MASTERS IN

FINANCE

MASTER´S FINAL WORK

DISSERTATION

SENTIMENT ANALYSIS OF FINANCIAL NEWS ON BLOOMBERG AND ITS IMPACT ON THE EURO STOXX 50

SOFIA KOROMPLI

OCTOBER - 2024

MASTERS IN
FINANCE

MASTER´S FINAL WORK
DISSERTATION

SENTIMENT ANALYSIS OF FINANCIAL NEWS ON BLOOMBERG AND ITS IMPACT ON THE EURO STOXX 50

SOFIA KOROMPLI

SUPERVISION:
PROFESSOR PEDRO RINO VIEIRA

OCTOBER - 2024

ABSTRACT

In today's fast-paced financial markets, where investors are flooded with huge amounts of news and data, understanding the role of market sentiment has become increasingly important. This work investigates the impact of financial news sentiment on the daily returns of the Euro Stoxx 50 index, covering the period from January 2022 to March 2024. By examining whether sentiment extracted from Bloomberg articles can predict stock returns, the study aims to shed light on how investor psychology, driven by news, interacts with market performance, especially when accounting for market volatility (VIX) and lagged returns. To explore this relationship, sentiment scores were calculated using the Loughran-McDonald Lexicon and combined with stock price data to conduct panel regressions. Both fixed effects and random effects models were considered, with the Hausman test used to select the appropriate model. Additionally, OLS regressions for each year were applied to further understand the dynamics between sentiment and returns over time. The results reveal that in 2022, sentiment had a significant positive effect on returns, suggesting that investor sentiment played a key role during periods of heightened market uncertainty. The findings indicate that while sentiment can affect market performance, its impact is largely contingent on volatility levels and broader market conditions. This study contributes to the growing body of literature on the interaction between sentiment and financial markets, emphasizing the variable nature of sentiment's influence across different market environments.

KEYWORDS: financial news; sentiment analysis; Eurostoxx 50; market efficiency; behavioral finance; stock returns

JEL CODES: G14; G41; C33; G12; G40

RESUMO

Nos mercados financeiros dinâmicos de hoje, onde os investidores são inundados com grandes quantidades de notícias e dados, compreender o papel do sentimento de mercado tornou-se cada vez mais importante. Este trabalho investiga o impacto do sentimento das notícias financeiras nos retornos diários do índice Euro Stoxx 50, cobrindo o período de janeiro de 2022 a março de 2024. Ao examinar se o sentimento extraído de artigos da Bloomberg pode prever os retornos das ações, o estudo busca esclarecer como a psicologia dos investidores, impulsionada pelas notícias, interage com o desempenho do mercado, especialmente ao levar em conta a volatilidade do mercado (VIX) e os retornos defasados. Para explorar essa relação, as pontuações de sentimento foram calculadas usando o Loughran-McDonald Lexicon e combinadas com dados de preços de ações para realizar regressões em painel. Foram considerados modelos de efeitos fixos e efeitos aleatórios, com o teste de Hausman utilizado para selecionar o modelo mais adequado. Além disso, foram aplicadas regressões OLS para cada ano, a fim de entender melhor as dinâmicas entre sentimento e retornos ao longo do tempo. Os resultados revelam que, em 2022, o sentimento teve um efeito positivo significativo nos retornos, sugerindo que o sentimento dos investidores desempenhou um papel crucial durante períodos de maior incerteza no mercado. Os achados indicam que, embora o sentimento possa afetar o desempenho do mercado, seu impacto é amplamente condicionado pelos níveis de volatilidade e pelas condições gerais do mercado. Este estudo contribui para o crescente corpo de literatura sobre a interação entre sentimento e mercados financeiros, enfatizando a natureza variável da influência do sentimento em diferentes ambientes de mercado.

Palavras-chave: notícias financeiras; análise de sentimento; Eurostoxx 50; eficiência de mercado; finanças comportamentais; retornos de ações

JEL CODES: G14; G41; C33; G12; G40

GLOSSARY

EMH – Efficient Market Hypothesis

AI – Artificial Intelligence

VIX – Volatility Index

OLS – Ordinary Least Square

RE Model – Random Effects Model

FE Model – Fixed Effects Model

STOXX – Euro Stoxx 50

NLP – Natural Language Processing

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ACKNOWLEDGMENTS

Completing this dissertation marks a significant milestone in my life, and I would like to take this opportunity to express my deepest gratitude to all those who supported me throughout this journey.

First and foremost, I would like to thank Professor Pedro Rino Vieira for his support, valuable guidance, and availability throughout the entire process. His insights and advice were essential to the completion of this work, and his encouragement helped me navigate the challenges along the way.

I also owe my deepest gratitude to my family for their constant support and for giving me the opportunity to pursue my academic goals. Their belief in me has been my greatest source of strength.

Finally, I want to thank my friends that I met during my Masters. They became more than friends - they became family. We shared unforgettable moments together, from long hours of studying to laughter, tears, and adventures while traveling. These memories will stay with me forever, and I am truly grateful for the bonds we created during this incredible journey.

1. INTRODUCTION

One of the most discussed and controversial topics in finance nowadays is stock market prediction (Jin et al., 2020). Experts have been trying to forecast the stock market and investigate which factors have an impact on it and to what extent. To achieve this goal, different theories have been developed and plenty of tools have been applied.

Living in the "Big Data era" suggests that technological advances, and particularly the sharp development of social networks and Artificial Intelligence Solutions, enables access to huge amounts of news data. Existing literature supports that integrating this public and private information into historical prices, can significantly enhance the predictability of stock market movements (Pagolu et al., 2016). In the field of finance, sentiment analysis has increasingly become a crucial tool for understanding and forecasting market reactions and investment decisions. The intersection of behavioral finance and computational linguistics, manifesting in sentiment analysis, offers sophisticated insights into market dynamics and investor behavior.

Sentiment analysis contributes to more informed and strategic investment decisions by translating subjective opinions expressed on social media and news platforms into quantifiable data (Malandri et al., 2018; Xing et al., 2019). Thus, the use of Artificial Intelligence (AI) algorithms reflects a significant shift from traditional data-centric approaches to a more nuanced understanding of market dynamics, influenced by public mood (Byrum, 2022).

News is a convenient way to grasp various aspects of the economy and how they relate to each other (Bybee et al., 2021). Financial news articles are used extensively by market participants, as they represent a key source of market information. Therefore, investors, brokers and other market agents are constantly changing their beliefs and actions, influencing not only their peers, but also stock price movements. This impact on stock performance is obtaining more and more academic attention since it cannot be appropriately interpreted by traditional financial theories.

The growth of Web 2.0 and the upward trend of social media have enhanced the spread of information and news, which is making the market more reliant on subjective opinions (Xing et al., 2018; Malandri et al., 2018). Hence, public mood impacts the stock-market indirectly through investors‘ attitude and behavior (Xing et al., 2018).

The sentiment of news articles and their impact on the stock market and price returns are being broadly explored in the field of finance (Li et al., 2014; Pagolu et al., 2016; Hasselgren et al., 2022). Nevertheless, the results of existing literature have been quite contradicting, making it hard to reach a common conclusion. Many papers focus predominantly on popular websites, as Twitter (Bollen et al., 2011; Pagolu et al., 2016), given its standard format and the convenience of APIs (Malandri et al., 2018)**,** leaving other social media platforms underexplored. Furthermore, some researchers are analyzing the US stock-market, while others focus on the European. Additionally, the delayed reaction on news releases also remains under-investigated. Despite the significant number of examinations on this subject, the correlation between investor sentiment and future returns is not clear yet.

Therefore, these gaps imply that there is potential for a more diversified approach in sentiment analysis on financial news, considering various social media platforms, geographical markets, and the timing of news releases. Such research could provide a more comprehensive understanding of the interplay between sentiment and market behavior, leading to more informed investment strategies.

The objective of the present paper is to determine the effect of sentiment of financial news, measured by a lexicon - based approach, on the performance of the Euro Stoxx 50. Based on existing literature (Fedyk, 2024) the Bloomberg terminal is selected as a relevant source to extract the financial news articles. Furthermore, in the current study the immediate impact of news sentiment on the STOXX is being explored.

The dissertation is divided into five main parts. In order to answer the research question *“How does sentiment derived from financial news announcements on Bloomberg affect the daily returns of the Euro Stoxx 50?”*, the efficient market hypothesis (Fama, 1970) as a foundational framework, but also relevant behavioral finance theories are going to be addressed in the first section. The second part of this thesis is a literature review on Sentiment Analysis in Finance. In the third part, Data and Methodology, the lexicon-based approach, based on the Loughran-McDonald Lexicon, is applied as a tool for sentiment analysis on the financial news articles gathered from Bloomberg, a highly significant platform (Fedyk, 2024), in order to calculate the sentiment scores. Heading to the statistical part, two regression models are built to test the impact of the sentiment scores on the index returns. Finally, the results are analyzed, and limitations are discussed, proposing suggestions for future research.

2. THEORETICAL FRAMEWORK

The theoretical framework of this thesis consists of an extensive Literature Review of three main concepts, the Efficient Market Hypothesis, Behavioral Finance Theories and Sentiment Analysis of Financial Data.

2.1 EFFICIENT MARKET HYPOTHESIS (EMH)

In this part, the Efficient Market Hypothesis is presented and defined, while it is shown how researchers‘ attention has shifted towards behavioral finance theories, due to anomalies persistent in the capital markets that cannot be rationalised.

2.1.1 DEFINITION OF EMH

Researchers and academics have been questioning the efficiency of financial markets for a long time. As a result, in 1960 the Efficient Market Hypothesis evolved as one of the strongest theories gaining more importance during the following three decades (Shiller, 2003; Shleifer, 2000). According to Fama (1970) and the EMH, „efficient“ markets are characterized by the presence of rational investors who are seeking to maximize their profits based on expected-utility characteristics. The hypothesis holds under the assumption that all relevant information is available for a sufficient number of investors and security prices „fully reflect“ all known information at any time (Fama, 1970; Shleifer, 2000; Malkiel, 2003). After Fama, other researchers tried to give a definition for market efficiency. In 2003, Malkiel defines efficient financial markets as markets in which investors do not have the possibility to earn above-average returns without accepting above-average risks. This is indicating that even inexperienced investors who purchase a variety of assets at market prices have the opportunity to earn a return comparable to that of qualified professionals (Malkiel, 2003).

2.1.2 INVESTOR RATIONALITY AND UTILITY THEORY

One of the key assumptions of EMH is the investor rationality, implying that all available information is fully reflected in asset prices. This is in alignment with Utility Theory, where investors maximize their expected utility by making decisions based on all known data and their individual risk preferences. Therefore, securities are valued with the net present value of their future cash-flows, discounted using their risk characteristics. This makes it impossible for any market participant to make abnormal profits and beat the market (Fama, 1970; Ţiţan, 2015). Security prices incorporate all available information immediately, since rational investors react quickly to news, by bidding up prices after good news and bidding them down after bad news. In that way, prices adjust to new levels (Shleifer, 2000).

2.1.3 MARKET EFFICIENCY IN RESPONSE TO NEWS

During his empirical work Fama (1970) introduces three forms of the efficient market hypothesis - weak, semi-strong and strong. Thereby, the three categories are addressing following questions:

a) Weak-form tests: *How well do prices reflect past market data?*

The first version of EMH supports that prices exhibit random walk and incorporate, at any moment, all the existing historical financial information. As a result, investors cannot obtain abnormal profits from investing in financial assets.

b) Semi-strong-form tests: *How quickly do security prices reflect public information announcements, including news?*

This form addresses the way markets incorporate new information into pricing. It implies that prices not just reflect all current market information but also adapt rapidly and without biases to include any other public news released. In this case, neither technical analysis nor fundamental analysis cater as tools to achieve the ideal asset allocation and therefore higher profitability, compared to investing in a random portfolio of financial assets.

c) Strong-form tests: *Do any stakeholders have private information that is not fully reflected in market prices?*

The strong form of EMH assumes that prices reflect all information in a market: past financial information (weak-form), recent public information (semi-strong-form), and any private information about a financial asset.

In 1991, Fama revised and updated the three categories. Hence, the weak-form tests are now concerned with the forecast of returns considering not only past returns, but also dividend yields, interest rates, voaltility, asset-pricing models, anomalies and seasonals. For the other two classifications the economist suggests title changes, namely from semi-strong-form tests to event studies and from strong-form tests to tests for private information.

The efficient market hypothesis argues that capital markets respond immediately to news announcements, without experiencing extended periods of overreaction or underreaction. As an extension of Fama's semi-strong-form, the two main empirical predictions of the EMH are: „the quick and accurate reaction of security prices to information", as well as „the non-information" (Shleifer, 2000). When new information about the value of a security is published, its price must quickly and accurately adjust to prevent latecomers from making profits. Prices are supposed to exhibit a moderate response to news without displaying any distinct patterns or shifts in direction following the initial reaction. Moreover, in the case of absent information, the price of an asset is expected to remain steady and unaffected by changes in supply and demand (Shleifer, 2000). Yet, Fama (1998) is recognising market anomalies as random and points out that overreaction of stock prices is as frequent as underreaction.

Malkiel (2003) associates the efficient market hypothesis with the concept of „random walk". The logic behind this idea is that in case information flows freely and is quickly incorporated into stock prices, future price movements will only be affected by future news and will not be tied to present price changes. However, news is inherently unpredictable, resulting to price changes that are also unpredictable and random.

2.1.4 CRITICISM AND LIMITATIONS OF EMH

From the early 90s the efficient market hypothesis started facing challenges in a theoretical as well as empirical level (Shleifer, 2000; Shiller, 2003). Herbert Simon (1955) has been one of the first academics to level criticism against the theory arguing that: „Our rationality is bounded and our acting is constrained." Nowadays, the concept of the EMH still remains controversial with numerous conflicting viewpoints (Ţiţan, 2015). Fama (1998) identified two main weaknesses of the EMH: access and availability. The challenge in keeping up with the rapid pace of changes, events and markets is substantial. Furthermore, news are distributed through a wide range of communication platforms. With respect to availability, at times, news data are open only to a certain group of people long before becoming accessible to others, causing information asymmetry in the stock-market. Bernard & Thomas (1989) argue that in the contemporary informational landscape, where people are dealing with numerous amounts of news articles, the distinction line between public and private information is often imprecise, and even public available information may not be directly and efficiently reflected in stock prices. For the aforementioned reasons, not everyone is able to access information at the same amount, time and frequency. Some articles are proving the quick reaction of asset prices on the first days after news announcements, validating the efficiency of financial markets, whereas others investigate a longer timeframe, for instance a couple of months, implying that price adjustments take place gradually (Fama et al., 1969). Investors' inattentiveness might be another factor that contributes to the delayed price reactions, or underreactions to event announcements, causing market inefficiency (Ţiţan, 2015).

2.2 BEHAVIORAL FINANCE

The debate has driven academic attention towards behavioral finance - an alternative view of financial markets, setting up new predictions (Baker & Nofsinger, 2010; Shleifer, 2000). The developed theories about the influence of human behavior on investing decision making appear not as an additional assumption, but rather as a contradictory perspective (Shiller, 2003). Behavioral finance is defined as a science that combines finance and cognitive psychology (Shefrin, 2002), trying to explain anomalies of the market as well as the financial decisions and behavior of investors (Kausar & Taffler, 2005; Ruppert, 2004). Zindel et al. strongly suggest that investors do not behave rationally and make mistakes in interpreting information. This is mainly due to emotions and biases of humans as well as so called heuristics (Zindel et al., 2014). Fama (1998) tried to understand how these biases of investors can lead to overreaction to particular events and underreaction to others. Thereby, the economist refers to two behavioral models, developed by Barberis, Shleifer and Vishny (1998) and Daniel, Hirshleifer, and Subrahmanyam (1997). The roots of the BSV model stem from cognitive psychology of two judgement biases: the representativeness bias of Kahneman and Tversky (1982) and conservatism, which is the slow reaction to new information, according to Edwards (1968). On the other hand, in DHS investors can be either uninformed or informed. In the second case, they suffer from overconfidence and/or self-attribution. The first one results in overreaction to private signals about an asset, whereas biased self-attribution leads to underreaction to public information (Fama, 1998). In 2005, Kausar & Taffler supported the DHS model with their results, proving that market underreaction to negative news is due to an initial overreaction, directly after public announcement of information. Implementing the heuristic of representativeness (Tversky et al., 1982) in market pricing, De Bondt and Thaler (1985) show that market participants overreact to both types of news (positive and negative). The authors underline that investors who suffered from losses in the past underestimate prices while winners tend to overestimate them.

Research results support that market sentiment offers a better understanding of capital markets and behavioral theories (Frugier, 2016). Inspite of the ongoing development in the field, one demanding task for behavioral finance remains to go beyond the criticism of the EMH and provide instruments that cater to operationalize investor behavior, considering their imperfect rationality. Sentiment analysis could shed light on this challenge as the application of AI enables the capture and convert emotions and opinions into numbers.

2.2.1 PROSPECT THEORY

Prospect theory was developed by Kahneman and Tversky (1979) as a critique of Expected Utility Theory, broadening Behavioral Finance Models. The renowned psychologists prove that market reactions to information are not always rational. According to this alternative descriptive model for decision making under risk and with a low number of outcomes, people tend to violate the tenets of Utility Theory with their preferences and choices. The two stages of the decision process include framing and valuation. During framing, individuals create a model of possibilities and outcomes that are crucial. For instance, a negative presentation (framing) of financial news results in negative sentiment, since the event will be perceived negatively, regardless of its nature. If news is framed positively, on the other hand, investors tend to be optimistic. During the second phase, the valuation, each prospect is being assessed so that a decision can be taken accordingly (Kahneman & Tversky, 1979; Tversky & Kahneman, 1986, 1992).

One of the first findings is that "people overweight outcomes that are considered certain, relative to those that are merely probable". This phenomenon is named certainty effect (Kahneman & Tversky, 1979). Another conclusion is that outcomes are perceived as gains and losses and their value is multiplied by decision weights instead of probabilities. Low probabilities are overweighted whereas moderate to high probabilities are underweighted, leading to a nonlinear transformation of the probability scale. The value function of Prospect Theory is generally "concave for gains, convex for losses, and steeper for losses than for gains". Thereby, gains and losses are defined in respect to a neutral reference point. This is affected by the expectations of the decision makers as well as the way prospects are presented to them. (Kahneman & Tversky, 1979; Tversky & Kahneman, 1986). In case sentiment of news announcements deviates from investor‘s expectations (for instance positive sentiment after a period of negative outcomes), the reference point will then be adjusted respectively.

The theory has been supported by several academics, such as Thaler and Shefrin & Statman. The latter, applied Kahneman and Tversky´s concept of framing decisions (1986) claiming that investors are predisposed to hold bad performing stocks for too long and sell well performing stocks too soon. The phenomenon was called disposition effect (Shefrin & Statman, 1985). In the following years, Tversky & Kahneman decided to review and update their theory. The new version, Cumulative Prospect Theory, has been extented for uncertain and risky prospects with any number of outcomes, using cumulative decision weights. The value and weighting functions are related to diminishing sensitivity and loss aversion.

2.2.2 OVER- AND UNDERREACTION TO NEWS

Some literature findings conclude that the market underreacts to information, while others argue the existence of overreaction. Through the attempt to explain these market anomalies, as shortcomings of the EMH, a theory, based on two psychological biases - overconfidence and biased self-attribution - was proposed (K. Daniel et al., 1998). The theory supports that investors overreact to private information signals and underreact to public information signals. Positive return autocorrelations result from constant overreaction, followed by long-run correction. The authors suggest that mispricing events lead to underreaction to new information. Moreover, Hong & Stein (1999) conclude that there must always be an overreaction in the long run when there is a short-term underreaction to news shocks.

Based on the development of Prospect Theory and the introduction of the concept „disposition effect", by Shefrin & Statman (1985), the work of Frazzini (2006) suggests that this effect induces underreaction to news, resulting in return predictability and postannouncement price shifts. Disposition effect is defined as the „tendency of investors to ride losses and realize gains“ and arised as a combination of prospect theory and mental accounting (Frazzini, 2006; Shefrin & Statman, 1985). Thereby, price trends depend on the nature of news (good or bad) and the investor’s reference price relative to the current price. Frazzini finds that assets with large unrealized gains underreact solely to positive news, while stocks with large unrealized losses underreact only to negative news. This happens due to the fact that disposition investors are hesitant to acknowledge the loss, while after positive news announcements, the active selling prevents quick stock-price adjustments.

In their literature review about short term overreaction caused by news sentiment, Chari et al. (2017) underline the role of financial news on the stock-market and investor decision making. The authors suggest that in the modern information - and media - rich environment the probability of overreaction is high, leaving space for speculators to achieve abnormal returns.

2.3 SENTIMENT ANALYSIS IN FINANCE

Building on the insights provided by Behavioral Finance and particularly Prospect Theory, it is essential to study how sentiment analysis has been applied in finance, in understanding the impact of market sentiment on asset prices and investor behavior. The following chapter reviews the existing literature on sentiment analysis, highlighting its role and impact in financial markets.

Jiao et al. (2020) argue that the usage of social media and the consumption of news is inconsistent with rational markets supporting that some investors are overconfident when interpreting news. The integration of sentiment data, especially from social media platforms like Twitter, has provided new insights into market dynamics. For instance, Tetlock‘s research on media sentiment, in the early 2000s, reveals its influence on market prices, indicating the significant role of media in financial markets. The author builds a measure of media content that seems to correspond to either negative investor sentiment or risk aversion. The findings confirm that: “High values of media pessimism induce downward pressure on market prices and unusually high or low values of pessimism lead to temporarily high market trading volume.” The study, *"Sentiment-aware volatility forecasting"*, has shown how integrating market sentiment with advanced neural networks can enhance volatility forecasting (Xing et al., 2019). Pagolu et al. (2016) developed a sentiment analyzer for Twitter, that can recognize and evaluate the type of sentiment hidden in each tweet. The findings expose a strong correlation between the rise and fall in stock prices to the public mood, communicated on Twitter through tweets and related to the corresponding company. Similarly, Nguyen et al. (2015) attempted to forecast stock price movement by means of the sentiment from social media. Their model containing public mood information outperformed 2.07% on average accuracy than the model without.

The recent paper "Front-Page News: The Effect of News Positioning on Financial Markets" explores the impact of news placement, specifically front-page news, on investor behavior and financial markets. This research highlights the critical role of media in shaping financial markets, underlining the power of news placement and presentation in driving investor behavior and market trends (Fedyk, 2024). Several other research papers have tested the relationship between news sentiment and stock market performance:

Hasselgren et al. (2022) use social media sentiment to improve investment decision making. For this purpose, the authors introduced a novel approach to factor social media metrics into the sentiment score to map the overall sentiment of users from individual tweets. The results identify a comparable trend between social media sentiment and the stock market performance of certain S&P 500 stocks.

Another study translates the news articles into market sentiment using the Harvard psychological dictionary and Loughran–McDonald financial sentiment dictionary, incorporating them into developed models to evaluate their forecasting power. The results of the two sentiment dictionaries do not differ significantly. The findings include that sentiment analysis enhances the prediction accuracy. Secondly, the models including sentiment analysis outperform the bag-of-words at stock, sector and index levels. On the other hand, the models using sentiment polarity did not perform well in all the tests, implying that the distinction between positive and negative dimensions does not offer relevant predictions (Li et al., 2014).

Niederhoffer's paper analyzes headlines of world events in the New York Times and ranks them into predefined semantic categories - from extreme-bad to extreme-good -

exposing that markets tend to overreact to bad news. Garcia (2013) argues that news sentiment has a strong predictive power of stock-market returns especially during recessions. Frugier (2016) shows that investor sentiment can be beneficial, and furthermore incorporates volatility into the model, while examining the performance of different portfolios formed by European stocks.

Based on previous literature, it can be suggested that integration of sentiments from social media could enhance the accuracy of stock market predictions. Nonetheless, some academics do not approve of the fact that sentiments from social media have predictive capabilities: Antweiler & Frank (2004) claim that the effect of messages on stock return is economically small, whereas the findings of Tumarkin & Whitelaw (2001) are consistent with the EMH, questioning the impact of news sentiment on stock market. Bollen et al. (2011), on the other hand, support that only certain mood dimensions, particularly “calm” and “happy”, have significant predictive value for Dow Jones Industrial Average (DJIA) changes.

Respective studies have also explored the relationship between sentiment analysis and portfolio management. Malandri et al. investigate the impact of public mood, gathered from social media and online news, on stock returns and asset allocation. Thereby, the authors collect historical price data from the New York Stock Exchange and combine it with sentiment data, utilizing three different machine learning models (Long Short-Term Memory Network, Multi-Layer Perceptron, and Random Forest Classifier). The five created portfolios were proven to outperform the equal-weighted (EW) portfolio, with LSTM being the most appropriate model. Consequently, incorporating public mood in the investment strategy is proved to be advantageous and can significantly improve stock performance. Xing et al. reveal similar results, proposing that all formed portfolios that have taken market sentiment into consideration achieve a better annual return than the benchmark. Jin et al. (2020) emphasize the significant impact of investors' emotional tendencies on stock predictions by combining sentiment analysis, advanced time-series decomposition, and deep learning techniques, showcasing a significant improvement in stock market prediction accuracy.

Devani & Patel (2022) collected financial news data and combined them with historical prices to predict the closing price, using the Long-Short-Term Memory model (LSTM).

One more research paper explores the role of financial news articles on three different textual representations, finding that noun phrases enhance the results, in comparison to bag of words. Applying a Support Vector Machine (SVM) algorithm, the authors prove that this model has a statistically significant impact on predicting stock prices, twenty minutes after the news announcement, compared to the linear regression model (Schumaker & Chen, 2009).

One more contribution to the growing literature is the article “Business news and business cycles“, which proves that news attention explains one fourth of the variation in stock-market fluctuation. The models that were used for this analysis include macroeconomic vector autoregression (VAR) and stock-market timing. The authors strongly support that news media offer a reflection of economic trends and represent investors‘ perceptions of the state of the market. Thereby, the news are divided into three categories, recurrent, seasonal, and emergent. Recurrent topics are those that receive media attention consistently, seasonal topics are for instance “presidential elections” or earnings forecasts while emergent topics arise from particular events and do not appear so often in the sample (Bybee et al., 2021).

The study of Lawrence et al. (2018) tests the effects of promotion earnings announcements on the stock markets through a field experiment. The findings suggest that on the announcement day promoted firms demonstrate an increase in abnormal returns in comparison to control firms. On the other hand, there is no proof for notable increases in trading volume or the amount of information by users subject to the promotion.

Another, recent, study explores for the first time in literature how companies use the voice to disclose information and how investors react to it (Sinh Thoi Mai, 2024). Using machine learning methods to measure passive voice in 10-K filings, the writer concludes that firms strategically avoid passive voice when publishing bad news. The motivation behind it is to moderate investors’ negative reactions by taking responsibility (using active voice) and reduce textual confusion and complexity. Moreover, investors seem to first overreact negatively to passive voice during 30 days after 10-K release. However, from the 30th day and up to two years after the release, investors reverse their overreactions and respond positively to passive voice. Besides, the less attention is paid to 10-K filings, the stronger is the relationship between passive voice and future stock return (Sinh Thoi Mai, 2024).

In brief, existing literature underlines the relevance of sentiment in predicting stock market returns and reveals that its application, especially on social media news, significantly influences market performance. Several studies show that news sentiment correlates with stock prices, with some predicting market movements being more accurately when public mood is taken into account. Notably, sentiment from media sources has been found to affect volatility and trading volume, with negative sentiment often causing downward pressure on prices. Moreover, integrating advanced machine learning techniques, such as Long Short-Term Memory Networks and Support Vector Machines, has further enhanced market prediction accuracy. However, some researchers argue that the economic impact of sentiment on stock returns is small or restricted, aligning more with the Efficient Market Hypothesis. Nevertheless, doubts and uncertainties about its consistency and significance across different contexts continue to persist.

Based on these findings and discussions presented in previous literature, and considering the gaps that still exist in this field, the following hypothesis is formulated:

H_1: There is an effect of financial news sentiment on the daily returns of the Euro Stoxx 50 index.

3. DATA AND METHODOLOGY

The following chapter aims to evaluate empirically the relationship between the sentiment scores of the financial news articles and the Euro Stoxx 50 index returns.

3.1 SAMPLE

Financial News Articles

To gather relevant financial news articles the Bloomberg terminal as one of the largest financial news platforms was chosen (Fedyk, 2024). The period of 1st January 2022 to 31st March 2024 was selected, to provide a substantial timeframe of more than two years that could capture a variety of market behaviors and significant economic events affecting the European stock market.

In order to ensure the relevance of the extracted news to the Euro Stoxx 50 index, a set of keywords was carefully selected, based on their significance to the European markets. Higher focus was given on the sectors, countries and companies with the largest weights within the index during the considered time period, according to Stoxx.com. In particular, the sectors with a weight of 5% or more include Financials, Consumer Discretionary, Materials, Information Technology, Consumer Staples, Healthcare and Energy. Furthermore, the countries with the highest number of members in the Euro Stoxx are France, Germany, Netherlands, Spain, Italy, Ireland, Belgium and Finland, whereas it was checked semiannually for changes within the two-year timeframe. Last but not least, the top 10 components of the index for the extended period are ASML HLDG, LVMH MOET HENNESSY, TOTALENERGIES, SAP, SANOFI, SIEMENS, L'OREAL, SCHNEIDER ELECTRIC, ALLIANZ and AIR LIQUIDE.

Therefore, keywords comprised "Euro Stoxx 50," “Europe”, “Equity market” as well as individual names of the major sectors and companies. Other filters included to select as a source Bloomberg only, as a language English and top ranked articles. These filters catered in selecting only the articles that have a more pronounced effect on the Euro Stoxx 50, among the huge amount of daily financial news. The news is covering a range of economic, geopolitical and seasonal events, including earnings forecasts, unemployment numbers, the inflation crisis, the Russian-Ukranian war, presidential elections and others.

The articles were downloaded and stored in a structured format, which included the article’s title and publication date. This procedure was crucial in managing the volume of data over the two-year period and allowed for efficient preprocessing in later stages.

Historical Prices and Volatility Index Data

Historical prices have been extracted from Yahoo Finance for the Euro Stoxx 50 index, in order to provide a comprehensive overview of the European stock market. The dataset has been applied for a period of two and a half years, in consistency with the period in which the news articles were extracted. Thereby, only trading dates were taken into consideration and for each, there are opening, high, low, closing prices as well as trading volumes included. However, the focus for the analysis was placed on the closing prices.

The Volatility Index (VIX) is applied as a control variable. The VIX, created by CBOE Global Markets, shows the market's expectation of 30-day volatility. The data is also extracted from Yahoo Finance, from the 01.01.2021 to the current date, and includes daily adjusted closing prices.

3.2 SENTIMENT ANALYSIS

For the sentiment analysis part, the Loughran-McDonald Lexicon was chosen, since it is specifically tailored for financial text analysis. This lexicon is particularly proper for the context of financial news, as it categorizes words into sentiments based on their typical use in financial reporting, distinguishing it from more general sentiment analysis tools, like the Harvard dictionary, that might not capture the nuances of financial language (Loughran & McDonald, 2011). For that reason, the lexicon is widely approved and used in literature for sentiment analysis of financial news (Li et al., 2014).

The process began with the extraction of text from financial news articles sourced by Bloomberg. As a next step, the articles were prepared by removing unnecessary characters, such as punctuation and numbers, normalizing all text to lowercase to maintain consistency, and then breaking down the content into manageable pieces or tokens.

Using Python's Natural Language Toolkit (NLTK), the prepared text was tokenized, splitting it into individual words. This step is crucial for the effective application of the sentiment lexicon. Additionally, stop-words - like "and", "the", and "is", which are prevalent in English but generally irrelevant to sentiment analysis - have been removed to refine the focus on more significant words.

After the preprocessing and cleaning of the data, the text was analyzed using the Loughran-McDonald Lexicon. This lexicon categorizes words as positive or negative within the context of financial discourse (Loughran & McDonald, 2011; Loughran & McDonald, 2016). It is particularly useful for this analysis because it correctly interprets the sentiment of words that, while possibly neutral in regular conversation, have specific connotations in finance. For example, "liability" typically carries a negative sentiment in financial contexts. For each article, the sentiment score was calculated by tallying the number of positive and negative words identified through the lexicon. The overall sentiment score for each article was determined by subtracting the total count of negative words from positive words, providing a net sentiment score that reflects the overall sentiment conveyed in the article. Additionally, for days on which multiple articles were extracted, the aggregate score was calculated by using the average function. The results fluctuate between -0.78 and +1.00, with -1.00 being the lowest and +1.00 the highest value that can be achieved (see Figure 1). These sentiment scores were then cleaned by replacing missing values with *NA* and converting the scores to numeric format. The scores were merged with the returns data according to matching dates.

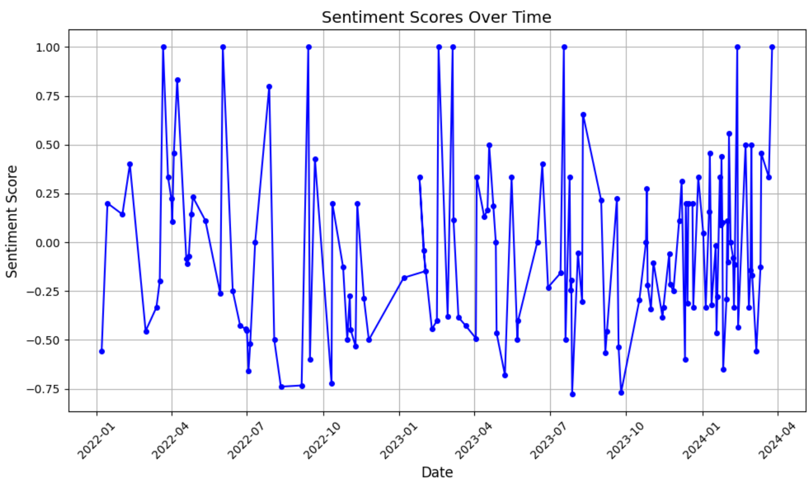


Figure 1- Sentiment Scores over time

3.3 Data Preparation and Preprocessing

The Euro Stoxx 50 index returns are selected as dependent variable in the model whereas sentiment scores are the independent variable. For higher explanatory power, the VIX, volatility index, is applied as a control variable. Lagged daily returns were also included in the model to control for the autocorrelation and momentum, capturing the effect of past market performance on current returns.

From the extracted closing prices, daily index returns were calculated based on the following formula:

Daily returns for the Euro Stoxx 50:

$$Daily\ Return_t = \frac{Adj\_Close_t}{Adj\_Close_{t-1}} - 1$$

, where

$\mathrm{Adj\_Close_t}$ is the adjusted closing price on trading day t,

$\mathrm{Adj\_Close_{t-1}}$ is the adjusted closing price on the previous trading day.

This calculation shows the daily return as a percentage change in the index value from one trading day to the next, in the form of decimal numbers.

Besides the daily returns, lagged variables, representing the return from the previous five days, were calculated to capture the relationship of past market behavior.

- Return_L1 is the return from one day prior,
- Return_L2 is the return from two days prior,
- Return_L3 is the return from three days prior,
- Return_L4 is the return from four days prior,
- Return_L5 is the return from five days prior.

Using the same formula, the VIX change rate was calculated as:

$$VIX\ Return_t = \frac{VIX.Adjusted_t}{VIX.Adjusted_{t-1}} - 1$$

To prepare the dataset for analysis all rows with missing values (NA) were removed. The closing price column was dropped from the dataset since it is no longer needed after the return calculations. Furthermore, empty strings in the sentiment dataset were replaced with NA values. Once the data is cleaned, the returns, sentiments and daily changes in market volatility are merged. As a last step of the data preparation and in order to enhance the model's explanatory power, the variable 'Year' is extracted from the Date column, enabling the analysis on a yearly basis.

3.4 STATISTICAL HYPOTHESES AND TESTS

As mentioned before, the research question of this study investigates wether sentiment scores derived from financial news have an effect on the Euro Stoxx 50 daily index returns. For this purpose, textual content operationalized by sentiment scores, is combined with historical prices through two regression models, Ordinary Least Square (OLS) and Panel Regression, to examine its impact on the performance of the index.

Therefore, the following null hypothesis is proposed:

H_0*: The sentiment score has no effect on the index returns.*
Alternative Hypothesis:

H_1*: The sentiment score has an effect on the index returns.*

To test this hypothesis, at the first step, each variable (Return, Sentiment Score and VIX) was tested for normality using the Jarque-Bera test. The results of the test suggest that all of the variables have p-values lower than 0.05 which indicates that they do not follow a normal distribution. For further analyses and particularly to test for correlation the Spearman rank correlation as a non-parametric method is applied.

The relationship between Sentiment Score and Return is displayed on the following scatter plot (Figure 2), along with the Spearman correlation coefficient and p-value. The fitted regression line in blue has a slight positive slope, indicating that on average, as the sentiment score is increasing, the returns are slightly increasing as well. The scatter of the points around the x-axis does not imply any upward or downward trend, meaning that the correlation of sentiment and returns is weak. The dispersion of the points around the regression line, is reflecting the variability in the returns, even in case of similar sentiment scores.

The results of the Spearman correlation test reveal an R of 0.19 and a p-value of 0.025. The coefficient implies that the correlation between sentiment score and returns is positive but weak, whereas the p-value indicates that the correlation is statistically significant at the 5% significance level.

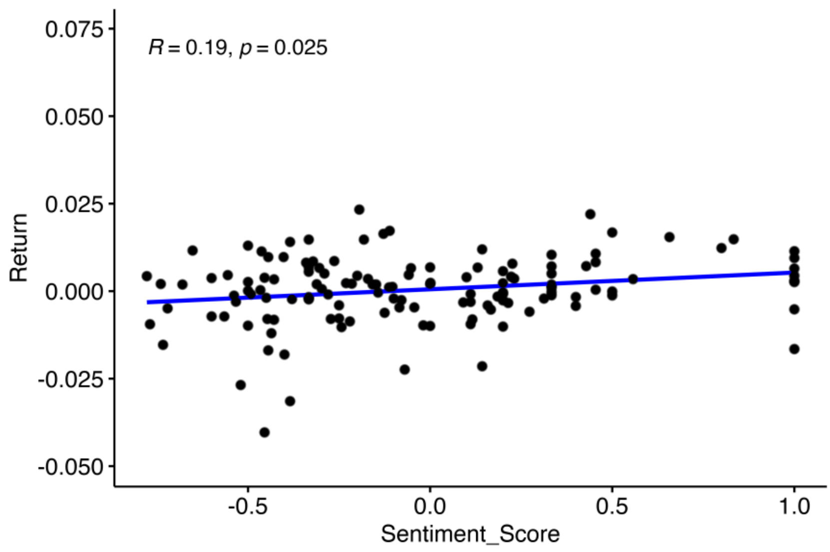


Figure 2- Scatter Plot

Tests for OLS regression model

To test for linearity, the RESET test was applied and according to its p-value of 0.8525, it is suggested that there is no need for higher-order terms as the assumption of linearity seems to hold.

The Breusch-Pagan test is appropriate to check for homoskedasticity in the residuals it shows no evidence of heteroskedasticity since its p-value is 0.1856 and therefore higher than 0.05. This means that one of the key assumptions of the OLS regression model is satisfied.

Another important assumption of the model is that residuals are not correlated across the observations. On account of this, the Durbin-Watson test is executed. The D-W statistic is very close to 2 (1.976486) and the p-value of 0.916 well above the threshold of 0.05, which implies that the residuals are independent and the assumption of no autocorrelation is also confirmed.

4. EMPIRICAL RESULTS AND DISCUSSION

In the following part, the descriptive statistics of all variables as well as the statistical results are presented in the tables below.

4.1 DESCRIPTIVE STATISTICS

The analysis in Table I summarizes the distribution of each variable.

A standard deviation of 0.012 indicates a moderate variability in daily returns of the Euro Stoxx 50. The skewness of 0.134 and kurtosis of 4.234 suggest a slight positive skew and heavy tails in the distribution of returns, respectively. This can be explained by the presence of outliers or extreme values.

Considering the sentiment score, from the mean of -0.033 we can conclude that during the observed period the sentiment was on average more negative than positive. The standard deviation (0.441) shows a fairly wide dispersion with values ranging from -0.778 to 1.000. The skewness (0.618) suggests a moderate positive skew, indicating the presence of more positive outliers in the sentiment data.

With the regard to the control variable VIX, with an average return close to zero it is clear that market volatility remained relatively stable. The standard deviation (0.061) reflects moderate fluctuations in volatility, with values between -0.144 and 0.244. The skewness (0.968) and kurtosis (1.491) suggest that large increases in volatility were more common than decreases.

Table I- Descriptive Statistics

Variable	Obs.	Mean	Median	Standard Deviation	Minimum	Maximum	Skewness	Kurtosis
Return	601	0.000	0.001	0.012	-0.050	0.074	0.134	4.234
Sentiment Score	134	-0.033	-0.107	0.441	-0.778	1.000	0.618	-0.203
VIX change	566	-0.001	-0.009	0.061	-0.144	0.244	0.968	1.491

Variables: Return = Daily Euro Stoxx 50 returns; Sentiment Score = sentiment scores of financial news articles on specific days; VIX change = Volatility Index change rate

4.2 REGRESSION MODELS

4.2.1 ORDINARY LEAST SQUARE REGRESSION

The effect of sentiment scores on index returns was first tested using OLS regression. The model has been applied in past literature, to test, for instance, how a pessimistic mood can predict market volatility and prices (Tetlock, 2005).

OLS regression without control variables

The model is used to test the direct relationship between the Sentiment Score and Return without controlling for other variables. Thereby, it estimates how much of variation in stock price returns can be explained by news sentiment. According to its coefficient, for each unit increase of the sentiment score, the return increases by approximately 0.48%. The coefficient is statistically significant, indicating a relevant impact of the sentiment score on returns. The Adjusted R-Squared with a value of almost 4%, underlines that the sentiment score explains only a small portion of the variance in returns. The F-Statistic of 6.537 and the corresponding p-value of 0.0117 reinforce that the model and thus the variable Sentiment_Score as a predictor, is statistically significant. Yet the relationship is not so strong, implying that there are other factors playing a more vital role. The results are visible on Table II.

Table II- OLS Results without Control Variables

Variable	Coefficient	p-value	Significance Level
Intercept	0.0005372	0.515	
Sentiment_Score	0.0047715	0.0117	**
Adjusted R^2		4%	
F-Statistic		6.537	
p-value		0.0117	

**** p<0.01, ** p<0.05, * p<0.1*

Variables: Intercept: the average return when sentiment score is equal to zero; Sentiment_Score of financial news articles

The null hypothesis (H_0) can be rejected as sentiment score does have an impact on daily returns.

OLS regression with control variables

The control variables VIX and return lags cater to accounting for other external factors affecting returns. The sentiment scores seem to have a weak effect on returns also in this case. The market volatility is the most significant predictor with a coefficient of -0.05839. This indicates that for each unit increase in the VIX, the daily return decreases by approximately 5.84%. The lagged returns are also significant, indicating that past market behavior has an impact on returns. The adjusted R-Squared suggests that 23.3% of the fluctuation in returns is explained by the model. The small p-value implies that the model is statistically significant and the variables have an appropriate explanatory power to predict the returns. The model has a satisfying explanatory power but the results (Table III) suggest that also other factors have an impact on daily returns.

Table III- OLS Results with Control Variables

Variable	Coefficient	p-value	Significance Level
Intercept	0.00001202	0.9878	
Sentiment_Score	0.003205	0.0640	*
VIX	-0.05839	2.97e-0.5	***
Return_L1	0.1644	0.0459	**
Return_L2	-0.1263	0.0666	*
Return_L3	0.07478	0.2619	
Return_L4	0.1892	0.0147	**
Return_L5	-0.03729	0.6150	
Adjusted R^2		23.3%	
F-Statistic		6.556	
p-value		1.47e-0.6	

**** p<0.01, ** p<0.05, * p<0.1*

Variables: Intercept: the average return when sentiment score is equal to zero; Sentiment_Score of financial news articles; VIX: Volatility Index; Return_L1: 1-day lag of returns; Return_L2: 2-days lag of returns; Return_L3: 3-days lag of returns; Return_L4: 4-days lag of returns; Return_L5: 5-days lag of returns

The null hypothesis (H_0) cannot be rejected as the effect of sentiment score on returns is no longer significant.

In the overall, it can be assumed that sentiment affects returns, but the effect is stronger when it is not accounted for market volatility and past market performance. VIX has a dominating impact on the stock returns, potentially mediating the effect of sentiment.

OLS Modeling for each year

To allow for a better understanding and an examination of how the effect of sentiment on returns varies over time, OLS regression is performed for each year separately, thereby accounting only for market volatility.

The Sentiment_Score has a positive and significant effect on returns in 2022. This suggests that in this year, news sentiment strongly influenced market performance. In 2023 the impact of sentiment scores on the stock performance diminished significantly and in the first quarter of 2024 the effect remained insignificant.

Table IV- OLS Results 2022

Variable	Coefficient	p-value	Significance Level
Sentiment_Score	0.007542	0.043223	**
VIX	-0.088726	0.000631	***
R^2		36.25%	
Adjusted R^2		32.39%	

**** p<0.01, ** p<0.05, * p<0.1*

Variables: Sentiment_Score of financial news articles; VIX: Volatility Index

Table V- OLS Results 2023

Variable	Coefficient	p-value	Significance Level
Sentiment_Score	0.0014205	0.558	
VIX	-0.0398559	0.064	*
R^2		6.18%	
Adjusted R^2		3%	

_p<0.01, ** p<0.05, * p<0.1_

Variables: Sentiment_Score of financial news articles; VIX: Volatility Index

Table VI- OLS Results 2024

Variable	Coefficient	p-value	Significance Level
Sentiment_Score	0.003925	0.2069	
VIX	-0.048774	0.0859	*
R^2		14.37%	
Adjusted R^2		8.25%	

**** p<0.01, ** p<0.05, * p<0.1*

Variables: Sentiment_Score of financial news articles; VIX: Volatility Index

4.2.2 PANEL REGRESSION

In the next step the model was tested for multicollinearity, linearity, homoskedasticity, autocorrelation and normality. The Variance Inflation Factor (VIF) value for all variables is clearly below 10, which indicates very low multicollinearity. Furthermore, the results of the Durbin-Watson test suggest that there is no significant autocorrelation in the residuals of the model. The RESET detects non-linearity between the residuals and the Breusch-Pagan test identifies the presence of heteroskedasticity. In addition, the residuals are not normally distributed according to the Jarque-Bera test.

As more robust, panel data models (fixed effects and random effects) are designed to handle the issues of heteroskedasticity by accounting for individual- or time-specific effects, which OLS does not address. This can improve the accuracy of results by controlling factors that vary across individuals or time periods, addressing the problem of non-linearity (Allison, 1994).

Thereby, both fixed effects and random effects models are estimated. Fixed effects (FE) models control for time-invariant characteristics of the entities, isolating the impact of the variables that change over time, such as Sentiment_Score, VIX, and past returns. In the case of panel regression without control variables, the effect of sentiment scores is isolated. Sentiment_Score has a positive and significant effect on returns in the fixed effects model, yet it explains a very small portion of the variation in returns (R-squared = 0.045868).

Table VII- FE Model Results

Variable	Coefficient	p-value	Significance Level
Sentiment Score	0.0046538	0.01367	**
R^2		4.59%	
F-statistic		6.24944	
p-value		0.013668	
**** p<0.01, ** p<0.05, * p<0.1*			
Variables: Sentiment Score of financial news articles			

Random effects (RE) models assume that the individual-specific effects are randomly distributed and thus, uncorrelated with the independent variables. Analogous to the FE model, Sentiment_Score has a positive and significant effect on returns, even the variation explained by the model is relatively low (R-squared = 0.046577).

Table VIII- RE Model Results

Variable	Coefficient	p-value	Significance Level
Sentiment Score	0.0047005	0.01121	**
R^2		4.66%	
Chisquared		6.43249	
p-value		0.011205	
**** p<0.01, ** p<0.05, * p<0.1*			
Variables: Sentiment Score of financial news articles			

The Hausman Test is used to choose whether the fixed or random effects model provided a better fit (Hausman, 1978). For this purpose, it is exploring if the individual effects have any correlation with the independent variables. The p-value of 0.7897 is much higher than 0.05, meaning that the Nullhypothesis cannot be rejected. This means that the individual-specific effects are not correlated with the independent variables making the random effects model more appropriate for this dataset.

In the instance of including control variables, the fixed effects model is preferred because the main assumption of the RE model - that individual specific effects are not correlated with the independent variables - might be violated. As a result, the estimates of the random effects model could be biased.

Since the dataset covers events during only three different years, the variation over time is limited and this makes it challenging to apply a random effects model. With such a restricted number of time periods, the model may struggle to distinguish between cross-sectional variation and time-specific variation.

On the other hand, the fixed effects model controls for all time-invariant characteristics of each entity, enabling a more reliable estimation of effects of time-varying factors such as Sentiment_Score, VIX, and lagged returns.

In the case of controlling for market volatility and past returns, the FE model implies a positive, yet weaker effect of sentiment on returns. The dominating and statistically significant effect of VIX demonstrates that increased volatility leads to decreased returns. The consideration of all control variables improves the explanatory power of the model, reaching an R-Squared value of 29%.

Table IX- FE Model Results with Control Variables

Variable	**Coefficient**	**p-value**	**Significance Level**
Sentiment Score	0.0031991	0.06270	*
VIX	-0.0581253	2.553e-05	***
Return_L1	0.1563275	0.05474	*
Return_L2	-0.1414657	0.03866	**
Return_L3	0.0764582	0.24570	
Return_L4	0.1955854	0.01073	**
Return_L5	-0.0346182	0.63747	
R^2		29.02%	
Adj. R^2		23.65%	
F-Statistic		6.949	
p-value		6.1811e-07	

**** p<0.01, ** p<0.05, * p<0.1*

Variables: Sentiment_Score of financial news articles; VIX: Volatility Index; Return_L1: 1-day lag of returns; Return_L2: 2-days lag of returns; Return_L3: 3-days lag of returns; Return_L4: 4-days lag of returns; Return_L5: 5-days lag of returns

5. CONCLUSION, LIMITATIONS AND FUTURE RESEARCH

To conclude, the correlation between news sentiment and investor behavior, as predicted by prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1986, 1992), can partly help explain why markets sometimes overreact or underreact to news events. The analysis of sentiment serves as a quantitative measure of the psychological and emotional drivers that behavioral finance outlines, providing insight into how investors interpret and react to information in the market. By measuring sentiment, we capture emotions and beliefs as well as the influence of optimism or pessimism on market movements, which can lead to deviations from rational pricing, such as market overreaction or underreaction.

The findings of this study align more closely with behavioral finance theories than with the Efficient Market Hypothesis (EMH). While EMH posits that markets are fully efficient and stock prices always reflect all available information, my results suggest that investor sentiment and market volatility introduce elements of inefficiency, particularly during periods of heightened uncertainty. In 2022, sentiment had a significant positive effect on returns, especially during a period characterized by market volatility. This could be explained by the behavioral tendencies of investors, as predicted by prospect theory, where emotional reactions to news events - particularly in uncertain environments - drive market movements.

The empirical analysis revealed that the Sentiment Score had a positive and significant correlation with returns in 2022, but its significance weakened in subsequent years (2023 and 2024), where the VIX (market volatility) became the dominant factor. Regression results showed that when controlling for VIX and lagged returns, the impact of sentiment diminished, indicating that volatility plays a more crucial role in explaining market returns. This finding underscores the importance of controlling for macro-level factors, as volatility tends to mediate the effect of sentiment on returns. The stronger effect of sentiment in 2022 could be attributed to the market's heightened sensitivity to news during times of significant global uncertainty, which amplified the influence of sentiment on investor behavior. This year was marked by heightened market volatility due to global occurrences such as the persistent effects of the COVID-19 pandemic and the Russia-Ukraine war. These events likely triggered behavioral tendencies in investors, as described by prospect theory, where emotional reactions to uncertain environments and significant news events drive market movements.

From a theoretical perspective, this study highlights the limitations of EMH in fully explaining market behavior and strengthens the case for behavioral finance. The results emphasize the role of investor psychology in driving market movements, particularly during volatile periods. The study also contributes to the growing body of literature by offering empirical evidence that supports the view that sentiment-driven behaviors, as outlined in behavioral finance, can lead to market inefficiencies.

However, this study faces several limitations. The period of time covered is relatively short, and the sentiment scores had several missing values, particularly on days with no news articles that met the selection criteria. Additionally, the model could be improved by incorporating more control variables, such as economic trends, major global events, or macro-level indicators, which might further clarify the dynamics between sentiment and returns. Another shortcoming of the study is the reliance on a single sentiment analysis dictionary, the Loughran-McDonald Lexicon. While this is a widely accepted tool for financial sentiment analysis, words and emotions are not always easy or absolute to operationalize. The complexity of language, especially in financial contexts, means that sentiment can be expressed in subtle and nuanced ways that a single dictionary might not fully capture. Future research could improve the analysis by incorporating multiple sentiment dictionaries or using more advanced natural language processing (NLP) techniques, such as machine learning-based sentiment models, to better capture the complexity of news sentiment and compare the results across different methods and dictionaries.

For future studies, it would be beneficial to explore how the interaction between sentiment and economic events- such as elections, climate changes, or global crises- affects market performance. Additionally, extending the time horizon and including long-term sentiment trends could provide a more comprehensive understanding of the role of sentiment in financial markets. Investigating the relationship between sentiment and sector-specific market reactions could also reveal deeper insights into how different industries react to news and volatility. Furthermore, it is strongly suggested to carry out cross-market studies comparing how sentiment influences asset performance across different regions or financial markets (for instance Asian vs. the US market). This could offer important insights into regional differences in market movement.

REFERENCES

Allison, P. D. (1994). Using panel data to estimate the effects of events. *Sociological Methods & Research*, *23*(2), 174–199.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, *59*(3), 1259–1294.

Baker, H. K., & Nofsinger, J. R. (2010). Behavioral finance: an overview. *Behavioral Finance: Investors, Corporations, and Markets*, 1–21.

Barberis, N. (1998). Shleifer, and Vishny, 1998,"A model of investor sentiment,". *Journal of Financial Economics, V49,* 307–343.

Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: delayed price response or risk premium? *Journal of Accounting Research*, *27*, 1–36.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Bybee, L., Kelly, B., Manela, A., & Xiu, D. (2021). Business news and business cycles. *The Journal of Finance*.

Byrum, J. (2022). *AI in Financial Portfolio Management: Practical Considerations and Use Cases* (pp. 249–270).

Chari, S., Hegde-Desai, P., & Borde, N. (2017). *A review of literature on short term overreaction generated by news sentiment in stock market*.

Daniel, K. D., Hirshleifer, D. A., & Subrahmanyam, A. (1997). A theory of overconfidence, self-attribution, and security market under-and over-reactions. *Self-Attribution, and Security Market Under-and Over-Reactions (February 19, 1997)*.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *The Journal of Finance*, *53*(6), 1839–1885.

De Bondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, *40*(3), 793–805.

Devani, D., & Patel, M. (2022). Stock market (BSE) prediction using unsupervised sentiment analysis and LSTM: A hybrid approach. *Proceedings of the 2nd*

International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2021, 79–87.

Edwards, W. (1968). Conservatism in human information processing. *Formal Representation of Human Judgment*.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383.

Fama, E. F. (1991). Efficient Capital Markets: II. *The Journal of Finance*, *46*(5), 1575–1617.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, *49*(3), 283–306.

Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, *10*(1), 1–21.

Fedyk, A. (2024). Front-Page News: The Effect of News Positioning on Financial Markets. *The Journal of Finance*, *79*(1), 5–33.

Frazzini, A. (2006). The disposition effect and underreaction to news. *The Journal of Finance*, *61*(4), 2017–2046.

Frugier, A. (2016). Returns, volatility and investor sentiment: Evidence from European stock markets. *Research in International Business and Finance*, *38*, 45–55.

Garcia, D. (2013). Sentiment during Recessions. *The Journal of Finance*, *68*(3), 1267–1300.

Hasselgren, B., Chrysoulas, C., Pitropakis, N., & Buchanan, W. J. (2022). Using Social Media & Sentiment Analysis to Make Investment Decisions. *Future Internet*, *15*(1), 5.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, 1251–1271.

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, *54*(6), 2143–2184.

Jiao, P., Veiga, A., & Walther, A. (2020). Social media, news media and the stock market. *Journal of Economic Behavior & Organization*, *176*, 63–90.

Jin, Z., Yang, Y., & Liu, Y. (2020). Stock closing price prediction based on sentiment analysis and LSTM. *Neural Computing and Applications*, *32*(13), 9713–9729.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263.

Kausar, A., & Taffler, R. J. (2005). Testing Behavioral Finance Models of Market Under- and Overreaction: Do they Really Work? *Organizations & Markets eJournal*.

Lawrence, A., Ryans, J., Sun, E., & Laptev, N. (2018). Earnings announcement promotions: A Yahoo Finance field experiment. *Journal of Accounting and Economics*, *66*(2–3), 399–414.

Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, *69*, 14–23.

LOUGHRAN, T., & MCDONALD, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187–1230.

Malandri, L., Xing, F. Z., Orsenigo, C., Vercellis, C., & Cambria, E. (2018). Public Mood–Driven Asset Allocation: the Importance of Financial Sentiment in Portfolio Management. *Cognitive Computation*, *10*(6), 1167–1176.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, *17*(1), 59–82.

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, *42*(24), 9603–9611.

Niederhoffer, V. (1971). The analysis of world events and stock prices. *The Journal of Business*, *44*(2), 193–219.

Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, 1345–1350.

Ruppert, D. (2004). Behavioral Finance. In *Statistics and Finance: An Introduction* (pp. 435–448). Springer New York.

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news. *ACM Transactions on Information Systems*, *27*(2), 1–19.

Shefrin, H. (2002). *Beyond greed and fear: Understanding behavioral finance and the psychology of investing*. Oxford University Press.

Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, *40*(3), 777–790.

Shiller, R. J. (2003). From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perspectives*, *17*(1), 83–104.

Shleifer, A. (2000). *Inefficient markets: An introduction to behavioural finance*. Oup Oxford.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 99–118.

Sinh Thoi Mai. (2024). Does passive voice matter for investors? *Hanken School of Economics, Department of Finance*.

Tetlock, P. C. (2005). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *SSRN Electronic Journal*.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, *4*(3), 199–214.

Ţiţan, A. G. (2015). The Efficient Market Hypothesis: Review of Specialized Literature and Empirical Research. *Procedia Economics and Finance, 32*, 442–449.

Tumarkin, R., & Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal*, *57*(3), 41–51.

Tversky, A., & Kahneman, D. (1986). The framing of decisions and the evaluation of prospects. In *Studies in Logic and the Foundations of Mathematics* (Vol. 114, pp. 503–520). Elsevier.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. In *Readings in Formal Epistemology* (pp. 493–519). Springer International Publishing.

Tversky, A., Kahneman, D., & Slovic, P. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge.

Xing, F. Z., Cambria, E., & Welsch, R. E. (2018a). Intelligent Asset Allocation via Market Sentiment Views. *IEEE Computational Intelligence Magazine*, *13*(4), 25–34.

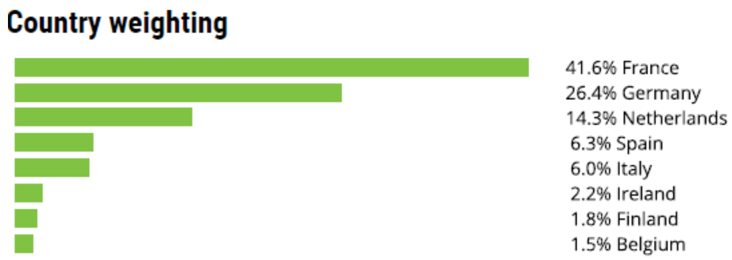
Xing, F. Z., Cambria, E., & Welsch, R. E. (2018b). Intelligent asset allocation via market sentiment views. *Ieee ComputatioNal INtelligeNCe MagaziNe*, *13*(4), 25–34.

Xing, F. Z., Cambria, E., & Zhang, Y. (2019). Sentiment-aware volatility forecasting. *Knowledge-Based Systems*, *176*, 68–76.

Zindel, M. L., Zindel, T., & Quirino, M. G. (2014). Cognitive bias and their implications on the financial market. *International Journal of Engineering and Technology*, *14*(3), 11–17.

APPENDICES

Appendix I- Euro Stoxx 50



Appendix II-Example 1 of Financial News Articles

Bloomberg   News Story

02/01/2023   10:27:41 [BN] Bloomberg News

Euro-Zone Inflation Eases as ECB Debate Over Rates Heats Up (2)

- Consumer prices rose 8.5% from year ago in January; est. 8.9%
- ECB is set to hike interest rates by a half-point on Thursday

By Jana Randow

(Bloomberg) -- Euro-area inflation slowed more than expected, suggesting a more heated debate to come at the European Central Bank over how much further interest rates must rise.

January's reading came in at 8.5%, Eurostat said Wednesday, less than economist estimates for a slowdown to 8.9%.

The third monthly retreat was driven by energy. But a gauge of underlying inflation that excludes volatile items like that held at an all-time high of 5.2%.

Appendix III-Example 2 of Financial News Articles

Bloomberg **News Story**

01/24/2024 16:04:15 [BN] Bloomberg News

Tech Stocks Shine in Europe After Blowout Updates From ASML, SAP

- ASML, SAP shares hit record high; lift sector rivals
- Investors say European tech shares are still not too expensive

By Allegra Catelli

(Bloomberg) -- Tech stocks led Europe's equity rally, as ASML Holding NV's bumper quarterly orders and SAP SE's forecast-beating results lifted their shares to record highs.

The gains added about €48 billion ($52.3 billion) to the market value of Europe's tech sector, which soared more than 5% after the region's top two tech companies signaled booming demand for their industries.

Chip equipment maker ASML – a bellwether for the industry's health – gained 10% on the day while SAP's cloud order backlog – an indicator of cloud revenue to be booked within next 12 months – saw robust growth in the fourth quarter. Its shares jumped more than 8%.

DISCLAIMER

This master thesis was developed with strict adherence to the academic integrity policies and guidelines set forth by ISEG, Universidade de Lisboa. The work presented herein is the result of my own research, analysis, and writing, unless otherwise cited. In the interest of transparency, I provide the following disclosure regarding the use of artificial intelligence (AI) tools in the creation of this thesis:

I disclose that AI tools were employed during the development of this thesis as follows:

- AI-based research tools were used to assist in literature review and data collection.
- AI-powered software was utilized for data analysis and visualization.
- Generative AI tools were consulted for brainstorming and outlining purposes. However, all final writing, synthesis, and critical analysis are my own work. Instances where AI contributions were significant are clearly cited and acknowledged.

Nonetheless, I have ensured that the use of AI tools did not compromise the originality and integrity of my work. All sources of information, whether traditional or AI-assisted, have been appropriately cited in accordance with academic standards. The ethical use of AI in research and writing has been a guiding principle throughout the preparation of this thesis.

I understand the importance of maintaining academic integrity and take full responsibility for the content and originality of this work.

Sofia Korompli – 15/10/2024