UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

High-Frequency Market Reactions to Unscheduled Stock-Specific News

An Empirical Analysis of the Intraday Market Dynamics of the Stockholm Stock Exchange

Olle Ekesryd & Tom Carlson

Supervisor: Magnus Andersson
Master's thesis in Finance, 30 hec
Spring 2022
Graduate School, School of Business, Economics and Law, University of Gothenburg, Sweden

Abstract

This study examines the effect of unscheduled stock-specific news on stock characteristics of the Swedish stock market and evaluates the opportunity of constructing a news trading strategy. It especially focuses on volume and volatility reactions between sixty minutes prior to and after the news releases. There are significantly large increases before the news releases, especially prominent for small cap stocks, indicating but not proving the presence of private information among informed investors being exploited. It also validates the hypothesis of the stock market not being perfectly efficient in relation to the Efficient Market Hypothesis. No significant correlations between pre and post-news returns were found, complicating the process of constructing a profitable trading strategy. However, with further improvements building upon the study, it may be possible in the future.

Keywords: unscheduled news, intraday, efficient market hypothesis, high-frequency trading, sentiment analysis.

Acknowledgements

We would like to thank our supervisor Dr. Magnus Andersson for supporting and challenging us week after week. The constructive feedback given has made this experience both educational and memorable, marking the end of a long academic journey.

Contents

1 Introduction

The introduction presents a background to news trading, motivates the importance of this study, as well as includes the purpose and main results.

1.1 Background

It is well-known that the Efficient Market Hypothesis (EMH) states that markets are efficient from an informative perspective, meaning that all stock prices already reflect the current available information. The prices of financial instruments are also said to be unbiased and already contain wisdom about future forecasts made by professional investors. Making excess profits should therefore not be possible utilizing known information as market prices change without delays. However, many have argued that prices are not consistent with company fundamentals, which requires a rethinking (Cutler, Poterba, and Summers 1989; Tetlock, Saar-Tsechansky, and Macskassy 2007; Zang and Skiena 2010).

News is one of the key factors that affects financial assets in terms of price, volume, and volatility (Quantra, 2021). A news trader is someone who makes investment decisions based on announcements related to economic reports, breaking news, or other reported events (Chen, 2021a). These announcements could be scheduled interest rate announcements or quarterly earnings reports from companies as well as unexpected black swan events (Picardo, 2021). The strategy is to take advantage of the market sentiment leading up to the announcement or to trade the market response afterwards, utilizing short-term volatility with alternatives to long or short the asset (Chen, 2021a).

In algorithmic news-based trading, algorithms i.e., pre-programmed, and automated trading instructions that account for variables such as price, timing, and volume are designed to interpret these news articles (Quantra, 2021; Chen, 2022). News can be scanned based on keywords that get quantified and interpreted as market sentiment, either being positive, negative, or neutral. The algorithm also considers the source and relevance of the news article to make trading decisions. However, there are limitations with this strategy, the data may not be processed if the new market-moving news would lie outside the pre-defined domain of search, leading to erroneous signals and trades (Quantra, 2021).

There seems to be a strong correlation between the number of news articles and the stock trading volume. Zhang and Skiena (2010) observed a correlation of 0.4 between logged normalized article counts and logged stock trading volume. There was also a stronger correlation between large cap compared to small cap stocks. When observing market sectors, biotech and military stocks had a much stronger correlation (>0.7) compared to electronic and IT stocks (<0.2).

1.2 Importance

Financial professionals and market participants need to read a large amount of information and stream news daily. The proportions surpass human capability to process and utilize such information in real-time, which motivates the use of deep learning approaches. We want to contribute to the already existing literature by evaluating the opportunity of using news analytic tools in the Swedish stock market, tied to unscheduled Swedish news. The majority of previous and similar studies have focused on scheduled news linked to either macroeconomics or quarterly reports, making this study unique. Other important aspects are that this study is one of few using solely news articles within the covid-19 pandemic, connected to a highly volatile market period (Harron and Rizvi, 2020), and also includes less liquid stocks. The results could be useful for constructing new trading strategies, utilized by both private and professional investors as portfolio managers.

1.3 Purpose

This paper aims to investigate how intraday and unscheduled news tied to Swedish companies affect the stock prices and if there are ways to predict them. In specific, the questions that will be answered are:

- (i) How do core stock characteristics such as price, volume, and volatility change around unscheduled news releases?
- (ii) Is there a significant correlation between pre and post-news returns that can be used for constructing a trading strategy?

1.4 Results

(i) Large fluctuations in various stock variables such as price, standard deviation, volume, and absolute return were observed around the news releases. About all of the studied parameters (except for large cap stocks) rapidly increased approximately 15 minutes before the news release and peaked at the publication time to then adjust 15 minutes subsequently. The rapid increase in volume and volatility before unscheduled news are indicating but does not prove that some investors utilize private information not available to others when investing.

(ii) In regards to the predictability of stock prices after news releases based on the pre and post-news return correlations, there were no observations indicating significant probabilities of it being possible, consistent with the Efficient Market Hypothesis and Random Walk Theory. This complicates the process of developing a profitable news trading strategy. However, it may be possible when using news analytic tools based on sentiment analysis, capable of sorting out irrelevant articles.

2 Literature Review

The literature review covers relevant articles related to this study, tied to how unscheduled news affect stock markets and how the knowledge could be utilized to develop trading strategies based on sentiment analysis.

2.1 Unscheduled news and market reactions

Groß-Klußmann and Hautsch (2011) analyzed how high-frequency intraday returns, volatility, volume, and liquidity movements were affected by unscheduled news arrivals. They used an algorithmic news engine (Reuters NewScope) that quickly provided information on whether the news was positive, neutral, or negative toward listed companies. Transaction data from 39 liquid stocks on the London Stock Exchange (LSE) were studied together with approximately 30 000 news related to the stocks.

One major finding showed that the trading activity significantly reacted to relevant news as both trading volume and volatility drastically went up around news arrivals. Another interesting finding displayed above-average movements in the majority of the studied variables more than six minutes before the news was published. This phenomenon is also known with scheduled news as quarterly reports where investors estimate the company sentiment beforehand based on prior knowledge to make profitable trades. The result indicated that some investors seem to get information more quickly than others which could be connected to private and insider information. They also attributed the pre-news reaction to clustering of news, meaning that a small news alert could turn into a full-blown news story within minutes like a snowball effect.

Holm and Rødde (2019) found the same relationship as Gross-Klussmann and Hautsch (2011) relating to increased volume and volatility around news releases but on the Oslo Stock Exchange (OSE), using 25 stocks in the OBX-index and almost 1000 news articles. They observed a larger increase in volume and volatility starting about ten minutes before the news releases and suspected one reason was due to information leakages discussed in the study by Kim and Verrecchia (1994). Other explanations concluded that the observed news could be based on other news published outside the sample, creating misleading data or that a reverse causality problem was present, e.g news articles could inform about volatility changes causing new reactions.

The results by Gross-Klussmann and Hautsch (2011) are also consistent with the findings from Dugast (2018) who observed spikes in trading volume and volatility around unscheduled news events. However, his approach differed as the study was based on a theoretical equilibrium model (no actual articles analysed) that assumed all trading agents were rational and with the ability to fully use dynamic strategies involving market orders, as well as cancellation and placement of limit orders. The study explained that investors get aware of news after a stochastic delay as they cannot monitor the market perfectly due to limited attention. Those who react quickly enough will naturally benefit from an informational advantage and potentially achieve excess returns. The study by Dugast (2018) further observed a positive covariance between trading volume and volatility explained by a volume spike, causing imbalanced order flows and eventually changes in asset prices. Bing and Cui (2021) partly built upon the research from Dugast (2018) and developed a Rational Expectation Equilibrium (REE) model that analyzed how public information affected market efficiency and liquidity with a special focus on investors imperfect monitoring of the market. They found that an increased precision of public information increased market liquidity and concluded that improvements in regulation could be made relating to information disclosure among listed companies to enhance market efficiency and liquidity. The suggestion of enabling a more accurate information disclosure on more media channels would benefit uninformed investors and decrease the gap to informed investors having private information.

2.2 Sentiment Analysis

Sentiment Analysis is a current research field tied to text mining methods, which computes sentiments and subjectivity of text. There are three different levels of classification: aspect-level, sentence-level, and document-level. Aspect-level aims to classify the sentiment that is divided, meaning that both a positive and negative sentiment would be present in the sentence, e.g. "the revenue increased but the profits decreased". The sentence-level classification identifies if there is an objective or subjective opinion in the sentence and then determines whether it is positive or negative. Document-level sentiment analysis involves classifying sentiments in whole documents consisting of one topic (Medhat et al., 2014).

There are various articles related to sentiment analysis, mentioning different applicable techniques in financial research as the random walk- and entropy-based algorithm. Tan and Wu (2011) designed an algorithm for a domain sentiment lexicon based on the random walk model, which assumes that the time series takes a random "step" from its last recorded position from one period to the next. They used three Chinese domain-specific data sets whereof one was linked to stock reviews and managed to vastly improved the identification accuracy of sentiment polarities.

Yu et al. (2013) investigated the intensity of emotion words to classify stock news sentiment by using a contextual entropy model. It measures the similarity and intensity between two words by comparing their context distribution with an entropy measure, i.e. amount of information in the text. The intensity varies among different words and can be given scores between -1 and 1, e.g. "soar" (0.9) would have a higher intensity than "rise" (0.5) and "collapse" (-0.9) and "fell" (-0.5) would have different scores as well. They claimed to improve classification performance compared to other models and proposed to investigate whether emotion words refer to the whole market or a specific company, further improving classification performance. Li et al. (2014) used 125 000 news articles as input and applied SVR (Support Vector Regression) as a machine-learning algorithm to analyze the stock prices of Chinese CSI 100 stocks. RMSEs

(Root Mean Squared Errors) were utilized for evaluating the performance and a three-month trading simulation was done indicating a 166.11% return without transaction costs and with an RME of 0.612. Mohan et al. (2019) developed various stock price predicting algorithms using news sentiment analysis based on deep learning models (a function that imitates the mechanisms of the human brain for finding patterns). 265 000 news articles were analyzed and used as input and MAPE (Mean Absolute Percentage Error) was the prediction accuracy measure. Four consecutive days of closing prices predicted the fifth day's stock price. The best performing algorithm achieved 80 percent lower MAPE on Apple, Amazon, American Airlines, Facebook, and Microsoft stocks over 50 days of testing compared to the baseline model ARIMA (Automated Regressive Integrated Moving Average). The so-called RNN-LSTM (Recurrent Neural Network-Long Short Term Memory) model also achieved eight percent lower MAPE with price and text polarity as input compared to only the stock prices.

Souma et al. (2019) used deep learning and RNN-LSTM as well but applied a novel way of defining news sentiment, i.e. polarity. They used the average DJIA 30 stock price of a one minute interval prior to and after the news was released to label the polarity as positive or negative. The high-frequency news and stock prices had a time precision in milliseconds and about 375 000 news was inputted to the deep learning model. Vicari and Gaspari (2020) applied RNN-LSTM to a binary classification, e.g. "does the stock go up or down?" with a cross-entropy loss function that calculates the difference between two probability distributions. They analyzed 25 daily news headlines from 2008 to 2016 (extended to 2020) with the aim of developing an algorithmic trading strategy. LSTM is motivated to be good from a linguistic perspective since its ability to memorize previous words in a sentence. The study removed the assumption that "positive financial sentiment"="positive words" and concluded that financial forecasting became utterly difficult as their deep learning model did not achieve significantly higher predictive accuracy than the flip of a coin (53%).

Beschwitz et al. (2019) used a different approach studying news analytics and algorithmic trading. They observed how the current available news analytic tools affected the stock market in high-frequency trading, in contrast to the majority of other papers focusing on the effect of the underlying news. RavenPack was used as the news analytic software from a leading company providing real-time sentiment and ESG data from articles. The software is able to determine the relevance to each company mentioned in each article from the Dow Jones Newswire, and whether the news is positive or negative. RavenPacks's subscribers get access to the analytics within milliseconds, allowing them to react faster than any human being. They concluded that news analytic software speeds up trading volume and stock price responses to articles, but reduces liquidity. Furthermore, inaccurate news analytics caused minor price distortions and the market impact was strongest for press releases.

3 Theory and Hypothesis

This section presents relevant theories used to develop a testable hypothesis relating to the purpose and research questions of the study.

3.1 Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) by Fama (1970) states that stock prices fully reflect the available market information, i.e., on average, competition causes instantaneous price fluctuations based on new information on intrinsic values. There are three forms of market efficiencies, weak, semi-strong and strong. The weak form states that asset prices have discounted all past relevant information, which means that e.g., technical analysis cannot give traders an edge in the market (Fama, 1970). On the other hand, new information can help identify over or undervalued assets based on fundamental analysis. The Semi-strong form is argued to be the most plausible scenario of the market in reality which suggests that all public information quickly causes asset prices to reach their equilibrium. Investors are said not to gain any advantage unless having access to material nonpublic information (MNPI), i.e., information that is not readily available to the public (Chen, 2021b). This implies that neither technical nor fundamental analysis would achieve superior gains compared to the index. The strong form argues that prices, as well as both public and private information, are reflected in the asset prices, creating a perfect and totally efficient market, and making it impossible to beat stock indices without having luck over a large time period. This study will test the semi-strong form as both price and public company information are analyzed.

3.2 Random Walk Theory

The logic behind EMH is based on the Random Walk Theory stating that stock prices have the same distribution and are independent of each other. Past movements or trends of stock prices are therefore assumed not to help predict future movements. Stocks are supposed to move in a random and unpredictable way making all attempts of achieving excess returns and beating the market almost impossible, similarly to the strong form of EMH (Smith, 2020). Peter Lynch, a senior hedge fund manager argued that the EMH is contradicting the Random Walk Theory and that students do not get taught about it. He stated that fluctuations in asset prices are not random if prices are rational and based on all available data as the EMH proposes. But if asset prices are rational, the Random Walk Theory is not valid which causes a contradiction (Lynch, 1989).

3.3 Mixture of Distribution Hypothesis

The Mixture of Distribution Hypothesis (MDH) by Clark (1973) and Harris (1987) implies that the joint distribution of volatility and volume is bi-variate normal conditional consequent to new information. In other words, the volatility and volume at a given interval are proportional to the arrival rate of information. According to this framework, price changes are therefore driven by news trading, whereas uninformed traders usually trade when they see large price movements.

3.4 Theory of Asymmetric Information

The economic theory of asymmetric information was developed in the 70s and 80s by George Akerlof, Michael Spence, and Joseph Stiglitz as a plausible explanation for market failures caused by an imbalance of information between buyers and sellers (The Noble Prize, 2001). Asymmetric information can furthermore generate adverse selection as e.g., CEOs, board of directors, and other "insiders" certainly know more than other shareholders about the profitability of the firm. Companies with lower than average profitability will therefore be overvalued and more prone to issuing own shares compared to highly profitable and undervalued firms when financing new projects. This will result in a tendency for low-profitability firms to grow more quickly, causing a stock market initially dominated by "lemons", i.e., overvalued assets, leading to large drops in share prices when uninformed investors discover their mistakes (Akerlof, 1970).

Media is said to mitigate asymmetrical information between investors as it enables a fair and simultaneous distribution of knowledge used to evaluate stocks (Fang and Peress, 2009). News especially resolves asymmetric information in small and illiquid stocks as the correlation between volume and absolute returns decreases more following news stories that consist of many newswire messages and earnings-related words (Tetlock, 2010).

3.5 Hypothesis

If the stock market is as efficient as the EMH states, the average volume and volatility would be close to constant before the unscheduled news is released (1), spike just when it is available to investors (2) and then quickly iterate to an equilibrium thereafter (3). This is under the condition that larger external news is not present in the observed event window. If not, there would exist market inefficiencies that could be used to achieve excess returns (including transaction costs) using trading algorithms fast enough to understand the market sentiment.

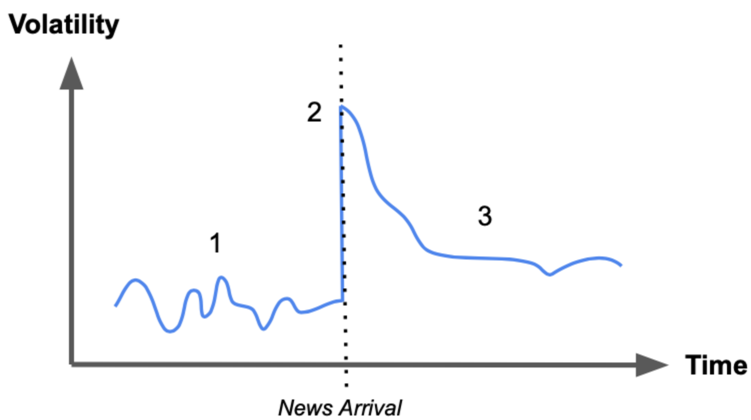

Figure 1: Illustration of hypothesis showing how the intraday volatility would change before (1), at (2), and after (3) the news arrival if it is consistent with the EMH.

4 Data and Methodology

This section presents and describes the basic data and data sources, as well as the methodology used in the study.

4.1 Data

There are two types of data, one of the prices for 4659 companies and derivatives listed in Sweden derived from the Swedish House of Finance (SHOF), and the other consists of approximately one million news articles published on Placera. However, due to certain criteria and time limits, the data was filtered down to 1097 stocks and 17 873 news articles.

4.1.1 Swedish House of Finance stock data

SHOF has data going back to 2010, however, only the years 2020 and 2021 were used in this project. Stock data is not available for every day (see *Appendix 1*), and information of various stocks is missing on the dates with available data. It is structured in JSON format (see *Table 1*) and only price, quantity, timestamp, and symbol were extracted from the JSON object. The timestamp provided precision in nanoseconds and the distribution of data by symbol and currency can be seen in *Appendix 2*.

type	example
currency	SEK
isin	G00OBL045030
market	FNSE
match number	10
message type	E
mpid	MSN
mpid count	AVA
order reference	2592195
price	429200
quantity	100
sell	false
seqno	4
SHOF id	147044
symbol	BULLNIOX2AVA4
symbol state	T
timestamp	22500021027009

Table 1: The data contained in one JSON object

4.1.2 Avanza news data

The news data were collected from Avanza's news platform Placera. It included just above one million news articles ranging from 2002 to 2022 that were extracted with a data mining script. The data contained information regarding the company, date with time in minutes, and the text in each news article could easily be divided into headlines and subtexts. A sample of news has been used to evaluate whether they are released simultaneously or if there are other news sources publishing earlier than Placera. In regard to international companies, there was a few news posted a couple of minutes earlier on Bloomberg that probably was translated to Swedish and later incorporated to Swedish news providers. However, this concern was solved as only Swedish stocks were included in the study.

4.2 Methodology

Downloading the stock data from SHOF was time-consuming as the daily amount of data contained about 250 MB, which had to be downloaded separately, explaining why only the years 2020 and 2021 were included in the study. The news articles also had to fulfill criteria, as being released intraday, i.e., when the market is open. Other excluded news did not inform about companies or was not in text format. The article count per month after the exclusion process can be seen in *Appendix 1*. The data retrieved from the two data sources had to be combined as the ticker from SHOF needed to match with the corresponding names from Placera. This was not a trivial task, since Placera had own names for the stocks, resulting in multiple acronyms for each ticker. Some additional news articles were removed as the stock data was not available for each article. To take the news clustering phenomenon mentioned by Groß-Klußmann and Hautsch (2011) into consideration, even more news were filtered out (30%) based on the criteria that only one news per day per stock could be included in the sample.

5 Results and Analysis

This section presents and analyzes plots of stock characteristics in a time window of sixty minutes prior to and after the news releases in the form of return, standard deviation, volume, and absolute return. The data was divided into four segments with decreasing average market cap per stock as the characteristics was expected to differ.

Market	Number of news	Number of stocks	Avg. number of news
Large Cap	5397	163	33
Mid Cap	3309	148	22
Small Cap	1454	94	14
Other	7713	692	11

Table 2: Overview of the news distribution between the Swedish markets segments. "Other" includes the markets; Spotlight Stock Market, First North Stockholm, Nordic SME Sweden, SPAC List Stockholm, NGM, Inofficiella (beQuoted), NGM PepMarket, Xterna listan (Avanza, 2022).

Large and statistically significant movements in various stock variables as price, standard deviation, volume, and absolute return were observed around the news releases. The volume and absolute return shared similar characteristics in terms of timing and shape of the plots. There were also noticeable differences in reactions comparing "Large Cap", "Mid Cap", "Small Cap" and "Other" with increasing magnitude. About all the studied variables (except for "Large Cap") rapidly increased approximately 15 minutes before the news release to reach its peak just when it got published and adjusted about 15 minutes thereafter. Something interesting but difficult to understand is the increase taking place again between 45 and 55 minutes after the news release, which is most prominent for the smaller market segments. As mentioned in the Data and Methodology chapter, the SHOF data has a precision in nano-second level, and the data from Placera has a minute level precision. This means that the combined upcoming results have a precision in minutes.

5.1 Average return

As neither a news analytic tool was used or own news sentiment algorithm was developed, the average return was observed to get an approximation of the distribution between positive and negative news. The fluctuations around the y-axis are relatively small for "Large Cap" indicating an even distribution between positive and negative news around the news release, "Mid Cap" looks similar but with slightly higher fluctuations. Both "Small Cap" and "Other" fluctuate much more and are rapidly increasing in value just before the news release, indicating but not proving a larger distribution of positive news compared to negative ones.

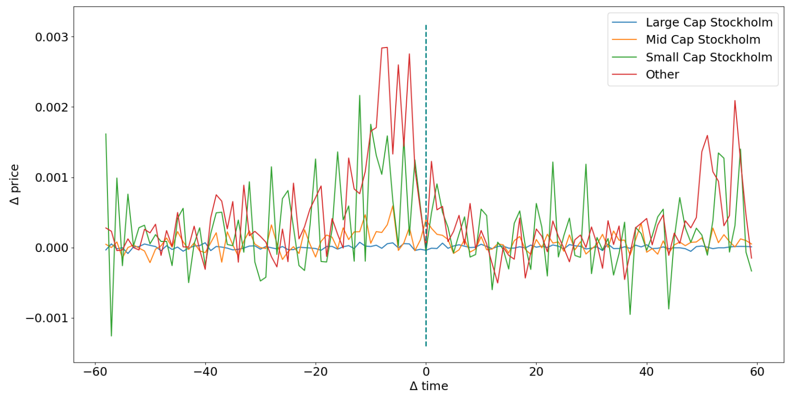

Figure 2: The x axis shows the time difference in minutes from the news release, ranging from -59 to 60. The y axis shows the average minute over minute returns. The dashed line represents Δ time=0.

5.2 Standard deviation

The standard deviation is noticeably different between the segments with rising values from "Large Cap" to "Other", something expected since a lower market cap is connected to higher volatility and vice versa. The slope of the plots is gradually increasing (except for "Large Cap") about 15 minutes before the news release and somewhat stabilizes 15 minutes afterward, "Other" spikes abnormally high about five minutes before publication.

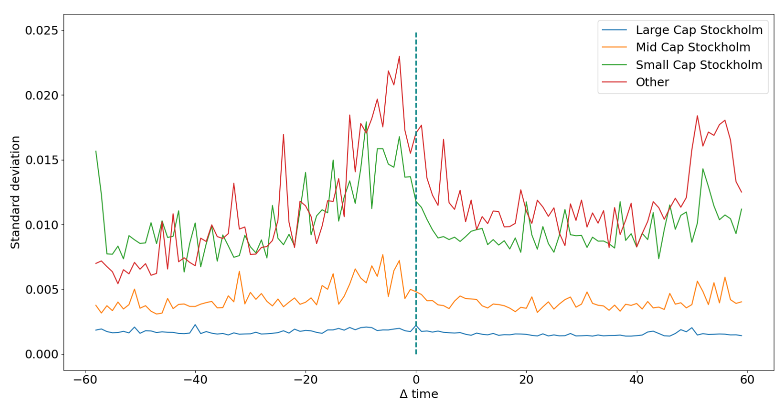

Figure 3: The x axis shows the time difference in minutes from the news release, ranging from -59 to 60. The y axis shows the standard deviation of the minute over minute returns. The dashed line represents Δ time=0.

5.3 Volume

All market segments are fluctuating in volume, but are relatively constant one hour to about 15 minutes before the news release to then rapidly increase thereafter until news publication (except for "Large Cap"). The volume is stabilizing about 15 minutes after the news release but increases again between 45 and 55 minutes, which is most prominent for "Other" and "Small Cap". The volume of "Other" and "Small Cap" is increasing much faster and peaks about 5 minutes before the news release whereas "Large Cap" and "Mid Cap" are spiking sharply just at the news release.

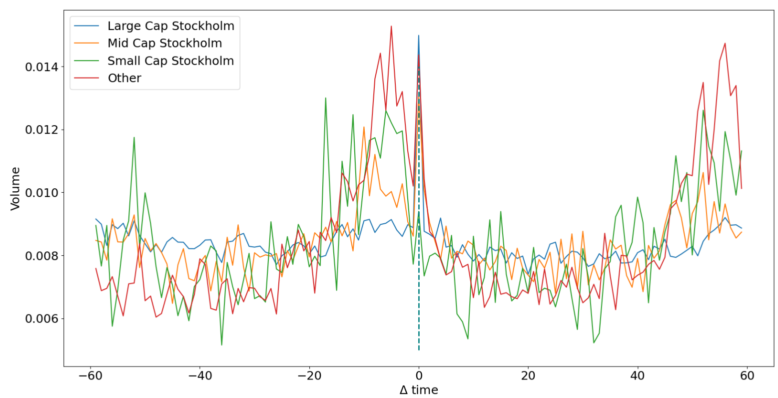

Figure 4: The x axis shows the time difference in minutes from the news release, ranging from -60 to 60. The y axis shows the average volume. The dashed line represents Δ time=0. The magnitude of the y axis is set so the sum of volumes equals 1.

5.4 Absolute return

The plot of absolute returns (volatility) shares similar characteristics to the volume plot and showcases aspects difficult to notice in the previously mentioned average return plot, which is not able to distinguish the effects of both positive and negative news on returns as they are balanced out. The absolute return for "Large Cap" is relatively constant before the news release, for "Other" it is increasing from 60 minutes to 5 minutes pre-news release, whereas for "Small Cap" and "Mid Cap" is relatively constant between 60 minutes and 15 minutes prior to the news releases. The absolute return is then rapidly increasing to reach its peak at the news release for "Large Cap" but a couple of minutes earlier for "Mid Cap", "Small Cap", and "Other".

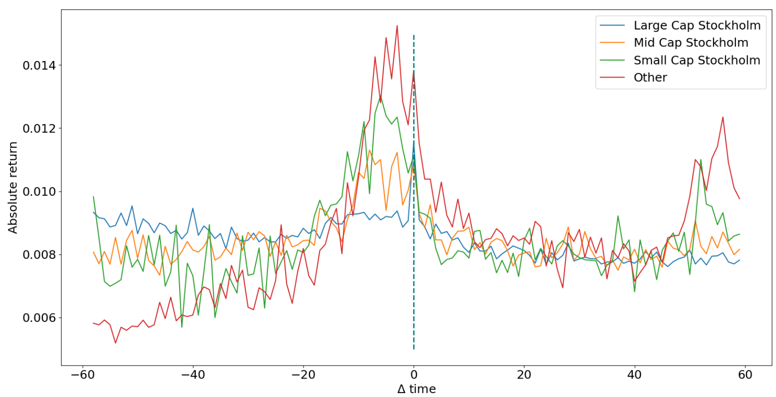

Figure 5: The x axis shows the time difference in minutes from the news release, ranging from -59 to 60. The y axis shows the average minute over minute absolute returns. The dashed line represents Δ time=0.

5.5 Stock price predictability

The cumulative return between -16 and -1, as well as -1 and +14 minutes was calculated to evaluate if the pre-news price development could predict the post-news price. A correlation different from zero between these returns could enable the prediction of stock prices (See *Table 3*), but the results indicate no correlation significantly different from zero. Plots for each of the four markets can be seen in *Appendix 3*.

Market	Correlation
Large Cap	0.043
Mid Cap	-0.153
Small Cap	-0.183
Other	-0.104

Table 3: Correlation between pre- and post news cumulative return. The cases where any of the cumulative returns exceeds an absolute value of 20% have been removed.

5.6 T-test

Upward movements before Δ time=0 were observed in the four plots illustrating average return, standard deviation, volume, and absolute return. One-sided t-tests were done, testing the hypothesis implying that the mean in the interval -60 to -20 minutes is equal to the mean in the interval -20 to -1 minutes. The p-values of these tests are shown in *Table 4*, indicating significant differences in the means between the two periods for 9 out of 16 measurements on a 1% significance level. The differences between the means indicates that information about the news articles are released to some traders prior to its official release.

	Average return	Standard deviation	Volume	Absolute return
Large Cap	$8.29 \cdot 10^{-2}$	$1.16 \cdot 10^{-1}$	$3.23 \cdot 10^{-3}$ *	$2.01 \cdot 10^{-2}$
Mid Cap	$1.83 \cdot 10^{-4}$ *	$5.53 \cdot 10^{-6}$ *	$5.23 \cdot 10^{-2}$	$4.59 \cdot 10^{-7}$ *
Small Cap	$1.28 \cdot 10^{-2}$	$4.87 \cdot 10^{-11}$ *	$2.25 \cdot 10^{-1}$	$4.93 \cdot 10^{-11}$ *
Other	$4.41 \cdot 10^{-23}$ *	$5.73 \cdot 10^{-78}$ *	$3.16 \cdot 10^{-13}$ *	$8.21 \cdot 10^{-55}$ *

Table 4: The table shows p-value for one sided t-test, with the hypothesis that the mean in the time period -60 min to -20 min is equal to the mean in the time period -20 min to -1 min. * indicates that there is a significant difference on a 1% significance level.

6 Discussion and Conclusions

This section summarizes the major findings discussed in relation to previous studies, theory and the proposed hypothesis. Research implications are mentioned and future research is suggested.

6.1 Discussion

The main findings are consistent with the results from Groß-Klußmann and Hautsch (2011), Holm and Rødde (2019) and Dugast (2018) regarding an observed increase in volume and volatility around the news releases. This supports the Mixture of Distribution Hypothesis stating that volume and volatility are correlated since they are driven by the same underlying information arrival process (Clark, 1973). However, neither article differentiated different market segments like this study which is able to distinguish low to high liquidity stocks, indicating interesting results. It seems like the less liquid the stocks are, the higher the increase in volume and volatility prior to the unscheduled news releases. "Other" and "Small Cap" even peaked higher 5 minutes pre-news release compared to at the release. The results for "Large Cap" are consistent with the findings from Groß-Klußmann and Hautsch (2011) and Holm and Rødde (2019) as a significant increases in volume were observed prior to the news releases, however the obvious increase takes place a couple of minutes later for "Large Cap". This observation is interesting since they observed the most liquid stocks on the London- and Oslo Stock Exchange.

News especially mitigates asymmetric information for small and illiquid stocks (Tetlock, 2010) and the rapid increase in volume and volatility could indicate that investors have more access to private information in this segment (Kim and Verrecchia, 1994; Bing and Cui, 2021). However, no noticeable increase was prominent for "Large Cap", indicating market efficiency from a strong form of EMH. This finding also contradicts the Random Walk Theory as the observed fluctuations do not seem to be random, but rather rational and based on available information about the stocks (Lynch, 1989).

No actual sentimental analysis was done in this study in comparison to previous studies mentioned, complicating the process of predicting the stock prices after unscheduled news releases. Although, one could theoretically argue that a significant correlation between price movements pre-news release compared to post-release together with an available news analytic tool, capable of distinguishing relevant news sentiment at "time=0" would enable opportunities finding patterns to gain on average excess profit through algorithmic trading. However, the findings show no significant correlations in any of the market segments, indicating low to no possibilities of beating the market, consistent with the EMH and Random Walk Theory. This finding corresponds with observations from previous studies (Groß-Klußmann and Hautsch, 2011; Holm and Rødde, 2019; Vicardi and Gaspari, 2020). Mohan et al. (2019) gained excess return without transaction costs but applied after-hours news and stock data, as well as longer trades, possibly explaining the difference.

An important aspect to consider regarding the pre-news reaction in this study is the possibility of a news clustering phenomenon being present mentioned by Groß-Klußmann and Hautsch (2011), i.e., there may be news included being small alerts that evolve to larger news within minutes, causing a slight time shift in the stock price reactions. However, this issue was controlled for by making sure no more than one news per day per stock were included in the sample. There is also a risk of a reverse causality problem affecting the results, meaning that volume and volatility changes could be affected by news writing about those variables. The fact that the data covers the covid-19 pandemic in 2020 and 2021 may have an effect on the results as it is connected to a more volatile market period (Harron and Rizvi, 2020) than previous years. One could argue that the signs of insider trading could be more prominent in this case, but it would be controlled for by extending the data usage period and by comparing the results. Moreover, as there were no articles available on Placera with second-level precision (SHOF provided nano-second data), some important details cannot be observed just around the news release, which would have been useful.

6.2 Conclusions

There are indications of inefficiencies in the stock market due to asymmetric information between informed and uninformed investors. Spikes in volume and volatility before unscheduled news (that should not be known to the public) could be explained by investors having access to private information not available to others. This concern is mostly observable for less liquid stocks in the smaller market segments and the reason can only be speculative. Is it easier to exploit insider information within those companies as they are not as public or regulated? However, the stock market is also showing proof of efficiencies as almost all returns already are incorporated into the stock prices when the news is released, making it utterly difficult to profit when trading based on them. The behaviour of "Large Cap" is also consistent with the hypothesis of the market being efficient in line with the EMH. This finding speaks against using news analytic tools for trading considering transaction costs and it may be even more difficult in the future when computer speed increases and more investors get access to the same technology. On the other hand, this only makes the market more efficient, but to reach an optimum, more action probably needs to be taken in form of regulations to prevent private information from affecting the market in the smaller market segments.

6.2.1 Future research

Something interesting to build upon this study would be using a news analytic tool that sorts out the most irrelevant articles, i.e., neutral articles that do not indicate whether a trade would be profitable or not. That would make it easier to evaluate the possibilities of developing an actual trading strategy. Instead of dividing the data into market segments based on market cap, it could also be done by comparing industries, e.g., biotech, bank, and IT. Another topic to investigate further could involve finding ways of measuring how private information from informed investors impacts the stock market in Sweden, but how to make that possible is unclear and may require a qualitative methodology.

7 References

Akerlof, G. A. 1970. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. The Quarterly Journal of Economics 84(3): 488-500.

Avanza. 2022. Aktielistan. Available at: https://www.avanza.se/aktier/lista.html (Accessed: 10 March 2022).

Beschwitz, B., Keim, D. B., and Massa, M. 2019. First to "Read" the News: News Analytics and Algorithmic Trading. Oxford University Press.

Bing, T., and Cui, Y. 2021. The Dynamic Effect of Public Information on Liquidity: From the Perspective of Limited Attention. Hindawi.

Chen, j. 2021a. News Trader. Investopedia.
Available at: https://www.investopedia.com/terms/n/news-trader.asp
(Accessed: 10 Nov 2021).

Chen, j. 2021b. Semi-Strong Form Efficiency. Investopedia.
Available at: https://www.investopedia.com/terms/s/semistrongform.asp
(Accessed: 20 Apr 2022).

Chen, j. 2022. Algorithmic Trading. Investopedia.
Available at: https://www.investopedia.com/terms/a/algorithmictrading.asp
(Accessed: 24 Feb 2022).

Clark, P. 1973. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. Econometrica 41: 135-155.

Cutler, D. M., Poterba, J. M., and Summers, L. H. 1989. What moves stock prices? Journal of Portfolio Manage- ment 15: 4–12. 93-105.

Dugast, J. 2018. Unscheduled News and Market Dynamics. The Journal of Finance 6: 2537-2586.

Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. The Journal of Finance 25: 383–417.

Fang, L., and Peress, J. 2009. Media Coverage and the Cross-section of Stock Returns. The Journal of Finance 5: 2023-2052.

Groß-Klußmann, A., and Hautsch, N. 2011. When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. Journal of Empirical Finance 18: 321-340.

Haroon, O., and Rizvi, S. A. 2020. COVID-19: Media coverage and financial markets behavior—A sectoral inquiry. Journal of Behavioral and Experimental Finance 27.

Harris, L., 1987. Transaction Data Tests of the Mixture of Distributions Hypothesis. The Journal of Financial and Quantitative Analysis 22: 127-141.

Holm, P. B., and Rødde, J. 2019. High-Frequency Market Reactions to Unscheduled Stock-Specific News. NHH Brage.

Kim, O., and Verrecchia, R. E. (1994). Market liquidity and volume around earnings announcements. Journal of Accounting and Economics 17: 41–67.

Li, Q., Wang, Tiejun., Gong, Q., Chen, Y., Lin, Z., and Sa-kwang, S. 2014. Media-aware quantitative trading based on public Web information. Decision Support Systems 61: 93-105

Lynch, Peter (1989). One Up On Wall Street. New York, NY: Simon & Schuster Paperback. p. 52.

Medhat, W., Hassan, A., and Korashy, H. 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal 5: 1093–1113.

Mohan, S., Mullapudil, S., Sammetal, S., Vijayvergial, P., and Anastasiul, D. C. Stock Price Prediction Using News Sentiment Analysis. 2019. IEEE computer society.

Picardo, E., 2021. How to Trade the News. Investopedia. Available at: https://www.investopedia.com/articles/active-trading/111313/how-trade-news.asp (Accessed: 10 Nov 2021).

Quantra, 2021. News Based Trading. Available at: https://quantra.quantinsti.com/glossary/News-Based-Trading (Accessed: 24 Feb 2022).

Souma, W., Vodenska, I., and Aoyama, H. 2019. Enhanced news sentiment analysis using deep learning methods. Journal of Computational Social Science 2:33–46.

Tan, S., and Wu, Q. 2011. A random walk algorithm for automatic construction of domain-oriented sentiment lexicon. Expert Systems with Applications 38: 12094-12100.

Tetlock, P. C. 2010. Does Public Financial News Resolve Asymmetric Information?. The Review of Financial Studies 9: 3520-3557.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. 2007. More than words: Quantifying language to measure firms' fundamentals. In Proceedings of 9th Annual Texas Finance Festival.

The Nobel Prize. 2001. Markets with Asymmetric Information. The Noble Prize. Available at: https://www.nobelprize.org/prizes/economic-sciences/2001/popular-information/ (Accessed: 10 May 2022).

Vicaril, M., and Gaspari, M. 2020. Analysis of news sentiments using natural language processing and deep learning. AI & SOCIETY 36:931–937.

Yu, L., Wu, J., Chang, P., and Chu, H. 2013. Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. Knowledge-Based Systems 41: 89-97.

Zhang, W; and Skiena, S., 2010. Trading Strategies To Exploit Blog and News Sentiment. Association for the Advancement of Artificial Intelligence.

Appendix 1: Stock and news data

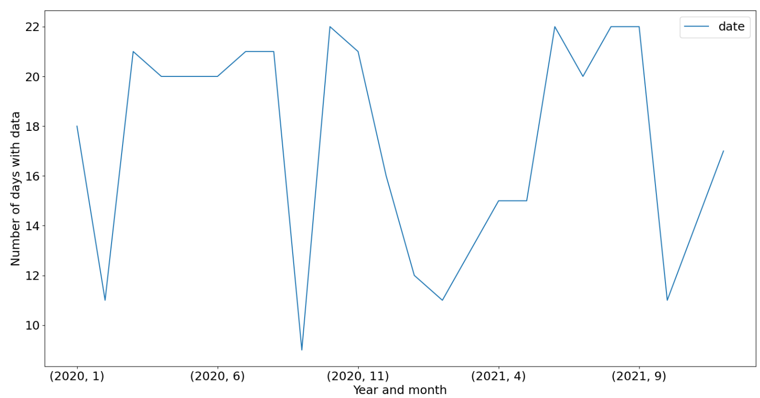

Figure 6: The x axis shows the months in 2020 and 2021. The y axis shows the number of days where data is available for each month.

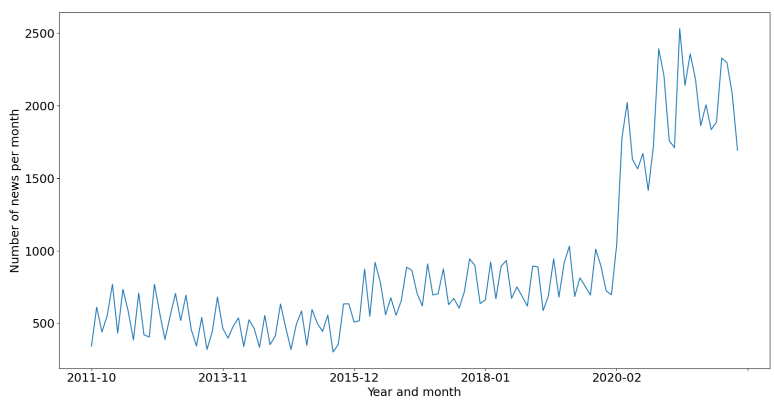

Figure 7: The x axis shows the months between October 2011 and January 2022. The y axis shows the number of news per month.

Appendix 2: Distribution of SHOF data

type	number of symbols
bull	727
bear	463
mini	844
turbo	172
tlong	111
tshrt	130
long	12
shrt	30
other	2170

Table 5: SHOF data by symbol for 30 december 2021

type	number of symbols
SEK	3035
EUR	699
DKK	695
NOK	205
ISK	25

Table 6: SHOF data by currency for 30 december 2021

Appendix 3: Scatter plots of returns

Each point represents an unscheduled news article

Large Cap

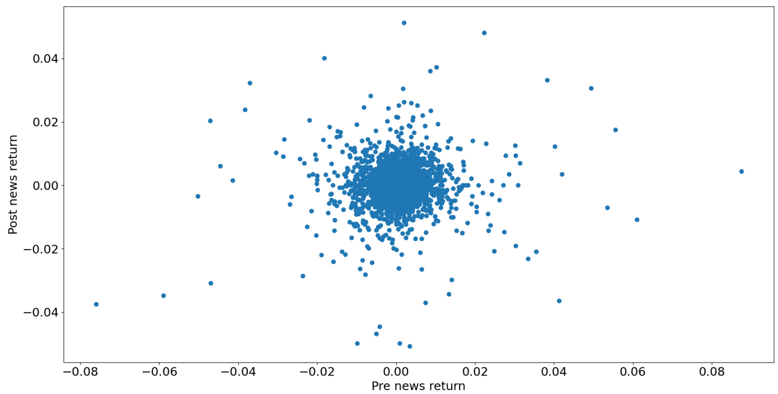

Figure 8: The x axis shows the cumulative returns -16 min to -1 min relative to the news article. The y axis shows the cumulative returns -1 min to +14 min relative to the news article.

Mid Cap

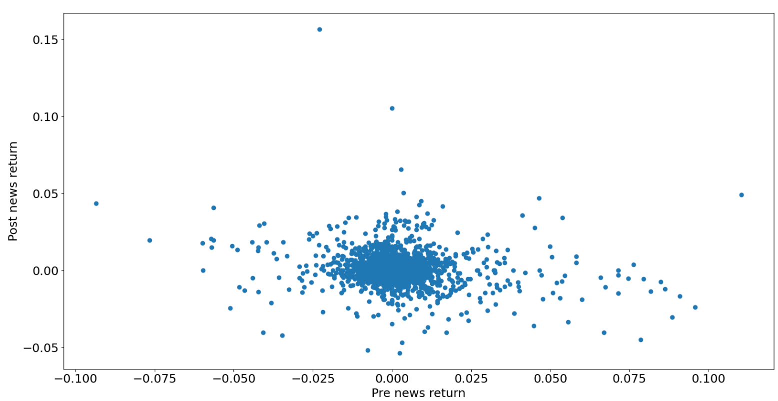

Figure 9: The x axis shows the cumulative returns -16 min to -1 min relative to the news article. The y axis shows the cumulative returns -1 min to +14 min relative to the news article.

Small Cap

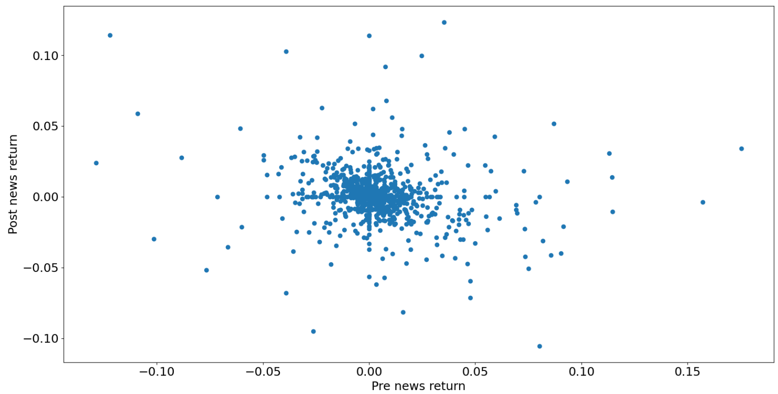

Figure 10: The x axis shows the cumulative returns -16 min to -1 min relative to the news article. The y axis shows the cumulative returns -1 min to +14 min relative to the news article. Each dot represent an article for the market small cap.

Other

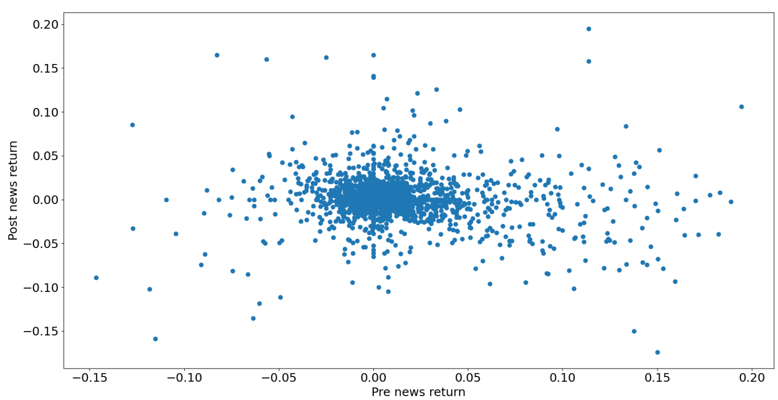

Figure 11: The x axis shows the cumulative returns -16 min to -1 min relative to the news article. The y axis shows the cumulative returns -1 min to +14 min relative to the news article.