ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Financial Economics

The relationship between the number of news items and stock's trading volume and volatility.

Name student: Adriaan Verstoep

Student ID number: 584850

Supervisor: Giovanni Coco

Second assessor: Dr. X. Ma

Date final version: 13-10-2024

Table of contents

Contents

Abstract

In this thesis the impact of news item frequency has on market volatility and trading volume is studied. This is research based on a paper of Barber and Odean (2008), there research was done on the S&P 500, based in the United States, this research was done in the Netherlands. This research could help investors make more informed decisions and it could help policymakers in creating new policies. Utilizing a dataset of 22 companies listed in the AEX index in 2022 and 2023, a strong positive correlation of 14925.44 extra shares traded with 1 extra news article published. It was found using a regression analysis between the number of news items published concerning a company and its trading volume. When a lagged variable of the number of news items was included, the explanatory power of the model was greater, indicating a more comprehensive understanding of the trading volume variation. In addition, it was observed that there is a weak negative correlation of -0.0002724 with 1 extra news published and the stock's volatility. Control variables, such as market capitalization were included in both regression models to mitigate potential cofounding effects. Despite these controls, part of data variance remains unexplained, suggesting the presence of other factors affecting stock volatility that would require further investigation. Overall, this research contributes valuable insights into the relationship between stock market dynamics and media coverage. Moreover, it underscores the necessity for further research to fully understand the dynamics of the stock market and media coverage.

Introduction

The influence of media on the stock market has been a subject of extensive investigation in the past. However, as the world continually evolves and trust in traditional media diverges more than ever before, the influence identified in earlier studies may not hold the same weight in the current context. This thesis aims to explore traditional media's role in financial markets and how its influence has changed, if changed at all. Recognition of the influence of traditional media on financial markets will help understand the behavior of investors in the stock market.

The effect of news items on stock volatility was examined by Baber and Odean (2008) using data from 1991 to 1996. They concluded that individual investors tend to buy attention-grabbing stocks that are featured in the news. As this study was conducted before the widespread use of social media, new studies are required to determine the influence of traditional media as this may have fundamentally changed.

The research described in this thesis investigated the influence of news items on trading volume and volatility of stocks registered in the AEX index. Volatility is defined as the degree that prices fluctuate and reflects the amount of risk and uncertainty associated with an asset. The AEX index comprising of the 25 biggest stocks of the Netherlands, providing a robust dataset with substantial news coverage, and therefore suitable for this research paper. The research question guiding this paper is:

What is the impact of the weekly frequency of news articles on volatility and trading volume of stocks in the AEX index?

This thesis will contribute to a deeper understanding of the Efficient Market Hypothesis by examining the role of media coverage on financial markets. Additionally, it will fill the gap in literature, as most existing papers focus on the S&P 500 index, the most well-known stock market in the world, which is based in the United States, and thereby enhancing the knowledge in of financial markets in a different region, i.e. The Netherlands, Europe. The findings of this study will also have practical implications for investors, enabling more informed investment decisions based on news impact. Furthermore, the results could provide policy makers with better insight into the effects of news on financial markets, guiding the development of appropriate regulations.

This thesis hypothesized that the frequency of news items would positively effect trading volume. Furthermore, it was hypothesized that the frequency of news items would positively affect stock volatility. Analysis revealed that frequency of news items had a positive correlation with trading volume. Contrary to the hypothesis, analysis revealed that frequency of news articles had a negative effect on a stock's volatility.

This paper will explore various studies on the influence of media. The hypothesis will be articulated in the theoretical framework section. Following this, the methodology will be outlined in two parts: first, the data collection and database construction will be discussed, and then the research methods will be elaborated upon. Subsequently, the results will be presented, analyzed, and discussed. The discussion section will address potential flaws and limitations of the study. The thesis will conclude with a summary of the findings and their implications.

Theoretical framework

The influence of media coverage on stocks trading volume and volatility

As mentioned above, Barber and Odean (2008) investigated the relationship between the frequency of a company's mentions in the news and its stock's volatility and trading volume. Their study analyzed data from a stockbroker, encompassing 78.000 different households and their stockbrokerage accounts from January 1991 through December 1996. Barber and Odean specifically focused on households and not institutional investors, as they hypothesized that the effect would be more pronounced with individual investors due to their limited access to stock market information. Their findings confirmed their hypothesis, concluding that attention-grabbing events significantly influenced investors' buying behavior, leading to a significant increase in the purchase of stocks that were prominently featured in the news.

Engelberg and Parsons (2011) investigated the causal relationship between news coverage and trading behavior. They utilized data from local news outlets from 19 different regions in the United States and a diverse set of stocks in the S&P 500 index, ensuring that the results were not driven by firm-specific factors, together providing a robust data set. It was concluded that the presence of or absence of local media coverage is strongly related to the probability and magnitude of trading. Specifically, their findings indicate that more local media coverage leads to a 75 percent rise in trading volume. Moreover, this study provides compelling evidence that trading behavior, measured by trading volume, is influenced by news coverage, even at the local level.

Byström (2016) also investigated the relationship between the number of published news items and the stock market volatility. The paper analyzed over 9 million reports written in Englisch or in Mandarin, also known as Chinese. This data was combined with data from stock indices, six of the stock indices were based in Englisch speaking regions, mainly the United States and United Kingdom, another six stock indices were base in the Mandarin speaking region, China and Hongkong. It was concluded there was a strong positive correlation between the number of articles published and the stock market's volatility.

Atkins et al. (2018) conducted a study on predicting the volatility of the stock market using data from the Reuters US news archive. Data from the NASDAQ index and the Dow Jones industrial average, two of the most prominent stock indices in the United States, from September 2011 through September 2012 were studied. Atkins et al. developed a predictive model utilizing news data, that predicts the direction of a specific stock's volatility with 56 percent accuracy, which is significantly higher than the 49 percent accuracy obtained when using a model solely based on closing prices. The findings of this study clearly demonstrate a correlation between news coverage and the stock market in the United States.

The influence of attention on stocks trading volume

Numerous studies, for example by Vlastakis and Markellos (2012) and Ding and Hou (2015), have been published on the influence of attention on the stock market. Indeed, media attention significantly increases attention to stocks, as potential investors read about and discuss these stocks with friends and family. The following section will review several key papers that explore the effect of attention on stock market behavior.

Vlastakis and Markellos (2012) investigated the relationship between information demand and investor behavior. In their study information demand was defined by the frequency of Google search engine queries. Their analysis covered the period from January 2004 to December 2009. Their findings indicate that information demand is significantly and positively correlated with both volatility and trading volume on stock in de NYSE and NASDAQ indices, both indices are based in the United States. Additionally, the researchers found that information demand increased when the stock market had higher returns, suggesting that investors seek more information during bullish market trends, bullish periods in the stock market are periods when the returns increase.

The attention effect has been extensively researched. Ding and Hou (2015) investigated the effect of the Search Volume Index (SVI), a measure of how often a stock is searched on the Google search engine, on turnover rate and stock liquidity. Analyzing frequency data of the S&P 500 in the period of January 2004 and December 2009. Their finding indicated that higher attention, as measured by SVI, contributes to a larger shareholder base, an increased turnover rate, as well as an increase in the relative number of shares available for trading to all the available shares. While this research did not examine the influence of attention on volatility, the observed increase in turnover rate would indicate that trading volume increases when the attention to a stock is increased.

Taken together, these studies of Baber and Odean (2008) and Engelberg and Parsons (2011) clearly indicate that there is a strong correlation between news articles and trading volume. Byström (2016) and Atkins et al. has proven there is correlation between the number of articles and stock's volatility. Vlastakis and Markellos (2012) and Ding and Hou (2015), have proven there is correlation between attention and investor behavior when trading on the stock market. However, as most of these studies were based on the stock exchange of the United States, the question remains how the Dutch stock exchange is influenced by news items in The Netherlands.

Hypothesis

Based on the literature mentioned above and my expectations I will formulate my hypothesis. Firstly, I hypothesize that an increase in the number of articles published about a company will lead to an increase in trading volume. This expectation is supported by the findings of Barber and Odean (2008) and Engelberg and Parsons (2011), who demonstrated that increased news coverage correlates with higher trading volumes. Similarly, Ding and Hou (2015) concluded that greater attention, as measured by search volume, results in increased trading volumes, suggesting that media coverage can stimulate this attention. Although these studies focus on markets in the United States, there is no compelling reason to believe that the results would diverge for the Dutch market. Secondly, I hypothesize that an increase in the number of news articles published about a company will lead to an increase in stock volatility. Atkins et al. (2011) demonstrated that news data can be used to predict stock market volatility, and Byström (2016) found a strong positive correlation between news volume and market volatility. It is possible that when news is published about a company, investors may reassess and reconsider their positions, leading to increased buying or selling activity and, consequently, higher volatility.

Data

Dataset

This study analyzes the weekly count of articles about the companies that are included in the AEX index. To achieve this, Factiva, a comprehensive database that contains thousands of articles from global news sources including magazines, newspapers, and newswires was utilized. Newswires are companies that collect and distribute news stories and information to various news outlets. They serve as a crucial source that ensures that media organizations have access to up-to-date information. Factiva allows users to search for articles using various filters, for example filters related to subjects, industries, and companies. However, Factiva has one limitation: it does not directly provide the weekly count of articles for the specified research period. To conquer this issue, headlines and dates of the relevant articles were extracted from Factiva. These data were then converted into a suitable format for this study, resulting in a database that includes the date of the week and the weekly count of articles written about each company.

The Amsterdam Exchange Index (AEX) is the primary and the most well-known index in the Netherlands. It consists of the 25 largest and most traded companies on the Euronext Amsterdam stock exchange. The index encompasses a variety of different sectors such as technology, finance, consumer goods, energy, and healthcare, thereby providing a comprehensive presentation of the Dutch stock market. This study examines the period of 2022 and 2023, immediately following the COVID-19 pandemic. This timeframe includes both a bearish period (2022), during which the AEX declined and a bullish period (2023), when the AEX increased. The inclusion of both bearish and bullish periods provides a balanced timeframe for this research, ensuring that the results are applicable to both market conditions. This is crucial as it allows for a comprehensive understanding of how news article frequency influences trading volume and volatility in different market conditions, thereby enhancing the generalizability and usefulness of these findings.

The data on trading volume and volatility was obtained from Yahoo finance. The information on the companies included in this study was extracted and compiled into a database specifically used for this research. While Yahoo finance directly provides data on the traded volume, the data of the volatility had to be calculated using the adjusted closing prices. Volatility was calculated using a 12 weak rolling period to accurately capture changes and reflect market conditions at the time, thus capturing the market's perception of risk following news events. The 12-week volatility was then annualized to ensure it is comparable with other assets.

Not all companies were suitable for this paper, three companies were omitted. The first company omitted was Koninklijke DSM N.V. due to its merger with the Swiss company Firmenich SA. The second company, EXOR N.V., was excluded because it transitioned from the Milan stock market to the Amsterdam stock market on 12 August 2022, and therefore was not part of the AEX index for the entire research period. The final company omitted was Unibail-Rodamco-Westfield, which left the Amsterdam stock market on 9 March 2022. These exclusions were necessary to ensure consistency and accuracy in the dataset, as including companies with incomplete data could skew the results and provide biased results. For the full list of companies that were included in this paper see appendix table 2.

Methods

Firstly, the data from Factiva and the data from Yahoo finance were matched, the matched data contains 2288 data points from 22 different companies within the AEX-index. The data spans the same 104 weeks period, from January 2022 to December 2023 for each company.

Using the matched data a regression analysis was conducted to examine the relationship between weekly count of news articles and trading volume and volatility was run. However, companies differ in market capitalization, also referred to as market cap, which could influence trading volume. To address this issue, a control variable was introduced. The control variable is a categorical variable, companies were divided in four groups based on their market cap:

Group 1: Companies with a market cap of 0 to 20 billion USD.

Group 2: Companies with a market cap of 20 to 50 billion USD.

Group 3: Companies with a market cap of 50 to 100 billion USD.

Group 4: Companies with a market cap of more than 100 billion USD.

These groupings were chosen to distinguish between companies with a smaller market larger market cap, as market cap differences can influence trading volume due to varying investor attraction and company characteristics, such as growth potential, liquidity, and risk. Details of the different companies and their market caps can be found in Appendix Table 2, with data sourced from CompaniesMarketCap (2024).

Results

In this section, the findings of my study are presented. This thesis looks at the correlation between frequency of news items and stock's trading volume and volatility in the Netherlands. The data, collected through Factiva and Yahoo finance, provide insights into stocks trading volume and volatility. Starting with a regression analysis of frequency of news items on stock's trading volume applied in three models, followed by a regression analysis of frequency of news items on stock's volatility applied in two models. The findings are discussed in detail, with tables included to enhance clarity and support the interpretations.

Trading volume

The initial models focus on analyzing the effects of weekly article count on trading volume. The hypothesis stated earlier, posited that an increase in weekly news item count would result in an increase in trading volume. In order to study the role of media attention on stock buying behavior, measured by trading volume, 3 different models were built. For Model 1 about trading volume, starting with the most basic model on the relationship between news item frequency and trading volume, it was found that the frequency of weekly articles has a strong positive influence on trading volume. Using Model 1 it was determined that with every additional news article published, the weekly trading volume increases by 245,566.4 shares, ceteris paribus. The R-squared of the first model is relatively low at 0.1635, indicating that more variables should be included in the model. The results of this model are significant at the 1 percent level. This data is represented in Table 1, Column 1.

Model 2, the results of which are shown in Table 1, Column 2, includes both the weekly count of articles and the lagged weekly count of articles. The lagged weekly count was included because news items from the previous week could still affect the stock market in the following week. It is important to note that the number of observations of stocks in Model 2 is reduced by 22 compared to Model 1. This reduction is a result of the use of a lagged variable requires omitting the first week of data for each company, as there is no preceding week available to provide the lagged data. Model 2 shows that both the weekly count of articles and the lagged weekly count have a strong positive impact on trading volume, with coefficients of 161,627.4 shares for the weekly count and 127,881.8 shares for the lagged weekly count. The results of this model are significant at the 1 percent level. The adjusted R-squared of this model is 0.1890, this is higher compared to Model 1, indicating that including the lagged variable improves the model's explanatory power.

Model 3, the last model for trading volume, includes the weekly count of articles, the lagged weekly count of articles, and a control variable for market capitalization (market cap). This control variable is categorical, as explained in the methods section of this thesis. As shown in Table 1, Column 3, the coefficients for the weekly count of articles and the lagged weekly count are 14,925.44 and 14,919.1, respectively, which are lower than in the first two models due to the inclusion of the market cap variable. The control variable for market cap has different coefficients for each group: the coefficient for small market cap is negative, indicating a negative relationship between market cap and trading volume; the coefficient for medium market cap is not significant at the 5 percent level and thus cannot be interpreted; and the coefficient for large market cap is also negative, indicating a negative relationship between market cap and trading volume. The adjusted R-squared of Model 3 is 0.2368, meaning 23.68 percent of the variance in the data is explained by this model. Given that Model 3 has the highest adjusted R-squared and all but one of the variables are significant, it is considered the best model for explaining the data.

These findings, as detailed in Table 1, suggest that media coverage, both current and lagged, positively influences trading volume. The negative coefficients for market cap indicate that larger companies might attract less incremental trading volume per article compared to smaller companies.

Table 1: Regression of trading volume and weekly news item count

	Model 1	Model 2	Model 3
Weekly count	245566.4**(0.000)	161627.4**(0.000)	14925.44**(0.000)
Lagged count		127881.8**(0.000)	14919.1**(0.000)
Market cap 1			-1.16e+7**(0.000)
Market cap 2			-2306414(0.050)
Market cap 3			-1.04e+7**(0.000)
Constant	9614205**(0.000)	8512378**(0.000)	1.34e+7**(0.000)
Observations	2288	2266	2266
Adj. R-squared	0.1635	```````	0.2368

Note. Standard errors are in parentheses; *p < 0.05, **p < 0.01

Formula of model 1: Trading volume = $\beta 0 + \beta 1 * Weeklycount + \varepsilon$

Formula of model 2: Trading volume = $\beta 0 + \beta 1 * Weeklycount + \beta 2 * Weeklycount\ -1 + \varepsilon$

Formula of model 3: Trading volume = $\beta 0 + \beta 1 * Weeklycount + \beta 2 * Weeklycount\ -1 + \beta 3 * Marketcap + \varepsilon$

Volatility

Besides the calculations on the effects of frequency of news items about a certain stock on stock volume, the data and techniques were also used to model stocks volatility. The results of these calculations are given in Table 2. The results of the regression analysis on the effect of weekly article count on volatility are presented here. The hypothesis stated earlier in this thesis was that the weekly article count has a positive impact on a stock's volatility. Starting with the most basic model for volatility, Model 4, it is found that weekly article count has a negative effect on a stock's volatility, as indicated by the negative coefficient of -0.0002315 for the weekly article count, as shown in Table 2, Column 1. This means that as the frequency of news items increases, the volatility of a stock decreases. These results are significant at the 1 percent level. However, the adjusted R-squared of 0.0031is very low, indicating that the model can only explain a small fraction of the variance in the data.

The second model, Model 5, measures the effect on volatility. As shown in Table 2, Column 2, this model includes the control variable for market capitalization. This model also indicates that the weekly count of articles has a negative impact on volatility, evidenced by a coefficient of -0.0002724. The control variable for market capitalization in the first two groups, small and medium market cap, is not significant at the 5 percent level and thus cannot be interpreted. However, the control variable for the largest market cap group shows a coefficient of 0.0260549, suggesting that when a company has a market capitalization exceeding $100 billion, the annual volatility increases by 2.60 percent. While the adjusted R-squared of this model remains relatively low, with 0.0053, it has improved compared to Model 4, and at least one of the groups in the control variable is significant. Therefore, Model 5 is considered the best at explaining the data for the volatility.

These results, detailed in Table 2, show that while the number of articles published has a slight negative impact on volatility, this effect is minimal. Additionally, the significant positive relationship between market cap of large companies and volatility indicates that larger companies experience higher volatility. However, the low adjusted R-squared value indicates that this model only explains a small fraction of the variance in the volatility.

Table 2: Regression of annual volatility and weekly news item count

	Model 4	Model 5
Weekly count	-0.0002315**(0.004)	-0.0002724**(0.002)
Market cap 1		0.0156299(0.051)
Market cap 2		0.088288(0.293)
Market cap 3		0.0260549**(0.009)
Constant	0.317907**(0.000)	0.3086941**(0.000)
Observations	2288	2288
Adj. R-squared	0.0031	0.0053

Note. Standard errors are in parentheses; *p < 0.05, **p <0.01

Formula of model 1: Annual volatility = $\beta 0 + \beta 1 * Weeklycount + \varepsilon$

Formula of model 2: Annual volatility = $\beta 0 + \beta 1 * Weeklycount + \beta 2 * Marketcap + \varepsilon$

Discussion

I studied the influence of the frequency of news items on the AEX index registered stock's trading volume and volatility. In conclusion, positive correlation was found between number of news articles and trading volume, negative correlation was found between number of news articles and volatility of a stock. For this study I used a dataset from Factiva. However, this dataset does not contain the weekly count of news items for the specific period researched in this thesis. Consequently, the headlines had to be manually converted into a weekly count. This process involved over 55000 different headlines, the method that was used to extract the data was labor intensive and prone to potential errors. Despite this, the large number of data points in the dataset should mitigate any inaccuracies, ensuring the overall reliability of the results.

This dataset consists of 2288 data points providing a substantial number of observations, with 2266 observations were used for building the different models. However, it includes only 22 big companies traded on the AEX stock exchange. Consequently, the results may not be generalizable to smaller companies or those not traded on the Dutch stock exchange. Therefore, the external validity of this study may be limited.

Furthermore, the adjusted R-squared for the models predicting the annualized volatility of the stocks were low, indicating that the model in this paper only explained a fraction of the variance in the data. This outcome was anticipated, since the volatility of a stock is influenced by numerous factors and the weekly count of articles may not be sufficiently specific variable to explain the volatility of a stock on its own. Using daily data might capture the direct effects of news articles and more effectively, potentially increasing the R-squared values and providing a clearer understanding of the impact of media coverage on stock volatility.

In the regression analysis of the count of articles and the trading volume, there may be endogeneity. Higher trading volumes could lead to more articles being written about a company, as increased interest in a company might prompt more coverage by authors. This potential reverse causality could bias the results, suggesting that the relationship between article count and trading volume is not purely unidirectional.

The only characteristic controlled for in this paper was market cap. The companies studied have different characteristics, therefore in future research should incorporate additional control variables, such as the sentiment of the news article or past returns. Including these variables could increase the adjusted R-squared and improve the model's ability to explain the data more accurately.

Due to the use of regression analysis, this thesis cannot establish a causal relationship between the variables. To determine causality, future research should be considered with different models beyond the Ordinary Least Squares (OLS) regression model.

Conclusion

Findings suggest a significant positive relationship between the number of news articles published in a week and the trading volume during that week. Although it is not proven that it is a causal relationship, a correlation is evident in the research period of the years 2022 and 2023. Additionally, it is determined that there is a small correlation between the number of articles published and the volatility, with the model explaining only a minor part of the variance in the dataset. These findings indicate that media coverage can influence investor behavior, potentially leading to increased trading activity. However, the limited explanatory power of the model for volatility suggests that other factors also play a significant role in determining stock price fluctuations. Future research should incorporate different models and additional control variables, such as sentiment analysis of articles and past returns, to enhance the understanding of the variables influencing the stock market. I recommend further research using different models and additional control variables to better understand the factors influencing the stock market. Nevertheless, this paper may give investors and policymakers a better understanding of the relationship between the news that is published and the stock market. This will help investors make more informed decisions and assist policymakers in developing further regulations considering the impact of media on the stock market.

Reference list

Atkins, A., Niranjan, M., & Gerding, E. (2018). Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science, 4*(2), 120-137. https://doi.org/10.1016/j.jfds.2018.02.002

Barber, B. M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. Review of Financial Studies, 21(2), 785–818. doi: 10.1093/rfs/hhm079

Byström, H. N. E. (2016). Language, news and volatility. *Journal of International Financial Markets, Institutions & Money, 42*, 79-96. *doi: 10.1016/j.intfin.2016.03.002*

CompaniesMarketCap. (2024). *Companies ranked by market cap*. Retrieved June 22, 2024, from https://companiesmarketcap.com/

Ding, X., & Hou, W. (2015). Retail investor attention and stock liquidity. *Journal of International Financial Markets, Institutions & Money, 37*, 12-26. https://doi.org/10.1016/j.intfin.2015.04.003

Engelberg, J., & Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance, 66*(1), 67-97. https://doi.org/10.1111/j.1540-6261.2010.01626.x

Vlastakis, N., & Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking & Finance, 36*(6), 1808-1821. *https://doi.org/10.1016/j.jbankfin.2012.02.007*

Appendix

Table 1 Companies in the AEX index that were included in this paper.

1. Adyen
2. Aegon
3. Ahold Delhaize
4. Akzo Nobel
5. ArcelorMittal
6. ASM International
7. ASML Holding
8. BE Semiconductor
9. Heineken
10. IMCD
11. ING Group
12. KPN
13. NN Group
14. Philips
15. Prosus
16. Randstad NV
17. RELX
18. Royal Dutch Shell
19. Signify NV
20. Unilever
21. UMG
22. Wolters Kluwer

Table 2 Companies of the AEX index and their market cap (in billions of US dollar)

Company name	Market cap (in billion USD)
Adyen	83.03
Aegon	9.99
Ahold Delhaize	35.52
Akzo Nobel	20.49
ArcelorMittal	29.04
ASM International	21.35
ASML Holding	325.22
BE Semiconductor	6.56
Heineken	64.45
IMCD	12.54
ING Group	52.33
KPN	12.76
NN Group	16.48
Philips	32.28
Prosus	261.92
Randstad NV	12.50
RELX	62.72
Royal Dutch Shell	89.76
Signify NV	5.72
Unilever	50.87
UMG	136.55
Wolters Kluwer	30.22

This table shows the different market caps at the beginning of 2022. The data is obtained from CompaniesMarketCap (2024).