Norwegian School of Economics
Bergen, Autumn 2023

Behavioral Impact of Short-Term Stock Price Trends on Equity Research

A Textual Analysis of Trend-Chasing Bias in Norwegian Equity Research Reports Covering the Oslo Stock Exchange

Torstein Øye Birgirsson & Sander Eide Dahling

Supervisor: Christian Langerfeld

Master Thesis, Economics and Business Administration

Major: Business Analytics & Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

We express our gratitude to our supervisor, Christian Langerfeld, for his contribution through valuable insight and thorough feedback. We would like to extend our gratitude to Carnegie, DNB Markets, Fearnley Securities and Pareto Securities for their collaborative efforts in providing us with an exciting dataset. Finally, we are grateful for the love and support we have received from our family and friends throughout the writing of this thesis.

Norwegian School of Economics

Bergen, December 2023

Torstein Øye Birgirsson

Sander Eide Dahling

Abstract

This thesis employs XGBoost and linear regression to reveal whether short-term stock price trends can explain the variance in textual sentiment in Norwegian equity research reports covering the Oslo Stock Exchange, investigating whether trend-chasing bias is present in equity research. The thesis is based on 2,350 equity research reports from the past 5 years covering the 25 largest companies by market capitalization listed on the Oslo Stock Exchange, published by Carnegie, DNB Markets, and Pareto Securities.

We present empirical evidence demonstrating an enhancement in the predictive efficacy of our model with the integration of short-term stock price trend indicators. Specifically, the incorporation of these indicators resulted in a 2.2% increase in the linear regression model's explanatory power compared to our reference model. The full model can account for 44.7% of the variance in textual sentiment. Further, the XGBoost model improves predictive accuracy over the linear model and returns the lagged sentiment, investment bank, financial leverage, and RSI to be the most important variables explaining sentiment, chronologically ordered by variable importance. The 3-month simple return and MACD prove to be similar in variable importance with traditional valuation metrics such as the P/E ratio and firm size. Thus, we find that stock price trend indicators improve the models capacity to explain the sentiment of an equity research report.

However, our findings cannot state that the given dependency is due to trend-chasing bias in Norwegian equity research. The textual sentiment is determined by numerous unobservable variables, making it likely that our model suffers from omitted variable bias, thus causing endogeneity issues. Further, we cannot determine if a change in textual sentiment is attributable to a measurable change in the perception of a company, or the fact that the reports summarize and relay market information.

Keywords – Equity Research, Target Price, Textual Sentiment, Trend-Chasing Bias

Contents

List of Figures

List of Tables

1 Introduction

Under market uncertainty, institutional and private investors refer to equity research analysts to provide valuable insight. These expert analysts provide buy, hold, and sell recommendations for securities with an associated target price, justifying their opinions through equity research reports. However, history has shown that even seasoned analysts struggle to beat the market and discern emerging market bubbles, as witnessed during the dot-com bubble and the global financial crisis (Tuckett, 2009). The prevalence of biases among financial analysts is well-documented in existing literature, yet the underlying cause remains a relatively unexplored area. The relevance of studying analyst bias is heightened as more people than ever are getting into the stock market (Aksje Norge, 2023) and there is an ongoing discussion on whether we are entering bubble territory (Forbes, 2023). Several studies refer to the practice of extrapolating short-term growth to long-term predictions as a root cause for bias in analysts' opinions (Barberis et al., 2018), often referred to as trend-chasing bias.

Previous research uses buy, hold, and sell recommendations in addition to target prices to identify analyst bias (Clarkson et al., 2020). However, a buy, hold, or sell recommendation is a vague classification of opinions, and target prices exhibit *stickiness*, reducing its utility to accurately identify analysts' current opinion on a stock (Bonini et al., 2010). We suspect that there might be more information conveyed in their writing than observed through their recommendations. Hence, we present an alternative approach by examining the presence of trend-chasing bias on the Oslo Stock Exchange through a sentiment analysis on Norwegian equity research reports through the following research question:

Do short-term price trends impact the textual sentiment in equity research reports from Norwegian investment banks on the Oslo Stock Exchange?

2 Background

2.1 Introducing Equity Research Reports

Equity research reports (ERR) provided by sell-side analysts establish price targets and buy/hold/sell recommendations for stocks, with the frequency of publication ranging from once to several times annually per covered company. Within an ERR, analysts discuss the potential of a given security and justify a price target, facilitating textual data reflecting the subjective opinions of analysts. Investment banks often have equity research departments. We refer to the analyst companies who make ERRs as *investment banks* throughout the thesis, but be aware that investment banking and equity research differ.

Put simply, analysts provide buy recommendations for companies they believe will perform well and sell recommendations for those expected to underperform. However, there is no universal standardized ranking system for recommendations, leading to discrepancies in interpretation. What one investment bank categorizes as a *buy* may be considered a *hold* by another, highlighting the subjective nature of ERRs. Due to these discrepancies, investment banks are obliged to define their recommendation structure in every ERR, usually under a *Disclaimer* section. Table 2.1 illustrates the recommendation structure employed by the investment banks within our dataset (Carnegie, 2021; DNB Markets, 2016; Pareto Securities, 2023a).

Analysts employ a multifaceted approach while establishing price targets and associated recommendations. Their methodology encompasses several valuation techniques, including cash flow analysis and multiples to assess a security's intrinsic worth. Moreover, as explicitly stated in Pareto Securities' ERRs, analysts may integrate "behavioral technical analyses of underlying market movements" (Pareto Securities, 2023b, p. 4). This valuation technique incorporates both behavioral and subjective opinions into the analytical process, which, while insightful, also facilitates potential bias in their recommendations.

There is no universal reporting standard for ERRs, causing each investment bank to have its own tailored format. Reports are usually crafted in a word processing software and exported to PDF. While this choice of format helps maintain the visual integrity of the report, it can present challenges when attempting to extract data from the document.

Rec.	**Definition**
Buy	Upside of at least 10% to the target price and with an attractive risk/reward profile
Hold	Neutral risk/reward profile or the stock is trading relatively near its target price
Sell	Unattractive risk/reward ratio as the stock is trading above its target price

(a) Carnegie

Rec.	**Definition**
Buy	Expected return greater than 10% within 12 months
Hold	Expected return between 0 and 10% within 12 months
Sell	Expected negative return within 12 months

(b) DNB Markets

Rec.	**Definition**
Buy	Expects total return to exceed 10% over the next 12 months
Hold	Expects total return to be between -10% and 10% over the next 12 months
Sell	Expects total return to be negative by more than 10% over the next 12 months

(c) Pareto Securities

Table 2.1: Recommendation Structure

ERRs can be obtained from both primary and secondary sources. Primary sources include investment banks and independent research providers, which often produce in-depth reports on specific companies or sectors for their clients. These reports are distributed to individual clients and institutional investors, such as hedge funds and mutual funds. Most ERRs are behind paywalls and financial institutions offer access to their reports through subscription services or proprietary platforms such as Bloomberg and Capital IQ. Secondary sources include financial news outlets, specialized financial websites, and online databases. A plethora of articles referring to a change in the target price of a security can be found in newspapers such as *Dagens Næringsliv* and *Finansavisen*. However, these sources may only provide summaries or excerpts of ERRs, not the full reports justifying their recommendation.

ERRs can be distributed as a traditional sell-side analysis or as commissioned research. Naturally, there have been concerns regarding the impartiality of commissioned reports (Gunvaldsen & Walmann, 2021), which the EU directive MiFID II tries to mitigate through investor transparency, requiring unbundling of the payments related to the research (Shearman & Sterling, 2023). We will solely focus on traditional sell-side analysis in this thesis.

2.2 Literature Review

2.2.1 Implicit Bias in Equity Research

Professional investors and equity research analysts have incentives to perform well and make accurate forecasts (Chevalier & Ellison, 1997). However, Buxbaum et al. (2021) find that stock analysts suffer from optimism bias, making their target prices inaccurate while demonstrating a clear preference for buy recommendations compared to sell recommendations. Bradshaw et al. (2013) found that 38% of target price predictions are met at the end of a 12-month forecast horizon, concluding that analysts, at best, have limited abilities to furnish accurate forecasts. The same study found that analysts on average expected a return rate 15% above the actual return with a target price forecast error averaging at 45%. This can partly be explained by the use of valuation heuristics instead of sophisticated models (Gleason et al., 2013).

Assuming target prices accurately reflect the intrinsic value of a stock, we should expect the target price to detect over- and underpricing in the market, consequently being negatively correlated with the market sentiment (Buxbaum et al., 2021). However, Clarkson et al. (2015) find a positive correlation between target prices and market sentiment, suggesting that analysts are biased. In their empirical study, they discuss optimism bias and the use of less sophisticated valuation methods as two potential explanations for the result.

On the other hand, studies have found evidence supporting analysts and institutional investors outperforming the market on a risk-adjusted basis, concluding that it is a result of skill and not luck (Bhootra et al., 2015). This result holds even when accounting for transaction costs with abnormal returns of 20 to 26 basis points relative to a representable benchmark, as documented by Puckett & Yan (2011).

While some analysts seem to outperform the market, Bonini et al. (2010) suggest that the accuracy of target price forecasts is limited and tainted by systematic biases. According to their findings, prediction errors exhibit "consistency, autocorrelation, non-mean reversion", and can be "substantial" (Bonini et al., 2010, p. 2), reaching up to 36.6%. Moreover, forecasting errors tend to escalate with the size of the firm and the predicted growth in stock prices. It was further identified that momentum and forecasting accuracy exhibit a negative correlation, implying that price trends introduce more bias in stock forecasting.

According to Shefrin & Belotti (2007), investors' expectations for future returns on the S&P 500 are dependent on the returns the previous year. This contradicts rational behavior as the previous year's return has limited predictive power for the subsequent year's returns. Shefrin explains this result through both individual and professional investors suffering from *recency bias*, albeit in distinct ways. Individual investors often fall prey to the *hot-hand fallacy*, erroneously assuming that past success guarantees future gains, while professional investors succumb to the *gamblers' fallacy* and the *law of small numbers*. They tend to place excessive faith in *mean reversion*, the idea that historical returns gravitate towards the long-term average. However, this tendency may not hold in the short term. Intriguingly, the target prices in this study suggest a belief in short-term reversals. After a year of positive returns, professionals are inclined to anticipate a subsequent year of negative returns, aligning with the gambler's fallacy as they aim to conform to the long-term mean. Fong (2014) explains how behavioral biases, such as *trend-chasing bias*, make individuals treat the stock markets as a casino.

Nofer (2015) found that following the recommendations from a large group of members in stock prediction communities on the internet yielded a trading strategy that outperformed institutional investors, making the *wisdom of crowds* better than professional recommendations. This result is emphasized by the study of Bodnaruk & Simonov (2015) and Hon-Snir et al. (2012) suggesting that financial expertise provides limited value, stating that they find no evidence of analysts outperforming or exhibiting lower behavioral biases than individual non-professional investors.

2.2.2 Sentiment Analysis in a Financial Context

Textual analysis is a field of study which aims to extract information from human generated textual data. Analyzing textual sentiment is a common way of decoding the general attitude of a given text based on the context and word choice used to convey information in a textual format (Pang & Lee, 2008). Sentiment analysis within finance leverages natural language processing techniques to assess the textual tone of financial data, news, and social media content (Peng, 2020). By analyzing textual sentiment, investors and financial institutions can make informed decisions about their investments. Positive sentiments often indicate confidence in a particular asset or market, while negative sentiments may signal caution or skepticism. Sentiment analysis challenges the efficient market hypothesis by suggesting that sentiment holds information about future market movements, potentially leading to short-term deviations from full market efficiency (Chowdhury et al., 2014).

Sentiment analysis has become an attractive research field within natural language processing (Cui et al., 2023), but whether sentiment analysis is useful in financial markets is not evident. On one hand, the literature finds evidence of sentiment analysis-based models outperforming the market (Nguyen et al., 2015), while on the other hand, we find literature referring to sentiment analysis as "virtually useless in financial forecasting" (Lai, 2023, p. 2). On the contrary, studies advocate for sentiment analysis as an informative tool within finance. Kearney & Liu (2014, p. 3) emphasize sentiment analysis as an "increasingly important approach" for behavioral finance and revealing inherent bias in stock forecasting, necessitating the need for further research utilizing sentiment analysis.

2.2.3 Detecting Stock Market Trends

Price trends can be used as part of an investment strategy, but there is no standardized method to depict a price trend (Investopedia, 2021). An intuitive way of depicting a trend is visual inspection. While visual inspection is useful, its efficacy diminishes for research purposes. Consequently, technical analysis emerges as a noteworthy approach. Research shows that visual inspection is inferior to technical analysis, arguing that technical analysis can capture small nuances indiscernible to the naked eye (Hojem & Ottenbacher, 1988). Technical analysis is a trading strategy based on studying statistical trends from price and volume data. Advocates of technical analysis believe that information about the future is embedded in price trends, utilizing this to predict price movements (Hayes, 2022).

Various technical indicators exist, including the average directional index, Bollinger Bands, and simple moving averages. However, among the most prominent indicators are the Moving Average Convergence Divergence (MACD) and the Relative Strength Index (RSI) (Chong & Ng, 2008; Frade, 2019). Since the inception of the RSI and MACD in the late 1970s, it has become popular amongst traders and analysts (Chong et al., 2014). Despite its adoption, the indicators have received limited scholarly scrutiny, resulting in inadequate literature assessing their reliability (Ülkü & Prodan, 2013). However, this is related to their predictive power, not their ability to define established trends.

Over the years, extensive literature has assessed the profitability of technical analysis. A comprehensive literature review conducted in 2004 revealed that approximately 63% of the examined studies presented corroborative findings affirming technical analysis' capacity to yield a positive *alpha*, while around 26% reported a negative *alpha* (Park & Irwin, 2004), which is a portfolio's performance compared to a representable benchmark (J. Chen, 2023). The remaining 11% of the studies did not yield statistically significant outcomes. However, despite the predominantly positive outcome, the paper also critiques some of the methodological approaches used, ultimately advocating for further research addressing the deficiencies in testing to provide any conclusive evidence regarding its profitability.

Han et al. (2013, p. 3) assessed technical analysis' profitability relative to the capital asset pricing model and the Fama-French three-factor model, revealing abnormal returns of "great economic importance". However, such strategies necessitate frequent transactions. Upon factoring in the associated trading costs, Bessembinder and Chan (1998) observed an incongruity: the inferred break-even cost was lower than the transaction costs prevalent in the market. This disjunction makes it economically unviable to obtain a consistent alpha. This is in line with the vast literature on technical analysis, exemplified by the seminal works of Fama and Blume (1966) and Jensen and Benington (1970), which arrived at the consensus that technical analysis does not provide substantive utility.

2.3 Theory

2.3.1 Efficient Market Hypothesis

In 1965, Paul Samuelson and Eugene Fama released one of the most essential hypotheses in financial literature, the efficient market hypothesis (EMH). The EMH argues that the price of a security reflects all information, entailing that no trading strategies can generate a consistent alpha (Jones & Netter, 2008). New information is quickly incorporated into the market, making the next price move unpredictable, acting as a *random walk*.

The EMH refers to three forms of strength: weak, semi-strong, and strong form. The weak form posits that the stock price reflects historical price data, consequently invalidating technical analysis' efficacy under the weak form EMH (Hudson et al., 1996). If the semi-strong form holds, the utility of fundamental analysis is invalidated, as this form suggests that all publicly available information is already reflected in the stock price.

Accordingly, investors can only derive abnormal returns from information not readily accessible to the public (Maverick, 2023). The strong form asserts that neither public nor insider information can be exploited to generate profits, positing that the current stock prices accurately reflect the true value of all stocks.

Extensive research offers empirical evidence of the EMH being violated in the short term (Naseer & Bin Tariq, 2015), documenting a "clear, strong, persistent, well-documented momentum effect" in returns (Shefrin and Belotti, 2007, p. 7). The core of technical and fundamental analysis is to find deviations from the weak and semi-strong form EMH, respectively, and exploit that information to make consistent profits (Teall, 2022).

2.3.2 Behavioral Biases

The hot-hand fallacy is a cognitive bias that involves the mistaken belief that a person who has experienced success with a random event, is more likely to continue that success in subsequent attempts (Gilovich et al., 1985). This fallacy suggests a perception of streaks or patterns in random sequences where, in reality, each event is independent of the previous ones. The hot-hand fallacy can thus be compared to *trend-chasing bias*, which is when investors believe past returns can predict future returns (Fong, 2014).

The *gambler's fallacy* is the mistaken belief that prior independent events in random processes impact future outcomes (Kenton, 2023). If a coin repeatedly lands on heads, an incorrect assumption is that tails are now more likely. Connected to this fallacy is the *law of small numbers* bias (Tversky & Kahneman, 1971), assuming that sample observations should conform to broader statistical probabilities. These fallacies reveal the human tendency to misjudge probabilities and seek patterns where they may not exist.

Recency bias is the tendency to overemphasize recent events relative to older events, assuming that current conditions persist (Lee et al., 2008; Tversky & Kahneman, 1973). This bias can cause overreactions to short-term fluctuations and suboptimal decisions.

Optimism bias is the tendency to overestimate positive events and underestimate negative events (Nikolopoulou, 2023). Optimism bias is well-documented in behavioral finance. Investors affected by optimism bias may believe that their investment portfolios are less susceptible to losses than objective assessments would suggest, potentially resulting in riskier financial decisions.

2.3.3 Trend Indicators

The MACD is a trend-following momentum indicator (Dolan, 2023) utilizing exponential moving averages (EMA). The indicator consists of the *MACD line* and the *signal line*, studying their relationship. The MACD line is given by the difference between the 26-period EMA and the 12-period EMA, while the signal line is the 9-period EMA of the MACD line (Chong et al., 2014). In a trending market, the faster 12-period EMA is more responsive to price changes compared to the 26-period EMA (Aspray, 2011), consequently entailing fluctuations. A buy indication occurs when the MACD line exceeds the signal line, and vice versa. The EMA of window length N is calculated accordingly with Chong et al. (2014), shown in Equation 2.1.

$$EMA_t(N) = \left[\frac{2}{N} * (P_t - EMA_{t-1}(N))\right] + EMA_{t-1}(N) \tag{2.1}$$

The RSI is another prominent movement oscillator used to detect price trends (Tretina, 2023). The indicator created by J. Welles Wilder measures both the speed and rate of changes in the price of a security on a scale from 0 to 100. In Wilder's traditional model, an RSI above 70 is used as an overbought signal, while an RSI below 30 indicates that the security is trading below its intrinsic value. A higher RSI indicates a stronger positive trend, while lower RSI values suggest the opposite. Utilizing the recent 14 trading days' average gains and losses, the RSI can be calculated accordingly with Equation 2.2.

$$RSI = 100 - \left[\frac{100}{1 + \left(\frac{Average\ Gain\ last\ 14\ days}{Average\ Loss\ last\ 14\ days}\right)}\right] \tag{2.2}$$

A simple, yet popular metric for depicting a trend, is *simple returns.* Simple returns provide a straightforward measure of the overall performance. By utilizing uncomplicated metrics, we tap into an approach aligning with the cognitive tendencies of mental shortcuts, allowing for a more instinctive interpretation of trends. Simple returns at time t are given by Equation 2.3, where P_t is the stock price at time t and $D_{t-1,\,t}$ is defined as total dividends paid out from $t-1$ to t.

$$r_t = \frac{(P_t - P_{t-1}) + D_{t-1,\,t}}{P_{t-1}} \tag{2.3}$$

2.3.4 Machine Learning

Machine Learning (ML), a subset of artificial intelligence, is a powerful tool for analyzing and interpreting complex datasets. It is particularly useful when uncovering patterns or relationships within data, such as investigating whether the sentiment of ERRs is influenced by market trends. In the context of ML, we often deal with a quantitative response Y and a set of predictors $X_1, X_2, ..., X_p$. The relationship between Y and X is expressed as $Y = f(X) + \epsilon$, where f is an unknown function that represents the systematic information X provides about Y, and ϵ is a random error term (Witten & James, 2013). This simple model provides the framework upon which most ML techniques are built.

Estimating the function f is crucial for two primary reasons: prediction and inference. For prediction, we use known inputs X to estimate an output Y, even when Y is not easily obtainable. This process often treats the estimated function $\hat{f}$ as a black box, focusing on its accuracy in predicting Y. For example, if X represents market trends and Y represents analyst sentiment, we could predict how sentiment might change based on observable market conditions (Witten & James, 2013).

Inference, on the other hand, is about understanding the relationship between Y and X. Here, the exact form of $\hat{f}$ tells us how different predictors are associated with the response. This is particularly important in our scenario, as we want to understand which aspects of market trends significantly influence ERR sentiment (Witten & James, 2013).

ML approaches to estimating f can be broadly categorized into parametric and non-parametric methods (Cox, 2006; Witten & James 2013). Parametric methods, such as Ordinary Least Squares (OLS) assume a specific form for f, such as a linear relationship, which simplifies the problem of estimating a set of coefficients. Non-parametric methods, Such as Extreme Gradient Boosting (XGBoost) do not make explicit assumptions about the form of f, offering more flexibility at the cost of increased computational complexity.

In this study, we utilize the OLS method due to its interpretability and minimal computational requirements, alongside an XGBoost model known for its proficiency in handling complex, non-linear relationships. This approach allows us to conduct a comparative evaluation of the models, focusing on their performance relative to their complexity and ease of interpretation.

2.4 Models

2.4.1 Linear Regression

Linear regression is a supervised ML model, which is primarily used for predicting a quantitative response. It is based on the assumption of a linear relationship between the dependent variable Y and one or more independent variables X. In its simplest form, known as simple linear regression, the model predicts Y as a linear function of a single predictor variable X, expressed as $Y = \beta_0 + \beta_1 X + \epsilon$, where β_0 is the intercept, β_1 is the slope coefficient, and ϵ represents the error term (Freedman, 2009).

OLS is the most common method used to estimate the coefficients β_0 and β_1 in linear regression. The OLS approach minimizes the sum of the squared differences between the observed values of Y and the values predicted by the linear model. This method yields unbiased, consistent, and efficient estimates under the classical linear regression assumptions, which include linearity, independence, homoscedasticity, and normality of the error terms (James H. Stock, 2020).

The strength and significance of the relationship between the variables are assessed using various statistical tests and metrics, such as the R-squared and p-value[1]. Linear regression is extensively used in various fields due to its simplicity and interpretability. However, its application is limited to situations where a linear relationship is a reasonable assumption. In cases where this assumption does not hold, other methods such as non-parametric models may be more appropriate.

2.4.2 Extreme Gradient Boosting

XGBoost is a refined implementation of gradient boosted trees, conceptualized by Chen & Guestrin (2016). It employs a non-parametric approach, known as gradient tree boosting, to sequentially construct decision trees aimed at correcting the residuals of previous trees (Witten & James, 2013). This process enhances the model's accuracy by focusing on unexplained variances from existing trees. It essentially combines the predictive power of a set of weak learners, namely regression trees, into a strong predictive model, which is called boosting (Sutton, 2005).

[1] These metrics are discussed in Section 4.3.1.

Initially, predictions for each observation are the mean of the response variables, with uniform weight distribution. These weights are dynamically adjusted in subsequent iterations, increasing for observations with larger prediction errors, thus directing the model's attention to these data points. XGBoost's iterative refinement of accuracy is shown in Equation 2.4

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta \cdot f_t(x_i) \tag{2.4}$$

where $\hat{y}_i^{(t)}$ denotes the prediction at iteration t, $f_t(x_i)$ is the prediction of the new tree, and η represents the learning rate (T. Chen & Guestrin, 2016).

A unique aspect of XGBoost is the integration of a regularized learning objective $\mathcal{L}(\phi)$, shown in Equation 2.5, encompassing a differentiable loss function $l(\hat{y}_i, y_i)$, which assesses the model's predictive accuracy, and a regularization term $\Omega(f)$ to penalize model complexity.

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{2.5}$$

$$where\, \Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{2.6}$$

The regularization term $\Omega(f)$, detailed in Equation 2.6, is a penalty on the number of leaves in a tree, represented by T, modulated by the γ parameter, and an L2 regularization term on the leaf weights, controlled by λ (T. Chen & Guestrin, 2016). This term penalizes the complexity of the model by discouraging large leaf weights and an excessive number of leaves, maintaining a balance between model complexity and generalization ability.

2.4.3 Hyperparameter Tuning in XGBoost

The effectiveness of the XGBoost algorithm relies on the fine-tuning of its hyperparameters, a process essential for achieving the right balance between model accuracy and complexity. Tuning is crucial to prevent overfitting and underfitting, ensuring the model generalizes well to new data (T. Chen & Guestrin, 2016). Essential to this process is the estimation of predictive loss, $\mathcal{L}(\phi)$, as outlined in Equation 2.5. This estimation, a second-degree approximation of the model's residuals, is fundamental when evaluating and adjusting the algorithm's performance.

An iterative testing of parameters is conducted during each tuning phase (Witten & James, 2013). Key among these is the *nrounds* (T) parameter, which determines the maximum number of trees in the model's ensemble. The optimal combination of parameters is identified by minimizing this loss function, thereby enhancing predictive accuracy. The Gamma (γ) parameter is pivotal in controlling the growth of these trees, setting a threshold for the minimum improvement in loss required for additional tree growth, thus preventing overcomplexity and overfitting.

The structure and composition of the trees are influenced by several parameters (Witten & James, 2013). The *max_ depth* parameter, for instance, regulates the depth of the trees. While deeper trees can model more complex patterns, they also risk overfitting; shallower trees, on the other hand, might be too simplistic and underfit. The *colsample_ bytree* parameter determines the fraction of features sampled for each tree, influencing the diversity of feature selection. The *subsample* parameter determines which fraction of the training data to use in constructing each tree, affecting the model's exposure to data patterns. The *min_ child_ weight* parameter determines a threshold for when to stop splitting the trees into further child nodes, which is crucial for determining the decision-making depth of each tree.

Lastly, the learning rate (λ) is a critical hyperparameter that influences the sequential addition of trees to the model using L2 regularization (Cortes et al., 2012). A lower learning rate ensures that each new tree has a smaller impact, leading to a more robust model. This necessitates a larger number of trees (T), highlighting the trade-off between model complexity and computational efficiency (Witten & James, 2013). This comprehensive tuning approach, encompassing a range of hyperparameters shown in Table 2.2, is instrumental in optimizing the XGBoost algorithm (XGBoost Developers, 2023).

Hyperparameter	**Description**
eta	Learning rate of the algorithm
gamma	Threshold for further splitting a tree leaf node.
max_depth	Maximum tree depth
min_child_weight	Splitting threshold for each tree node
subsample	Subsample ratio of the training instances
colsample_bytree	Subsample ratio of columns
nrounds	Number of decision trees

Table 2.2: XGBoost Hyperparameters Description

3 Data

3.1 Data Sources

3.1.1 Equity Research Reports

Through collaborative efforts with Carnegie, DNB Markets, and Pareto Securities, we have compiled a dataset comprising 2,350 ERRs spanning the past five years. The dataset covers the 25 largest companies on the Oslo Stock Exchange (OSE), as determined by market capitalization on the 5th of September 2023.

When selecting investment banks for our research, we prioritized broad equity research coverage of the firms on the OSE. Carnegie, DNB Markets and Pareto Securities are all among the market leaders in Norwegian equity research, with extensive coverage of the largest companies on the OSE. To access and download the ERRs, we entered into individual agreements with each of the aforementioned investment banks, gaining access to their research portals, namely Carnegie Edge, DNB Alpha, and Pareto Securities Research Portal. The data was manually downloaded and imported into our database.

The dataset extends from January 1st, 2018, to September 4th, 2023. Selecting a five-year time frame was motivated by the practice of investment banks to remove ERRs from their research portals after five years. For instance, DNB Alpha lacks reports before Q4 2018, and although Carnegie and Pareto Securities have some ERRs older than five years, the number of reports significantly diminishes in comparison to recent fiscal quarters[2].

Notably, Fearnley Securities also gave us access to their research portal, but based on the aforementioned criteria for choosing investment banks, we made a strategic decision to exclude their ERRs from our analysis. Fearnley Securities, known for their expertise in the maritime and energy sector, only covers 11 out of the 25 companies we have selected. Including their reports could potentially skew our results, given their sector-specific focus and the uneven coverage of the listed companies. This decision was crucial to maintain the integrity and representativeness of our study, ensuring that our analysis reflects a more balanced and comprehensive view of the market.

[2]Detailed further in Figure 3.1 and Section 3.4.

Ensuring a representative and robust selection of companies on the OSE with wide coverage was essential when selecting companies to include in our analysis. Integral to our selection criteria was market capitalization. By focusing on the 25 largest companies on the OSE, we aimed to capture a representative snapshot of the stock exchange's movements. These companies represent 75.96% of the total OSE market capitalization (Euronext, 2023).

The inherent trade-off between the scope of data and its representativeness was a significant consideration. While a larger number of companies might offer more data points, it is paramount that the data accurately reflects broader market trends. We observed that many smaller companies on the OSE had limited coverage compared to their larger counterparts. This disparity could introduce biases, especially if smaller sample sizes led to an increased number of outliers. By emphasizing larger companies, we not only ensured a more consistent and reliable dataset but also mitigated potential biases.

We have omitted certain reports from DNB Markets pertaining to DNB and Kongsberg Gruppen due to conflicts of interest. DNB released 29 reports on their own performance creating a situation where their assessment may be influenced by their own stake in the matter, potentially compromising objectivity. Furthermore, DNB acted as an advisor in the potential listing of Kongsberg Digital in 2022, introducing a similar conflict of interest. These conflicts of interest could undermine the impartiality and reliability of two of the reports written in this time period, thus necessitating their omission. In total, the number of reports included in our analysis amounted to 2,319.

The complete set of chosen firms used as a basis for this thesis is presented in Table 3.1. Our selection process was underpinned by the desire for analytical robustness. The utility of choosing this set of firms lies in its ability to offer meaningful insights, and by selecting companies with substantial market capitalization and ample data coverage, we are confident that our set of selected companies provides a comprehensive and accurate reflection of the general use of ERRs in Norway. We posit that the validity of our findings remains consistent despite the significant variance in the number of reports across companies. Some entities have an extensive collection of over 157 reports, while others possess a minimal count of 7. It is important to note that while there are outliers in terms of report volume, the majority of companies exhibit a comparable number of associated reports.

Ticker	Company	Start Date	End Date	Observations
ADE	Adevinta	09.01.2020	01.09.2023	88
AKER	Aker ASA	14.05.2018	16.06.2023	54
AKRBP	Aker BP	15.01.2018	14.07.2023	132
AUTO	Autostore	06.07.2022	18.08.2023	35
BAKKA	Bakkafrost	07.05.2018	23.08.2023	104
DNB	DNB	01.02.2018	12.07.2023	94
EQNR	Equinor	08.02.2018	27.07.2023	129
GJF	Gjensidige	26.01.2018	14.07.2023	101
KOG	Kongsberg Gruppen	07.02.2018	13.07.2023	123
LSG	Lerøy Seafood	27.02.2018	24.08.2023	96
MOWI	Mowi	14.02.2018	28.08.2023	157
NHY	Norsk Hydro	16.02.2018	22.08.2023	122
ORK	Orkla	08.02.2018	14.07.2023	149
SALM	Salmar	15.05.2018	25.08.2023	107
SCHA	Schibsted	08.02.2018	18.07.2023	121
SRBNK	SpareBank 1 SR-Bank	08.02.2018	10.08.2023	53
STB	Storebrand	07.02.2018	14.07.2023	138
SUBC	Subsea 7	02.03.2018	26.07.2023	52
TEL	Telenor	01.02.2018	21.07.2023	106
TOM	Tomra	25.04.2018	17.07.2023	86
VAR	Vår Energi	14.07.2022	25.07.2023	34
YAR	Yara International	08.02.2018	20.07.2023	119
FRO	Frontline	24.04.2019	25.08.2023	50
HAFNI	Hafnia	02.02.2023	28.08.2023	7
WAWI	Wallenius Wilhelmsen	14.05.2018	16.08.2023	62

Table 3.1: Set of Selected Companies

3.1.2 Financial Data

For our research, we sourced stock data from Yahoo! Finance. Yahoo! Finance was selected due to its extensive database, ease of access, and reliability of its data (Yahoo!, 2023). Regarding financial metrics, such as leverage and P/E ratio, we sourced historical data from a Bloomberg Terminal, as this data is not readily available through public APIs (Bloomberg L.P., 2023).

Our primary focus was on the stock prices for this set of companies over a period spanning from 01.01.2016 to 04.09.2023. This amounted to 43,803 data points. The data was extracted using the *tidyquant* package in R. We set our start date for this data collection to be 2 years ahead of our main analysis timeframe, as it allowed us to extract figures which require lagged return data in our analysis.

3.2 Data Cleaning

3.2.1 Scraping the PDF Reports

The ERRs were formatted as PDFs, which is not internally structured for data extraction. The text within the PDFs was scraped to convert it to a usable format, removing extraneous formatting to maintain data integrity. To accommodate Scandinavian letters in company names, the PDF encoding was adjusted to UTF-8. While this does not affect sentiment analysis, as the ERRs are written in English, it enhances the readability of the data, where company names serve as a sorting token. The encoded textual data was added to our main corpus for further processing.

As the PDFs lack metadata, scraping the raw textual data in our corpus for key information, such as target price and recommendation, is important. One of the complexities of the data cleaning process was the lack of uniformity in the report structures. Each company have their unique formatting style for their reports, making it infeasible to employ a universal cleaning script. Consequently, the scraper was tailored to accommodate the specific style of each investment bank, ensuring the precision of data extraction.

An essential piece of information for our analysis was the report date, which is not available in a given reports metadata. We observed that most reports contained a statement similar to: "This report was completed and disseminated 14 Oct 2022: 17:40 CET". To capture the date, a function was written that scanned the first 20 lines of the PDF, extracting any recognizable date. The extracted dates were standardized, and their occurrences counted. In instances of multiple identified dates, the date with the highest frequency was selected. In the event of a tie, the observed date closest to the current date was selected to ensure relevant and accurate data. For other data points, such as recommendation and target price, a similar approach was employed.

The reports contained mandatory sections which investment banks are legally bound to include. These sections elucidate the legal constraints of their recommendations. Given their consistent nature across reports, these sections were excluded. This decision was made to prevent bias in our analysis, focusing on the unique and pertinent content of each report. Upon completion of the data extraction and cleaning process, the refined data and the sanitized PDF text were stored in a data frame.

3.2.2 Cleaning the Financial Data

The data gathered from Yahoo! Finace was directly loaded into a data frame for further processing. The Bloomberg Excel add-on was used to automate the process of downloading the necessary fiscal data for each firm, and converted it into a structured data format. It is worth noting that data from reputable sources like Yahoo! Finance and Bloomberg often comes well-structured and largely free from glaring discrepancies. However, it is a cardinal rule in data analysis to never trust data at face value. Hence, even though the data was formatted to our expectations upon download, we reviewed the data thoroughly. We found no large discrepancies in the downloaded data. We adjusted our stock prices for stock splits and dividends, as these events can introduce discontinuities in time series data. By adjusting for such events, we ensured that our dataset maintained its continuity and comparability throughout the chosen analysis period.

Ticker	**Company**	**Stock Split Date**	**Split Ratio** (κ_t)
TOM	Tomra	27.05.2022	2:1
VAR	Vår Energi	15.02.2023	11:10

Table 3.2: Stock Split Information

Yahoo! Finance incorporates stock split adjustments in its historical price data. Conversely, previously published ERRs do not undergo similar updates. Consequently, it becomes imperative to rectify the provided target prices in accordance with the stock split ratio. As a result of this adjustment, the target prices for Tomra and Vår Energi, prior to the stock split date, are now suitably transformed in order to keep the ratio the target prie and share price equal equal to when the report was published. The stock split ratios can be found in Table 3.2 and the adjusted target price $\hat{P}'_{a,t}$ is calculated as given by Equation 3.1

$$\hat{P}'_{a,t} = \frac{\hat{P}_{a,t}}{\kappa_t} \tag{3.1}$$

where $\hat{P}_{a,t}$ denotes the target price of investment banks a in time t and κ_t the split ratio of a stock[3].

[3]We treat the split ratio κ_t of n:m as $\frac{n}{m}$.

3.3 Final Dataset

3.3.1 Sentiment Score

To assess the textual sentiment, we employed a lexicon-based approach[4]. This method relies on a predefined dictionary or lexicon, where each word is associated with a polarity score. The sentiment is determined by aggregating the scores of its constituent words, as shown in Equation 3.2

$$s_q = \frac{\gamma_{q_i}^{pos} - \gamma_{q_i}^{neg}}{\gamma_{q_i}^{pos} + \gamma_{q_i}^{neg}} \tag{3.2}$$

where the textual sentiment score s of report q_i is set as s_q. This score is defined as the fraction of the difference between total positive tokens $\gamma_{q_i}^{pos}$ and total negative tokens $\gamma_{q_i}^{neg}$, and the sum of total positive and negative tokens. This produces a normalized sentiment score defined as a real number within the closed interval $[-1,\ 1]$.

A notable advantage of this approach is its independence from labeled training data. We base our sentiment score on the work done by Feuerriegel et al. (2015) with some slight modifications. As the ERRs contain tables and graphs, we opted out of standardizing the score based on word count. Although we lose some ability to measure the strength of the sentiment in a report, we believe that the negative effects of including a high degree of noise in our score would outweigh the benefits. We found that the number of loaded[5] words were highly correlated with the number of total words in a report, which tells us that we keep some ability to measure sentiment strength.

3.3.2 Target Ratio

To compare target prices, we employ the relative measure *target ratio* ρ_t, shown in Equation 3.3, drawing inspiration from De Vincentiis (2010). By dividing each target price $\hat{P}'_{a,t}$ in time t by the stock price one day prior to the ERR being published P_{t-1}, we allow for a consistent assessment of target prices regardless of varying stock levels, and ensure that new information contained in the ERR does not impact the intraday stock price.

$$\rho_t = \frac{\hat{P}'_{a,t}}{P_{t-1}} \tag{3.3}$$

[4] We go into detail about this implementation in Section 4.1.2.

[5] We define loaded words as the sum of positive and negative words in a report, i.e $\gamma_{q_i}^{pos} + \gamma_{q_i}^{neg}$.

3.3.3 Explanatory Variables

We incorporate the *RETURN_3M*, *MACD*, and *RSI* as explanatory variables, adeptly capturing short-term market trends. We decided to only use a 3-month simple return as time frames such as 1-week and 12-month returns cause multicollinearity issues, correlating with the RSI and the 3-month return respectively.

3.3.4 Control Variables

Controlling for firm-specific differences impacting opinions toward a stock, thus affecting the sentiment score, we have added several control variables. Note that some of the variables are log-transformed to address issues of skewness and heteroscedasticity, ensuring a more accurate representation of their influence on the dependent variable (Ford, 2018).

Firm size has in relevant literature been used as a proxy for the extent of publicly available information about a company (Das et al., 1998). Larger firms tend to have a more extensive and accessible pool of data, which can impact the accuracy of forecasts. Moreover, larger companies often garner more attention from investors, leading to a greater depth of coverage (Fortin, Roth, et al., 2007). In line with Al-Awadhi et al. (2020), we control for firm size through the market capitalization of a firm, *log_market_cap*.

Existing literature finds a significant correlation between firm age and beta, indicating that younger companies tend to exhibit higher levels of risk and volatility (Chincarini et al., 2020). This, in turn, can influence target price estimations, as beta is a critical factor in determining a stock's expected return. Furthermore, younger companies, often in need of external financing to sustain growth, may be incentivized to exceed earnings expectations to attract investor attention and secure necessary capital (Coad et al., 2016; Loderer & Waelchli, 2010). We account for the temporal aspect of a company's post-IPO tenure through the natural logarithm of years since IPO-date, *log_time_on_OSE*.

Addressing potential bias stemming from unobserved time-invariant factors is important. Variations in reporting standards or other idiosyncrasies can influence the dependent variable. By introducing *analyst* as a dummy variable for investment bank a, we effectively account for unobserved heterogeneity. The dummy variables *analystDNB* and *analystPARETO* are incorporated, making Carnegie the reference group.

Next, we incorporate a dummy variable *dividend* for dividend payments within the last 30 days. Capstaff et al. (2004) document significant abnormal returns on the OSE following a dividend announcement, providing support for dividend signaling theory in Norway.

The price-to-earnings (P/E) ratio is a valuation multiple of the share price divided by the earnings per share, providing a measure of how much a company is valued relative to their earnings. While reading the ERRs in our dataset, it becomes evident that P/E is an essential part of their target price, explicitly using P/E and forecasted P/E to value companies. Thus, we incorporate the control variable *log(pe_ratio)*.

Dechow & You (2020) find that increased financial leverage causes equity research analysts to assume higher returns implied through their target price. As this might impact their assessment of the stock, hence the textual sentiment, we control for *financial_leverage*.

Volatility influences the range of target prices, reflecting the inherent uncertainty in stock performance (Cho, 2014). Employing a 1-year volatility estimate coincides with the forecast horizon in ERRs, thus accounting for short to medium-term market dynamics. While sophisticated models like GARCH(1,1) can account for excess kurtosis in stock returns, the impact on our volatility estimates would likely be marginal. Thus, we opt for a simplistic approach. The control variable *volatility* is calculated through three steps. The first step calculates the continuously compounded daily return u_i, as this is preferred for stock volatility (Harper, 2023). The second step estimates the daily standard deviation σ_n on a rolling 30-day basis, while the third and final step annualizes it accordingly with Höhler & Lansink (2021). The three steps are presented in Equation 3.4, 3.5, and 3.6.

$$u_i = ln\left(\frac{P_i}{P_{i-1}}\right) \tag{3.4}$$

$$\sigma_n = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(u_{n-i} - \overline{u})^2} \tag{3.5}$$

$$\sigma_a = \sigma_n * \sqrt{252} \tag{3.6}$$

Considering the recent 5-year period characterized by heightened market uncertainty surrounding the COVID-19 pandemic, impending rate hikes, war, and recessionary fears, the VIX-index serves as a great variable to account for the extraordinary volatility. Further, literature finds that investment banks tend to issue more optimistic earnings forecasts and buy recommendations under market uncertainty, as measured by the VIX (Chang & Choi, 2017). Thus, incorporating the control variable *VIX* ensures that the model adequately captures the nuanced effects, providing a more robust and reliable analysis.

Analyst opinions tend to exhibit a degree of *stickiness* over time, driven by a reluctance to abandon beliefs (Bouchaud et al., 2019). Studies find that investment banks exhibit an excessive adherence to long-term trends, leading to a lag in their response to short-term shifts (Filiz et al., 2021). Thus, we control for the sentiment from the last ERR, *sentiment_1lag*, and the *target_ratio_1lag*, accounting for stickiness.

3.3.5 Presenting Model Dataset

Following the identification and definition of the dependent, explanatory, and control variables, we present the full dataset which forms the basis of our analytical framework. Table 3.3 illustrates the complete array of data employed in our empirical analysis[6].

Variable Name	**Definition**
Dependent Variables	
sentiment	Our calculated sentiment score
target_ratio	Target price divided by share price at $t-1$
Explanatory Variables	
RSI	Relative Strength Index
MACD	Moving Average Convergence Divergence
RETURN_3M	3 month simple return of stock
Control Variables	
log(market_cap)	Market cap of company
log(time_on_OSE)	Years since IPO
analyst	Dummy variable for investment bank
dividend	Dummy variable for dividend paid last 30 days
log(pe_ratio)	Price to earnings ratio
financial_leverage	Debt divided by equity
volatility	1-year annualized rolling volatility estimate
VIX	Volatility index
sentiment_1lag	Sentiment in previous ERR
target_ratio_1lag	Target price divided by share price for ERR in $t-1$

Table 3.3: Variables in the Model Dataset

[6]A set of summary statistics for all variables can be found in Appendix A.

3.4 Descriptive Statistics

Figure 3.1 shows the distribution of ERRs by fiscal quarter. As the number of reports from 2018 Q3 and before are not as well represented in our dataset, we see that the investment banks tend to remove reports which are more than five years old from their database which is available to customers. DNB seem to be more strict in their removal policy, as we find no ERR reports which is older than five years.

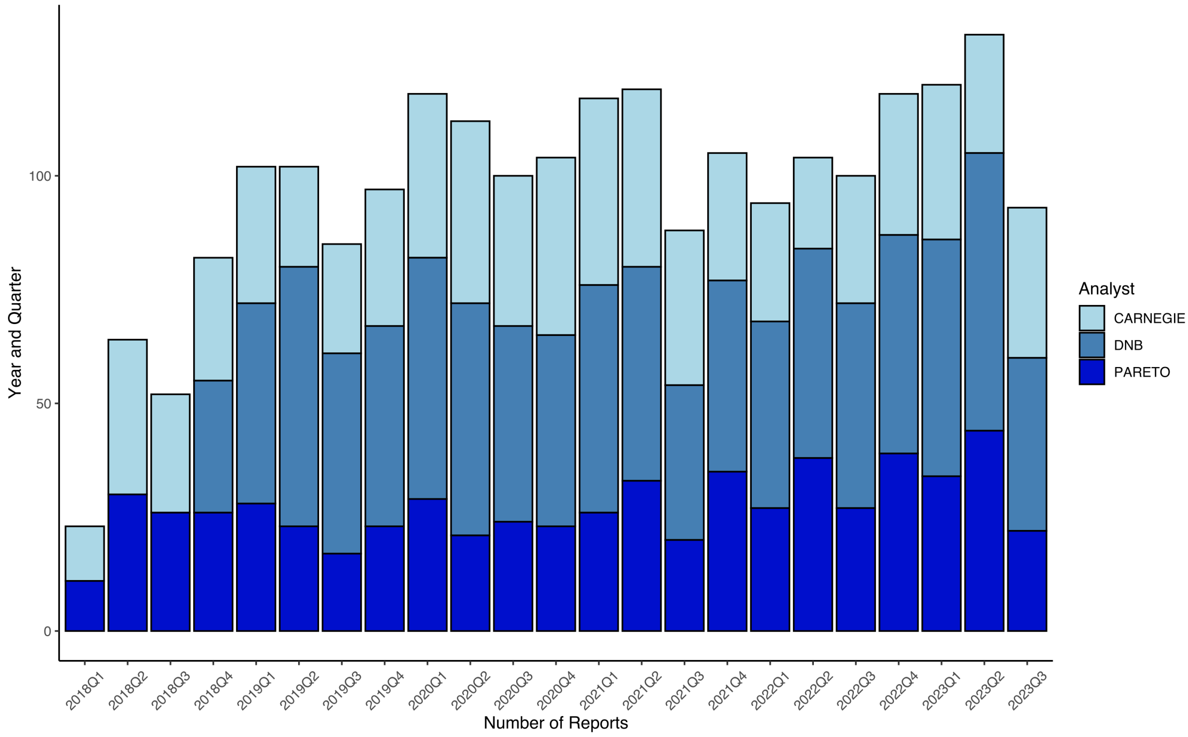

Figure 3.1: Number of Equity Research Reports Published per Quarter

For the sentiment score, defined in Section 3.2, we expect to find a distribution which is comparable to a Gaussian distribution based on the assumptions of the central limit theorem. In Figure 3.2 we find that the sentiment score is normally distributed with a mean of around -0.14, although this distribution changes if we differentiate between the investment banks. We find that each investment bank follows quite different distributions, with DNB and Pareto having comparable means, while Carnegie exhibits a much more positive distribution in terms of sentiment. This could be explained by varying policies or attitudes within each investment bank, which could be confirmed by analyzing other similar metrics. If these results are not reproducible, it could point to a weakness in our sentiment score. Additionally, the negative mean sentiment score could indicate that even though the majority of recommendations are *BUY*, the sentiment of the reports tend towards the negative end of the scale. This negative bias could be explained by the choice of sentiment dictionary, as the LM dictionary contains 6.75 times more negative words than positive words, which could skew the average sentiment of the reports.

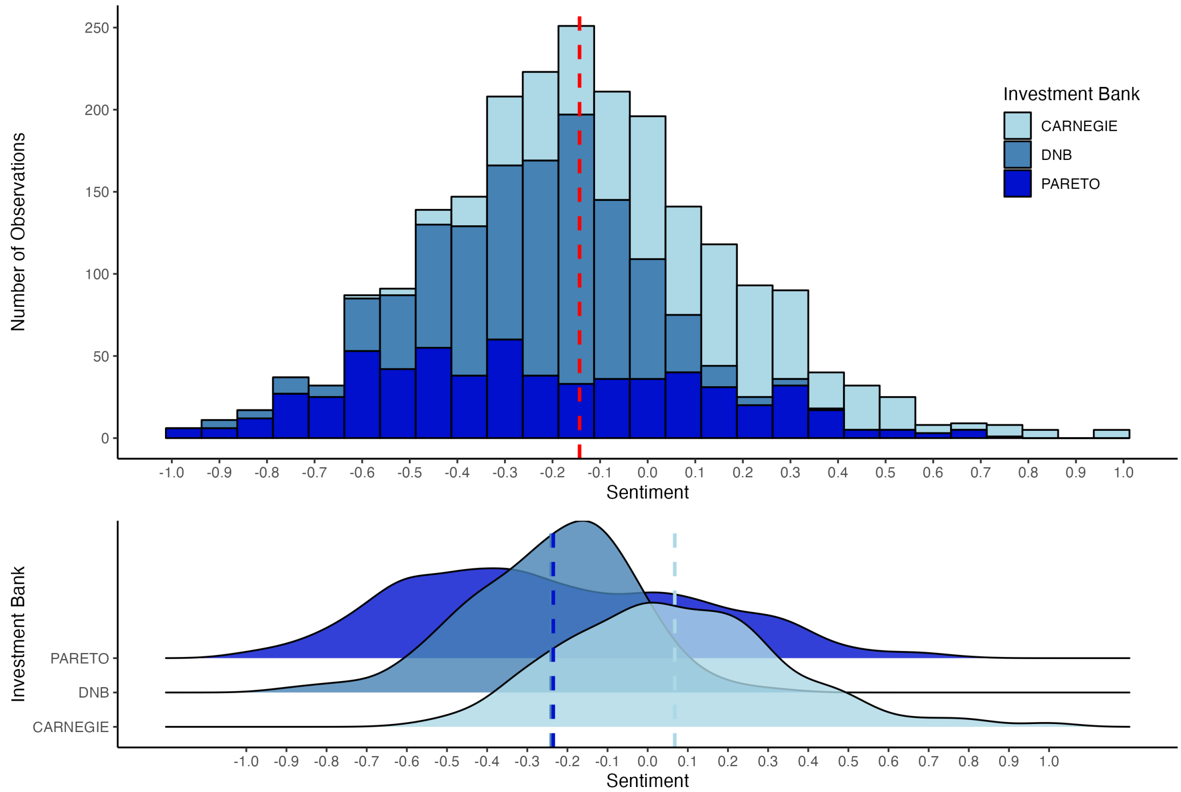

Figure 3.2: Textual Sentiment Score Distributions by Investment Bank

In Figure 3.3, we analyze the change over time of recommendation distributions from three investment banks. Statistical analysis reveals distinct recommendation patterns: Carnegie exhibits a significantly elevated *BUY*-ratio, driven by a relatively low *HOLD*-ratio, and a moderate *SELL* inclination. DNB, characterized by a higher prevalence of *HOLD* and *SELL* recommendations, shows a pronounced sentiment distribution, seen in Figure 3.2. Pareto demonstrates a dominant *HOLD*-ratio with fewer *BUY* and *SELL* suggestions, correlating with a flatter sentiment distribution in Figure 3.2, indicating a more responsive reporting approach compared to the other two companies. These distributions show us that there is a high probability of there being structural differences in how the reports are written between each investment bank, which could be the root cause behind the difference in distribution and mean of the sentiment scores.

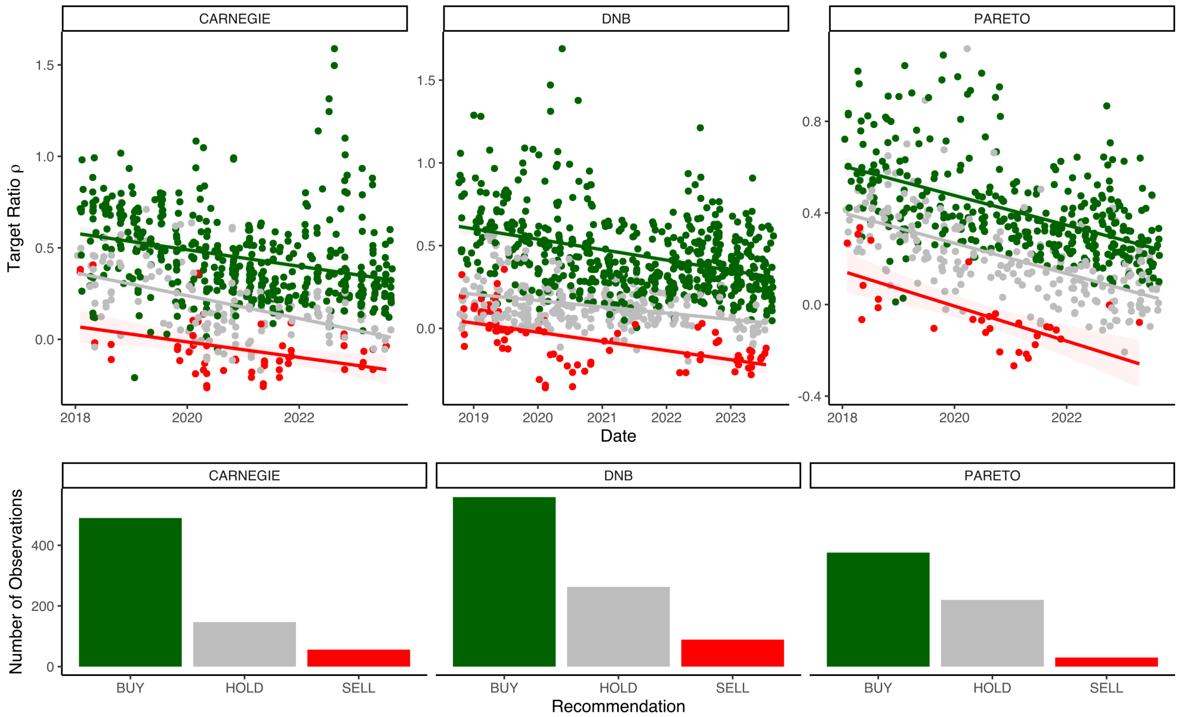

Figure 3.3: Target Ratio and Recommendation Distributions Between Investment Banks

4 Methodology

4.1 Textual Data Analysis

4.1.1 Textual Data Processing

During our data cleaning phase, all words were converted to lowercase to standardize the textual data. This uniformity ensures the reduction of potential discrepancies that might arise from capitalizations. To further refine the dataset and emphasize meaningful content, common stop-words were removed. This step is pivotal in minimizing noise and allowing a focus on the more context-specific words present in the reports. Additionally, any superfluous white space and punctuation were eliminated to streamline the data.

These reports make use of tables to a great extent, which are filled with text and numerical data. To maintain our emphasis on textual analysis, all numbers were excluded from the dataset. Financial terminologies, such as *EBITDA*, which often serve as column or row names in tables, found their context disrupted due to the removal of surrounding numbers. While such cleaning steps inevitably impact basic metrics like word count, a more profound implication emerges in sentiment analysis. Specifically, the removal of numbers can distort the perceived relationship between words. Words that were originally separated by a series of numbers might now appear adjacent, leading to potential misinterpretations as a bigram or a coherent sentence. Such alterations can introduce inaccuracies when conducting bigram or sentence-level sentiment analysis, as the modified structure might not truly represent the original context in the raw textual data.

In addressing the challenges posed by the extensive use of tables and financial terminologies in the reports, the Bag-of-Words (BoW) model was employed for sentiment analysis. This model, known for its simplicity and efficiency, was particularly suitable given the altered context of words post-data cleaning. BoW's focus on word frequency allowed for a fundamental understanding of the textual content within the reports. Despite its limitations in capturing linguistic nuances, BoW's decent accuracy and ease of implementation made it a valuable tool in this context, ensuring meaningful sentiment analysis amidst the altered textual structure.

4.1.2 Dictionaries

In the domain of financial text analysis, the selection of an appropriate sentiment dictionary is crucial for obtaining accurate and contextually relevant sentiment scores from textual data. In this area, a sentiment dictionary is defined as a compendium of words each tagged with a designated sentiment value. The choice of dictionary significantly impacts the outcomes of sentiment analysis.

Among available options, the Loughran and McDonald Sentiment Dictionary is well-suited for financial text analysis. This dictionary is distinctively designed for financial discourse, capturing the unique terminologies and nuances inherent in this field, as highlighted by Loughran & McDonald (2011). This specialization contrasts with general-purpose linguistic dictionaries, such as the General Inquirer (GI) and DICTION, which may inaccurately interpret the sentiment of financial terms. For example, terms typically perceived as neutral or positive in a financial context, such as 'tax' and 'liability', are often incorrectly labeled as negative in these general dictionaries (Li, 2010). This mislabeling can distort sentiment analysis, compromising the reliability of the research.

The Loughran and McDonald dictionary is also notable for its detailed sentiment categorization, encompassing seven distinct sentiment types. This granularity facilitates a more nuanced sentiment analysis. In our study, we focus on the dictionary's allocation of words to positive and negative sentiments, as these are most aligned with our research objectives. However, it is important to note that the dictionary contains a disproportionate number of negative words (2,345) compared to positive words (347). This imbalance could potentially introduce a skew in the sentiment analysis results, leaning more toward negative interpretations. This aspect is carefully considered in our analysis to ensure a balanced and accurate representation of sentiment in financial texts.

4.1.3 Tokenization Process

In the tokenization process of our textual data analysis, a critical step was converting the reports into document-term matrices (DTM). This transformation is essential for quantitative analysis, as it converts textual content into a structured format suitable for computational methods. A mathematical formulation of this matrix is presented in

Equation 4.1, where $x_{i,p}$ represents the term frequency of term n_p in report q_i. In this matrix, i is defined as the index of a given report, and p is defined as the index of the term. The DTM is therefore defined as a matrix of $[n_p \times q_i]$. As our dataset contains 2,319 reports and 31,517 unique terms, our matrix has the dimensions of $[2319 \times 31517]$.

$$DTM = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \dots & x_{1,p} \\ x_{2,1} & x_{2,2} & x_{2,3} & \dots & x_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{i,1} & x_{i,2} & x_{i,3} & \dots & x_{i,p} \end{bmatrix} \tag{4.1}$$

Our approach does not include *lemmatization*, a common practice in text processing where words are reduced to their base or dictionary form. This decision was guided by the use of the Loughran and McDonald sentiment dictionary, which already accounts for various grammatical endings to the word stem (Loughran & McDonald, 2011). As this dictionary is tailored for financial texts, it includes different forms of words relevant to our analysis. This aspect means that our methodology, though it bypasses the lemmatization step, does not lose the precision or depth often attributed to this process.

4.2 Model Data

4.2.1 Data Cleaning & Processing

Preparing the data for further analysis involves encoding numeric values accurately, removing missing values, and either eliminating strings or converting them into categorical factors. These preprocessing steps resulted in a reduction of the dataset by 3.8%. Consequently, our refined model dataset now comprises 2,230 observations. Some missing values (NAs) were created in the lagged variables, specifically in the lagged sentiment. Although we chose to exclude these observations, it is presumed that because of the lagged nature of the variable, the dataset retains a considerable amount of the information initially derived from the removed reports.

4.2.2 Data Partitioning

Cross-validation (CV) is essential in ML and statistical modeling for ensuring model robustness and generalizability. It combats overfitting by assessing model performance on unseen data, providing a realistic measure of predictive accuracy. This process is vital in contexts with limited or varying data, as it confirms that the model's conclusions are not mere artifacts of specific datasets. Overall, CV enhances the credibility and reliability of model findings, making it a fundamental step in model evaluation. The difference in various CV techniques usually stems from the trade-off between estimation uncertainty and computational time.

The CV method chosen for our linear regression model is a *Leave-One-Out Cross-Validation* (LOOCV) approach. This CV approach is computationally intensive, yet offers low bias even when the sample size is relatively small. LOOCV creates two subsets of the complete set of data of length n, where one subset contains a single observation (x_1, y_2) and the other contains the rest of the observations $(x_2, y_2), ..., (x_n, y_n)$ (Witten & James, 2013). The statistical learning model is then trained on $n-1$ observations, which are consequently used to predict $\hat{y}_1$ for the left-out observation. These steps are repeated n times, where the resulting performance metrics are aggregate measures of *testMSE*[7] and R^2. A visual representation of the subsetting procedure is shown in Figure 4.1.

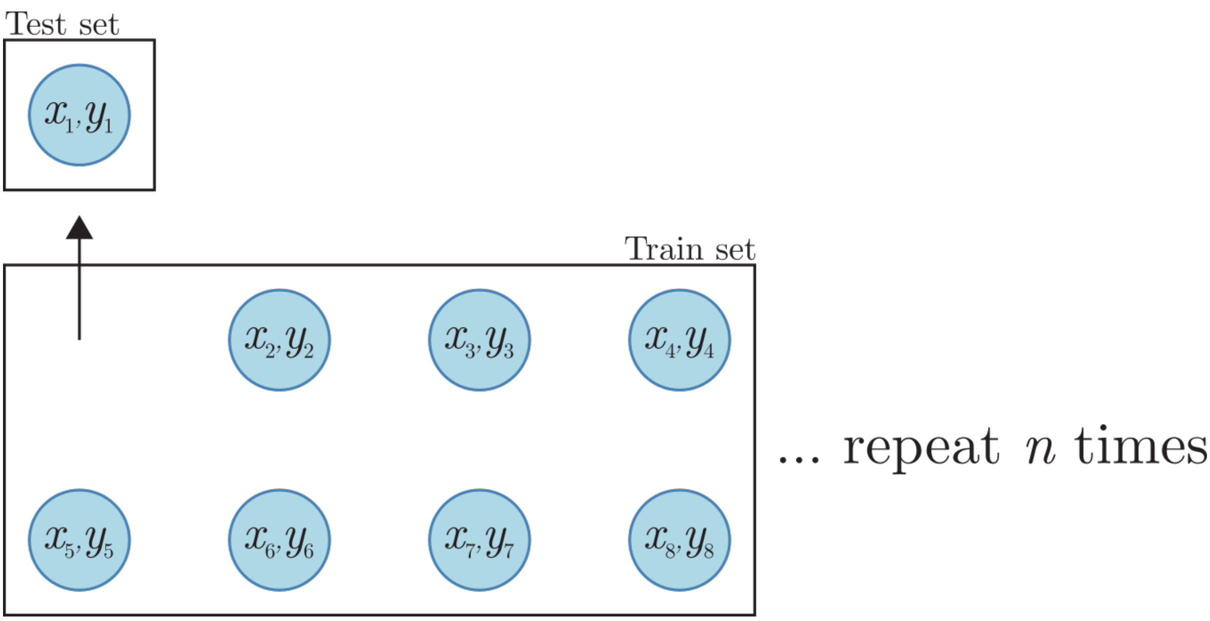

Figure 4.1: Schematic of LOOCV

[7] *RMSE* is presented in Equation 4.2.

While developing the XGBoost model, a k-fold cross-validation (k-CV) method was employed to validate its performance. Unlike the computationally intensive LOOCV, k-CV strikes a balance between computational efficiency and the reliability of the validation process. This CV method is particularly suitable for larger datasets or more computationally intensive ML models, although this is dependent on the chosen value for k, as a k-CV with $k = n$ is equivalent to a LOOCV method. This CV method can be applied to a dataset with n observations by dividing the data into subsets of size $\frac{n}{k}$. The training of the model is conducted k times, where each instance uses $k - 1$ subsets, or folds, for the training purpose. The last fold is reserved for validation.

Specifically, if the dataset is divided into k folds, then each iteration, $(k-1)\cdot\frac{n}{k}$ observations are used for training, and the remaining $\frac{n}{k}$ observations are used for validation. This process is sequentially repeated until each fold has been used once as the validation data (Witten & James, 2013). This approach is particularly advantageous in the context of an XGBoost model, which can be sensitive to the structure of the training data. By rotating the validation set through the entire dataset, k-CV ensures a representative measure of a model's performance across the whole set of data.

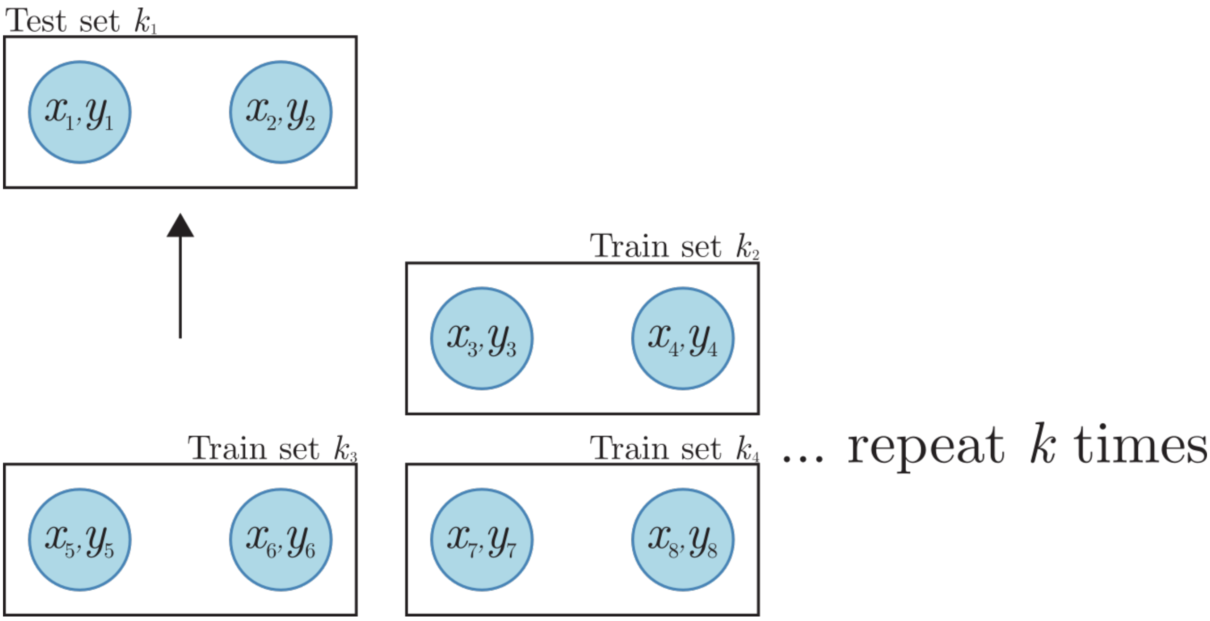

Figure 4.2: Schematic of k-Fold CV

Moreover, the choice of k is an important consideration in k-fold CV. A higher value of k generally results in a more accurate estimate of the model's performance at the cost of increased computational burden. Common choices for k include 5 or 10, balancing the trade-off between computational efficiency and validation accuracy. A graphical illustration of the k-fold cross-validation process, showcasing the division of data into folds and the rotation of these folds through the training and validation phases, is presented in Figure 4.2. The number of folds for the XGBoost model was set at $k = 5$.

4.2.3 Data Balancing

Data balancing is a critical process in ML that involves adjusting the distribution of different classes in a dataset to prevent model bias. When a dataset is imbalanced, the model may become overly attuned to the majority class, leading to poor generalization and predictive performance on under-represented classes (Jadhav et al., 2022). By ensuring a balanced distribution, data balancing helps in creating more robust ML models.

In this particular set of data, the challenge of imbalance can be more pronounced due to the temporal nature and potential autocorrelation within the data. In dealing with these kinds of panel data, it is important to consider how a random subset of an imbalanced dataset could lead to bias in the predictions of an ML model trained on the subset. The distribution of our dataset with respect to the *analyst* column is shown in Table 4.1.

Investment Bank	**Carnegie**	**DNB Markets**	**Pareto Securities**
N	675	911	626

Table 4.1: Observations by Investment Bank

In analyzing the distribution among these analysts, it is observed that while there is no extreme skewness necessitating up- or downsampling, the disparity is noteworthy. Consequently, we employed a variation of the k-CV sampling technique in the form of the stratified k-CV method. This approach guarantees that each analyst segment of the panel data retains a proportionate representation of the various classes. Such a methodical sampling strategy is essential to maintain the integrity and reliability of the data analysis, ensuring that the sample accurately reflects the population's characteristics across different time intervals. This enhances the validity of any inferential statistics derived from the dataset, as it mitigates the risk of sampling bias and improves the generalizability of the findings.

4.3 Model Performance

4.3.1 Assessing Predictive Accuracy and Inference

In order to address the dual objective of our modeling approach, encompassing both inferential analysis and evaluation of predictive performance, it is imperative to delineate the specific metrics employed for assessing the performance of the models.

When discussing inference in our OLS models, we will use a p-value significance threshold of 5%. *R-squared* (R^2) is a metric used in regression analysis to represent the proportion of variance observed in the dependent variable explained by the independent variables (Casella, 2002). R^2 values range from 0 to 1, where higher values indicate a better fit for the model. This coefficient is used to assess the explanatory power of a model, helping to determine how well it captures the underlying patterns in the data.

A notable weakness of R^2 is its tendency to increase with the addition of more variables to the model, regardless of whether those variables are truly relevant or meaningful. This can lead to overfitting, where the model becomes overly complex and starts capturing noise rather than the underlying trend in the data. Consequently, an inflated R^2 might give a misleading impression of model effectiveness. To address this, adjusted R^2 is often used as it adjusts for the number of variables in the model, providing a more accurate reflection of its explanatory power for models with a different number of predictors.

In XGBoost models, the focus of analysis shifts to various metrics, different from those used in OLS models. For instance, OLS models emphasize the R^2 and p-value, but XGBoost, as a gradient boosting method, utilizes alternative indicators. A key metric in XGBoost is the feature importance score. This score is pivotal for understanding each feature's role and impact within the model. It reveals the usefulness and value of each feature in building the model's boosted decision trees. The importance score of a feature is determined by assessing how significantly it enhances the model's performance metrics, such as accuracy or purity when included in the trees (T. Chen & Guestrin, 2016). However, it is important to note that feature importance in XGBoost is typically leveraged for enhancing the predictive capability of the model, rather than for inferential purposes. It essentially serves as a measure to identify and utilize the most influential variables for creating the most effective predictive model.

To evaluate how accurately our models can predict outcomes, we will use the *Root Mean Square Error* (RMSE). RMSE calculates the average difference between our predicted values and the actual observed values. By squaring these differences, RMSE ensures that all errors contribute positively to the overall measure. RMSE ranges from zero to infinity, indicating no error to maximum error, respectively. An important aspect of RMSE is that its value changes in proportion to the magnitude of the data it measures. For instance, an RMSE of 0.25 in our analysis implies that the average prediction error of our model is 0.25 units away from the actual sentiment scores. RMSE is calculated as shown in Equation 4.2

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{4.2}$$

where $\hat{y}_1, \hat{y}_2, ..., \hat{y}_i$ are the predicted values of a dependent variable and $y_1, y_2, ..., y_n$ are observed dependent values in a dataset. The number of observations is defined as n.

4.3.2 Hyperparameter Tuning

The hyperparameters of the XGBoost model were tuned using the grid search algorithm, which expands all possible variations of a set of hyperparameters into a large grid matrix. This expansion is done without taking the effect of each parameter on the model into consideration, which makes it a computationally demanding tuning algorithm (Alibrahim & Ludwig, 2021). The model is trained on each set of parameters in this expanded grid and the tune which minimizes the loss function presented in Section 2.4.3 is chosen as the best tune. There are alternatives to the grid search approach, such as random search which only runs the model on a small but representative sample of the full grid, which is necessary for larger models. Running this hyperparameter tuning algorithm on our data produces the set of optimal parameters for our specific model, which is found in Table 4.2.

Hyperparameter	**Value**
nrounds	200
max_depth	7
eta	0.05
gamma	0.2
colsample_bytree	0.5
min_child_weight	3
subsample	0.7

Table 4.2: XGBoost Tuned Hyperparameters

5 Analysis

5.1 Inference Analysis

5.1.1 Sentiment Scores and Target-to-Share Price Ratios

By training an OLS model on our set of data, we seek to get an understanding of how short-term price trends affect the sentiment and behaviour of the ERR analysts. The first step in this process involves exploring the relationship between textual sentiment and target prices, as this could have implications on how we interpret the financial implications behind the sentiment score we have created. In total, four simple OLS models were created, where the explanatory variables differ between them. The explanatory variables used are the ratio between Target Price and Share Price in t, the ratio between Target Price and Share Price in $t-1$ and the percentage change in target price from $t-1$ to t where t is defined as a set of points in time. Lastly, a full model is trained with all variables included. The regression summaries are shown in Table 5.1.

	Dependent variable: sentiment			
	Full Model	Target Ratio	Target Ratio Lag	% Change in Target Price
	(1)	(2)	(3)	(4)
target_ratio	-0.126^{***}	0.087^{***}		
	(0.045)	(0.024)		
target_ratio_1lag	0.236^{***}		0.080^{***}	
	(0.046)		(0.024)	
target_price_change_percent	0.521^{***}			0.397^{***}
	(0.055)			(0.048)
Constant	-0.190^{***}	-0.173^{***}	-0.171^{***}	-0.151^{***}
	(0.010)	(0.010)	(0.010)	(0.007)
Observations	2,230	2,230	2,230	2,230
R^2	0.046	0.006	0.005	0.030
Adjusted R^2	0.044	0.006	0.005	0.029
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 5.1: Sentiment Against Target Price Metrics

The regression summaries from Table 5.1 show that an increase in target price is generally associated with a more positive sentiment score. Their p-values are close to zero, which makes us confident that our assumption of correlation holds. As the unit of measurement is not the same across the explanatory variables, the specific coefficients are not easily comparable. The adjusted R^2 values are between 0.5% and 4.4%, which tells us that the variations in target price only account for a small part of the overall variation in sentiment in time t. Based on these results, we assume that the sentiment score of an ERR gives an indication of the banks attitude towards the price development of the covered security.

This large spread in adjusted R^2 is also quite interesting in itself as it aligns with the theory of stickiness (Bouchaud et al., 2019) in target prices. If target prices are sticky, a lot of the fluctuations in the target ratio might stem from stock price fluctuations rather than choices made by the analyst. Given that the *Target Ratio* is mathematically defined as the ratio between target price and share price, it is evident that a static target price does not necessarily translate to a stable *Target Ratio*, especially considering the independent nature of the *Share Price* variable. In contrast, the *Percentage Change in Target Price* is exclusively under the control of the analyst, albeit within the constraints imposed by corporate and regulatory frameworks. As the *% Change in Target Price* has a relatively high explanatory power in adjusted R^2 compared to the values found in the models using the *target_ratio* and *target_ratio_1lag*, we find that the change in target price over time is able to account for more of the sentiment variance relative to the target ratio metrics in a given point in time t.

This aspect of the model points towards the fact that analysts may exhibit optimism bias towards companies which experience upward revisions in target prices, yet do not seem to sustain this bias over time. This paints a picture of the sentiment score being a measure of change in attitude, rather than a snapshot of the attitude towards a stock in time t. From these results we will assume that a positive sentiment score can be attributed to positive attitudes towards the price development of a given stock.

5.1.2 Evaluating Sentiment Response

As we found that a shift in sentiment correlates with a corresponding directional change in the target price, we can move on to exploring how the textual sentiment of ERRs responds to stock market trend indicators. The trend indicators are used as explanatory variables in Models 2 to 4, in addition to a full model in Model 1, using all variables as explanatory variables. The regression results are presented in Table 5.2.

	Dependent variable: sentiment			
	Full Model	Return 3M	MACD	RSI
	(1)	(2)	(3)	(4)
RETURN_3M	0.253***	0.354***		
	(0.050)	(0.038)		
MACD	0.004		0.012***	
	(0.004)		(0.003)	
RSI	0.003***			0.005***
	(0.001)			(0.001)
Constant	−0.280***	−0.152***	−0.145***	−0.387***
	(0.039)	(0.007)	(0.007)	(0.027)
Observations	2,230	2,230	2,230	2,230
R^2	0.049	0.038	0.005	0.037
Adjusted R^2	0.048	0.038	0.005	0.037
Note:			*p<0.1; **p<0.05; ***p<0.01	

Table 5.2: Sentiment Against Trend Indicators

In Model 1, RSI displays a statistically significant positive coefficient, suggesting a meaningful positive relationship with sentiment. This significant coefficient at the 0.01 level indicates that as RSI values increase, sentiment is likely to increase as well, with this relationship being robust and unlikely due to coincidences. The coefficient for *RETURN_3M* remains significant and positive, underscoring its continued positive influence on sentiment. The *MACD* demonstrates a positive and significant relationship with sentiment at the 0.01 level in Model 3, yet in the full model this coefficient is no longer significant. The constant terms across the models remain negative and statistically significant, indicating a persistent baseline shift in sentiment when all other variables are at zero. The explained variance in sentiment by these models, as evidenced by the adjusted R^2 values, suggests a moderate explanatory power, with Model 1 continuing to explain the largest proportion of variance in sentiment.

These results show that the return of a given stock in the time period leading up to a report explains much more of the variation in sentiment than the more advanced trend indicators such as MACD. The market movement leading up to an ERR has some effect on the sentiment of said report, yet with the basic models applied here it is difficult to draw definitive conclusions about whether the analysts are responding to these trends or if they are capturing the effects of a third or more unknown variables.

The effects of stock price trends on ERR sentiment are isolated further by adding the set of control variables presented in Section 3.3.4 to the regression. Control variables account for alternative explanations and potential external variables that might otherwise distort the relationship between the independent and dependent variables. By implementing these variables, we can more accurately separate impact of the trend indicators on sentiment.

We train several OLS models, the summaries of which are presented in Table 5.3. Model 1 contains the full model with all explanatory variables, whilst Model 2 to 4 isolates each explanatory variable. Model 5 to 6 are used as benchmarks for our other models. Model 5 contains only the control variables, and are used to measure the explanatory power of the model without the trend indicators. Model 6 is trained with the *target_ratio* as the dependent variable, as we want to see how a similar model using another metric of analyst sentiment would perform under similar circumstances. The summaries of the last series of regressions show that the adjusted R^2 hovers around 41%-43% for Models 1 through 4, indicating that the variables included account for just under half of the variance in sentiment. Additionally, there is not a large discrepancy between the R^2 and the adjusted R^2, which tells us that the models are not overly complex.

In Model 1 we find that most of the control variable coefficients are significant with a p-value below 0.01, although there are some control variables, namely *VIX*, *log(pe_ratio)* and *log(time_on_OSE)*, which are not considered significant in the full model. This is similar to the results from Model 5.

Although the coefficients for the trend indicators are considered significant with a p-value below 0.01 in Model 2 to 4, they behave differently when combined in Model 1. In the full model, only the three-month return and RSI are considered significant. The MACD seems to be overshadowed by the simple return and RSI. An explanation as to why MACD are significant in Model 3, but not in Model 1 is that there might be a strong connection between market trends and sentiment, and when returns and RSI are excluded the MACD indicator captures some of these effects on the sentiment. Yet, when they are combined in the full model their effects are isolated and are shown to be insignificant.

	Dependent variables:					
	sentiment					target_ratio
	Full Model	Return 3M	MACD	RSI	Reference Model	Full Model (TR)
	(1)	(2)	(3)	(4)	(5)	(6)
log(market_cap)	−0.020***	−0.021***	−0.022***	−0.020***	−0.022***	−0.004
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.003)
log(time_on_OSE)	0.003	0.005	0.003	0.002	0.004	−0.002
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.004)
analystDNB	−0.184***	−0.185***	−0.179***	−0.182***	−0.179***	−0.004
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.006)
analystPARETO	−0.175***	−0.176***	−0.168***	−0.172***	−0.168***	−0.003
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.007)
dividend	0.053***	0.046***	0.045***	0.052***	0.041***	−0.012
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.008)
log(pe_ratio)	0.005	0.006	0.010*	0.006	0.009	−0.002
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.003)
financial_leverage	−0.009***	−0.009***	−0.008***	−0.009***	−0.008***	0.001**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
volatility	−0.070*	−0.044	−0.108***	−0.088**	−0.095**	0.002
	(0.039)	(0.039)	(0.039)	(0.038)	(0.039)	(0.020)
VIX	−0.001	−0.001	−0.001**	−0.001	−0.002**	0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0004)
sentiment_1lag	0.420***	0.418***	0.439***	0.427***	0.439***	
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	
target_ratio_1lag						0.845***
						(0.010)
RETURN_3M	0.090**	0.214***				−0.095***
	(0.040)	(0.032)				(0.021)
MACD	0.001		0.010***			−0.005***
	(0.003)		(0.003)			(0.002)
RSI	0.003***			0.004***		−0.002***
	(0.001)			(0.0004)		(0.0003)
Constant	0.432***	0.597***	0.653***	0.407***	0.643***	0.265***
	(0.159)	(0.157)	(0.158)	(0.158)	(0.159)	(0.083)
Observations	2,230	2,230	2,230	2,230	2,230	2,230
R^2	0.451	0.439	0.432	0.449	0.428	0.790
Adjusted R^2	0.447	0.436	0.429	0.446	0.425	0.789
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 5.3: Main Regression Model Summaries

In Model 6 it is harder to draw inference from the various coefficients in the model, as the lagged version of *target_price* is highly correlated with the dependent variable. This is not inherently a sign of a bad model if the goal is to predict the dependent variable, yet it does make it harder to understand the relationship between variables or the effect of the explanatory variables on *target_price*. The various coefficients are either small or have an opposite sign of what is expected from our research in comparison to our other models and hypothesis. It can therefore be inferred that the target price for a stock at time t is predominantly influenced by its historical pricing at $t-1$, indicating a significant temporal autocorrelation in the target price establishing process.

5.1.3 Model Robustness

In order to trust the results from the presented models, we need to evaluate some of the most pivotal OLS assumptions in order to ensure their validity. By scrutinizing the robustness of our model against violations of these assumptions, we aim to increase the reliability of our results. Specifically, our analysis of the models validity focuses on the unbiased nature of the sentiment coefficient in our regression models. In alignment with the Gauss-Markov theorem, our assessment adheres to four fundamental conditions for establishing an unbiased estimator (James H. Stock, 2020). These include ensuring that the model is linear in parameters, based on a randomly sampled dataset, devoid of perfect collinearity among the independent variables, and characterized by a zero conditional mean. Our investigation extends to the examination of the homoskedasticity of residuals, which, if confirmed, not only reinforces the unbiased nature of our estimator but also elevates it to the status of the best linear unbiased estimator (BLUE).

Linearity in parameters is a critical assumption in OLS regression, which posits that the relationship between the independent variables and the dependent variable can be described by a straight line, or more generally, that the parameters of the model are linearly related to the outcome. To asses whether the assumption holds, the trend indicators were plotted against the sentiment score, which can be found in Appendix B. A visual inspection tells us that there is a weak linear relationship between the 3-month return, RSI and sentiment, and close to no linear relationship between the MACD indicator and sentiment.

For the inferences drawn from this model to be relevant to an entire population, the data must constitute a random subset of that population. In this instance, we compiled reports from three large investment banks on the 25 largest companies on the OSE. Although we can assume that this gives us a representative sample of this specific population, it is not a random sample in the sense that it represents *all* investment banks and *all* companies on the OSE. The focus on big firms and a select few investment banks makes it harder to generalize our findings across the population of investment banks, as larger companies often differ from smaller ones in various aspects, and the practices of larger investment banks may not reflect those of smaller ones. According to Mercer et al. (2017), confounders might be unevenly distributed among the whole population of ERRs, which makes it harder to draw general conclusions from our models. While the study offers valuable perspectives within its scope, extrapolating these findings to the entire OSE requires caution.

To assess the assumption of no perfect multicollinearity, an examination of the correlation matrix for the model variables was created, which can be found in Appendix C. This matrix indicates that the independent variables *RETURN_3M*, *MACD*, and *RSI*, do not exhibit collinearity at a level of concern for OLS estimations. Specifically, the absence of correlation coefficients at or near the extremes of 1 or -1 among these predictors confirms the non-violation of this assumption. Control variables included in the model also demonstrate correlations below critical levels, further substantiating the model's adherence to the no perfect multicollinearity assumption.

The zero conditional mean assumption assumes that the expected value of the regression error term, given any value of the independent variable, is zero. In practise, this means that there are no positive or negative bias in the errors when predicting sentiment across the sentiment scale. In order to test this assumption, plotted the residuals against the fitted values in our models, which can be found in Appendix D. Here we expect to see the residuals evenly distributed above and below the horizontal line at zero, which they seem to do. Thus, we can assume that this condition holds.

Heteroscedasticity refers to a condition in statistical models where the variability of the residuals is not consistent for all values of the independent variables. To assess the presence of heteroscedasticity in our models, we conducted a Breusch-Pagan test (Breusch & Pagan, 1979). The results of this test yielded p-values of 0 for all models, suggesting the presence of heteroscedasticity in the data, as p-values of 0 indicate a significant deviation from the assumption of homoscedasticity. This could indicate OVB being present in our models, which was expected as a consequence of the nuanced nature of textual data.

Our robustness analysis of the OLS regression model reveals that while key assumptions like linearity, absence of perfect multicollinearity, and zero conditional mean are met for the majority of our exploratory variables, there are limitations to this model that could impact our interpretation of its outputs. The presence of heteroscedasticity and non-random sampling of data from big firms and investment banks restrict the generalizability of our findings. Consequently, while the model provides valuable insights within its scope, caution is advised in extrapolating the results to the OSE. To summarize, we cannot assert that these models are BLUE.

5.2 Prediction

5.2.1 OLS Model

The testRMSE measures presented in Table 5.4 are extracted from the OLS models trained using the LOOCV method. We run 5 models through this algorithm, one reference model without any trend indicators, a full model with all indicators, and three separate models with each trend indicator. These OLS models will be used as a benchmark in order to measure the performance of the XGBoost model.

Metric	**Model**	**Reference**	**Full Model**	**Return 3M**	**MACD**	**RSI**
RMSE	OLS	0.23971	0.23575	0.23786	0.23880	0.23567

Table 5.4: RMSE Outputs From OLS

In our analysis, we observe a statistically significant enhancement in model performance across all tested models relative to the reference model. This improvement suggests that the incorporation of trend indicators increases the predictive accuracy of our model concerning the textual sentiment score. The best models in terms of RMSE are the *RSI* model and the *Full Model*. In comparison to the *Reference* model, these models have a 0.04 lower RMSE, which equates to a 1.7% increase in performance.

5.2.2 XGBoost Model

In Table 5.5 we find the testRMSE values extracted from XGBoost models trained using k-fold CV with $k = 5$. In order to compare these metrics to the benchmark OLS models, we train five separate XGBoost models on the same forms as the OLS models presented in Section 5.2.1.

Metric	**Model**	**Reference**	**Full Model**	**Return 3M**	**MACD**	**RSI**
RMSE	XGBoost	0.23351	0.22751	0.23296	0.23233	0.22879

Table 5.5: RMSE Outputs From XGBoost

The testRMSE figures from the XGBoost models indicate a consistent superiority in performance over the OLS benchmarks, corroborating our initial hypotheses. This trend is evident in the evaluation of model performance metrics, where models incorporating trend indicators consistently surpass the baseline reference model in efficacy. Notably, the *Full Model* emerges as the most proficient, evidenced by its RMSE of 0.22751, an enhancement in accuracy by approximately 3.5% relative to the OLS model with the lowest RMSE. Additionally, the XGBoost model, employing solely the RSI as a predictor, also demonstrates commendable performance levels. The results indicate that the XGBoost models, particularly the *Full Model*, yield lower RMSE values, suggesting improved prediction accuracy. This finding aligns with the expectation that more complex models like XGBoost, which are capable of capturing non-linear relationships, can potentially offer enhanced predictive capabilities compared to linear models like OLS.

In addition to looking at prediction error terms, it is pertinent to examine how the selection and weighting of predictor variables in the XGBoost models contribute to these observed differences in performance. Our focus will now shift towards identifying key predictors and assessing their relative importance within the XGBoost framework. This evaluation aims to provide insights into the features that are most influential in driving the model's predictive accuracy. By doing so, we can gain a deeper understanding of the model's behavior and guide more effective feature selection in similar predictive modeling contexts. The calculated internal variable importance is shown in Figure 5.1. We extracted the variable importance from two models, namely the *Full Model* and the *Reference* model.

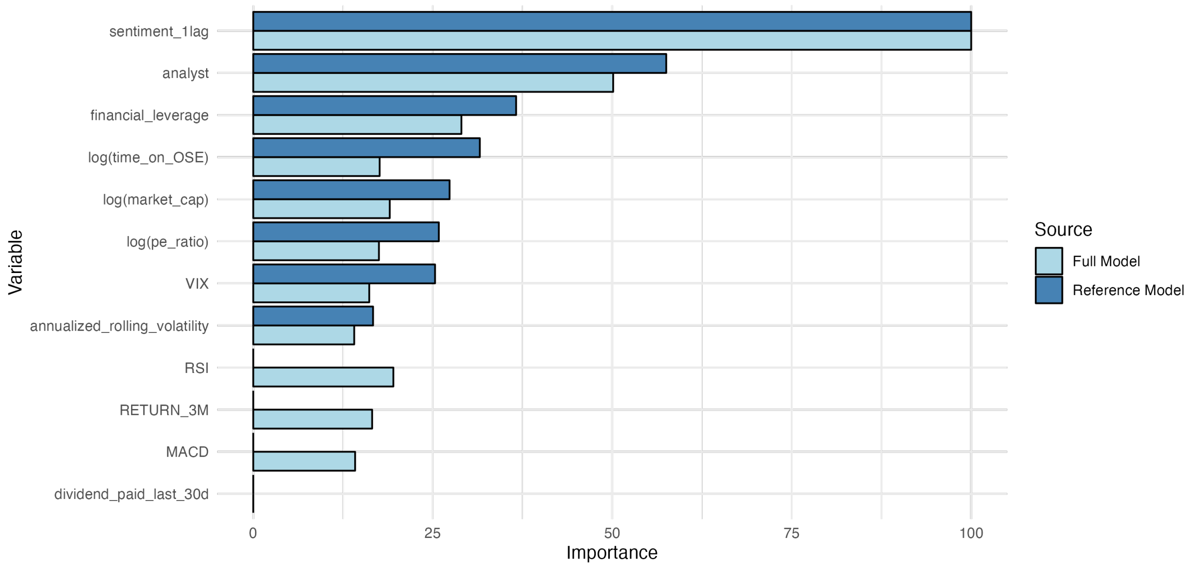

Figure 5.1: XGBoost Variable Importance

Figure 5.1 tells us that the most important factor in determining textual sentiment is the lagged value for said sentiment. This shows that the attitude towards a given stock is sticky, as the sentiment report in $t-1$ seems to highly correlate with the sentiment of a report in t. We also find that the investment bank an analyst works for has a lot of impact on the sentiment, which is most likely a result of varying internal guidelines in a given company. The variables continue to be the most important even after we introduce the trend indicators in the full model. This aligns with the assumptions of sticky attitudes and seems to be essential if we want to predict the numerical value of sentiment.

After the lagged sentiment and investment bank we notice a significant fall-off in importance. The financial leverage of a given company is deemed to be the third most important variable in this model. High leverage can signal a greater risk of default or financial distress, which in turn negatively impacts stock prices (Cai and Zhang, 2011), potentially causing negative ERR sentiment. Conversely, manageable leverage often indicates a stable financial position as it can increase returns, likely eliciting positive sentiments in ERRs.

Variables such as the logarithm of companies' P/E-ratio, market capitalization and time on OSE seem to have a significant impact on prediction, albeit less than the previously mentioned metrics. These variables describe a wide array of company attributes, such as earnings, valuation of equity and its tenure of being a publicly traded company. It is reasonable to assume that these are factors analysts take into account when determining the future performance of a stock. The volatility of a stock and the VIX are the variables which are deemed to be least important for prediction. This can be attributed to the fact that analysts use a six-to-twelve month time-frame when judging target prices. Due to the law of large numbers they may assume that these metrics of uncertainty will even out over time, given that their expected value is estimated to be 0. Recent dividend payout seems to have no impact on sentiment, which could be attributed to effective expectation management from the companies paying out dividends.

When looking at the *Full Model*, we see a shift in the importance of certain variables. RSI is the fourth most important variable, right below financial leverage. This suggests that market movements in the last 14 trading days account for as much predictive power as the leverage of a company. In addition to this, the MACD and three-month return seems to be comparable to the market capitalization and P/E-ratio of a company. Although we cannot conclude that these trend indicators capture the predictive effects, it seems like the textual sentiment in ERRs is affected by market movements in the period leading up to the release of a report. When combined with the fact that the full model has a lower error when predicting new observations than the reference model, it adds up to a conclusion indicating that these market indicators do impact the textual sentiment in the ERRs we have analyzed.

6 Discussion

6.1 Sticky Sentiment and Target Prices

Model 6 of Table 5.3 reveals a strong effect of the lagged target ratio, explaining 78.9% of the variation in the current target ratio. This finding underscores the sticky nature of the target ratio, aligning with the insights presented by Bonini et al. (2010). Similarly, we find a strong dependency between the lagged textual sentiment and current sentiment in Model 1 of Table 5.3. However, this effect is notably weaker compared to the lagged target ratio, confirming our belief that target prices are more sticky than sentiment, providing a more accurate estimate of their opinions on a given stock at time t.

Altering the target price can be seen as a more drastic measure as it is very visible. Customers are notified about the change in target price, and newspapers might write an article about the change, making the change seem more drastic. Given that an ERR on a company usually is written by the same analyst in an investment bank, a change in target price implies a modification of beliefs from the previous report, suggesting a prior error in judgment or the release of important unanticipated information. Analysts are to a larger extent held accountable for changes in target price rather than changes in textual sentiment, potentially causing analysts to be more reluctant to change their target price, evidently causing stickiness in target prices.

The key takeaway is that textual sentiment is less sticky relative to target prices. This prompts a deeper investigation into sentiment analysis, a domain where expressions may be more reflective of analysts' opinions at a given time. This brings us back to our research question of whether short-term price trends impact the textual sentiment in ERRs.

6.2 Implications of Findings

In Model 1 presented in Table 5.3, it was observed that the utilization of the technical indicator MACD did not produce a statistically significant impact on the sentiment expressed in ERRs, consistent with the expectations under the weak form EMH. This result can be attributed to the nature of the MACD, which is based on the most recent 26 trading days, whereas the reports primarily focus on projecting performance over the subsequent year. This suggests a possible indication that analysts may be unbiased and adhere to the weak form of the EMH, thereby questioning the efficacy of technical analysis as outlined by Fama & Blume (1966) and Jensen & Bennington (1970).

However, a notable discovery emerged in the form of a statistically significant coefficient for the 3-month return and RSI. This finding aligns with the research of Fong (2014), providing support for the potential trend-chasing bias arising from extrapolating short-term growth. Initially unexpected under the weak form EMH, the trend metrics imply that analyst sentiment may be influenced by short-term price trends, corroborating the findings of Bonini et al. (2010). While this result alone does not necessarily challenge the EMH, it does raise inquiries about the potential implications of investment recommendations and target price assessments. Assuming the weak-form EMH holds, this result is in line with Clarkson et al. (2015) and Buxbaum et al. (2021), suggesting that analysts may be susceptible to an optimism bias in bull-runs influenced by short term returns and trend-chasing bias.

As presented in Section 2.2.1, Shefrin & Belotti (2007) documents a difference in behavior amongst professional analysts and individual investors, succumbing to the gamblers' fallacy and hot-hand fallacy, respectively. However, we find that professional analysts coincide with individual investors, falling for the hot-hand fallacy. This is documented through the variable importance results from the XGBoost model, presented in Figure 5.1, showing RSI, MACD, and the 3-month return as some of the most influential factors on sentiment. Interestingly, this finding indicates that professional analysts do not behave differently than individual non-professional investors, coinciding with the findings of Bodnaruk & Simonov (2015) and Hon-Snir et al. (2012).

The trend indicator based on the shortest time frame, RSI, explains more of the variance in sentiment compared to the MACD, implying *recency bias*, supporting the findings of Shefrin & Belotti (2007). These findings illustrate how the human mind tends to oversimplify complex situations, succumbing to the psychological tendency to exaggerate the significance of recent events. Whether confronted with positive or negative news, people naturally assume that it will directly translate into improved or worsened performance in the subsequent year. Yet, they may overlook the EMH assertion that stock prices have already adjusted based on the information, potentially exaggerating the significance of news or price trends in their recommendations. This potential bias underscores the need for further exploration into the decision-making processes of equity analysts and their susceptibility to short-term market fluctuations in shaping recommendations and target price assessments.

On the other hand, it is important to recognize that following a strong or negative performance leading up to a report, it is natural to expect a positive or negative sentiment, respectively. This is expected because ERRs discuss recent events in their report, thus affecting the sentiment score. However, under the EMH, the stock price already reflects the given information, rendering the subsequent price movement unpredictable and resembling a *random walk*. Consequently, this should result in a more neutral tone in the analysis, reflecting the uncertainty of the next price move and yielding an overall more neutral sentiment. Our research findings reveal a notably positive and statistically significant coefficient for the 3-month return. This suggests that analysts may suffer from optimism bias or trend-chasing bias, that they do not believe the weak form EMH holds, or they use less sophisticated models as explained through the findings of Clarkson et al. (2015) and Gleason et al. (2013).

Analysing the sentiment scores across financial reports from Carnegie, DNB, and Pareto reveals patterns providing insights into their reporting practices. Carnegie, the reference group in Table 5.3, systematically exhibits a more positive sentiment compared to DNB Markets and Pareto Securities. This finding implies that Carnegie either have a different reporting standard or tendencies towards optimism bias. The utilization of DNB and Pareto as control variables further illuminates the systematic differences in sentiment, emphasizing the need to consider varying reporting standards among these entities. Pareto's long-tailed distribution encompassing both very negative and very positive sentiments, detailed in Figure 3.2, contrasts with DNB's conservative stance. Notably, even with the recognition of the Loughran-McDonald dictionary's strong tendency to classify words as negative, Carnegie consistently exhibits a positive sentiment score. This finding, reflected in their sentiment score distribution, underscores the distinctiveness of Carnegie's reporting style. In summary, this analysis does not only highlight systematic differences in sentiment across the investment banks, but also variations in reporting standards, potentially revealing a higher likelihood of optimism bias within Carnegie.

According to Shefrin & Belotti (2007), we would expect our trend indicators to have a negative coefficient, due to the excessive faith in mean reversion. However, we find the opposite. Shefrin referred to professional investors and analysts succumbing to the gamblers' fallacy, but we find that the analysts tend to become similar to individual investors and suffer from the hot-hand fallacy instead. This finding is intriguing as it portrays professionals as no different from individual non-professionals, supporting Bodnaruk & Simonov (2015) and Hon-Snir et al. (2012). Note that the MACD is not significant, and the RSI is close to 0. This result would imply that the analysts are unbiased and adhere to the weak-form EMH.

The XGBoost model returned *RETURN_3M*, *RSI*, and *MACD* as some of the most important variables in explaining sentiment, presented in Figure 5.1. Under the weak-form EMH, the trend indicators should not impact a target price forecast. However, the XGBoost results imply that the 3-month simple return, RSI, and MACD are more important in shaping sentiment than traditional valuation techniques and firm-specific factors. Although the relationship between textual sentiment and target ratio tested in Table 5.1 is modest, the relationship between trend indicators and textual sentiment in ERRs is evident, potentially revealing trend-chasing bias in Norwegian equity research.

As shown in Figure 3.3, the majority of buy recommendations, relative to hold and sell, is evident. This is the same finding Buxbaum et al. (2021) documented, a clear preference and overweight of buy recommendations amongst stock analysts, signalling optimism bias. However, trend-chasing bias cannot be identified through this discovery alone.

So far, we have discussed the implications of our findings given that the EMH holds. However, as mentioned in Section 2.3.1, markets do not seem to exhibit efficiency in the short term (2015). This fact may alter our discussion, as analysts might consider behavioral biases and market inefficiencies as part of their valuation techniques, referring to Pareto Securities' statement on behavioral technical analyses, presented in Section 2.1. As discussed, we find that sentiment in ERRs follows trends, either contradicting the EMH or supporting bias in analyst forecasts. One explanation we have not discussed yet is the possibility that analysts use momentum as input, adhering to the findings of Shefrin & Belotti (2007); Winners tend to follow winners and losers tend to follow losers.

6.3 Financial and Economic Implications

Stating that analysts are biased and that their performance is based on luck rather than skill entails significant implications and must not be taken lightly. The assumption of trend-chasing bias in ERRs suggests that the financial landscape may inadvertently mirror institutionalized gambling, as explained through the gambling analogy of Fong (2014) in Section 2.2.1. This analogy emphasizes the potential consequences for less informed investors. Customers are putting their savings at risk based on the opinions of equity research analysts, who have no prerequisite of beating the market under the assumption of bias. This assumption becomes increasingly important when discussing investment management, where professionals receive a portion of the funds for their service. If they do not have any prerequisite to outperform the market, it could be viewed as morally wrong to take fees for a service rooted in luck rather than skill.

Although this thesis refrains from drawing parallels to institutionalized gambling, we find tendencies of trend-chasing bias in their sentiment. The importance of further research on analyst bias and the potential ramifications cannot be overstated.

6.4 Limitations

Our findings fall short of causal inference regarding bias in analyst forecasts. Textual sentiment depends on many unobservable variables, making it likely that the model suffers from OVB, causing endogeneity issues. Throughout the discussion, we have assumed that the sentiment score accurately identifies the attitude towards a stock, as described under Section 5.1.1. Some of the interpretations related to the EMH are only valid under this assumption. Even though we find a dependency between the sentiment and target ratio, stronger empirical evidence is needed to support the assumption.

In acknowledging the limitations of our study, it is important to highlight the omission of distinctions between negative and positive trends. Extensive literature indicates that analyst behavior differs during bull and bear markets (Hanna et al., 2020; Kim & Nofsinger, 2007). Additionally, research suggests a propensity among analysts to defer downgrades while potentially expediting upgrades (Ho et al., 2018). The lack of a nuanced exploration into how positive and negative trends may cause different biases and have a different impact on the sentiment constraints our analysis. Recognizing and incorporating such nuances could enhance the depth and accuracy of our findings.

While examining the OLS and XGBoost model in Chapter 5.2, a relatively high RMSE was revealed, aligning with our expectations given the inherent complexity of predicting sentiment. While the models serve as valuable tools for inference and understanding the factors influencing sentiment in ERRs, caution is advised against relying on them for predictive purposes. Recognizing the challenges in forecasting sentiment, we emphasize the utility of our models for insightful analysis rather than predictions.

6.5 Further Research

To advance our research, further refinement of our model is imperative. The complexities inherent in capturing all potential effects on sentiment demand a comprehensive approach that considers variables beyond those initially examined.

In our thesis, we have identified factors influencing the ERR sentiment. Going further, exploring whether sentiment serves as a reliable predictor of bias is of interest. We recommend building upon our findings to understand whether sentiment can be used to detect bias in target price forecasts, as measured by Das et al. (1998), shown in Equation 6.1 where *1y* denotes one year and target price is defined as $\hat{P}_{a,t}$.

$$BIAS = \frac{\hat{P}_{a,t} - P_{t+1y}}{P_{t+1y}} \tag{6.1}$$

In this thesis, we have solely focused on bias among professional analysts. We recommend performing similar research for non-professional investors to understand whether they behave differently from professional analysts, aiming to seek if the findings of Bodnaruk & Simonov (2015) and Hon-Snir et al. (2012) applies to the OSE.

Additionally, exploring the impact of sentiment on financial markets presents a compelling avenue for research, examining whether sentiment fluctuations contribute to market dynamics. Finally, we acknowledge a third dimension overlooked in our study, the companies covered. Delving into this aspect could unveil potential discrepancies or unique patterns tied to specific entities, enhancing the depth and applicability of our findings.

7 Conclusion

This thesis has explored the impact of short-term stock price trends on the textual sentiment in Norwegian equity research reports on the Oslo Stock Exchange, discerning whether Norwegian investment banks succumb to trend-chasing bias.

The OLS model revealed that the trend indicators, while controlling for firm-specific differences, can explain 2.2% more of the variance in textual sentiment compared to the models without the trend indicators, making our full model exhibit an explanatory power of 44.7%. The short-term trend indicators prove to be at least as important as fundamental metrics about the company under scrutiny. This finding suggests that trend-chasing bias is just as important as some traditional valuation techniques in shaping sentiment.

We find that the 3-month simple return, RSI, and MACD are important to explain variance in the sentiment of equity research reports through the XGBoost model, returning RSI as the fourth most important variable. These findings suggest that trend indicators has a measurable effect on textual sentiment. The most important variable proves to be the lagged sentiment value, emphasizing the stickiness inherent in sentiment, albeit lower than stickiness in target prices. Further, we note discrepancies in textual sentiment between the investment banks, which can be explained by either differences in reporting standards, differing likelihood of trend-chasing bias, or the fact that market inefficiencies are accounted for in their projections.

Despite these results, our findings do not provide any causal inference of trend-chasing bias being present in Norwegian equity research. Detecting bias through sentiment is an intricate topic, and it is hard to distinguish whether the results of our analysis are due to the presence of bias or whether the analysts do not believe in the weak form efficient market hypothesis in the short run, thus accounting for market inefficiencies in their projections. We have a small subsample of investment banks, which might not be representative of the population. Thus, we advise caution when extrapolating the results to the entire population of investment banks. Nevertheless, our implementation of sentiment analysis in studies relating to bias in equity research could serve as a valuable tool in subsequent research endeavours.

References

Aksje Norge. (2023). Nordmenn og aksjer - andre kvartal 2023. https://aksjenorge.no/wp-content/uploads/2023/07/AksjeNorge-Aksjestatistikk-2kvartal-2023.pdf

Al-Awadhi, A. M., Alsaifi, K., Al-Awadhi, A., & Alhammadi, S. (2020). Death and contagious infectious diseases: Impact of the covid-19 virus on stock market returns. *Journal of behavioral and experimental finance*, *27*, 100326.

Alibrahim, H., & Ludwig, S. A. (2021). Hyperparameter optimization: Comparing genetic algorithm against grid search and bayesian optimization, 1551–1559.

Aspray, T. (2011). Demystifying the MACD. Forbes. https://www.forbes.com/sites/tomaspray/2011/10/22/demystifying-the-macd/

Barberis, N., Greenwood, R., Jin, L., & Shleifer, A. (2018). Extrapolation and bubbles. *Journal of Financial Economics*, *129*(2), 203–227.

Bessembinder, H., & Chan, K. (1998). Market efficiency and the returns to technical analysis. *Financial management*, 5–17.

Bhootra, A., Drezner, Z., Schwarz, C., & Stohs, M. H. (2015). Mutual fund performance: Luck or skill? *International Journal of Business*, *20*(1), 52.

Bloomberg L.P. (2023). Bloomberg terminal.

Bodnaruk, A., & Simonov, A. (2015). Do financial experts make better investment decisions? *Journal of financial Intermediation*, *24*(4), 514–536.

Bonini, S., Zanetti, L., Bianchini, R., & Salvi, A. (2010). Target price accuracy in equity research. *Journal of Business Finance & Accounting*, *37*(9-10), 1177–1217.

Bouchaud, J.-P., Krueger, P., Landier, A., & Thesmar, D. (2019). Sticky expectations and the profitability anomaly. *The Journal of Finance*, *74*(2), 639–674.

Bradshaw, M. T., Brown, L. D., & Huang, K. (2013). Do sell-side analysts exhibit differential target price forecasting ability? *Review of Accounting Studies*, *18*, 930–955.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, 1287–1294.

Buxbaum, M., Schultze, W., & Tiras, S. L. (2021). Target price optimism, investor sentiment, and the informativeness of analysts' target prices. *Kelley School of Business Research Paper*, (19-26).

Cai, J., & Zhang, Z. (2011). Leverage change, debt overhang, and stock prices. *Journal of Corporate Finance*, *17*(3), 391–402.

Capstaff, J., Klaeboe, A., & Marshall, A. P. (2004). Share price reaction to dividend announcements: Empirical evidence on the signaling model from the oslo stock exchange. *Multinational Finance Journal*, *8*(1/2), 115–139.

Carnegie. (2021). Disclosures and disclaimers. https://www.carnegie.se/app/uploads/2021/09/stillfront.pdf

Casella, G. (2002). *Statistical inference (2nd edition)*. Thomson Learning.

Chang, J. W., & Choi, H. M. (2017). Analyst optimism and incentives under market uncertainty. *Financial Review*, *52*(3), 307–345.

Chen, J. (2023). Alpha: What It Means in Investing, With Examples. Investopedia. https://www.investopedia.com/terms/a/alpha.asp

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chevalier, J., & Ellison, G. (1997). Risk taking by mutual funds as a response to incentives. *Journal of political economy*, *105*(6), 1167–1200.

Chincarini, L. B., Kim, D., & Moneta, F. (2020). Beta and firm age. *Journal of Empirical Finance*, *58*, 50–74.

Cho, J.-S. (2014). The effect of the accrual component of earnings volatility on security analysts' target price forecast performance. *Investment Analysts Journal*, *43*(79), 51–56.

Chong, T. T.-L., & Ng, W.-K. (2008). Technical analysis and the london stock exchange: Testing the macd and rsi rules using the ft30. *Applied Economics Letters*, *15*(14), 1111–1114.

Chong, T. T.-L., Ng, W.-K., & Liew, V. K.-S. (2014). Revisiting the performance of macd and rsi oscillators. *Journal of risk and financial management*, *7*(1), 1–12.

Chowdhury, S. G., Routh, S., & Chakrabarti, S. (2014). News analytics and sentiment analysis to predict stock price trends. *International Journal of Computer Science and Information Technologies*, *5*(3), 3595–3604.

Clarkson, P., Nekrasov, A., Simon, A., & Tutticci, I. (2015). Target price forecasts: Fundamental and non-fundamental factors. *Available at SSRN*, *2104433*.

Clarkson, P., Nekrasov, A., Simon, A., & Tutticci, I. (2020). Target price forecasts: The roles of the 52-week high price and recent investor sentiment. *Journal of Business Finance & Accounting*, *47*(9-10), 1365–1399.

Coad, A., Segarra, A., & Teruel, M. (2016). Innovation and firm growth: Does firm age play a role? *Research policy*, *45*(2), 387–400.

Cortes, C., Mohri, M., & Rostamizadeh, A. (2012). L2 regularization for learning kernels. *arXiv preprint arXiv:1205.2653*.

Cox, D. R. (2006). *Principles of statistical inference*. Cambridge University Press.

Cui, J., Wang, Z., Ho, S.-B., & Cambria, E. (2023). Survey on sentiment analysis: Evolution of research methods and topics. *Artificial Intelligence Review*, 1–42.

Das, S., Levine, C. B., & Sivaramakrishnan, K. (1998). Earnings predictability and bias in analysts' earnings forecasts. *Accounting Review*, 277–294.

De Vincentiis, P. (2010). Accuracy and bias of equity analysts in an environment characterised by higher disclosure: Empirical evidence from the italian market. *EMFI Working Paper*.

Dechow, P. M., & You, H. (2020). Understanding the determinants of analyst target price implied returns. *The Accounting Review*, *95*(6), 125–149.

DNB Markets. (2016). Terms. https://www.dnbalpha.com/terms/

Dolan, B. (2023). MACD Indicator Explained, with Formula, Examples, and Limitations. Investopedia. https://www.investopedia.com/terms/m/macd.asp

Euronext. (2023). Product directory. https://live.euronext.com/nb/products/equities/list

Fama, E. F., & Blume, M. E. (1966). Filter rules and stock-market trading. *The Journal of Business*, *39*(1), 226–241.

Feuerriegel, S., Wolff, G., & Neumann, D. (2015). Information processing of foreign exchange news: Extending the overshooting model to include qualitative information from news sentiment. *Available at SSRN 2603435*.

Filiz, I., Judek, J. R., Lorenz, M., & Spiwoks, M. (2021). Sticky stock market analysts. *Journal of Risk and Financial Management*, *14*(12), 593.

Fong, W. M. (2014). Trend-chasing. *The Lottery Mindset: Investors, Gambling and the Stock Market*, 46–76.

Forbes. (2023). Is the stock market in a bubble right now? https://www.forbes.com/sites/qai/2023/07/07/is-the-stock-market-in-a-bubble-right-now/

Ford, C. (2018). Interpreting Log Transformations in a Linear Model. University of Virginia Library. https://library.virginia.edu/data/articles/interpreting-log-transformations-in-a-linear-model

Fortin, R., Roth, G., et al. (2007). Analyst coverage of small cap firms in a period of brokerage firm retrenchment. *Journal of Business & Economics Research (JBER)*, *5*(12).

Frade, V. (2019). *Technical indicators. a comparative study* [Doctoral dissertation].

Freedman, D. (2009). *Statistical models: Theory and practice*. Cambridge University Press.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive psychology*, *17*(3), 295–314.

Gleason, C. A., Bruce Johnson, W., & Li, H. (2013). Valuation model use and the price target performance of sell-side equity analysts. *Contemporary Accounting Research*, *30*(1), 80–115.

Gunvaldsen, C. L., & Walmann, H. (2021). *The difference in bias and inaccuracy between commissioned and traditional sell-side equity research: Evidence from nordic investment banks after mifid ii implementation* [Master's thesis].

Han, Y., Yang, K., & Zhou, G. (2013). A new anomaly: The cross-sectional profitability of technical analysis. *Journal of Financial and Quantitative Analysis*, *48*(5), 1433–1461.

Hanna, A. J., Turner, J. D., & Walker, C. B. (2020). News media and investor sentiment during bull and bear markets. *The European Journal of Finance*, *26*(14), 1377–1395.

Harper, D. (2023). Using Historical Volatility To Gauge Future Risk. Investopedia. https://www.investopedia.com/articles/06/historicalvolatility.asp

Hayes, A. (2022). Technical Analysis: What It Is and How to Use It in Investing. Investopedia. https://www.investopedia.com/terms/t/technicalanalysis.asp

Ho, T. Q., Strong, N., & Walker, M. (2018). Modelling analysts' target price revisions following good and bad news? *Accounting and Business Research*, *48*(1), 37–61.

Höhler, J., & Lansink, A. O. (2021). Measuring the impact of covid-19 on stock prices and profits in the food supply chain. *Agribusiness*, *37*(1), 171–186.

Hojem, M. A., & Ottenbacher, K. J. (1988). Empirical investigation of visual-inspection versus trend-line analysis of single-subject data. *Physical Therapy*, *68*(6), 983–988.

Hon-Snir, S., Kudryavtsev, A., & Cohen, G. (2012). Stock market investors: Who is more rational, and who relies on intuition. *International Journal of Economics and Finance*, *4*(5), 56–72.

Hudson, R., Dempsey, M., & Keasey, K. (1996). A note on the weak form efficiency of capital markets: The application of simple technical trading rules to uk stock prices-1935 to 1994. *Journal of banking & finance*, *20*(6), 1121–1132.

Investopedia. (2021). Identifying Market Trends. Investopedia. https://www.investopedia.com/articles/technical/03/060303.asp

Jadhav, A., M. Mostafa, S., Elmannai, H., & Karim, F. K. (2022). An empirical assessment of performance of data balancing techniques in classification task. *Applied Sciences*, *12*(8).

James H. Stock, M. W. W. (2020). *Introduction to econometrics, global edition* (4th ed.). Pearson Education Limited.

Jensen, M. C., & Benington, G. A. (1970). Random walks and technical theories: Some additional evidence. *The Journal of finance*, *25*(2), 469–482.

Jones, S. L., & Netter, J. M. (2008). Efficient capital markets. *The Concise Encyclopedia of Economic*, *15*.

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, *33*, 171–185.

Kenton, W. (2023). Gambler's Fallacy: Overview and Examples. Investopedia. https://www.investopedia.com/terms/g/gamblersfallacy.asp

Kim, K. A., & Nofsinger, J. R. (2007). The behavior of japanese individual investors during bull and bear markets. *The Journal of Behavioral Finance*, *8*(3), 138–153.

Lai, I. (2023). Sentiment analysis is virtually useless in financial forecasting. *Available at SSRN 4545418*.

Lee, B., O'Brien, J., & Sivaramakrishnan, K. (2008). An analysis of financial analysts' optimism in long-term growth forecasts. *The Journal of Behavioral Finance*, *9*(3), 171–184.

Li, F. (2010). The information content of forward-looking statements in corporate filings—a naıve bayesian machine learning approach. *Journal of Accounting Research*, *48*(5), 1049–1102.

Loderer, C. F., & Waelchli, U. (2010). Firm age and performance. *Available at SSRN 1342248*.

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.

Maverick, J. (2023). The Weak, Strong, and Semi-Strong Efficient Market Hypotheses. Investopedia. https://www.investopedia.com/ask/answers/032615/what-are-differences-between-weak-strong-and-semistrong-versions-efficient-market-hypothesis.asp

Mercer, A. W., Kreuter, F., Keeter, S., & Stuart, E. A. (2017). Theory and Practice in Nonprobability Surveys: Parallels between Causal Inference and Survey Inference. *Public Opinion Quarterly*, *81*(S1), 250–271.

Naseer, M., & Bin Tariq, D. Y. (2015). The efficient market hypothesis: A critical review of the literature. *The IUP Journal of Financial Risk Management*, *12*(4), 48–63.

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.

Nikolopoulou, K. (2023). What Is Optimism Bias? | Definition & Examples. Scribbr. https://www.scribbr.com/research-bias/optimism-bias/

Nofer, M., & Nofer, M. (2015). Are crowds on the internet wiser than experts?–the case of a stock prediction community. *The Value of Social Media for Predicting Stock Returns: Preconditions, Instruments and Performance Analysis*, 27–61.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, *2*(1–2), 1–135.

Pareto Securities. (2023a). Disclaimer and legal disclosures. https://paretosec.com/compliancedownloads/disclaimer.pdf

Pareto Securities. (2023b). Time to go long - up to buy [Published equity research report on Adevinta].

Park, C.-H., & Irwin, S. H. (2004). The profitability of technical analysis: A review. *AgMAS Project Research Report*.

Peng, N. Y. (2020). How NLP has evolved for Financial Sentiment Analysis. Medium. https://towardsdatascience.com/how-nlp-has-evolved-for-financial-sentiment-analysis-fb2990d9b3ed

Puckett, A., & Yan, X. (2011). The interim trading skills of institutional investors. *The Journal of Finance*, *66*(2), 601–633.

Shearman & Sterling. (2023). Mifid ii: An update on the rules for unbundling of research.

Shefrin, H., & Belotti, M. L. (2007). Behavioral finance: Biases, mean-variance returns, and risk premiums. *CFA Institute Conference Proceedings Quarterly*, *24*(2), 4–12.

Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. *Handbook of statistics*, *24*, 303–329.

Teall, J. L. (2022). *Financial trading and investing*. Academic Press.

Tretina, K. (2023). What Is The Relative Strength Index (RSI)?. Forbes. https://www.forbes.com/advisor/investing/relative-strength-index-rsi

Tuckett, D. (2009). Addressing the psychology of financial markets. *Economics*, *3*(1), 20090040.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological bulletin*, *76*(2), 105.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207–232.

Ülkü, N., & Prodan, E. (2013). Drivers of technical trend-following rules' profitability in world stock markets. *International Review of Financial Analysis*, *30*, 214–229.

Witten, D., & James, G. (2013). *An introduction to statistical learning with applications in R*. Springer Publication.

XGBoost Developers. (2023). XGBoost Parameters — xgboost 2.0.2 documentation. https://xgboost.readthedocs.io/en/stable/parameter.html

Yahoo! (2023). Historical financial data. https://finance.yahoo.com/

Appendices

A Table Summary of Final Dataset

Statistic	N	Mean	St. Dev.	Min	Max
sentiment	2,230	−0.143	0.310	−1.000	1.000
target_ratio	2,230	0.320	0.261	−0.359	1.690
sentiment_1lag	2,230	−0.141	0.313	−1.000	1.000
target_ratio_1lag	2,230	0.326	0.265	−0.359	1.690
market_cap (NOKm)	2,230	118,537	167,784	3,802	1,289,717
time_on_OSE	2,230	17.014	5.665	0.405	23.650
dividend	2,230	0.114	0.318	0	1
pe_ratio	2,230	44.160	98.973	1.953	1,288.444
financial_leverage	2,230	4.525	5.256	−1.000	29.074
volatility	2,230	0.330	0.160	0.087	1.545
VIX	2,230	21.511	8.225	10.160	76.450
RETURN_3M	2,230	0.018	0.167	−0.597	0.813
MACD	2,230	0.020	1.973	−14.264	13.485
RSI	2,230	51.326	12.929	7.890	88.049

Table A.1: Summary Table of Model Dataset

B Assumption of Linearity

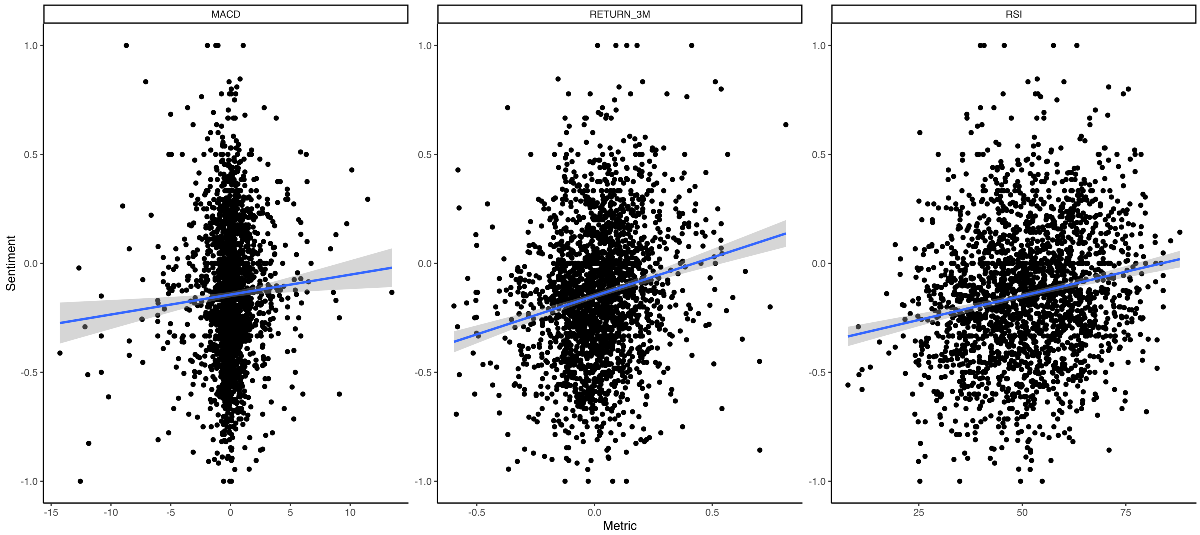

Figure B.1: Scatterplots of Stock Trend Indicators Against Textual Sentiment Score

C Correlation Matrix

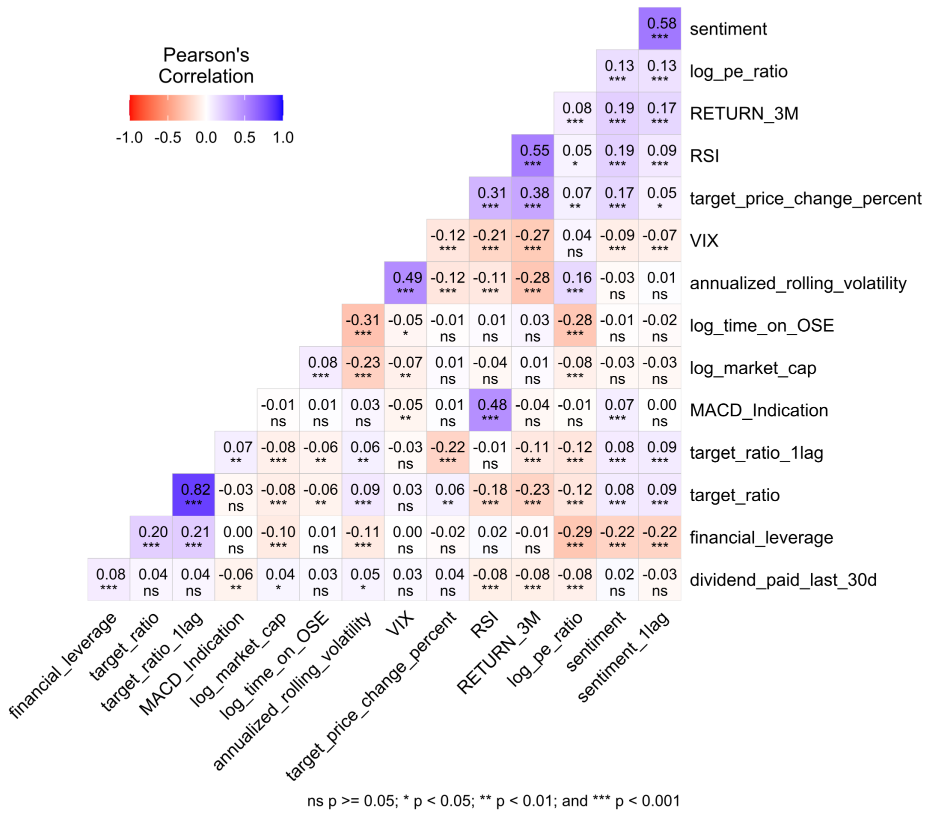

Figure C.1: Pearson's Correlation Matrix for Model Data

D Assumption of Zero Conditional Mean

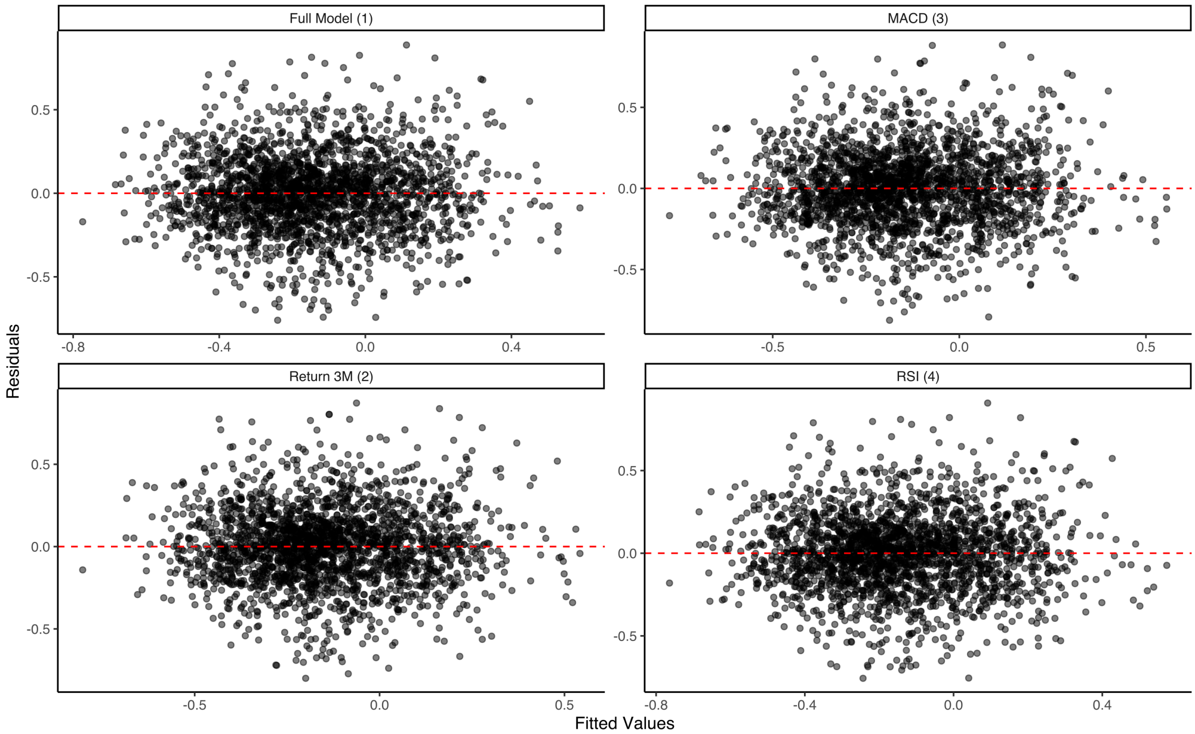

Figure D.1: Scatterplots of Residuals Against Fitted Values for Model 1 to 4