Stock Price Prediction using Financial News Sentiment Analysis with Pre-trained Large Language Models

Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Economics & Business
Specialization: Financial Economics

Author:	Aydin Aghaliyev
Student number:	617229
Supervisor:	Ruben de Bliek
Second reader:	Clint Howard
Finish date:	August 15, 2024

Abstract

Financial news are known to affect asset prices. In this paper we attempt to extract the sentiment information from financial news and use them to predict stock prices and create a trading strategy based on this. This is done using a recently developed Large Language Model FinBert, which has been shown to be performant for sentiment analysis tasks. Our analysis shows that the sentiment values calculated using FinBert are positively correlated with stock returns. Furthermore, the trading strategy built on the calculated sentiment values shows encouraging results. However, our research also highlights the shortcomings of this approach and offers possible directions for future improvements.

Contents

List of Tables

List of Figures

1 Introduction

The fact that financial news affect the asset prices is a trivial observation. Markets react both to general market-wide news, such as announcements about monetary policy change (Kim et al., 2004), and to asset specific news, such as earnings reports and acquisition announcements (Li et al., 2014). A recent example of such news is the Alaska Airlines Flight 1282 incident, where technical problems with Boeing planes revealed during this incident caused the company's shares to lose 9% of their value (Topham & correspondent, 2024). According to the idea of the fully efficient financial markets, such news should immediately be incorporated into the stock price. However, due to real world frictions such price changes are often not instant, though still quite fast thanks to the modern high frequency algorithmic trading techniques. Therefore, an algorithm that can quickly extract information from such news articles could be effective in creating competitive short-term investment strategies. The recent emergence of Large Language Models (LLMs) that can effectively quantify the information given in human written text creates opportunities for potentially more accurate methods of news information extraction. This extracted information can then be used in algorithmic trading applications in order to take advantage of the anticipated price changes.

While previous research already tries to answer some important questions in this topic there remains more areas of research that should be looked into more in detail. One such area is the specific model used for the sentiment analysis. It's true that Lopez-Lira and Tang (2023), recognizing the importance of correct prompting for such models, tried to utilize the pre-trained context of ChatGPT by instructing it with specific prompts. Although this method produced some positive results as discussed previously, it highlights the main disadvantage of such methods: this is basically a workaround trick to be able to use the model for a different purpose than what it was originally intended for. A model that has been specifically pre-trained for sentiment analysis task can overcome this limitation, in addition to potentially being more accurate by including more specialized information for sentiment analysis purposes.

A state-of-the-art example of such specifically pre-trained LLM is FinBert, first developed by Araci (2019) in their attempt to specialize a general text generation model BERT for financial contexts. BERT is a bidirectional generative text model first introduced by Devlin et al. (2019) which has shown strong performance in text related tasks. FinBert builds upon the BERT architecture by further training and fine-tuning it using vast amounts of financial texts. In their paper Araci (2019) demonstrates the strong performance of FinBert in tasks related to understanding of context within financial texts. This makes this model a perfect starting point for building a sentiment analysis based investment strategy. In this paper we aim to measure the feasibility as well as the performance of this approach. Therefore, the main research question of this paper arises:

Can we derive an efficient investment strategy using financial news sentiment analysis with FinBert LLM?

Additionally, we would also like to test the performance of this strategy against ones derived using previous methods of sentiment analysis. Therefore, the second research question of this paper arises:

How does an investment strategy using financial news sentiment analysis with FinBert LLM compare to other news sentiment based strategies?

Building upon the work of Lopez-Lira and Tang (2023) we will use similar data sources for stock returns. The daily stock returns of companies listed on NYSE, NASDAQ, and AMEX will be obtained through the R package `tidyquant` with the sample period of year 2018 to 2024. For our sentiment analysis we will use FinBert model that has been pre-trained on the Financial PhraseBank from Malo et al. (2014), with precalculated weights provided in the software repository of FinBert (*ProsusAI/finBERT*, 2024). As our source of financial news we will use the amazing API provided by *TickerTick - the broadest stock news* (n.d.) which gives us access to financial news articles up to 10 years old with additional meta-data marked, such as the ticker of the companies being talked about in the article, which allows us to efficiently derive historical sentiment values for the companies of our interest. We first test for this calculated sentiment's predictive power using a simple linear regression. The coefficients derived from the regression will be used to examine for any correlations between sentiment and stock returns. Afterwards, to construct the portfolio we run the sentiment analysis model on the headlines each day before the market's opening time. Based on the previous day's sentiment results we construct a long-short portfolio. At the end of our sample period the resulting returns are examined to assess the performance of the algorithm. Other performance indicators such as the Sharpe ratio are also calculated to gain more insight into the results. Finally, the results are compared with those of other more conventional portfolio construction strategies.

Regarding our first research question we expect to find significant correlation between sentiment and stock returns. Therefore, we also expect to be able to construct a portfolio based on this sentiment value which will have significant positive returns. When it comes to our second research question, it is less clear what the results might be, but based on previous research on LLMs, it can be reasonably predicted that our FinBert based approach will outperform previous approaches that use more conventional sentiment analysis models. However, the magnitude of this improvement and the practical efficiency of the derived algorithm is still a subject of speculation before the full analysis is conducted.

2 Literature Review

The relationship between financial news and stock price movements has captured a considerable amount interest in financial research. Markets are known to react to both macroeconomic news and asset-specific announcements. Kim et al. (2004) demonstrated that macroeconomic news announcements significantly impact the US bond, stock, and foreign exchange markets. Similarly, Li et al. (2014) found that specific events such as earnings reports and acquisition announcements could lead to substantial price movements in individual stocks. These findings show the potential of utilizing the information present in financial news for predicting stock prices using automated processing.

Integrating behavioral finance theories with conventional financial models has become a growing interest in recent studies. Since its introduction by academics such as Kahneman and Tversky (1979), behavioral finance has focused on the psychological aspects of investor behavior, including herd behavior, loss aversion, and overconfidence. These elements frequently cause market inefficiencies, which can be taken advantage of to generate unusual profits. The increasing research on behavioral biases has led to a reassessment of conventional asset pricing models, with sentiment indicators being included as a significant explanatory factor.

In attempt to extract some of this information from financial text researchers have tried calculating the overall sentiment within the text using various machine learning techniques. A good definition of textual sentiment analysis is given by Pang and Lee (2008) in their paper. Here the authors define sentiment analysis as the task of identifying the polarity (positive, negative, or neutral) of a given piece of text, which often involves extracting opinions, emotions, and subjective expressions. This paper provided a comprehensive view of known methods of sentiment analysis.

The very first attempt at performing such analysis is done by (Pang et al., 2002) in their paper titled ”Thumbs Up? Sentiment Classification Using Machine Learning Techniques”. This paper focused on classifying movie reviews as positive or negative using machine learning methods. This was pioneering work in this field, and set stage for future research on sentiment analysis.

Researchers have also investigated possibility of using more domain specific algorithms for analyzing the sentiment of texts within a specific subject, such as financial texts. Early approaches to sentiment analysis in finance often relied on sentiment dictionaries and bag-of-words models. For example, Tetlock (2005) utilized a sentiment dictionary to analyze the tone of news articles and its impact on stock prices, finding that negative news sentiment is correlated with downward price movements. However, these methods often lacked the sophistication to capture the nuanced meanings of financial texts.

The developments in the area of machine learning and natural language processing (NLP) have resulted in advancements in the area of sentiment analysis. New techniques such as vector Machines (SVM) and Long Short-Term Memory (LSTM) networks started outperforming previous methods. Khadjeh Nassirtoussi et al. (2014) have shown the effectiveness of SVM-based models in predicting stock market trends using financial news analysis. Similarly, Chen et al. (2013) have demonstrated that LSTM based models can be used to capture temporal relationships within the news data, showing promising opportunities for usage in stock price prediction.

More recently, NLP has been revolutionized by the development of Large Language Models (LLMs). These models are cutting-edge neural network architectures that have been trained to understand and generate human-like text by training them on large amount of textual data. Such models, like ChatGPT and BERT leverage their complex architecture, specialized techniques, and the vast amounts of training data to understand the context, semantics, and syntax in a significantly more sophisticated manner than previous models. This makes them perfect for many complex applications requiring a nuanced understanding of textual contexts, including sentiment analysis.

BERT (Bidirectional Encoder Representations from Transformers) developed by Devlin et al. (2019) has marked a significant milestone in the progress of LLMs as it allows models to understand context bidirectionally. By training BERT on financial texts Araci (2019) has introduced FinBert, which has significantly improved its understanding of context for texts in the financial domain. This ability of FinBert to understand financial jargon and context has made it a valuable tool to sentiment analysis of financial news.

Despite these advancements, there are still gaps in research. The use of ChatGPT and BERT for predicting stock price movements was explored in the paper by Lopez-Lira and Tang (2023). The study showed a big potential in usage of LLMs in the area of financial investments. However, it also pointed out the limitations of such an approach. In particular a need for more model flexibility and task-specific training became evident. This suggests that while general purpose models can be effective, task-specific models like FinBert might offer opportunities for significant improvements in the performance of the sentiment analysis approach for financial applications.

The previous research demonstrates that there have been rapid developments in the area of financial sentiment analysis, especially with the introduction of LLMs. However, there is a need for further research to fully exploit the capabilities of these models. This study builds upon the FinBert model to develop and test investment strategies based on the sentiment analysis of financial news. This approach is expected to give us insights into the full power of utilization of LLMs in financial contexts and their potential in practical applications.

3 Data

3.1 Sources

Our primary dataset comes from the API provided by *TickerTick - the broadest stock news* (n.d.). This dataset has close to 8 million news headlines for all companies listed in US stock market (around 10,000 tickers) and hundreds of top startups. The data is obtained using web-scraping from 80 different top online news sources. Each row of data includes the title of the news article, the URL to the full source, the timestamp of the article publication, and the ticker of primary company being talked about in the article. Full structure of the data and one sample row can be seen in Table 1. Although the dataset includes data starting from 2013 up to 2023, only the data starting from 2018 was analyzed in this paper. This is because before 2018 the frequency of the data is not enough to calculate the news sentiment on the daily basis, which is essential for our analysis methods.

Table 1: Sample data from TickerTick

Columns	Sample
`title`	”Moser Wealth Advisors LLC Invests $14.13 Million in Apple Inc. (NASDAQ:AAPL)”
`url`	”https://www.marketbeat.com/instant-alerts/nasdaq-aapl-sec-filing-2023-11-24/”
`unix_timestamp`	1700817211
`id`	”-6541169029012568585”
`tickers_direct`	[”aapl”]
`tickers_indirect`	null
`description`	”Moser Wealth Advisors LLC acquired a new position in Apple Inc. (NASDAQ:AAPL - Free Report) in the 2nd quarter, according to its most recent 13F filing with the Securities & Exchange Commission. The firm acquired 72,822 shares of the iPhone maker’s stock, valued at approximately $14,125,000. Ap”

To assess the performance of the news sentiment analysis, we use stock prices obtained using the R package `tidyquant`, which in turn uses Yahoo Finance as its primary source (Dancho & Vaughan, 2024). We will use daily adjusted closing prices for each of the stocks of interest. This daily data will be utilized in both the predictive linear regression assessment and the long-short strategy methods.

3.2 Cleanup

As TickerTick only provides raw web-scrape data, it needs to be cleaned up before using it for our analysis. Furthermore, only the parts of the dataset that have sufficient frequency of data to perform a daily analysis must be kept. Additionally, the data columns that will not be used in our analysis were dropped. To achieve all these goals 3 main filtering steps were taken:

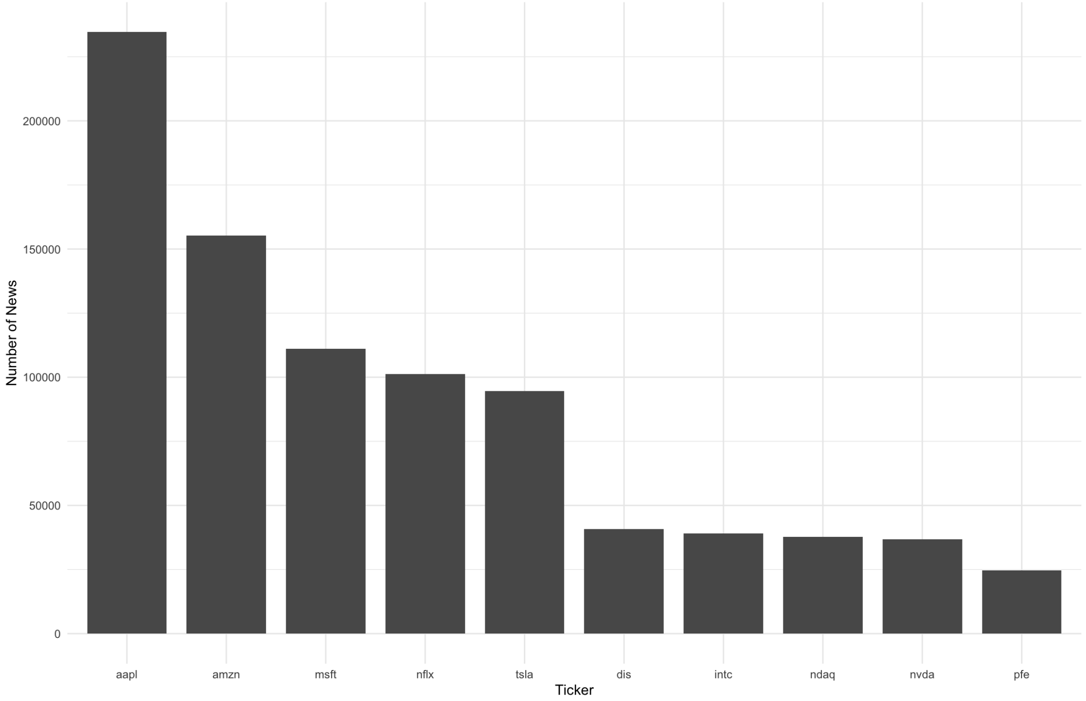
1. Only data columns for `title, url, unix_timestamp,` and `ticker` were kept.

2. Only data points with single direct ticker were kept. This was done so that the sentiment values can be clearly applied to represent a single stock.

3. Only news for the 10 most popular companies were kept, measured by the amount of news articles in the dataset. The list of the 10 company tickers in descending order is: `aapl, amzn, msft, nflx, tsla, dis, intc, ndaq, nvda, pfe`. This step removes the companies with infrequent data and allows us to focus on companies with most observations to reach more significant results.

This resulted in 10 data tables, one per each company, with number of observations from about 230,000 for `appl` to 24,000 for `pfe`. Figure 1 shows the number of news for all companies.

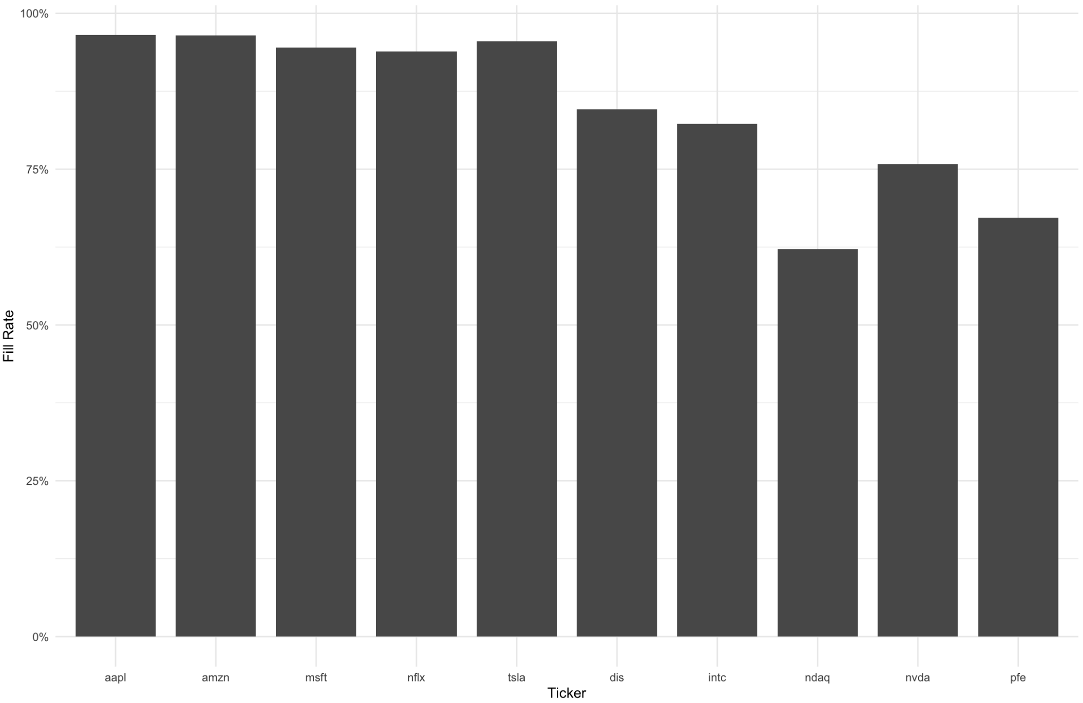
Figure 1: Bar Graph of Number of News per Company



However, the news articles are not distributed uniformly across the whole time period of interest.

Therefore, there are no news articles present for some of the trading days in the year. In this paper we call the percentage of days of the year that have at least one news article present the fill rate. Ideally, we would have a fill rate of $\sim$100% for all the listed companies. However, we do not get this number for our data, although most of the companies come quite close. Figure 2 demonstrates the fill rate for the companies in our dataset.

Figure 2: Bar Graph of Fill Rate per Company



To see the exact distribution of the amount of news articles throughout the time period per company refer to the figure 5 in appendix B.

The resulting data has the news headline, the timestamp of the news publication and a ticker of the company being primarily spoken about for each news article in our dataset. This is an ideal format to pass on the FinBert as input to calculate the sentiment values. The timestamp and ticker values let us organize the sentiment data as a panel, giving us an opportunity to use it as an independent variable in both linear regressions and investment algorithms.

4 Methodology

4.1 FinBert

To assess the recent developments in the area of natural language processing for finance we will be using a large language model to perform our analysis. This will contrast with previous researchers methods, such as their usage of sentiment dictionaries and bag-of-words methods. For this purpose we have chosen FinBert, first introduced by Araci (2019), as the model of our interest. At its core, FinBert utilizes ageneral LLM architecture names BERT. BERT is consists of multiple layers of bidirectional transformers. These layers give the model the ability to effectively quantify natural language text. To achieve this the model is first trained on a vast corpus of general text, which enables it to understand lexical and grammatical meanings of sentences. The resulting model is further trained on a large corpus of financial text, such as financial statements and news. After this step the model adapts to the specific financial terminology and phrasing of the financial texts. Final training step involves fine-tuning the model. This is done using datasets with specifically labeled data for financial natural language processing tasks. This step improves the model's accuracy and optimizes its performance. To finalize the architecture of FinBert a classification layer is added, which extracts the stored sentiment value of input sentences inside the model to determine the sentiment score.

These steps produce a model that has some advantages over both the previous non-LLM models and general text LLM models. The central advantage of this approach is that FinBert is tailored specifically for financial texts, making it more accurate in understanding financial jargon and the sentence structure of financial texts. Indeed, Araci (2019) in their paper finds that FinBert outperforms previous sentiment analysis models, even ones previously considered more performant such as LSTM. However, this paper couldn't conclusively state whether the additional finance specific training significantly improved the performance of the model in practical tasks. For this reason, the paper suggests that testing FinBert directly on market return data, similar to what is being investigated in this paper, might be needed to yield more conclusive results.

4.2 Sentiment Analysis

To perform the sentiment analysis all the news article headlines have been passed as the input of FinBert. The output is one sentiment score value per headline. This score is a value in the $[-1:1]$ range, with -1 representing very negative sentiment, 0 neutral sentiment, and 1 very positive sentiment. Every day in our time period of interest is assigned a value representing the mean of sentiment scores of all news articles that have been published on that day. In case there are no articles published on a specific day, it is assigned a value of 0. The result is panel data with time series of daily sentiment score for each of the 10 companies. Table 2 displays the mean and variance of the daily sentiment scores per company. For equations in the paper this variable will be represented as $Sentiment_{i,t}$, where it represents the sentiment value of i-th company at time point t.

Table 2: Descriptive Statistics of the Daily Sentiment Scores

The asterisks indicate the result of t-test with H_a: true mean is not equal to 0. ***, **, and * mean p-values less than 0.01, 0.05, and 0.1 respectively. No asterisks mean insignificant result.

Ticker	Mean	Standard Deviation	Max	Min	Observations
aapl	-0.013***	0.1510972	-0.9505135	0.9182228	1484
amzn	-0.022***	0.1517809	-0.9608383	0.9327924	1484
msft	0.022***	0.1710038	-0.9487871	0.9023985	1484
nflx	0.0028	0.1633279	-0.9502887	0.9316728	1484
tsla	-0.015***	0.1794122	-0.9473629	0.8758304	1484
dis	-0.029***	0.2083443	-0.9644195	0.8874178	1484
intc	0.065***	0.2352224	-0.9577429	0.903126	1484
ndaq	0.048***	0.1281475	-0.9572893	0.921414	1484
nvda	0.05***	0.2115265	-0.9599633	0.9350883	1484
pfe	0.1***	0.2424687	-0.9643098	0.9352533	1484

4.3 Analyzing the predictive power of the calculated sentiment

4.3.1 Linear Regression Analysis

We are going to base our a linear regression model on the one used by Lopez-Lira and Tang (2023) as shown in Equation 1. However, as we want to examine the effect of historical sentiment values at every point in time, we will add lags of $Sentiment_{i,t}$ as well. The resulting model with K number of lags is described by the equation 2.

$$r_{i,t+1} = a_i + b_t + \beta \cdot Sentiment_{i,t} + \varepsilon_{i,t+1} \tag{1}$$

$$r_{i,t+1} = a_i + b_t + \sum_{k=0}^{K} \beta_k \cdot Sentiment_{i,t-k} + \varepsilon_{i,t+1} \tag{2}$$

Here the dependent variable $r_{i,t+1}$ is the stock return of the i-th company for the next day. $Sentiment_{i,t-k}$ represents the calculated sentiment values of i-th company from k days in the past. a_i and b_t represent firm and time fixed effects respectively. These will account for observable or unobservable time-invariant company characteristics and time-specific characteristics that might influence the stock returns. Standard errors are clustered by both by date and company.

However, a test for non-stationarity of the data has to be performed before estimating the model and the necessary adjustments have to be made in case it is. First step is to plot the returns and the sentiment scores to visually look for signs of non-stationarity. Both the graphs for stock returns and sentiment scores are shown in Figures 6 and 7 in appendix B respectively. Afterwards, as our data is in panel form, we will perform a Levin-Lin-Chu Unit-Root Test (Levin, Lin, & James Chu, 2002) on both the stock return and sentiment score variables to determine if the data is non-stationary.

To determine the optimal value for K Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) will be used. Models with different values for K will be fitted and their AIC and BIC values will be compared to determine this number. Models up to $K = 30$ will be examined to cover the maximum scope of a month.

Additionally, we will perform a Breusch-Pagan Test on the selected model to test for heteroskedasticity in our data. In case heteroskedasticity is detected we will use robust standard errors in our estimations.

Finally, with the appropriate parameters and estimation methods determined, the model described by Equation 2 is going to be estimated. The resulting model will also be utilized for further analysis by including it within one of the investment strategy algorithms.

It is also useful to assess whether the addition of fixed effects to our model significantly improves it. To do this, we will estimate the model using pooled data and perform a Likelihood Ratio test between the fixed effects and pooled variants.

4.3.2 Investment Portfolio Strategy Analysis

To further analyze the effectiveness of the calculated sentiment for investment purposes, we also implement a long-short strategy based on the sentiment values. The algorithm of this strategy is described below:

On each day, for all stocks analyzed if a stock has a sentiment value defined on that day enter a long or a short position based on the sentiment value and close the position on the next day. Enter a long position if the sentiment value of the stock is larger than $\mu + \theta \cdot \sigma$ or short the stock if its sentiment value is less than $\mu - \theta \cdot \sigma$. Otherwise, do not perform any trades on the stock. Here θ is a number within the range $[0 : 1]$ and μ and σ are the mean and the standard variation of the sentiment values of all stocks with the values defined on the current day accordingly. Multiple values of θ will be tested to achieve the highest returns and Sharpe ratio. We will call this strategy the **$\boldsymbol{\theta}$-Threshold strategy**, where θ will be replaced with the specific value used.

The strategy is rebalanced daily. The investigation of this algorithm gives an opportunity to treat the calculated sentiment as a raw indicator of how well the stock is doing. To quantify the performance of the algorithm we will calculate the cumulative returns and the Sharpe ratio of the resulting portfolio. We will perform this analysis for several values of θ, as well as calculate the same metrics for two other portfolios for comparison: an S&P 500 portfolio, and an equal weighted portfolio. The performance of the algorithm will give us insights into the general usefulness of the calculated sentiment for investment purposes.

5 Results

5.1 Linear Regression Analysis

Before we estimate the model from Equation 2, we need to perform a number of tests. We first test for non-stationarity using the Levin-Lin-Chu Unit-Root Test. The results of this test on both the dependent and independent variables is demonstrated in the Table 3. As the p-values of the test for both variables is 0, we can confidently assume that none of the variables in our model are non-stationary. This means our model needs no modifications to combat non-stationarity.

Table 3: Results of the Levin-Lin-Chu Unit-Root Test

Variable	Z-value	p-value
$r_{i,t+1}$	-124.03	0.0
$Sentiment_{i,t}$	-53.552	0.0

Next we calculate the AIC and BIC for all models with values of K up to 30 to determine how many of lags is best suited for our model. The full results of this analysis can be seen in Table 9 of Appendix A. This analysis shows that AIC is minimized when $K = 1$, while BIC is minimized when $K = 0$. Therefore, we will be estimating two models: one with no lags and one with the first lag of $Sentiment_{i,t}$.

Finally, before estimating the models we determine whether we need to use robust standard errors by looking for heteroskedasticity. For this we will perform a Breusch-Pagan Test. The results of this test on our model are shown in Table 4. As the p-value is 0, this clearly indicates that there is heteroskedasticity in our data, therefore robust standard errors should be used.

Table 4: Results of the Studentized Breusch-Pagan Test

Number of Lags	BP	df	p-value
$K = 0$	19.502	1	0.0
$K = 1$	25.676	2	0.0

After taking into the account the results of the tests, we will be estimating the models with $K = 0$ and $K = 1$ with robust standard errors. The full forms of both models are described below:

Model 1 $$r_{i,t+1} = a_i + b_t + \beta Sentiment_{i,t} + \varepsilon_{i,t+1}$$

Model 2 $$r_{i,t+1} = a_i + b_t + \beta_1 \cdot Sentiment_{i,t} + \beta_2 \cdot Sentiment_{i,t-1} + \varepsilon_{i,t+1}$$

Table 5 demonstrates the results of the estimations of both models.

Table 5: Results of the Linear Regression with $K = 0$ and $K = 1$

	Dependent variable:	
	$r_{i,t+1}$	
	Model 1	Model 2
$Sentiment_{i,t}$	0.00463***	0.0046028***
	(0.001008)	(0.0010142)
$Sentiment_{i,t-1}$		0.000153
		(0.0009265)
Observations	14830	14820
R^2	0.4639	0.4641
Adjusted R^2	0.4039	0.4042
F-statistic	7.735***	7.737***

Note: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in parentheses.

Furthermore, the results of the Likelihood Ratio Tests between the models and their pooled variants are also presented in Table 6. This analysis shows that indeed the fixed effects are an important part of both models as they significantly contribute to their predictive power.

Table 6: Results of Likelihood Ratio Tests

Comparison	χ^2 statistic	df	p-value
Model 1 fe. vs pooled	9227.5	1491	0.0^{***}
Model 2 fe. vs pooled	9227.8	1490	0.0^{***}

Note: *p<0.1; **p<0.05; ***p<0.01.

As seen in Table 5, the $Sentiment_{i,t}$ variable seems to have significant coefficients with value of around 0.0046 in both Model 1 and 2. However, the first lag of this variable, $Sentiment_{i,t-1}$ does not seem to have any significant effects, meaning its inclusion in the model is not relevant. Both models show relatively right values for R^2 and significant results for the F-test. All of this shows strong evidence for the fact that the sentiment scores calculated by FinBert significantly correlate with the returns of the stocks. This already indicates the sentiment score might be a good base for development of a trading algorithm. To definitively test this idea, will conduct portfolio simulations in the next section.

5.2 Investment Portfolio Strategy Analysis

Figure 3: Cumulative Returns of the θ-Threshold Strategy over Time

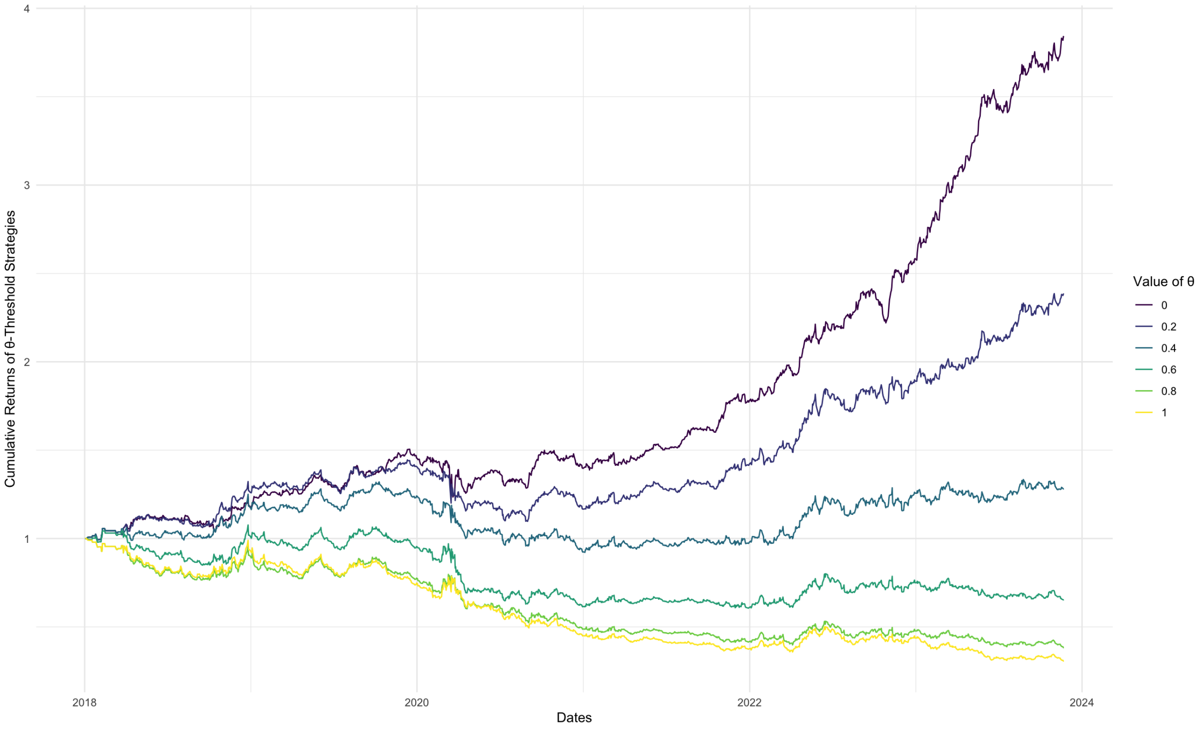


Table 7: Cumulative Returns and Annualized Sharpe ratios of θ-Threshold Strategies

Value of θ	Cumulative Return	Annualized Sharpe ratio
0	3.8367	1.8612
0.2	2.3816	1.0941
0.4	1.2820	0.3394
0.6	0.6544	-0.2814
0.8	0.3835	-0.6719
1	0.3074	-0.7753

After the regression analysis showed positive results, we move on to construct the portfolio using the strategy described above as the θ-Threshold strategy. We used 6 different values for θ and evaluated the cumulative returns and Sharpe ratio of the strategies with corresponding values. These results can be seen in Figure 3 and Table 7. This shows that the strategy with 0 for the value of θ performs the best with good cumulative returns and Sharpe ratio. As this strategy is substantially better performing than those with higher values of θ, we will only consider this one for further comparisons.

Figure 4: Cumulative Returns of Different Portfolios

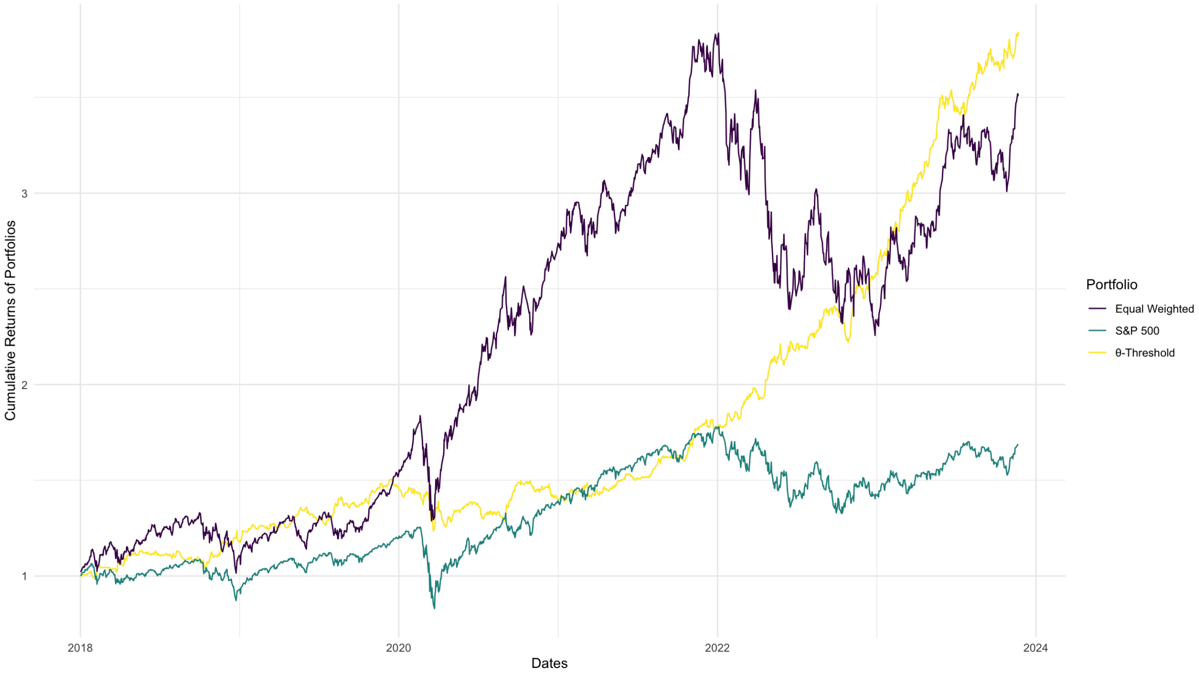


Table 8: Cumulative Returns and Annualized Sharpe ratios of Different Portfolios

Portfolio	Cumulative Return	Annualized Sharpe ratio
Equal Weighted	3.5125	0.9209
S&P 500	1.6834	0.5298
θ-Threshold	3.8367	1.8612

Next, we compare the results of the θ-Threshold portfolio with two other portfolios: a portfolio constructed by holding the stocks with equal weights and an S&P 500 portfolio. The cumulative returns and Sharpe ratios of these portfolios are shown and compared in Figure 4 and Table 8. This demonstrates that our strategy clearly outperforms S&P 500, both in terms of returns and Sharpe ratio. However, the picture is less clear when comparing to the equally weighted portfolio. The total returns of both strategies are quite close, although our strategy still performs better. But looking at the Sharpe ratios, our strategy shows much better results, having double the Sharpe ratio of the equally weighted portfolio. This shows that, although our strategy might not have much better returns, it has a substantially better risk profile, making it potentially better performing while having significantly less value at risk.

6 Conclusion

This thesis focuses on exploring the potential of leveraging the power of FinBert, a pre-trained large language model tailored for financial sentiment analysis, to predict stock prices and develop effective investment strategies. The primary subject of discussion was whether the sentiment values calculated by FinBert by analyzing financial news could be used as a reliable indicator for stock price prediction and serve as a basis for a profitable trading strategy.

The first of our key findings was that the future stock returns are positively correlated with the calculated sentiment values. Positive return values are generally associated with positive sentiment values. This was determined by estimating a linear regression with fixed effects on future returns using the sentiment value as an independent variable. This supports our hypothesis that the sentiment values extracted from the financial news is a meaningful indicator of stock market behavior.

Our second finding was about the proposed trading strategy that uses the calculated sentiment values as its main indicator. This strategy would generally go short with the stocks with lower sentiment values, while going long with stocks with higher sentiment values. Over the testing period this strategy showed generally positive returns, suggesting sentiment analysis could indeed be used to inform investment decisions. However, more mixed performance of the strategy over some parts of the investigated time period also highlighted some limitations and possible future improvement domains of the model.

6.1 Limitations

While the findings are encouraging, some limitations of this approach have become evident. Firstly, the analysis only focuses on a rather small set of 10 stocks for its analysis. Although such focus is widespread in the related literature, a wider analysis could significantly improve the quality of the research. However, this limitation is partly caused by the nature of the data, as a source of high quality financial news which covers a large time period well enough to be investigated using our approach is hard to get a hold of. Secondly, our model's results could be influenced by the noise in the news data, especially factors such as sensationalist reporting and irrelevant information. Lastly, the study didn't account for market liquidity and transaction costs, which could affect the profitability of the proposed investment strategy in real-world scenarios.

6.2 Future Research

There are several directions future research could focus on. One such direction is inclusion of other variables, such as data from social media sources and macroeconomic indicators. These additions could greatly improve the predictive power of our model. Furthermore, using other machine learning techniques beyond linear regressions, such as deep learning models or ensemble methods could provide more sophisticated models and increase the predictive power.

In conclusion, this thesis has shown that financial news sentiment analysis using FinBert can indeed provide good insights for predicting stock price movements and be used as a base for investment strategies. While there are challenges and limitations, the promising results provide a strong foundation for future exploration in this emerging field of financial research.

References

Araci, D. (2019, August). *FinBERT: Financial Sentiment Analysis with Pre-trained Language Models.* arXiv. Retrieved 2024-06-23, from `http://arxiv.org/abs/1908.10063` (arXiv:1908.10063 [cs])

Chen, H., De, P., Hu, Y. J., & Hwang, B.-H. (2013, December). *Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media* [SSRN Scholarly Paper]. Rochester, NY. Retrieved 2024-07-13, from `https://papers.ssrn.com/abstract=1807265` doi: 10.2139/ssrn.1807265

Dancho, M., & Vaughan, D. (2024). *tidyquant: Tidy Quantitative Financial Analysis.* Retrieved from `https://business-science.github.io/tidyquant/`

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, May). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.* arXiv. Retrieved 2024-07-13, from `http://arxiv.org/abs/1810.04805` (arXiv:1810.04805 [cs]) doi: 10.48550/arXiv.1810.04805

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. Retrieved 2024-08-14, from `https://www.jstor.org/stable/1914185` (Publisher: [Wiley, Econometric Society]) doi: 10.2307/1914185

Khadjeh Nassirtoussi, A., Aghabozorgi, S., Ying Wah, T., & Ngo, D. C. L. (2014, November). Text mining for market prediction: A systematic review. *Expert Systems with Applications*, *41*(16), 7653–7670. Retrieved 2024-07-13, from `https://linkinghub.elsevier.com/retrieve/pii/S0957417414003455` doi: 10.1016/j.eswa.2014.06.009

Kim, S.-J., McKenzie, M. D., & Faff, R. W. (2004, July). Macroeconomic news announcements and the role of expectations: evidence for US bond, stock and foreign exchange markets. *Journal of Multinational Financial Management*, *14*(3), 217–232. Retrieved 2024-04-05, from `https://www.sciencedirect.com/science/article/pii/S1042444X03000501` doi: 10.1016/j.mulfin.2003.02.001

Levin, A., Lin, C.-F., & James Chu, C.-S. (2002, May). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, *108*(1), 1–24. Retrieved 2024-07-27, from `https://www.sciencedirect.com/science/article/pii/S0304407601000987` doi: 10.1016/S0304-4076(01)00098-7

Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., & Chen, Y. (2014, September). The effect of news and public mood on stock movements. *Information Sciences*, *278*, 826–840. Retrieved 2024-04-07, from `https://www.sciencedirect.com/science/article/pii/S0020025514003879` doi: 10.1016/j.ins.2014.03.096

Lopez-Lira, A., & Tang, Y. (2023, September). *Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models.* arXiv. Retrieved 2024-04-07, from `http://arxiv.org/abs/2304.07619` (arXiv:2304.07619 [cs, q-fin]) doi: 10.48550/arXiv.2304.07619

Malo, P., Sinha, A., Takala, P., Korhonen, P., & Wallenius, J. (2014, April). Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts. *Journal of the American Society for Information Science and Technology.* doi: 10.1002/asi.23062

Pang, B., & Lee, L. (2008, January). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, *2*, 1–135. doi: 10.1561/1500000011

Pang, B., Lee, L., & Vaithyanathan, S. (2002, May). *Thumbs up? Sentiment Classification using Machine Learning Techniques.* arXiv. Retrieved 2024-08-08, from `http://arxiv.org/abs/cs/0205070` (arXiv:cs/0205070) doi: 10.48550/arXiv.cs/0205070

ProsusAI/finBERT. (2024, April). Prosus AI. Retrieved 2024-04-08, from `https://github.com/ProsusAI/finBERT` (original-date: 2019-10-30T10:20:43Z)

Tetlock, P. C. (2005, March). *Giving Content to Investor Sentiment: The Role of Media in the Stock Market* [SSRN Scholarly Paper]. Rochester, NY. Retrieved 2024-07-13, from `https://papers.ssrn.com/abstract=685145` doi: 10.2139/ssrn.685145

TickerTick - the broadest stock news. (n.d.). Retrieved 2024-06-23, from `https://tickertick.com/`

Topham, G., & correspondent, G. T. T. (2024, January). Boeing shares fall after door panel blown out of plane mid-flight. *The Guardian.* Retrieved 2024-04-05, from `https://www.theguardian.com/business/2024/jan/08/boeing-shares-fall-door-plane-mid-flight-spirit-aerosystems-737-max-9`

Appendices

The appendices include tables and figures that would have been too cumbersome to include in the main text. Appendix A displays the full results of the AIC and BIC analysis, conducted to determine the most optimal number of lags for our model. Appendix B displays three figures. The first figure shows the distribution density of news articles over time for each company in our analysis, useful to determine the optimal time period to investigate in the paper, as well as to look out for other possible irregularities in the data. The last two figures show the graphs of stock returns and sentiment values of the companies. This visual representation is a convenient tool to look for non-stationarity in the variables.

A Tables

Table 9: Results of AIC and BIC Analysis

Number of $Sentiment_{i,t}$ lags	AIC	BIC
0	-67091.83	-67053.81
1	-73292.80	-61932.82
2	-73241.38	-61882.41
3	-73181.83	-61823.87
4	-73136.47	-61779.52
5	-73078.55	-61722.61
6	-73020.44	-61665.51
7	-72964.12	-61610.20
8	-72905.81	-61552.91
9	-72846.36	-61494.47
10	-72790.81	-61439.93
11	-72730.75	-61380.89
12	-72672.40	-61323.55
13	-72615.15	-61267.32
14	-72580.69	-61233.88
15	-72528.52	-61182.73
16	-72473.60	-61128.82
17	-72441.24	-61097.48
18	-72387.47	-61044.73

19	-72331.40	-60989.68
20	-72278.58	-60937.88
21	-72224.54	-60884.87
22	-72176.02	-60837.36
23	-72122.30	-60784.67
24	-72070.53	-60733.92
25	-72016.68	-60681.09
26	-71968.29	-60633.73
27	-71922.88	-60589.34
28	-71870.09	-60537.58
29	-71811.66	-60480.18
30	-71754.76	-60424.30

B Graphs

Figure 5: News Distribution Density per Company

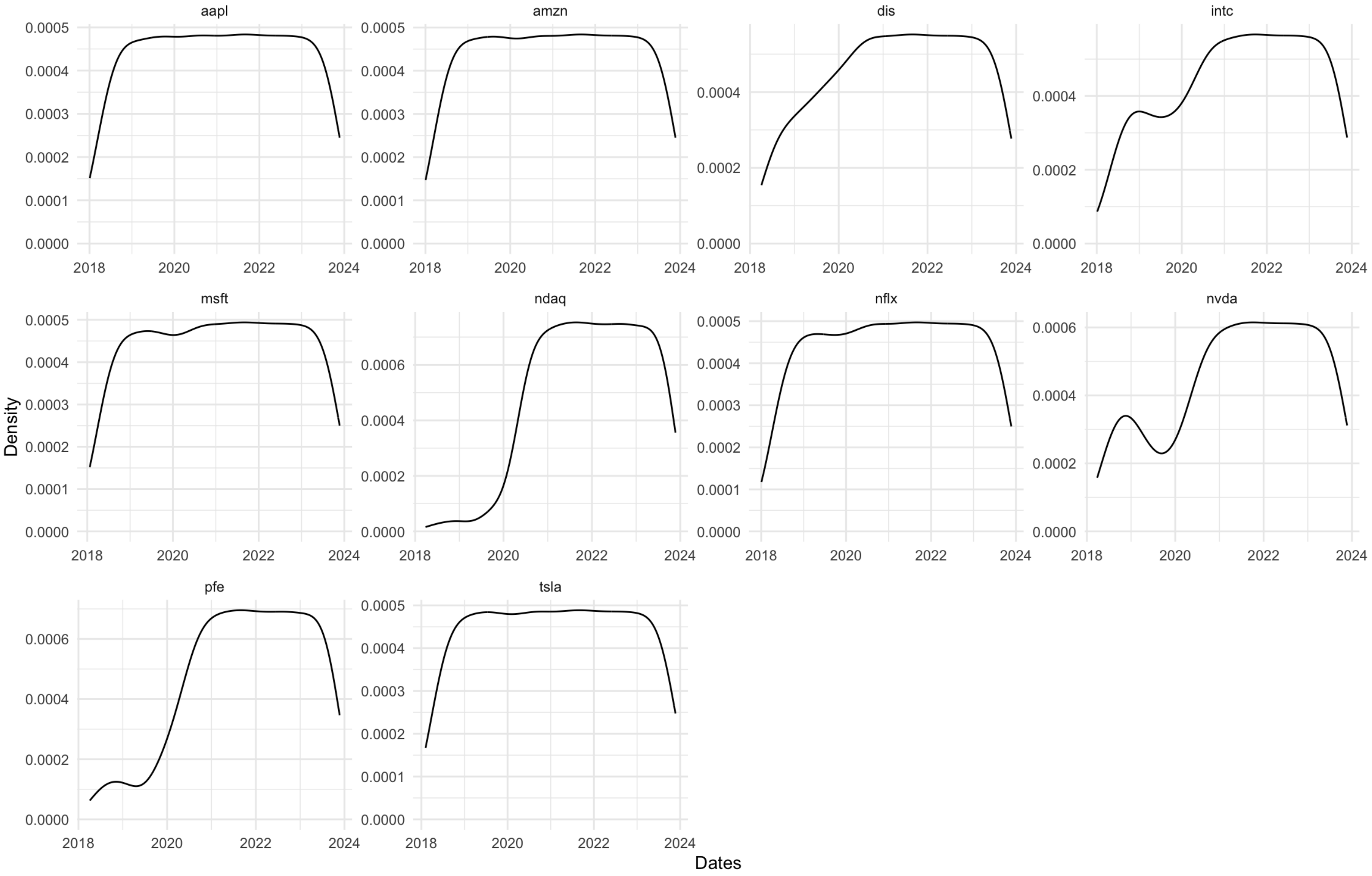


Figure 6: Stock Returns per Company

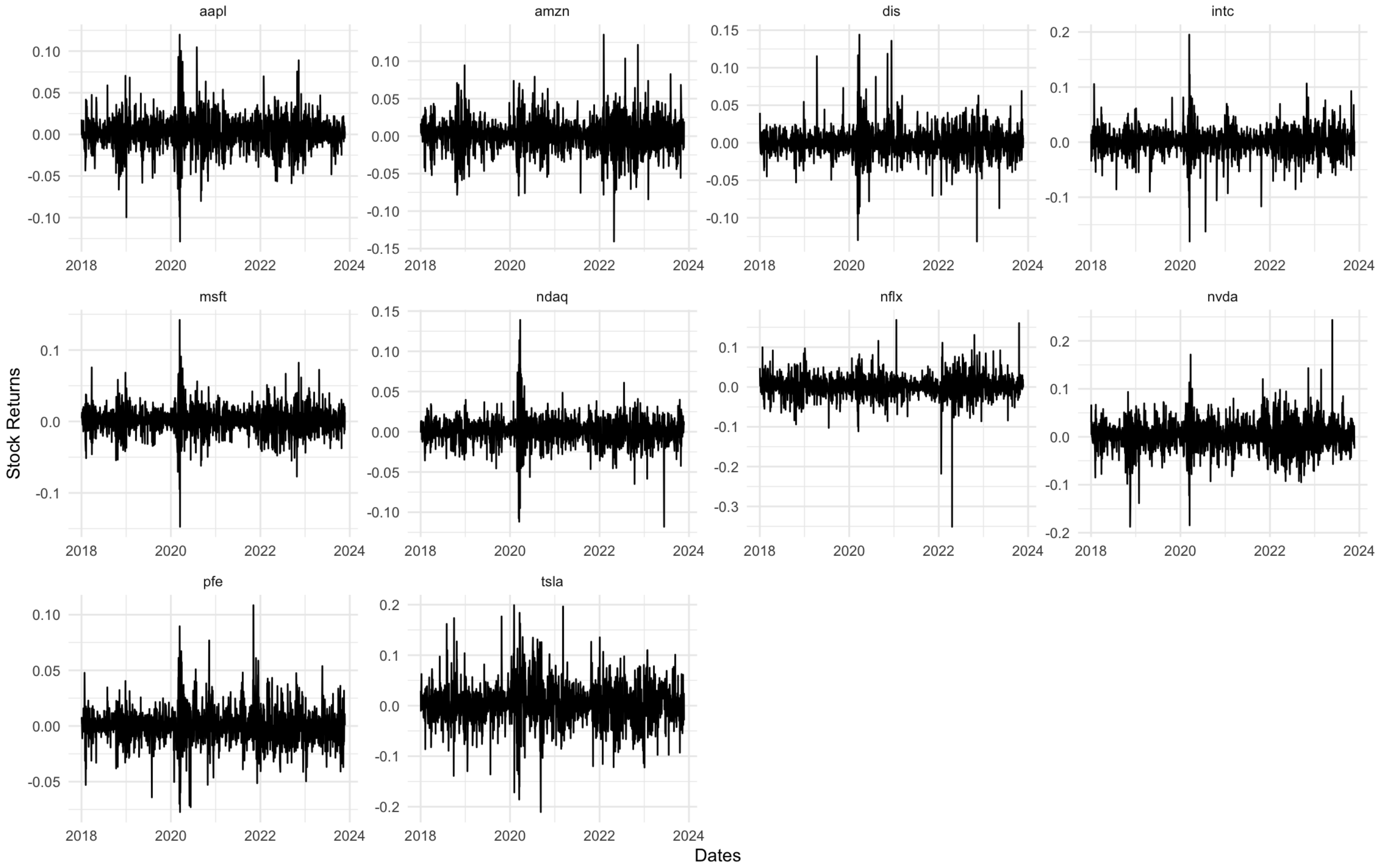
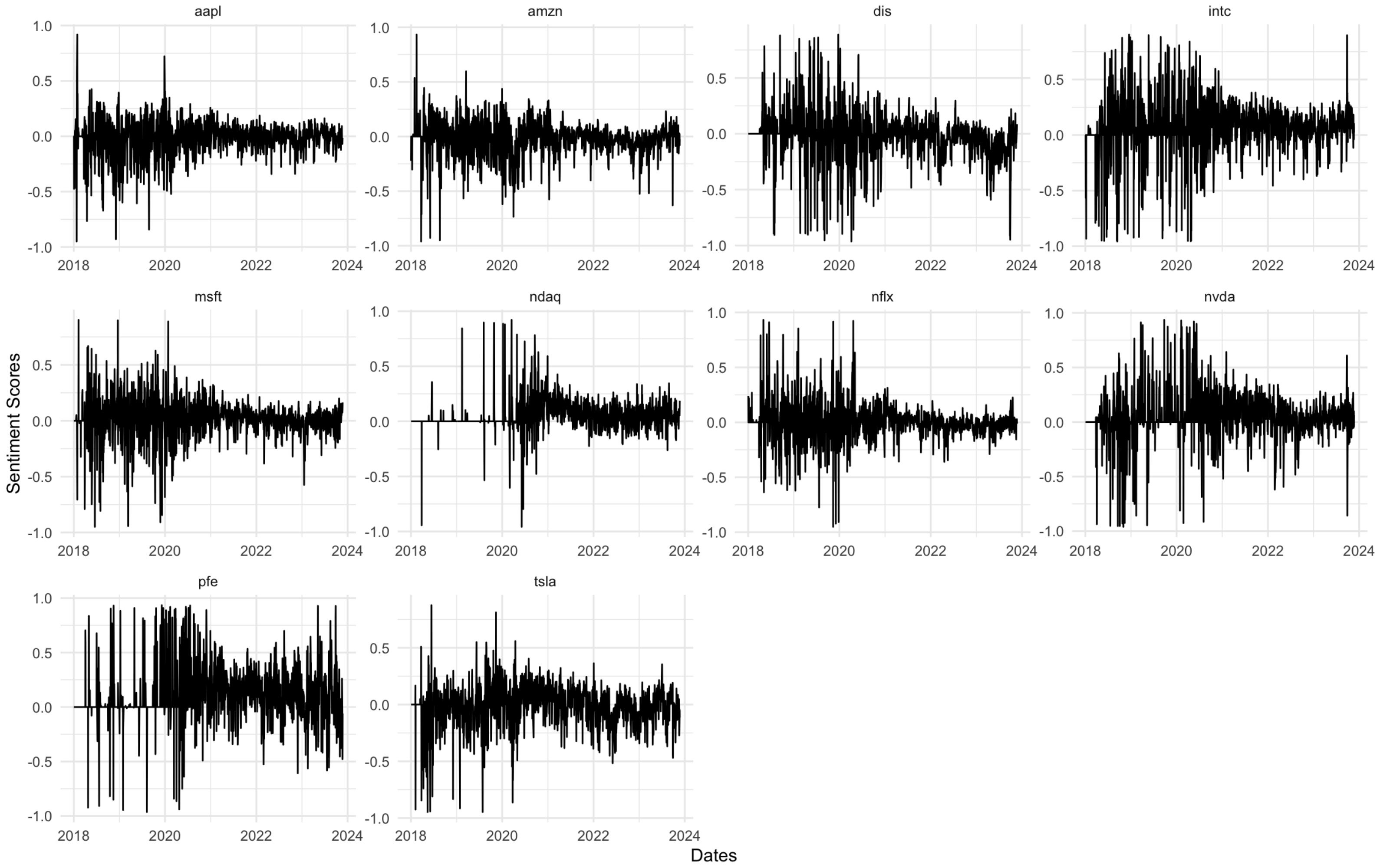


Figure 7: Sentiment Scores per Company


aapl
amzn
dis
intc
msft
ndaq
nflx
nvda
pfe
tsla
Sentiment Scores
Dates