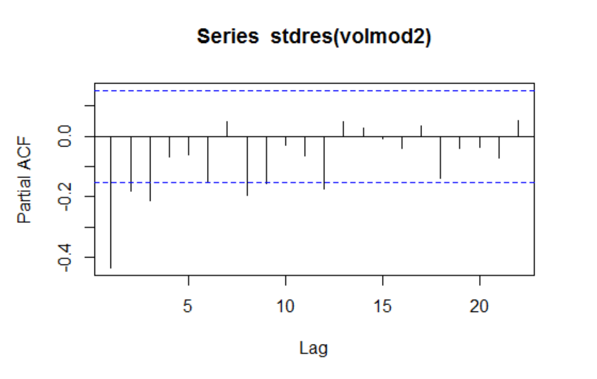
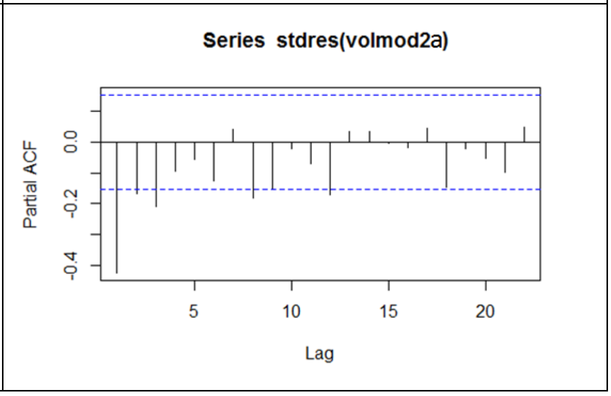
Second half	
	
	

Table 19: ACF- and PACF plots of standardised residuals, models on trading volume.

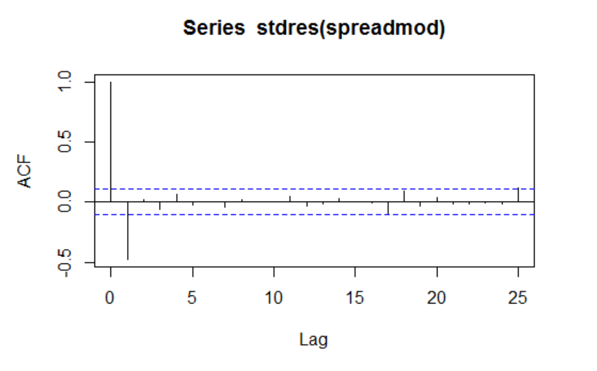
The plots show clear indications of autocorrelation problematics in all six models on trading volume. Both the exclusion and the inclusion models look similar in regard to these problems. Regressions of ε_t against ε_{t-1} give highly significant coefficients and R^2 values of 0.1541, 0.2045 and 0.1881 for the three exclusion models respectively, further confirming the autocorrelation. Due to the similarities between the models within each time period, I expect identical results for the inclusion models. Ljung-Box tests and Durbin-Watson tests also strongly support rejection of the null hypothesis of zero autocorrelation for all models. For the second half, autocorrelation is only significant in the first lag according to the Durbin-Watson test, but the Ljung-Box test rejects zero autocorrelation overall for the 25 first lags.

Several researchers have reported on difficulties related to the strong autocorrelation patters that news data exhibits, and therefore I might not be able to avoid this problem completely in my kind of study (see for example Kalev et al. (2004)). To account for this problem, I therefore consequently report Newey-West standard errors, which are robust to autocorrelation and heteroscedasticity. This decision blew up all coefficients' standard errors, and reduced their statistical significance. But all coefficients which were statistically significant at a 5% level or lower, remained statistically significant after this correction, except for Δfinpress.tot in the second half inclusion model which dropped from a p-value of 0.0477 to a p-value of 0.09484. Note that *all* references to standard errors in this paper refers to Newey-West standars errors, except on page 29 where it explicitly says *OLS standard errors*.

5.5.2. The model on intraday price spread

To determine serial correlation in the residuals, I evaluate ACF- and PACF plots of the standardised residuals for the model on intraday price spread. These are found in table 20 below.

ACF- AND PACF PLOTS OF STANDARDISED RESIDUALS



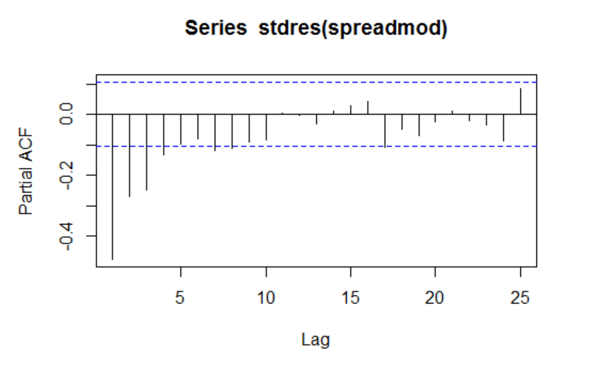


Table 20: ACF- and PACF-plots of standardised residuals, model on intraday price spread.

Visually, the autocorrelation problem appears to be smaller in the intraday price spread model compared to the trading volume models. However, it is still present. A regression of ε_t against ε_{t-1} gives a highly significant coefficient and an R^2 values of 0.2297. A Durbin-Watson test on the standardised residuals show a statistical significant autocorrelation only on lag 1. A Ljung-Box test, on the other hand, rejects the null hypothesis of zero autocorrelation both on lag 1 and overall on the first 25 lags. As with the models on trading volume, I therefore report Newey-West standard errors, which are robust to autocorrelation and heteroscedasticity.

5.6. Normality

To shed light on the normality of the residuals, I evaluate histograms of the standardised residuals and calculate the skewness and excess kurtosis for each model's residuals.

5.6.1. The models on trading volume

Table 21 below presents histograms of the standardised residuals with a normal distribution curve, along with skewness and excess kurtosis measures.

The skewness and kurtosis levels are slightly high for the first half models, but overall they look satisfactory for the models on trading volume.

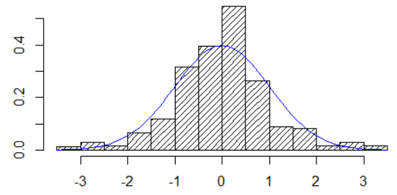
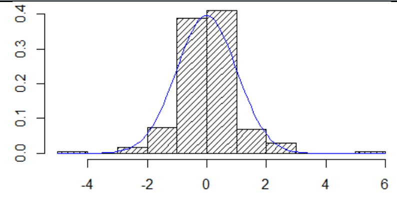
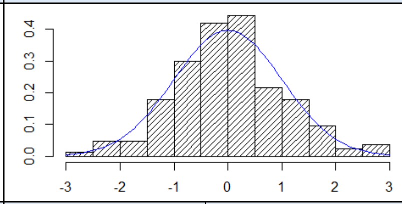
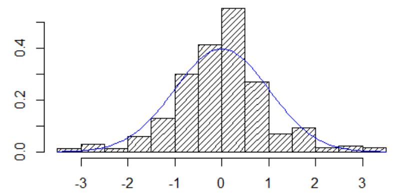
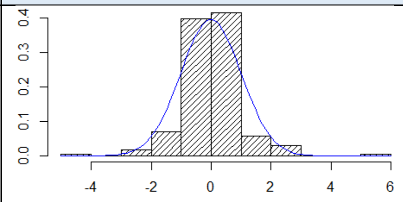
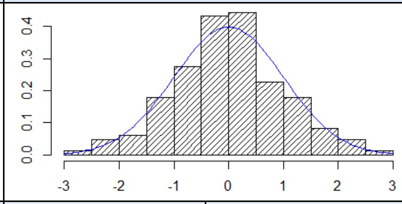
NORMALITY IN THE STANDARDISED RESIDUALS					
EXLUSION-MODELS					
Full period		First half		Second half	
					
Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
0.1055044	1.569006	0.4727054	5.575476	0.1402317	0.0545023
INCLUSION-MODELS					
Full period		First half		Second half	
					
Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
0.08160669	1.550014	0.4949291	5.639825	0.06861225	0.07467866

Table 21: Normality in the standardised residuals, models on trading volume.
Histograms: Y-axis: density, X-axis: standardised residuals. Curve: normal distribution. Kurtosis value is excess *kurtosis.*

5.6.2. The model on intraday price spread

See table 22 below for histograms of the standardised residuals with a normal distribution curve, along with skewness and excess kurtosis measures, for the model on intraday price spread.

The normality measures look satisfactory for the model on intraday price spread.

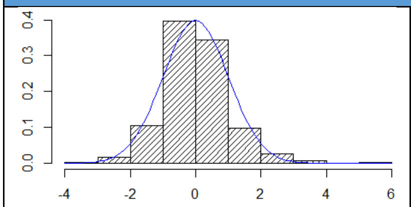
NORMALITY IN THE STANDARDISED RESIDUALS		
	Skewness	Kurtosis
	0.4978474	2.221313

Table 22: Normality in the standardised residuals, model on intraday price spread.
Histogram: Y-axis: density, X-axis: standardised residuals. Curve: normal distribution. Kurtosis value is excess *kurtosis.*

5.7. Conclusion of the regression diagnostics, and other potential issues

I have in this section shown that the assumptions regarding the OLS on time series data hold to a fair degree, except for the problems concerning autocorrelation in the residuals, which is accounted for by consequently reporting Newey-West standard errors.

Another potential issue with my data is the lack of detail when looking at weekly data. I already discussed why I went for weekly data on page 24 and 25, and it had to do with simplicity, as well as being an atempt to account for other problems discovered with detailed data. A related issue to news being released on non trading days, is that they also are released outside of trading hours, and this I had no way to control for. Even the most detailed function in Atekst does not sort articles per minute, so I cannot identified which articles were released within trading hours and which were not. This means that even with a daily alignment, I would regress news from after trading hours with trading volume from earlier that same day.

I also already discussed the issue with the search word Norwegian not being unique (page 19). This was solved by using *Norwegian + fly**, but there is no doubt that also this search might be imprecise. There is also a big chance that the search includes articles not about Norwegian, but about other airlines or companies where Norwegian is only mentioned for the purpose of comparison. If these articles are irrelevant or not is not a straightforward question. However, as "articles" is meant to be a measure of the general media intensity for the company, my opinion is that they are in fact still relevant. Also, if there are "relevant" articles in the financial press revealing fundamental financial information about SAS, and it uses Norwegian as a comparison, it is arguable that these articles could affect the trading volume of Norwegian as well.

6. Conclusion

I have in this paper looked at the relationship between media intensity and the Norwegian share's trading volume and intraday price spread. I have found evidence that the investors in Norwegian are attention-driven, in that trading volume increases when the company is in the news. This is consistent with other research, if we consider the investor profile of Norwegian, which has a relatively high number of non-institutional investors. Nonetheless, there is also evidence that financial news stories have higher importance than tabloid news stories, which is consistent with other research on rational vs. sentiment-based investors. But the novelty of my research is looking at ratios between financial- and non-financial press articles, and there are some indications in my findings that the investors in Norwegian have difficulties screening out the relevant from the irrelevant news when the financial press is drowning in a high general media intensity. Finally, I notice that the investors were more sentiment- and attention-driven when the company was smaller. This is also in line with other research that bigger companies tend to be more transparent and have a higher degree of rational investors.

When it comes to intraday price spread, I cannot find evidence of statistically significant relationships with the news variables, but a strong relationship with the volume variable. I infer from this that the changes in the intraday price spread is due to spill-over from the effects found in trading volume. I.e. trading volume is affected by the media intensity, and this, in turn, leads to changes in the intraday price spread. It might also be that my news variable is too noisy and not enough detailed to pick up on the connections between news and intraday price spread.

My conclusions are robust to the general assumptions for OLS on time series data, and the problems with autocorrelation is accounted for by reporting Newey-West standard errors and coefficient significance based on these assumptions. My results must, however, still be viewed with caution, due to the lack of details in the news variable, and the possibility of reverse causation and reciprocal relationship between the dependent- and independent variables.

Based on the findings in this paper, it looks like it will be advantageous for future research to include a wider news variable than strictly financial news wires. Further studies on financial news sources in relation to non-financial ones could give a deeper understanding of investor behaviour and behavioural finance. Future research on the same topic could include a selection of companies, potentially divided into categories. If the capacity permits, it could also look at daily data rather than weekly, and move Saturday's and Sunday's news stories to the following Monday. Research like this could give us a deeper insight into the press' full role in information conveyance, and whether this varies between industries.

Sources

Alanyali, Merve, Helen Susannah Moat, and Tobias Preis. 2013. "Quantifying the Relationship Between Financial News and the Stock Market." *Scientific Reports* 3:3578. doi: 10.1038/srep03578

http://www.nature.com/articles/srep03578#supplementary-information.

Barber, Brad M., and Terrance Odean. 2008. "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." *The Review of Financial Studies* 21 (2):785-818.

Bernard, Victor L., and Jacob K. Thomas. 1990. "Evidence that stock prices do not fully reflect the implications of current earnings for future earnings." *Journal of Accounting and Economics* 13 (4):305-340. doi: http://dx.doi.org/10.1016/0165-4101(90)90008-R.

Berry, Thomas D., and Keith M. Howe. 1994. "Public Information Arrival." *Journal of Finance* 49 (4):1331-1346.

Birz, Gene, and John R. Lott Jr. 2011. "The effect of macroeconomic news on stock returns: New evidence from newspaper coverage." *Journal of Banking & Finance* 35 (11):2791-2800. doi: http://dx.doi.org/10.1016/j.jbankfin.2011.03.006.

Bloomberg. Ownership Summary for Norwegian Air Shuttle ASA. Bloomberg LP, accessed: 12.02.2017.

Chan, Wesley S. 2003. "Stock price reaction to news and no-news: drift and reversal after headlines." *Journal of Financial Economics* 70 (2):223-260. doi: http://dx.doi.org/10.1016/S0304-405X(03)00146-6.

Choi, Hyunyoung, and H. A. L. Varian. 2012. "Predicting the Present with Google Trends." *Economic Record* 88:2-9. doi: 10.1111/j.1475-4932.2012.00809.x.

Clark, Peter K. 1973. "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices." *Econometrica* 41 (1):135-155. doi: 10.2307/1913889.

Cornell, Bradford. 2013. "What Moves Stock Prices: Another Look." *Journal of Portfolio Management* 39 (3):32-38.

Da, Z. H. I., Joseph Engelberg, and Pengjie Gao. 2011. "In Search of Attention." *The Journal of Finance* 66 (5):1461-1499. doi: 10.1111/j.1540-6261.2011.01679.x.

Deephouse, David L. 2000. "Media Reputation as a Strategic Resource: An Integration of Mass Communication and Resource-Based Theories." *Journal of Management* 26 (6):1091-1112. doi: 10.1177/014920630002600602.

Ederington, Louis H., and Jae Ha Lee. 1993. "How Markets Process Information: News Releases and Volatility." *Journal of Finance* 48 (4):1161-1191.

Engelberg, Joseph E., Adam V. Reed, and Matthew C. Ringgenberg. 2012. "How are shorts informed?: Short sellers, news, and information processing." *Journal of Financial Economics* 105 (2):260-278. doi: http://dx.doi.org/10.1016/j.jfineco.2012.03.001.

Engle, Robert F., and Victor K. Ng. 1993. "Measuring and Testing the Impact of News on Volatility." *Journal of Finance* 48 (5):1749-1778.

Epps, Thomas W., and Mary Lee Epps. 1976. "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis." *Econometrica* 44 (2):305-321. doi: 10.2307/1912726.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2009. "Detecting influenza epidemics using search engine query data." *Nature* 457 (7232):1012-1014. doi:

http://www.nature.com/nature/journal/v457/n7232/suppinfo/nature07634_S1.html.
Glosten, Lawrence, Ravi Jagannathan, and David E. Runkle. 1993. "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." *Journal of Finance* 48 (5):1779-1801.
Groß-Klußmann, Axel, and Nikolaus Hautsch. 2011. "When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions." *Journal of Empirical Finance* 18 (2):321-340. doi: http://dx.doi.org/10.1016/j.jempfin.2010.11.009.
Harris, Milton, and Artur Raviv. 1993. "Differences of Opinion Make a Horse Race." *The Review of Financial Studies* 6 (3):473-506.
Ikenberry, David L., and Sundaresh Ramnath. 2002. "Underreaction to Self-Selected News Events: The Case of Stock Splits." *The Review of Financial Studies* 15 (2):489-526.
Joulin, A., A. Lefevre, D. Grunberg, and J.-P. Bouchaud. 2008. "Stock price jumps: News and volume play a minor role." *Wilmott Magazine* 36:1-7.
Kahneman, Daniel. 1973. "Attention and effort." *Prentice-Hall series in experimental psychology*:x, 246 p.
Kahneman, Daniel, and Amos Tversky. 1984. "Choices, values, and frames." *American Psychologist* 39 (4):341-350. doi: 10.1037/0003-066X.39.4.341.
Kalev, Petko S., and Huu Nhan Duong. 2011. "Firm-specific news arrival and the volatility of intraday stock index and futures returns." In *The Handbook of News Analytics in Finance*, edited by Gautam Mitra and Leela Mitra, 271-288. John Wiley & Sons, Ltd.
Kalev, Petko S., Wai-Man Liu, Peter K. Pham, and Elvis Jarnecic. 2004. "Public information arrival and volatility of intraday stock returns." *Journal of Banking & Finance* 28 (6):1441-1467. doi: http://dx.doi.org/10.1016/S0378-4266(03)00126-2.
Karpoff, Jonathan M. 1986. "A Theory of Trading Volume." *Journal of Finance* 41 (5):1069-1087.
Katsyri, J., T. Kinnunen, K. Kusumoto, P. Oittinen, and N. Ravaja. 2016. "Negativity Bias in Media Multitasking: The Effects of Negative Social Media Messages on Attention to Television News Broadcasts." *PLoS One* 11 (5):e0153712. doi: 10.1371/journal.pone.0153712.
Kim, Oliver, and Robert E. Verrecchia. 1991. "Trading Volume and Price Reactions to Public Announcements." *Journal of Accounting Research* 29 (2):302-321. doi: 10.2307/2491051.
Lillo, Fabrizio, Salvatore Miccichè, Michele Tumminello, Jyrki Piilo, and Rosario N. Mantegna. 2014. "How news affects the trading behaviour of different categories of investors in a financial market." *Quantitative Finance* 15 (2):213-229. doi: 10.1080/14697688.2014.931593.
McCombs, Maxwell E., and Donald L. Shaw. 1972. "The Agenda-Setting Function of Mass Media." *The Public Opinion Quarterly* 36 (2):176-187.
Michaely, Roni, Richard H. Thaler, and Kent L. Womack. 1995. "Price Reactions to Dividend Initiations and Omissions: Overreaction or Drift?" *Journal of Finance* 50 (2):573-608.
Michaely, Roni, and Kent L. Womack. 1999. "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations." *The Review of Financial Studies* 12 (4):653-686.
Mitchell, Mark L., and J. Harold Mulherin. 1994. "The Impact of Public Information on the Stock Market." *Journal of Finance* 49 (3):923-950.

Newman, Nic, Richard Fletcher, David A. L. Levy, and Rasmus Kleis Nielsen. 2016. Reuters Institute Digital News Report 2016. Reuters Institute for the Study of Journalism.

Ranco, Gabriele, Darko Aleksovski, Guido Caldarelli, Miha Grčar, and Igor Mozetič. 2015. "The Effects of Twitter Sentiment on Stock Price Returns." *PLoS ONE* 10 (9):1-21. doi: 10.1371/journal.pone.0138441.

Rozin, Paul, and Edward B. Royzman. 2001. "Negativity Bias, Negativity Dominance, and Contagion." *Personality & Social Psychology Review (Lawrence Erlbaum Associates)* 5 (4):296-320.

Sidorov, Sergei P., Paresh Date, and Vladimir Balash. 2013. "GARCH Type Volatility Models Augmented with News Intensity Data." In *Chaos, complexity and leadership 2012*, edited by Santo Banerjee and Sefika Sule Erçetin, 199-208. New York: Springer.

Sidorov, Sergei P., Andrey Revutskiy, Alexey Faizliev, Eugene Korobov, and Vladimir Balash. 2014. "GARCH Model with Jumps: Testing the Impact of News Intensity on Stock Volatility." In *Proceedings of the World Congress on Engineering 2014*. WCE 2014, July 2 - 4, 2014, London, U.K.

Souza, Thársis T. P., Olga Kolchyna, Philip C. Treleaven, and Tomaso Aste. 2015. "Twitter Sentiment Analysis Applied to Finance: A Case Study in the Retail Industry."

Sprenger, Timm O., Philipp G. Sandner, Andranik Tumasjan, and Isabell M. Welpe. 2014. "News or Noise? Using Twitter to Identify and Understand Company-specific News Flow." *Journal of Business Finance & Accounting* 41 (7-8):791-830. doi: 10.1111/jbfa.12086.

Tauchen, George E., and Mark Pitts. 1983. "The Price Variability-Volume Relationship on Speculative Markets." *Econometrica* 51 (2):485-505. doi: 10.2307/1912002.

Taylor, Shelley E. 1991. "Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis." *Psychol Bull* 110 (1):67-85.

Tetlock, Paul C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance* 62 (3):1139-1168. doi: 10.1111/j.1540-6261.2007.01232.x.

Tetlock, Paul C. 2010. "Does Public Financial News Resolve Asymmetric Information?" *The Review of Financial Studies* 23 (9):3520-3557.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy. 2008. "More Than Words: Quantifying Language to Measure Firms' Fundamentals." *The Journal of Finance* 63 (3):1437-1467. doi: 10.1111/j.1540-6261.2008.01362.x.

Thompson, Robert B., Chris Olsen, and J. Richard Dietrich. 1987. "Attributes of News about Firms: An Analysis of Firm-Specific News Reported in the Wall Street Journal Index." *Journal of Accounting Research* 25 (2):245-274. doi: 10.2307/2491017.

Vega, Clara. 2006. "Stock price reaction to public and private information." *Journal of Financial Economics* 82 (1):103-133. doi: 10.1016/j.jfineco.2005.07.011.

Wooldridge, Jeffrey M. 2013. *Introductory econometrics : a modern approach*. 5th ed. Mason, OH: South-Western Cengage Learning.

Yechiam, Eldad, and Guy Hochman. 2013. "Losses as modulators of attention: review and analysis of the unique effects of losses over gains." *Psychological bulletin* 139 (2):497-518. doi: 10.1037/a0029383.

Yu, Xiang, Gautam Mitra, and Keming Yu. 2013. "Impact of news on asset behaviour: Return, Volatility and Liquidity in an Intra-Day Setting." Brunel University.

Zyglidopoulos, Stelios C., Andreas P. Georgiadis, Craig E. Carroll, and Donald S. Siegel. 2012. "Does media attention drive corporate social responsibility?" *Journal of Business Research* 65 (11):1622-1627. doi: 10.1016/j.jbusres.2011.10.021.

Appendix A

KPSS TEST OF THE VARIABLES AT LEVEL (R command: kpss.test)			
Variable	**KPSS Level**	**p-value**	**Interpretation**
volume	5.097	< 0.01	Unit root
spread	0.74008	< 0.01	Unit root
articles	0.412	0.07198	Stationary
finpress.tot	2.2907	< 0.01	Unit root
tabloids.tot	0.43209	0.06332	Stationary
tabloids.finpress	0.99337	< 0.01	Unit root
nonfin.fin	2.6891	< 0.01	Unit root
oseax	2.8259	< 0.01	Unit root
oseax.spread	0.76279	< 0.01	Unit root
Critical value for KPSS level = 0.463 (significance level = 5%) (Kwiatkowski et al. 1992).			

ADF TEST OF THE VARIABLES AT LEVEL (R command: adfTest)			
Variable	**Dickey-Fuller Stat.**	**p-value**	**Interpretation**
volume	-2.9831	< 0.01	Stationary
spread	-2.0652	0.03972	Unit root
articles	-5.5882	< 0.01	Stationary
finpress.tot	-2.7339	< 0.01	Unit root
tabloids.tot	-3.7723	< 0.01	Stationary
tabloids.finpress	-5.1078	< 0.01	Stationary
nonfin.fin	-4.3598	< 0.01	Stationary
oseax	-1.9595	0.04915	Unit root
osea.spread	-2.0565	0.04049	Unit root
Critical value for DF stat. = -2.88 (significance level = 5%, N = 250) (Fuller 1976).			

DIFFERENCED DATA				
	KPSS		**ADF**	
Variable	**KPSS Level**	**p-value**	**Dickey-Fuller Stat.**	**p-value**
volume	0.035469	> 0.1	-19.5922	< 0.01
spread	0.021367	> 0.1	-19.4524	< 0.01
articles	0.010964	> 0.1	-19.6404	< 0.01
finpress.tot	0.016961	> 0.1	-23.9658	< 0.01
tabloids.finpress	0.009656	> 0.1	-22.1000	< 0.01
oseax	0.028436	> 0.1	-19.7094	< 0.01
oseax.spread	0.018421	> 0.1	-17.4552	< 0.01

Fuller, W.A. 1976. *Introduction to statistical time series*: Wiley.

Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin. 1992. "Testing the null hypothesis of stationarity against the alternative of a unit root." *Journal of Econometrics* 54 (1):159-178. doi: http://dx.doi.org/10.1016/0304-4076(92)90104-Y.

Appendix B

SUMMARY: EXCLUSION MODELS ON TRADING VOLUME – STANARDISED COEFFICIENTS				
Full period: t_1-t_{340}				
	Coefficient	**p-value**	**F statistics**	53.96
(intercept)	$-1.4797e^{-16}$	1	**Multiple R^2**	0.3918
	(0.01617)		**Adjusted R^2**	0.3846
Δarticles	0.31650	$4.00e^{-08}$ ***		
	(0.05630)			
Δfinpress.tot	0.20909	0.00093 ***		
	(0.06257)			
Δoseax	0.20472	0.00840 **		
	(0.07722)			
Δspread	0.32658	$3.36e^{-06}$ ***		
	(0.06908)			
First half: t_1-t_{173}				
	Coefficient	**p-value**	**F statistics**	29.93
(intercept)	-6.6891E-17	1	**Multiple R^2**	0.4161
	(0.03522)		**Adjusted R^2**	0.4022
Δarticles	0.39835	0.00007 ***		
	(0.09785)			
Δfinpress.tot	0.13750	0.02528 *		
	(0.06092)			
Δoseax	0.05296	0.34292		
	(0.05568)			
Δspread	0.43876	0.00009 ***		
	(0.10920)			
Second half: t_{174}-t_{340}				
	Coefficient	**p-value**	**F statistics**	51.28
(intercept)	-4.4942E-17	1	**Multiple R^2**	0.5587
	(0.02690)		**Adjusted R^2**	0.5479
Δarticles	0.25477	0.00782 **		
	0.09460			
Δfinpress.tot	0.22703	0.00079 ***		
	(0.06634)			
Δoseax	0.29926	$2.20e^{-06}$ ***		
	(0.06094)			
Δspread	0.39206	$4.36e^{-06}$ ***		
	(0.08245)			

SUMMARY: INCLUSION MODELS ON TRADING VOLUME – STANARDISED COEFFICIENTS

Full period: t_1-t_{340}

	Coefficient	p-value
(intercept)	-1.4538E-16 (0.01616)	1
Δarticles	0.30123 (0.05860)	$4.68e^{-07}$ ***
Δfinpress.tot	0.16812 (0.07507)	0.02578 *
Δoseax	0.19667 (0.08008)	0.01456 *
Δspread	0.32775 (0.06955)	$3.60e^{-06}$ ***
Δtabloids.finpress	-0.06638 (0.07984)	0.40633

F statistics	43.50
Multiple R^2	0.3944
Adjusted R^2	0.3853

First half: t_1-t_{173}

	Coefficient	p-value
(intercept)	-7.0408E-17 (0.03534)	1
Δarticles	0.37588 (0.10577)	0.00049 ***
Δfinpress.tot	0.08387 (0.07494)	0.26467
Δoseax	0.03943 (0.05329)	0.46042
Δspread	0.44217 (0.10937)	0.00008 ***
Δtabloids.finpress	-0.08285 (0.07210)	0.25213

F statistics	24.17
Multiple R^2	0.4198
Adjusted R^2	0.4024

Second half: t_{174}-t_{340}

	Coefficient	p-value
(intercept)	-4.5208E-17 (0.03031)	1
Δarticles	0.22458 (0.08259)	0.00726 *
Δfinpress.tot	0.13280 (0.07607)	0.08276 .
Δoseax	0.29020 (0.05877)	$1.96e^{-06}$ ***
Δspread	0.38607 (0.07778)	$1.75e^{-06}$ ***
Δtabloids.finpress	-0.15668 (0.06774)	0.02199 *

F statistics	43.24
Multiple R^2	0.5732
Adjusted R^2	0.5599

Appendix C

SUMMARY: MODEL ON INTRADAY PRICE SPREAD – STANARDISED COEFFICIENTS		
	Coefficient	**p-value**
(intercept)	1.7108E-17 (0.02156)	1
Δarticles	0.00846 (0.05630)	0.8806
Δfinpress.tot	0.04362 (0.06430)	0.4980
Δoseax.spread	0.34147 (0.04884)	$1.48e^{-11}$ ***
Δvolume	0.41533 (0.06090)	$4.25e^{-11}$ ***

F statistics	42.93
Multiple R^2	0.3389
Adjusted R^2	0.3310

Appendix D

DESCRITPIVE STATISTICS: SUMMARY OF THE VARIABLES					
Variable	**N**	**Minimum value**	**Mean value**	**Maximum value**	**Std. deviation**
Full period					
volume	340	13 405 708	325 195 503	1 285 962 285	287 383 359
articles	340	61	294.3	2 726	258.0
finpress.tot	340	0.0089	0.0693	0.1842	0.0299
oseax	340	3 079 337 643	20 412 858 504	45 897 092 095	7 864 145 749
spread	340	0.0003	0.0354	0.0818	0.0125
tabloids.finpress	340	0.1111	0.9761	5.0000	0.7389
First half					
volume	173	13 405 708	116 001 225	945 122 371	135 119 668
articles	173	63	265.7	1 756	234.6
finpress.tot	173	0.0089	0.0607	0.1842	0.0316
oseax	173	3 079 337 643	23 163 931 445	45 897 092 095	9 249 491 587
spread	173	0.000328089	0.038037964	0.081799486	0.0147
tabloids.finpress	173	0.1111	1.0953	5.0000	0.8570
Second half					
volume	167	77 124 536	541 753 799	1 285 962 285	247 168 353
articles	167	61	323.9	2 726	300.8
finpress.tot	167	0.0207	0.0781	0.1508	0.0269
oseax	167	4 773 072 769	17 540 977 746	31 747 500 855	4 753 639 324
spread	167	0.0139	0.0325	0.0624	0.0097
tabloids.finpress	167	0.1591	0.8352	5.0000	0.6245