Norwegian School of Economics
Bergen, Spring 2023

Central Bank Sentiment Analysis and Asset Prices

Using Machine Learning and Natural Language Processing to Conduct Sentiment Analysis for Predicting Stock Prices in a Norwegian Financial Context

Julian Cæsar Andersen & Magnus Glad Thorshaug

Supervisor: Christian Langerfeld

Master thesis, Economics and Business Administration
Major: Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

This thesis seeks to investigate the relationship between text sentiment in central bank communication and stock market returns in a Norwegian context using a machine learning approach. We collect textual data from Norges Bank (the Norwegian Central Bank) consisting of monetary policy evaluations spanning the last 24 years and apply a multinomial inverse regression (MNIR) to create a positive and negative sentiment dictionary. For performance comparisons, we employ a set of naïve methods, one of which is developed by Kirkeby and Larsen (2021) and one developed by ourselves. Results indicate that there is no significant relationship between the sentiment of Norges Bank and stock returns at Oslo Børs (Oslo Stock Exchange) using test data. Out-of-sample, significant results were only found using a negative sentiment dictionary constructed by ourselves. Counterintuitively, these results indicate that a more negative sentiment leads to higher stock returns. We theorise that not finding significant results with the MNIR-dictionary can be contributed to a few factors. Loss of generality could explain parts of our results. Only using single terms to capture sentiment means that some significance might be lost in the process. We also discuss the possibility that our model mainly captures attributes in the financial landscape that leads to higher or lower stock returns, rather than capturing actual sentiment. Future research into this field using variations in data or methodology can be successful in further investigations of the relationship between central bank communication and asset prices.

Acknowledgements

Our master's thesis was written at the conclusion to our Master of Science in Economics & Business Administration, Business Analytics major, at the Norwegian School of Economics in 2023.

In our thesis we use machine learning and natural language processing in the domain of textual data analysis to produce sentiment analysis on communication from the Norwegian Central Bank. We use the sentiment produced from the machine learning method in an empirical model and compare its performance to more manual approaches to sentiment dictionaries. Our goals were to conduct sentiment analysis in a Norwegian context without relying on manually produced dictionaries, and to understand whether there is a consistent dynamic between sentiment in central bank communication and stock prices.

This topic piqued our interest due to the macroeconomic circumstances surrounding the time in which this paper was written, where record-breaking inflation is being combatted by central banks globally, causing fallouts in financial markets. It has been incredibly rewarding to work on a topic we are passionate about, and we feel that this analysis has enriched our understanding of data science, the macroeconomic environment, and financial markets.

We express our sincere gratitude to our supervisor, Christian Langerfeld, for guidance and feedback throughout the process. We extend our thanks to the teaching staff at NHH for providing a rich education which has culminated in this thesis.

Julian Cæsar Andersen *Magnus Glad Thorshaug*

Bergen, May 2023

NHH

Contents

Table of Figures

1. Introduction

On August 26, 2022, Federal Reserve Chairman Jerome Powell held a highly anticipated speech which would set the tone in financial markets for the remainder of the calendar year and was followed by a multi-trillion-dollar fallout in the financial sector. Financial markets were expecting a rate hike from the Federal Reserve, the central bank of the United States of America, to be announced at the speech. However, the Standard & Poor's 500 index (S&P 500), the US' single largest stock market index, was up in the few days prior to the event, as investors hoped for a pivot from the Federal Reserve (Cox, 2022). At the time, factions within financial markets believed that the Federal Reserve would decelerate or even reverse rate hikes in order to prevent a recession in the US economy. When Federal Reserve Chairman Jerome Powell instead declared that there was "pain on the horizon", confirming that rate hikes would continue, markets crashed, and they would continue to trend downwards as investor sentiment became more negative, and fears of a deep recession became greater. The S&P 500 ended the 2022 calendar year down 20%, and global equity markets lost $33 trillion in value, the largest drop since the 2008 Global Financial Crisis (Goodkind, Horowitz & Goldman, 2022). Whilst movements in equity markets can rarely be attributed to a single factor, Jerome Powell's speech appeared to be an economic condemnation for investors and serves as evidence that financial markets listen closely to central bank communication during periods of inflation, in crises and during changes in monetary policy.

The objective of this thesis is twofold. First, we seek to examine the relationship between central bank communication and asset prices in equity markets using sentiment analysis and machine learning (ML). Sentiment analysis is a Natural Language Processing (NLP) approach which attempts to identify the 'tone' or degree of positivity, or negativity, of a piece of text (Tetlock, 2007). Our hypothesis is that if the sentiment of central bank communication can be accurately captured, such a metric will correlate positively with asset prices. As our succeeding literature review reveals, controlling for monetary policy has been a missing piece in prediction of asset prices. Second, we seek to apply sentiment analysis in a Norwegian Finance & Accounting context. Whilst some attempts have been made to capture metrics of sentiment using Norwegian text, these are typically limited to counting the occurrences of a selected few terms manually produced by the researcher, as we highlight in our literature review. Instead, this thesis attempts to produce a robust Norwegian sentiment dictionary, akin to the dictionary developed by Garcia, Hu and Rohrer (2020). If successful, our analysis will improve the accessibility of sentiment analysis in Norwegian contexts, providing a sentiment dictionary widely applicable for Finance & Accounting textual research, possibly improving the accessibility of sentiment analysis and textual analysis at large in any non-English context. This thesis would also introduce a new metric indicating the sentiment of central bank communication during prediction tasks or empirical & econometric studies, which might be superior to controlling for the interest rate. We decided to write this thesis out of interest in the relationship between central banks and financial markets, as well as the potential contributions to the field of textual research and predictive analytics.

Our thesis paper is divided into 5 parts. We start by reviewing relevant literature in a somewhat chronological order, showing the development of the field of textual analysis and its introduction to machine learning and Natural Language Processing, as well as textual analysis studies performed in a Norwegian context. Here we also present empirical studies working with monetary policy metrics which inspired us to pursue sentiment analysis. Next, we describe our data, consisting of press releases from Norges Bank (Norwegian Central Bank), as well as the data preprocessing steps required to obtain our final dataset. Third, we outline the core of our methodology, the robust Multinomial Inverse Regression (MNIR) framework developed by Garcia Hu and Rohrer (2020), as well as naive methods for comparison. Fourth, we present our results, the resulting sentiment dictionary and reveal if we have successfully created a sentiment measure which correlates significantly with asset prices in Norway. Finally, we discuss the validity of our findings as well as their implications for the field of textual data analysis and empirical econometric studies.

2. Literature Review

Empirical studies and prediction tasks have attempted to incorporate the effect of central bank monetary policy on asset prices with mixed results. An empirical study measured the effect of interest rate policies on the stock prices of 272 US banks, finding that the magnitude of the change was important in explaining the variance, as well as the change in monetary policy compared to the expected change (Ghazanfari, Rogers & Sarmas, 2007). In other words, there was a correlation between interest rate levels and stock prices, but there was also a 'shock'-factor to the changes in stock prices. Whilst the discrete levels of interest rates correlated with bank stock prices, the paper works only with firms in the banking sector, arguably the sector that sees the strongest direct impact from monetary policy interest rates. Another study working with stock prices using a larger sample of industries in emerging markets found only a moderate effect of interest rates on stock price changes in a prediction setting, with currency exchange rates and trading volume being the strongest predictors (Sabri, 2004). However, the study was conducted prior to the 2008 Global Financial Crisis and for emerging markets, so the results might not be reproducible in a contemporary Norwegian setting. Other papers echo the causal effect between interest rates and stock prices in crises for specific industries; as parts of large sets of macroeconomic variables; but with caution to their interpretability (Mouna & Anis, 2017; Çakmaklı & van Dijk, 2016; Tahir, Gul & Qazi, 2019). A paper on Japanese stock forecasting used lagged stock returns and interest rates as prediction variables and found both to be useful, but struggled with a lack of variance in interest rate levels (Aono & Iwaisako, 2011). This raises an issue with controlling for interest rates in most empirical studies, as the policy remains unchanged for several years at a time. We also hypothesise that an interest rate which is stopped after a series of hikes or reductions will send signals to financial markets about economic conditions. For example, the central bank might communicate that they will pause rate hikes as inflation has started to normalise. So, in a substantial selection of papers, results are highly mixed. This highlights the importance of the contents of central bank communication as opposed to only controlling for the level of an interest rate at a given time.

In the field of finance and accounting, Tetlock's paper on investor sentiment paved the way for much of today's literature (Tetlock, 2007). In his paper, Tetlock quantitatively measures interactions between a popular Wall Street Journal column and the stock market. The findings indicate that negative or positive sentiment in the column correlated with asset prices, an assumption that is central to this thesis. Tetlock uses the Harvard-IV dictionaries in his sentiment analysis. These are general-language dictionaries, not specifically intended for finance and accounting topics (Loughran & McDonald, 2011). They consist of positive and negative words, as classified by psychologists. In his paper, Tetlock constructs a media pessimism measure. He finds that high levels of media pessimism induce downward pressure on market prices. Unusually high levels of pessimism predict a higher trading volume in the markets. Loughran and McDonald further developed the research on sentiment analysis in the field of finance and accounting (Loughran & McDonald, 2011). The Harvard-IV dictionaries were commonly used in sentiment analysis at this time. However, Loughran and McDonald argue that the Harvard-IV dictionary does not necessarily translate as well when used in specific fields such as finance since words in the English language can often have multiple meanings. Instead, they use over 50 000 firm-year 10K-filings to create their own dictionary. They show that a substantial part of the words in the dictionaries from Harvard is misclassified when it comes to a financial context. Almost as much as 75% of the words classified as negative in the Harvard dictionaries are not considered negative in Loughran and McDonalds financial dictionary. They conclude by saying that researchers in the field of finance should exercise caution when relying on word classification schemes not made for finance and accounting contexts. This conclusion is mirrored in Gentzkow, Kelly and Taddy (2019), however, here the method of using dictionaries in natural language processing is criticised for having potentially low power compared to more sophisticated methods. These methods can include, but are not limited to, generative models, multinomial inverse regressions and word embeddings. These methods all have their strengths and the task at hand often dictates what method is favoured. However, combining techniques is also a perfectly valid option (Gentzkow, Kelly & Taddy, 2019). Further, Gentzkow, Kelly and Taddy (2019) argue that dictionaries can be very useful in situations where prior information is strong. This would typically be in a situation where there is no training data to fit a supervised model. A dictionary-based method would then be a strong option as long as the information captured in this dictionary is also seen in the new text data. However, these dictionary-based methods will soon be outperformed by more sophisticated methods due to rapid expansions in the domain of machine learning and related statistical methods. (Gentzkow, Kelly & Taddy, 2019). Perhaps the strongest argument for using methods based on machine learning or other statistical methods, is that they are free of any bias present when dictionaries are produced by humans. The dictionaries will naturally be biased by the texts they are trained on but are free from human preconceptions of the topic. On the other hand, Loughran and McDonald (2020) argue that even though machine learning methods might find a set of words able to predict, for example stock returns, they might not capture the actual sentiment. Rather, they might identify attributes that produce positive or negative outcomes.

The increased availability of computing capacity matched with the shortcomings we listed regarding manually produced dictionaries led researchers to explore the usage of machine learning (ML) in the field of textual analysis and sentiment analysis. Taddy introduced the multinomial inverse regression (MNIR)-model, the central method of our paper, to the field of textual data analysis (Taddy, 2013). In essence, Taddy's method relies on using distributions of phrase counts to predict a variable, or a supervisor, assigning coefficients to phrases. This methodology was built on by the works of Garcia, Hu and Rohrer in a paper highly influential to our own research (Garcia, Hu & Rohrer, 2020). In their paper, they use quarterly earnings calls of companies and their stock prices to create sentiment dictionaries by using MNIR with the stock prices as the supervisor. They also apply cross-validation to their method to make it more robust, by using the dictionary to produce sentiment and to predict stock prices out-of-sample in company 10-Ks. By doing this, Garcia, Hu and Rohrer were able to out-perform the predictions of Loughran & McDonald and produce a sentiment dictionary with more breadth and objectivity for the Finance & Accounting field of research (Garcia, Hu & Rohrer, 2020). Whilst the papers of Loughran & McDonald and Garcia, Hu & Rohrer produce sentiment dictionaries which can be utilised by any researcher, likely with a high degree of success, there are some issues. First, they are best suited to a specific Finance & Accounting context (discussing financial results) as they are trained or produced from textual data relating to earnings calls and quarterly results. For the same reason, the language used is most commonly found from corporate executives, and the dictionaries might not perform well when adapted to more casual settings such as online investor forums. Finally, and notably to this thesis, the dictionaries are in English, and are not applicable in a non-English speaking context unless carefully translated manually, which is both time-consuming and subject to the same biases we highlighted for other manual dictionaries. Instead, the work of Garcia, Hu & Rohrer should entice researchers to produce sentiment dictionaries for their own purpose in whichever context they decide. With sufficient textual data and a chosen supervisor, researchers can now perform sentiment analysis without being limited by off-the-shelf solutions that are only for English language, or by the biases they expose their research to by producing such a dictionary themselves, something we will attempt to take advantage of in our analysis.

Textual data analysis has seen limited attention in a Norwegian context. A substantially relevant paper by Kirkeby and Larsen attempted to create a category specific sentiment measure using a manually produced sentiment dictionary in Norwegian news data (Kirkeby & Larsen, 2021). They were successful in developing a sentiment measure which correlated with other macroeconomic indicators, and their sentiment dictionary is applicable in Norwegian contexts though it is subject to the aforementioned biases of manually produced dictionaries. Larsen also attempted to measure category-wise uncertainty by counting the occurrences of uncertainty-terms category-wise in news data, identifying multiple sources of uncertainty (Larsen, 2020). Ter Ellen, Larsen & Thorsrud later attempted a similar strategy to create an “uncertainty”-measure by simply counting the number of occurrences of the term “usikkerhet” (uncertainty) in Norges Bank communication, finding correlations between the uncertainty measure as well as other macroeconomic indicators (Ter Ellen, Larsen & Thorsrud, 2022). Larsen and Thorsrud also investigated the occurrence of specific topics in Norwegian news data and their correlation to stock returns, further showing that NLP-methods are highly applicable in non-English-speaking contexts (Larsen & Thorsrud, 2021). As evidenced in the literature, several attempts have been made to conduct sentiment analysis in a Norwegian context, but these all rely on very simple methods such as counting the number of occurrences for a specific term, or by using manually produced sentiment dictionaries subject to human bias. There is a gap in the literature to create a more objective and widely applicable sentiment dictionary for Norwegian sentiment analysis.

Central bank communication has been the focus of several past studies, but this has typically been limited to the Federal Reserve of the United States or other English-speaking entities. A paper working on FOMC (the Federal Reserve Board) statements used a natural language processing algorithm to capture the surprise-component of the monetary policy statement (Doh, Song & Yang, 2022). Here, surprise is identified as variation between the expectations of markets/listeners from monetary policy statements and the actual statements. They find that when there is a surprise present and there is an issuing of quantitative tightening, stock markets decline. Two of the aforementioned authors had previously investigated the tone of statements’ effect on the prices of bonds, and found particularly strong correlations in medium-term expectations (Doh, Kim & Yang, 2021). Effectively, the authors of these two papers display that machine learning and NLP can capture quantitative data from qualitative information issued by central banks in a manner that correlates with asset prices in developed economies. Another study using statements from the Eastern Caribbean Central Bank (ECCB) used text mining tools to create a readability index, which they found to correlate with the accumulation of foreign assets by the ECCB (Caterini, 2020). In other words, as the complexity of the ECCB statements increased, the accumulation of foreign assets was found to decrease, which the authors explain with a lowered credibility in the context of emerging economies. These papers are all successful in an application of machine learning and NLP when working with central bank communication and were able to extract components which correlate with other macroeconomic indicators or the prices of assets. This is promising for our own research objectives, as we seek to apply similar ML methods to extract similar components of central bank communication, but in a Norwegian context rather than an English-speaking one.

The literature we have examined evidences a certain chronology of events. Empirical studies which attempted to measure the effect of central bank monetary policy on other macroeconomic indicators or asset prices had largely mixed results. This motivated researchers to use more qualitative information, developing sentiment analysis methods which relied on dictionaries outlining negative and positive terms. However, these dictionaries were manually produced and subject to bias, and here researchers found an application for machine learning to produce more objective and unbiased methods of extracting quantitative metrics from text. This is also shown to have been successfully applied on central bank statements from the FOMC and the ECCB, showing that ML can be used to show how central bank communication causes fallouts in financial markets. We also identify that there exists a large gap in sentiment research in a Norwegian context, as the work that has been produced, whilst prudent, is mostly reliant on simple methods and manually produced dictionaries. This thesis' contribution consists of an attempt to fill the gap in the literature by applying ML to sentiment analysis in a Norwegian Finance & Accounting context, working with central bank communication, something that has not been executed previously. It might also serve to further develop our understanding of the dynamics between central bank communication and financial markets in developed economies by attempting to extract the sentiment of central bank statements which might serve as an addition to prior research attempting to correlate components of monetary policy text with other indicators.

3. Data & Pre-Processing

In this section we provide a description of our data, a set of Norges Bank publications. We will be frequently referring to the term ‘corpus’ which can be understood as the equivalent to a dataset but for textual data. More accurately, it is a set of text documents. Next, we also outline the pre-processing steps required for data cleaning and preparation prior to applying our methodology.

3.1 Describing the Data

Our corpus consists of 186 Norges Bank monetary policy evaluations (MPEs) published from 1999 to 2023 (Norges Bank a., 2023). Each MPE contains a decision about the ‘Styringsrente’ (interest rate), the Norwegian equivalent of the Federal Funds Rate in the United States or the European Central Bank interest rate. They also contain an evaluation of economic circumstances including unemployment, inflation, uncertainty, and so on, as the reasoning behind their monetary policy decision. In other words, they provide the context and background for the Norges Bank’s monetary policy decision. Each document consists of approximately 2-3 pages with a mean word count of 661 across all years. The frequency of MPE publications varies somewhat, possibly in-line with the aggressiveness of the interest rate changes. Norges Bank issued 8 or 9 MPEs per year in the period 1999-2011 except in 2008 in which they held 10. 6 publications were issued per year in 2012-2015, and 8 publications were issued per year in 2018-2022. Along with each MPE is published a highly condensed monetary policy decision press release. Whilst these documents overlap somewhat, the press releases will mostly mention only the final monetary policy decision, along with 1-2 sentences about the reasoning for this, and are thus much more limiting as a corpus. Each MPE is dated the same as the public presentation in which the interest rate decision is announced. These public presentations or press meetings are the pieces of communication typically discussed and cited in the media by news organisations. The language used in the MPEs is considerably more formal and academic compared to the language used in the public presentations. However, we believe this corpus is still highly suited for the purpose of our analysis, as keywords such as “inflation”, “crisis” or “uncertainty” will still appear in both mediums. As press meeting footage or transcripts are not readily available for early dates in our corpus, we assume that the content covered is reflected in the topics covered in the publications. The corpus is publicly available in the Norges Bank archive of monetary policy meetings (Norges Bank a., 2023).

The MPEs vary in length measured by word count, which has some implications for our analysis. Word count here is calculated as the number of strings separated by whitespace (the number of individual words). We hypothesise that these differences can be attributed to a change in reporting standards over time. This is because the publications contain different combinations and bundles of documents and press releases. In the periods 1999-2002 Norges Bank published the MPEs as an introduction to the press conference (presentation). Then, in 2003, they ceased publishing the MPEs separately and instead published one combined document for the press releases and the MPEs, which might lengthen or shorten the documents. This reporting standard was kept until 2008 where they started to publish the MPEs separately again, under a new name dubbed 'Hovedstyrets begrunnelse for rentebeslutningen' (the Board's reasoning for the monetary policy decision). In 2011 they again changed their reporting standard as they started to involve inflation reports in their publications. Inflation reports were released quarterly. On the dates of inflation report releases Norges Bank would still publish an MPE, but on the dates where inflation reports were not published they would instead release a document containing the changes since the previous inflation report. These 'change'-documents follow a similar format to MPEs in terms of language and length but in less of an academic outline, instead relying on notes. They continued to do this until 2013 where they again started publishing MPEs with each press release. Norges Bank would continue to publish separate MPEs until present day, but the name of the document was changed to 'Monetary Policy Evaluations' in 2019. So, the reporting standards have been altered significantly over time which will affect the contents of each document as well as their length and format. We explore this theory further in Part 5. It is also possible that MPEs change in length due to macroeconomic circumstances. Figure 2 shows that MPEs increased in length during the middle of 2002 with which we are unable to correlate any macro-events. There was again a spike in the length of the documents in 2012 which might be related to the Euro-crisis (Johnsrud, June 2012). We also identify a spike in 2020 which might relate to the Coronavirus-pandemic. Our initial hypothesis was that press releases would be longer during the 2008 Global Financial Crisis and in 2022 which saw rapid inflation and central bank response, but there appears to be no correlation between these events and the lengths of the MPEs. This leads us to believe that reporting standards are a considerable driver in the change of document length. During some periods, remarks and points from the press conferences and introductions are included in the press release, while in other periods they are omitted. This happens exogenously from there being periods of inflation or monetary policy adjustments. Regardless, this causes an analytical issue as the sentiment of the communication might correlate with the text length which we explore further in Part 5.

We mentioned previously that the reporting standards have shifted throughout the years, which demands that we make decisions on which documents to include in our corpus and which documents to omit. The aim is to always collect, in essence, the same document for each publication date. For publication dates ranging from 1999-2002 we collect the documents titled 'Innledning til pressekonferanse' (introduction to press conference). For 2003-2007 we collect the press releases which were greatly extended in length and detail while Norges Bank ceased to publish the separate MPEs. For 2008-2010 we use the MPEs titled 'Hovedstyrets begrunnelse for rentebeslutningen'. For 2011 and partly in 2012 we use the documents titled 'Ny informasjon siden pengepolitisk rapport' (New information since monetary policy report) on dates where there is no MPE, and we use the MPE on the dates where they are released, which coincides with the dates where a monetary policy report is published. From 2013 to 2019 we use the MPEs titled 'Hovedstyrets vurdering' (the Board's evaluation) and the MPEs titled 'Pengepolitiske Vurderinger' (Monetary Policy Evaluations) from 2020 to present day. In this way, we are using the same fundamental document for all the dates. Whilst the documents vary somewhat in length, the purpose of the MPE stays the same for each document, which is to provide the public with an evaluation of macroeconomic circumstances and reasoning for increasing, decreasing or leaving the interest rate unchanged. Each document is downloaded manually and imported using the rvest-package's web scraping capabilities in R programming (Wickham, 2022). In some instances, the MPEs are uploaded as a Portable Document Format (PDF). For these documents, the PDF is downloaded manually and imported using the pdftools-package in R (Ooms, 2023). This constitutes our primary corpus and the text documents we will use for NLP and cross-validation.

3.2 Preprocessing Steps

We conduct a series of pre-processing steps in order to clean and prepare the data for model fitting and sentiment analysis. First, we remove some repetitive sentences such as introductory remarks, disclaimers and generic statements about the mission of Norges Bank which are present and equal in a large number of documents. We also remove other items such as contact details. Next, we convert all the letters to lowercase to prevent duplicate terms from being detected. For example, ‘Usikkerhet’ and ‘usikkerhet’ (uncertainty) should be interpreted as the same term, not two different terms. We also remove all punctuation for similar reasons as converting capital letters to lower. We replace line breaks with whitespace. All numbers are removed as we do not want dates and the mentioning of specific metrics and rates to be included in the model. These can be controlled for separately, and our analysis attempts to capture the sentiment of the presentation’s effect on stock returns, as opposed to the stock return effect of a specific interest rate level. Finally, we remove stopwords using a Norwegian list of stopwords from the tm-package in R (Feinerer, 2023). This leaves us with the raw text of each press release and press conference in a cleaned manner for sentiment analysis and machine learning applications.

Finally, we need to construct a document-term-matrix (DTM) as part of our data cleaning and preparation. Our corpus is a set of n documents. Each document has a set of unique unigrams p (unique words, terms). The DTM counts the occurrence of each term in each document. In other words, it creates a list of every term in the corpus on one axis, then an identification number of each document on the other axis, hence constituting a matrix. The values are then the occurrence of the term in the intersecting document. We require that word lengths need to be no shorter than 3 letters and no longer than 23. This choice is arbitrary, but it was identified as a reasonable limit to prevent very short, meaningless words while also filtering out editing errors or words that are conjoined. We also specify that a term will not be included in the document term matrix if it occurs in fewer than 20 documents or if it occurs in more than 70 documents out of 186. This way we avoid terms that only appear in a select few documents which would be inapplicable for a more generalised corpus, and we also avoid terms which appear in most or all documents, which would not have any meaningful association or be correlated with any stock price shock (if the term exists in all documents, there is no variance). This results in 343 unique terms, with 19684 non-sparse entries and 23840 sparse entries, sparse meaning close to 0 counts. Our DTM contains the terms and data which will be used for our NLP method as outlined in the methodology-part.

4. Methodology

In this section we will outline our methodology. We will attempt a selection of methods for capturing the sentiment of the MPEs and attempt to correlate the sentiment measure with the returns of the Oslo Børs Exchange-index (OBX). The core of our method is the usage of multinomial inverse regression (MNIR) and its performance will be evaluated against a few naive methods generated by us and by the works of Kirkeby and Larsen (2021).

4.1 Multinomial Inverse Regression

MNIR was developed by Taddy, and further built on by the works of Garcia, Hu and Rohrer for the purpose of textual data analysis (2013; 2022). Our notation will be consistent with that of Garcia, Hu and Rohrer (2022). MNIR is, first and foremost, a machine learning model. This entails that the model attempts to predict a dependent variable using independent variables, with all receiving a coefficient or a "loading" denoted as z in terms of their predictive power and the direction of their effect. To translate that to this context, we will use the OBX returns R_j as the dependent variable, the supervisor, and we will use the term frequencies ij of the DTM as predictors. Each term will receive a loading which will either be negative or positive depending on whether they are useful (frequently present) exclusively in predicting a negative or positive OBX return observation on the same date. MNIR uses a lasso-style penalty, specified by, to include or omit relevant terms. The model fitting method accepts two hyperparameters: and n. We require that $\gamma = 0.1$ and $n\lambda=1$. Upon executing the model regression and model estimation, the MNIR algorithm will omit some terms which are deemed irrelevant due to its penalisation parameter, and it will attach coefficients to the term frequencies in our corpus based on their correlation with OBX returns. This yields our sentiment dictionary with an overview of the positivity and negativity of all p n-grams. We then use this dictionary to produce predictions on the training set and the test set in order to evaluate the validity of the model in terms of its predictive power and overfitting. Put in simple terms, we investigate which terms are used in MPEs when the OBX returns are up and which terms are used in MPEs when the OBX returns are down, with some restrictions on what terms may or may not be included depending on how often they appear and in how many documents they appear. The resulting sentiment dictionary is an item of high importance to this thesis and will be discussed in Part 6 as it is of key interest whether the method has produced a robust and plausible dictionary of negative and positive terms in a Norwegian context.

In order to prevent overfitting, we make a train/test-split of our dataset, which is common practice in the domain of predictive analytics and machine learning. This is done using random sampling. At random, MPE publications are sorted either into the training data, dubbed “Group A”, or the test data, dubbed “Group B”. We will only be working with documents present in Group A for the purpose of training and estimating the MNIR model, which means that terms occurring in Group B will not be included in the sentiment dictionary unless they also appear in Group A. The MNIR algorithm is fitted on term frequencies in Group A, and the resulting dictionary is used for calculating sentiment scores and predicting OBX returns in Group B. This is referred to as cross-validation.

4.2 Naïve Methods

We will also employ a series of naive methods, which rely on a simpler measurement of sentiment in the MPEs using dictionaries developed by ourselves for the purpose of this analysis, and by Kirkeby and Larsen (2021). In their paper, Kirkeby and Larsen developed a relatively small sentiment dictionary for the purpose of measuring sentiment in Norwegian news data. Whilst their corpus is different to that of this thesis, we believe that the dictionary might be applicable in both contexts as they were also working with news regarding finance and economics. The Kirkeby and Larsen dictionary (KL) consists of 73 negative words and 70 positive words (Kirkeby and Larsen, 2021, p 20). They also produced a simple way of measuring uncertainty by counting several variations of the word “usikkerhet” (uncertainty) in a text. We will also employ this list in our analysis to produce the same uncertainty measure for MPEs, as this might also be a valid metric to help explain shocks in OBX stock returns. Finally, we produce our own simple list of words which intuitively would be used in negative contexts in MPEs, such as “war”, “fear of recession” and “concerning” (Appendix H). These lists and dictionaries comprise a series of more simple and naive methods, all of which are intended to capture a sentiment metric which might correlate with OBX stock returns and will be evaluated against the dictionary resulting from the MNIR method.

After specifying the lists of terms and sentiment dictionaries, we use these methods to compute the sentiment of the Norges Bank MPEs. To compute the sentiment we will use a bag-of-words approach in which we count the number of occurrences for positive terms, deduct the number of occurrences for negative terms and divide by the total word count of the document. For the dictionaries and lists which are intended to capture a measure of uncertainty and a measure of negativity we will simply count the number of occurrences and divide by the total word count. This follows the methodology of Kirkeby and Larsen in their work (Kirkeby and Larsen, 2021). We follow the notation of Kirkeby and Larsen, which defines that sentiment S, uncertainty U and negativity N scores for document i are computed as:

$$S_j = \frac{(\#positive\ terms - \#negative\ terms)_j}{wc_j} \quad (1)$$

$$U_j = \frac{\#uncertainty\ terms_j}{wc_j} \quad (2)$$

$$N_j = \frac{\#negativity\ terms_j}{wc_j} \quad (3)$$

where wc_j is the total word count for the MPE j.

For the dictionary computed by MNIR we will also attempt to make use of the information gathered by the model fitting which specifies the loading of each individual term. In other words, each term will be negative or positive to varying degrees, and we will also attempt to produce sentiment which incorporates this information and controls for the fact that some terms will have a stronger shock than others when used by Norges Bank in the MPE. Here we follow the notation of Garcia, Hu and Rohrer with some slight variations to suit the readability of our own thesis (Garcia, Hu and Rohrer, 2022). The MPEs j are divided in a 50/50 split into a training set and a test set. We build a dictionary with a set of terms p from the DTM δ which only contains the terms from the MPEs in the training set. The term frequencies tf_{ij} are the frequencies of the terms p expressed as a vector. Then, the sentiment Z for document j is:

$$Z_j = \sum_{i \in \delta_i} (\frac{tf_{ij}}{wc_j}) \tag{4}$$

4.3 Empirical Model

Following the measurement of sentiment scores in the MPEs we will estimate our empirical model which is consistent with the works of Garcia, Hu and Rohrer (2022). We test the validity of all the sentiment measures computed by running linear regressions, using the OBX returns R on a specific date t as the dependent variable, and the sentiment score S or Z, with Z denoting the sentiment score computed with the MNIR-developed dictionary. Our empirical model can be specified as:

$$R_{jt} = BS'_{jt} + \gamma X'_{jt} + \epsilon_{jt} \tag{5}$$

where X' are control variables and constitutes our error term. Note that we will use both sentiment scores S and Z, negativity measure N and uncertainty measure U in separate regressions, all denoted here by S' in our empirical model to simplify. We are mainly interested in the -coefficient and its statistical significance, as it will indicate whether we have successfully developed a sentiment measure which correlates with stock returns. Additionally, our literature review suggests that there might exist a ‘shock’-component to the dynamic between central bank communication and stock returns (Ghazanfari, Rogers and Sarmas, 2007). To attempt to capture this, we test an additional regression model relying on the change in sentiment from the previous MPE, effectively creating a sentiment delta. Our hypothesis here is that higher changes in sentiment might be what correlates with stock returns. For these models, the specification is:

$$R_{jt} = \beta \Delta S'_{jt} + \gamma X'_{jt} + \epsilon_{jt} \tag{6}$$

where denotes the change in sentiment for this MPE from the previous document. These empirical models will be estimated and compared in terms of their statistical significance, requiring a cut-off of a p-value of 0.05 for statistical significance at the 95% confidence level for the variables to be considered significant in explaining variance in stock returns.

5. Results

We apply four different methods for measuring the effects of central bank communication on stock prices. First, the machine learning, NLP-method of the MNIR model where we look at sentiment scores of the documents in the corpus. Using the word loadings from our machine learning model to characterise the sentiment of the words in our corpus, we make a positive and a negative sentiment dictionary. We take a closer look at these dictionaries and the words they contain. Second, we copy the uncertainty-method of Kirkeby and Larsen (Kirkeby and Larsen, 2021) and use word counts to see if variations of the word "usikkerhet" (uncertainty) can provide significant results. We apply another of Kirkeby and Larsen's methods by using their positive and negative sentiment dictionary. (Kirkeby & Larsen, 2021). Lastly, we produce our own list of negative words and use word counts to measure shifts in the stock market similarly to the Kirkeby and Larsen method. We also calculate a delta for the four measures. This delta looks at the change in the scores for the three measures between press releases of the monetary policy evaluations (MPEs). The goal of these deltas is to measure shocks in the stock markets.

5.1 MNIR Dictionaries

In this section we present a list of the most frequent positive and negative terms resulting from the MNIR model (Appendix G). The positive terms list consists of 65 words in total. Words referencing months (February, October, November) appear relatively often and might indicate a generally higher sentiment in these months. However, we are not in any position to draw any conclusions based on this. Similarly, we find words referencing countries (Sweden, USA, United Kingdom) which might indicate positive relationships, even though we are not able to make conclusions here either. Then there are other words you would expect to find in a positive dictionary in a financial or macroeconomic setting such as "betydelig" (significant), "større" (bigger) and "vekstutsiktene" (growth prospects). Our intuition tells us that these terms might be used in a positive context. Many of the words in the list are words most of us would not be able to classify as either positive or negative. "Komiteen" (the committee), "indikerer" (indicates), "nettverk" (network) and "virkningene" (the effects) are words that have no inherent positive or negative meaning but depending on context might be used in either situation. Lastly, we also see a few words we would not expect to see in a positive context. “Usikkerheten” (the uncertainty) “stramt” (tight) and “lavt” (low) are all words that alone fit better in a negative setting. This is a strength of the MNIR-method, as it is less subject to human bias and will objectively identify terms which are used negatively and positively. However, saying that the unemployment rate is low is undoubtedly something positive and might explain why we find some of these words in a positive term list. This also indicates that whenever the central bank refers to the employment rate, a majority of the times it is in a positive context. To conclude, we have seen that the MNIR provides a list of positive words including some that are not easily categorised by humans as well as words that fit more into our notion of what positive sentiment is.

We will now have a look at the negative terms from the MNIR model (Appendix G). The list contains 85 words in total, and we see a similar pattern as with the positive terms list where some words initially fit better in the expected context, while some are quite the opposite. “Faren” (the danger), “nedgangen” (the decline) and “kostnadsveksten” (cost increase) all fit in as negatively loaded words. There is also a plethora of words in this list that can hardly be classified as negative or positive without context by humans. “Grunnlag” (basis), “forholdene” (the conditions) and “vurderingen” (the assessment) are all examples of this. These words do not indicate any positive or negative sentiment but are all nonetheless examples of words more associated with falling stock returns in our model. We also find words like “god” (good) on this list, which is more positive than negative in most situations. However, in sentences like “the good times are over”, the meaning changes drastically, which might explain why it is considered a negative word here. We can see that similarly to the positive terms, the negative terms consist of both expected negative words but also words that show a lot of ambiguity depending on what contexts it is used in.

5.2 MNIR Sentiment Score

The MNIR model is built on the training data which consist of 50% of our collected data. The remaining 50% is then used as test data to find our results. This split was done using random sampling. Based on the output of the MNIR, we create a sentiment dictionary. Each of the words in the dictionary were associated with a positive or negative score indicating the sentiment. The positive and negative word lists can be seen in Appendix G. By using word counts for positive and negative words and adjusting for total word count in the document, we make a sentiment score for each document. These sentiment scores are used in a simple regression to predict our dependent variable, stock returns at the Oslo Stock Exchange. The regression output seen in the first column in Table 5 shows that the results are not significant using the test set. Using the training set however, provides significant results at the 0.01 level. Furthermore, the R-squared (0.004) indicates that only 0.4% of the variation in stock returns can be explained from the sentiment score (0.413 for the training set). The coefficient at -0.160 indicates that a one unit increase in sentiment score leads to a decrease in stock returns by 0.16. Meaning that a higher sentiment score (more positive) leads to lower stock returns. This goes against our intuition that a positive sentiment in the central bank communication would lead to higher stock returns. However, as previously mentioned these results are not significant. Looking at the sentiment delta regression in Table 9, the coefficient is at -0.007, much lower than the sentiment score itself but also with the negative sign. The R squared is approximately zero and the result is not significant. To summarise, we see that only the training set provides significant results and not the test set. The results we get with the test set are also opposite of what we would expect in terms of the effect that the sentiment has on stock returns.

We find that document length correlates with sentiment as seen in Figure 9. Meaning that as the length of a document increases, so does the sentiment score. While making the model, we controlled for this by dividing sentiment score by the word count of the document. Still, we see that there is a correlation between the two. In Figure 4, we can see how the sentiment score changes over the period 1999-2023 with the MNIR words. Most notably, the sentiment score becomes higher over the years, similarly to the document length seen in Figure 2. In the last decade of the data, the sentiment of the MPEs is generally higher than zero and visibly higher on average than the previous decade. Different levels in sentiment could be caused by changes in publication standards or changes in the monetary policy board, among other things. There is a spike in the sentiment around the time of the COVID-19 outbreak, simultaneously stock returns drop. Initially, this would be counterintuitive to our predictions but this could be due to the increasing length of the documents in recent years. As we have seen, longer documents lead to higher sentiment scores which again could lead to lower stock returns, even though this is not a significant result. There are some negative implications of having a correlation between document length and sentiment. For example, if a long document is mostly negative in sentiment, we might not get a correct negative score solely because of the length. This might prove troublesome particularly in later years where we expect negative sentiment due to COVID-19 and high inflation levels. Incidentally, due to these events, we also expect document length to increase as there are more events to be considered when deciding on policy rates.

Table 1 Regression Results MNIR

Empirical Model: MNIR Sentiment Score

	Dependent variable:	
	OBX Returns Normalised	
	Group a	Group b
	(1)	(2)
Sentiment Score MNIR	0.012***	-0.001
	(0.001)	(0.001)
Constant	0.020***	0.018***
	(0.001)	(0.001)
Observations	93	93
R^2	0.413	0.002
Adjusted R^2	0.406	-0.009
Residual Std. Error (df = 91)	0.014	0.014
F Statistic (df = 1; 91)	63.949***	0.207
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table description: Regression results from empirical model following the computation of sentiment scores using the MNIR dictionary and loading-adjusted z-values. Group a constitutes the training set and group b constitutes the test-set.

Table 2 Regression Results MNIR Bag-of-Words

Empirical Model: MNIR Sentiment Score using bag-of-words		
	Dependent variable:	
	OBX Returns Normalised	
	Group a	Group b
	(1)	(2)
Sentiment Score MNIR bag-of-words	0.420***	-0.041
	(0.068)	(0.060)
Constant	0.027***	0.017***
	(0.002)	(0.002)
Observations	93	93
R^2	0.293	0.005
Adjusted R^2	0.285	-0.006
Residual Std. Error (df = 91)	0.016	0.014
F Statistic (df = 1; 91)	37.726***	0.470
Note:	*p<0.1; **p<0.05; ***p<0.01	

5.3 KL-Dictionary

Next we interpret the results from the KL dictionaries, using the sentiment and uncertainty word lists' measures in our empirical model. When used in a linear regression, the coefficient is 0.407 as seen in Table 8. A one-unit increase in the uncertainty-score leads to stock returns going up by 0.407. However, the uncertainty score is not significant. The R-squared statistic is 0.003. Meaning that 0.3% of the variation in stock returns can be explained by the uncertainty score. The regression output for the uncertainty-delta in Table 9 shows that the coefficient is -0.014 and the R-squared is close to zero. Results here are also not significant. The KL dictionary also provides results that are not significant. The coefficient is -0.160, meaning that a higher sentiment score leads to lower stock returns, something that we also saw with the MNIR method and goes against our intuition of a higher sentiment leading to higher stock returns. The KL-delta is not significant and with a coefficient of -0.007, meaning that we detect no shocks in the stock market using this method as seen in Table 9. Using the methods of Kirkeby and Larsen (2021) provides no significant results for our data set both in terms of their positive/negative dictionaries and their uncertainty-method.

5.4 Negative Dictionary

The last method we apply in measuring the response in the stock market from central bank communication is our own negative term dictionary consisting of 23 words (Appendix H). Similarly, to the previous method, we count the number of times a word from the list appears in each text and divide by the total number of words in the text to get a negative index score. As with the uncertainty-score, we apply the negative score to a simple linear regression, and we see that it performs significantly better as evidenced in Table 8. The R-squared is 0.021, meaning that 2.1% of the variation in the stock returns can be explained by this variable. The result is significant at a 5% level. A one unit increase in the negative index score raises stock returns by 0.633. In other words, more negative words used leads to higher stock returns. In Table 9 the coefficient for the regression with the delta is 0.246, much higher than for the other delta regressions. However, similarly to the other delta coefficients it is not significant. In conclusion, we see that the only method (except for results from training data) that has provided significant results is our negative dictionary. Even though earlier results have not been significant, we see a slight trend in the results that higher (lower) sentiment leads to lower (higher) stock returns.

5.5 Summary

To summarise, in this section we have inspected the positive and the negative frequent terms list closer. Both lists include words that we expect to find in either a positive or negative context. We also see that both lists include many words hardly classified as either negative or positive by human eyes. However, the context of the words often dictates their meaning. We have also looked at the results of the four different methods applied to measure reactions in the stock market based on central bank communications. When comparing the four methods, only the last method using the negative index word list provided significant results. The score from MNIR was significant for the training data but not for the test data as was the case for both KL-dictionaries. The negativity measure was statistically significant. As we have seen though, these results indicated that more negative expressions lead to higher stock returns where we initially would expect the opposite. Reasons why this might be the case will be further discussed in the next section. We also observe a correlation between sentiment and document length, a factor that might have some negative implications for our results as sentiment should be independent of document length.

Table 3 Empirical Model Results Bag-of-Words

Empirical Models: Sentiment Score on OBX Returns

	Dependent variable:				
	OBX Returns Normalised				
	MNIR Bag Group a	MNIR Bag Group b	KL	Uncertainty	Negativity
	(1)	(2)	(3)	(4)	(5)
MNIR Sentiment Score	0.420^{***}	-0.041			
	(0.068)	(0.060)			
KL Sentiment Score			-0.160		
			(0.186)		
Uncertainty Score				0.407	
				(0.534)	
Negativity Score					0.633^{**}
					(0.317)
Constant	0.027^{***}	0.017^{***}	0.020^{***}	0.018^{***}	0.014^{***}
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)
Observations	93	93	186	186	186
R^2	0.293	0.005	0.004	0.003	0.021
Adjusted R^2	0.285	-0.006	-0.001	-0.002	0.016
Residual Std. Error	0.016 (df = 91)	0.014 (df = 91)	0.016 (df = 184)	0.016 (df = 184)	0.016 (df = 184)
F Statistic	37.726^{***} (df = 1; 91)	0.470 (df = 1; 91)	0.736 (df = 1; 184)	0.582 (df = 1; 184)	3.985^{**} (df = 1; 184)
Note:					$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

6. Discussion

In this final section we discuss the validity of our results and outline the implications of our work and findings on the field of study, as well as provide some recommendations for future research.

6.1 Validity of the Results

In this section we discuss the validity of our results. Both approaches to measuring sentiment score outlined in our methodology yield the same conclusions. Whilst the coefficient is statistically significant for Group A and positive, indicating that an increase in sentiment leads to an increase in OBX returns, both coefficients are negative and statistically non-significant when working with the test-sample of Group B as seen in Table 6 and Table 7. The reason why we observe these results is because Group A consists of the dataset which we trained the MNIR model from. In other words, we produced the sentiment dictionary using the corpus from Group A and OBX returns on dates of MPE publications in Group A. The risk of this approach is that the MNIR model is subject to overfitting, which is why we employ a train/test-split of our data to investigate whether we have accurately captured a sentiment score that can predict OBX returns. Evidently, the sentiment dictionary we have produced from MNIR fails to predict stock returns out-of-sample, and it appears that we have not succeeded in producing a dictionary that is generalisable for capturing sentiment in other documents for a Norwegian Finance & Accounting context. Similarly, we observe that the sentiment score and the uncertainty score produced from the KL-dictionaries were also non-significant in predicting OBX returns as seen in Table 8. The only measure that was statistically significant was that which was produced from a list of "negative" words hand-picked by us for the purpose of this analysis. However, the sign of the coefficient is contrary to our expectations, indicating that an increase in the negativity score leads to an increase in OBX returns. It appears here that we have identified key terms, but not the context they are used in. Statistically speaking, phrases like "war", "price inflation" and "recession" are used more frequently in MPEs which coincide with an increase in OBX stock prices. It appears here that a loss of generality is detrimental to our understanding of central bank communication and stock market dynamics, as we need to account for the context in which specific terms are used to measure sentiment more accurately. For example, it might be the case that “price inflation” is discussed both positively and negatively, which we might account for by instead counting phrases containing “price inflation is higher than expected” as negative whilst “price inflation is lower than expected” as positive. Evidently, loss of generality is an issue in our analysis and approach to capturing sentiment, and our model appears overfitted on the training data.

There might not be evidence to suggest that there exists a significant shock in OBX returns on the dates of MPE publications. Whilst we do observe that, aggregated across all years, the date of the MPE publication sees the highest absolute shock in OBX returns when working with a 6-day event window as seen in Figure 3, this difference in return shock is relatively small, equating to about 0.001 higher absolute returns than the day before and about 0.0005 higher absolute returns than the day after. If OBX returns do not in fact correlate with MPE publications, sentiment scores will also not manage to correlate with OBX returns. Building on our event study analysis, we run statistical measures, regressing the OBX returns on the event days around the MPE publications (Appendix D). For most years, the publication date of the MPE has low coefficients, and the coefficients are rarely statistically coefficient when working with a cut-off value of 1.96 for a t-statistic, as seen in Table 4 and Table 5. In fact, only the year 1999 sees a statistically significant shock on the dates of MPE publications. It might appear that the underlying shock in OBX returns is not present for MPE publications. On the other hand, it is likely that it is highly present on certain dates and not present for many others. Intuitively, MPE publications will only cause a shock in financial markets in times of uncertainty or periods of monetary policy changes, such as a tightening in the budget, or an increase in interest rates. In a period of relative financial stability, the MPEs will not even be covered in the media, whilst it will be highlighted in great detail during periods of inflation, crises and when recession fears are high. So, it is possible that there exists shock in financial markets as a result of MPEs, but it might not be frequent enough to elicit statistical significance across the entirety of our dataset.

It is worth exploring why our analysis fails at achieving the results produced by the works of Garcia, Hu and Rohrer in a different context. Whilst we have obtained statistical significance in the training sample and no statistical significance in the test sample, they were able to produce robust correlations both in, and out-of-sample (Garcia, Hu & Rohrer, 2022). One explanation is that we are working with a much smaller corpus and fewer observations for OBX returns. Whilst they were working with 144,383 unique articles spread across 87,198 business days, we are working with 186 unique documents spread across 186 unique days. This thesis is working with far fewer observations and we increase the probability of having fitted our sentiment dictionary on random noise in the OBX returns. Another explanation is that we have worked solely with unigrams whilst Garcia, Hu and Rohrer worked with bigrams and trigrams as well, meaning that they used phrase counts of 3 terms and 2 terms instead of just single terms. To explain, they produced sentiment dictionaries consisting of two-and-two, and three-and-three phrases in addition to the unigram similar to ours. We noted that a loss of generality might be an issue present in our dictionary, and that making attempts to capture the context of a key term might be important to better capture sentiment. Another explanation might be that the underlying OBX return shock on the dates of MPE publications might not be strong enough to produce a robust sentiment dictionary. Whilst we observed some higher absolute returns on the dates of MPE publications as displayed in Figure 3, the comparative paper is working with far higher deviations on the date of earnings calls, above 2% on the date of earnings calls compared to a mean below 1% (Garcia, Hu & Rohrer, 2022, p. 32). The final explanation we wish to highlight is that the dynamic between the earnings calls used in the comparative paper and stock returns is potentially highly different from that of Norges Bank MPEs. In their paper, Garcia Hu and Rohrer use company-specific articles and returns to produce sentiment dictionaries. In this context, it is quite definitive that negative terms in an earnings call will result in negative returns in the associated company's stock price, but for Norges Bank, the relationship might not be as clear cut. For example, a Norges Bank MPE signalling high inflation is negative, and might cause some companies' stocks to fall, but for companies in the banking sector it might lead to increased returns, as a hike in interest rates leads to higher revenues for banks. Another notable example is that if Norges Bank communicates that there is an energy crisis on the horizon, most companies' stock prices would respond negatively, but Equinor, a heavyweight on the OBX index, would surge in value as they benefit from the increased cost of oil. By and large, we believe that the difficulties we have experienced in reproducing the robust results of Garcia, Hu and Rohrer in a Norwegian context can be attributed to underlying issues in the data and the reliance on one single index as the 'general' response of financial markets.

The points mentioned above raise the question of whether we have truly captured the sentiment score of the MPEs using the MNIR-method. As we saw in our literature review, Loughran and McDonald (2020) argues that the MNIR-method might be good at identifying attributes that produce positive or negative outcomes. However, it might not capture the sentiment of the situation. Our MNIR dictionaries include many words that alone do not seem to fit into either a positive or negative sentiment context. One explanation could be that they only capture some characteristics or attributes in the macroeconomic landscape that affects stock markets and that sentiment has little to nothing to do with it. This is a potential strength of an MNIR-produced dictionary, but it might result in the actual measure not truly being equivalent to that of a sentiment score. As seen in Figure 4 and Figure 5, the sentiment score is largely negative despite some large shocks in the first half of the time horizon of the corpus, whilst it is almost always positive in the latter half of the time horizon (Appendix F). This is contrary to our intuition, seeing as we expected at least a series of negative observations for the sentiment score in 2021 and 2022 when Norges Bank issued aggressive rate hikes and negative press conferences in the light of record-high inflation. One might argue that this is due to the fact that the OBX stock price has, by and large, grown rapidly in recent years, but this does not explain why the sentiment score has a large positive shock during COVID-19, when OBX returns were significantly down. Based on this, there is evidence to suggest that the MNIR-dictionary has not accurately captured a true measure of sentiment and might be creating term loadings endogenously.

6.2 Tuning Parameters

When fitting the MNIR-model to produce our sentiment dictionary we make several decisions in terms of hyperparameter specifications. First, when producing the DTM we are presented with a few choices for the inclusion and exclusion of terms. We are able to change the bounds of the DTM, which effectively excludes terms that are present above a certain number of documents and/or below a different certain number of documents. The goal here is to strike a balance between including terms that are important enough to be present in more than just a few documents, but not as frequent as being present in most or all documents, as those terms would be void of any significance and wouldn’t signal any particular topic or key term. If we specify a narrower interval in terms of which terms to allow, we are left with a sparser DTM, but possibly a more relevant selection of words. If we specify a wider interval we will include a larger number of terms but possible less relevant ones too. We are also able to place constraints on word lengths. We experienced that our strictness in these parameters by and large didn’t change the end-results up to a certain point.

If the constraints are too loose, the sentiment dictionaries are filled with meaningless terms such as prepositions and conjunctions. Second, we are required to specify a gamma-parameter when fitting the MNIR-model. This gamma-parameter refers to the gamma-lasso function of the MNIR method (Taddy, 2013). The higher the gamma, the stricter the penalisation parameter in the function, and the terms will see lower loadings or they might be erased entirely if not found to be important. Several parameter specifications were tested manually ranging from 0.001 to 10000, and these changes largely did not change our end-results. Whilst the coefficients were somewhat smaller and larger for each change in the gamma-parameter, the coefficient for the test-set of the corpus was still negative and non-significant. As a consequence of this, we inspect the dictionary output resulting from tweaking these parameters. Upon inspection, we found that excluding terms present in fewer than 20 and more than 70 documents, excluding words shorter than 3 letters and longer than 23, and choosing a gamma of 0.1 yields the most robust sentiment dictionary intuitively, though it does not produce any significant results no matter what specification we decide.

6.3 Implications & Future Recommendations

The results of our analysis holds some implications for the field of textual analysis and natural language processing. We have shown that the MNIR framework developed by Taddy and extended by Garcia, Hu and Rohrer can be used for producing sentiment dictionaries in a non-English-speaking context (Taddy, 2013; Garcia, Hu & Rohrer, 2022). However, the role of manually produced dictionaries cannot be discounted just yet. Researchers stand to benefit from employing a multitude of methods for computing sentiment scores when working with textual data. Furthermore, a loss of generality might be detrimental in sentiment analysis, as we have seen the importance of contextual terms in a dictionary for computing sentiment scores. Researchers should strive for a deep understanding of their corpus and it would also be beneficial to work with bigrams and trigrams. It is also important that researchers work with a corpus that is consistent in terms of its lengths and contents, as document lengths and reporting standards may correlate with document sentiment. Researchers should also note that the MNIR-method for computing dictionaries might be useful for other purposes than just computing sentiment. In this thesis, we have been working under the assumption that it is the degree of negativity or positivity in an MPE which is driving stock return shocks in the OBX, but it might instead be that it is a surprise-factor, or a level of fear for example. In other terms, any underlying element in speech or text that is driving OBX returns might be measured with MNIR, or at least, it will detect words used in these contexts. Whilst we have been unable to produce statistically significant results in this thesis, the usage of MNIR opens the path for computation of sentiment scores in non-English-speaking contexts, and for computation of other scores and metrics in speech and text which might drive a plethora of macroeconomic indicators.

7. Conclusion

This thesis sought to apply multinomial inverse regression and sentiment analysis to a Norwegian financial context in order to better understand the dynamic between central bank communication and stock returns. We did this by performing a train/test-split on a corpus of monetary policy evaluations published by Norges Bank, training the MNIR-model on the training data to retrieve a sentiment dictionary, and investigating for correlations in the test-data. For performance comparison, we also employed the dictionaries and measures of Kirkeby and Larsen (2021). Our results are mixed. Whilst we were able to retrieve a positive correlation in the training set, our sentiment dictionary did not produce significant correlations out-of-sample, and neither did the KL-dictionary or uncertainty-measure. Instead, a measure for negativity, a handpicked list of perceived negative words, yielded significant results. We have discussed the validity of our results, soliciting questions around the loss of generality as well as the underlying statistical presence of any shock in our empirical model for the dates of the publications. Finally, we have outlined the implications of our analysis for the field of textual data analysis, citing that MNIR creates the possibility for the creation of a range of scores and metrics in English, as well as non-English contexts.

We started our thesis with two research objectives. We wanted to apply sentiment analysis in a Norwegian context using a less biased dictionary produced by MNIR, and we wanted to investigate the dynamic between Norges Bank communication and stock returns. We have succeeded in both our research objectives as we have managed to produce a sentiment dictionary with a list of positive and negative words that are largely plausible. Whilst we were unable to retrieve statistically significant results upon the application of said dictionary for sentiment analysis, we were able to extract some insights from this. We found that, as the length of the document increases, sentiment increases, and we also conclude that the dynamic is more complicated than positive sentiment corresponding to higher stock returns in financial markets, as what is positive for one specific firm might be negative for another's results. Our hypothesis was that MNIR could be used to produce a Norwegian Finance sentiment dictionary, and that a sentiment score would correlate positively with OBX returns. We are unable to accept this hypothesis, as we cannot reject a null-hypothesis of sentiment scores not affecting OBX returns when using cross-validation. Still, we believe that MNIR lowers the barrier for more researchers to conduct sentiment analysis without relying on English translations and without exposing their research to human bias, but also that MNIR can be utilised to extract other scores and metrics in text depending on the specification of the model and theorised relationship. Where our thesis has been unsuccessful, others might find success working with alternative corpora or by exploring variations in the methodology.

References

Aono, K., & Iwaisako, T. (2011). Forecasting Japanese Stock Returns with Financial Ratios and Other Variables. *Asia-Pacific Financial Markets, 18*(4), 373–384. https://doi.org/10.1007/s10690-010-9135-z.

Çakmaklı, C., & van Dijk, D. (2016). Getting the most out of macroeconomic information for predicting excess stock returns. *International Journal of Forecasting, 32*(3), 650–668. https://doi.org/10.1016/j.ijforecast.2015.10.001.

Cox, J. (2022, August 26). Powell warns of 'some pain' ahead as the Fed fights to bring down inflation. *CNBC,* 8.

Doh, T., Kim, S., & Yang, S. (2021). How you say it matters: Text analysis of FOMC statements using natural language processing: Agricultural and business conditions, tenth federal reserve district. *Economic Review - Federal Reserve Bank of Kansas City, 106*(1), 25-40. https://www.proquest.com/scholarly-journals/how-you-say-matters-text-analysis-fomc-statements/docview/2542755670/se-2

Doh, T., Song, D. and Yang, S. X. (2022). Deciphering Federal Reserve Communication via Text Analysis of Alternative FOMC Statements. *Federal Reserve Bank of Kansas City Working Paper Forthcoming*. http://dx.doi.org/10.2139/ssrn.4135736

Norges Bank (2023). Rentebeslutninger. *Pengepolitikk*. Retrieved from https://www.norges-bank.no/tema/pengepolitikk/Rentemoter/?tab=newslist.

Feinerer, I. (2023). *Text Mining Package*. *CRAN.*

Ghazanfari, F. P., Rogers, H. C., & Sarmas, P. (2007). The effect of the Federal Reserve interest rate policies on the returns of commercial banks stocks. *Journal of Financial Services Marketing, 11*(4), 349–359. https://doi.org/10.1057/palgrave.fsm.4760044.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as Data. *Journal of Economic Literature, 57*(3), 535–574. https://doi.org/10.1257/jel.20181020

Giacomo Caterini. (2020). La comunicazione della Banca Centrale dei Caraibi Orientali: un'analisi testuale (On the communication of the Eastern Caribbeans Central Bank: A textual analysis). *Moneta e credito, 73*(289), 57–82. https://doi.org/10.13133/2037-3651_73.289_4.

Goodkind, N., Horowitz, J., Goldman, D. (2022, December 30). Goodbye 2022 - and good riddance. Markets close out their worst year since 2008. *CNN Business*, 10.

Johnsrud, I. (2022, 12 June). Slik analyserer de eurokrisen. *E24*. Retrieved from https://e24.no/norsk-oekonomi/i/g7p8r1/slik-analyserer-de-eurokrisen.

Kirkeby, S. J., & Larsen, V. H. (2021). House price prediction using daily news data. *Staff Memo*.

Mouna, A., & Anis, J. (2017). Stock Market, Interest Rate and Exchange Rate Risk Effects on non Financial Stock Returns During the Financial Crisis. *Journal of the Knowledge Economy, 8*(3), 898–915. https://doi.org/10.1007/s13132-015-0301-4.

Larsen, V. H., & Thorsrud, L. A. (2022). Asset returns, news topics, and media effects. *The Scandinavian Journal of Economics*, *124*(3), 838–868. https://doi.org/10.1111/sjoe.12469

Loughran, T., & McDonald, B. (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance*, 66, 35–65.

Loughran, T. & McDonald, B. (2020) Textual Analysis in Finance, *University of Notre Dame*. http://dx.doi.org/10.2139/ssrn.3470272.

Ooms, J. (2023). *pdftools: Text Extraction, Rendering and Converting of PDF Documents*. https://docs.ropensci.org/pdftools/.

Rashid Sabri, N. (2004). Stock Return Volatility and Market Crisis In Emerging Economies. *Review of Accounting & Finance, 3*(3), 59–83. https://doi.org/10.1108/eb043408.

Taddy, M. (2013). Multinomial Inverse Regression for Text Analysis. *Journal of the American Statistical Association, 108*(503), 755–770. https://doi.org/10.1080/01621459.2012.734168.

Tahir, M., Gul, A., & Qazi, S. (2019). THE CAUSALITY BETWEEN MACROECONOMIC FACTORS ON STOCK RETURNS: A STUDY OF PAKISTAN EQUITY MARKET. *International Journal of Information, Business and Management, 11*(4), 212–222.

Ter Ellen, S., Larsen, V. H., & Thorsrud, L. A. (2022). Narrative Monetary Policy Surprises and the Media. *Journal of Money, Credit and Banking*, *54*(5), 1525–1549. https://doi.org/10.1111/jmcb.12868.

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance,* 62, 1139-1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x.

Wickham, H. (2022). *rvest: Easily Harvest (Scrape) Web Pages*. https://rvest.tidyverse.org/.

8. Appendices

8.1 A. OBX Returns in the Time Frame of our Dataset

Figure 1 OBX Returns

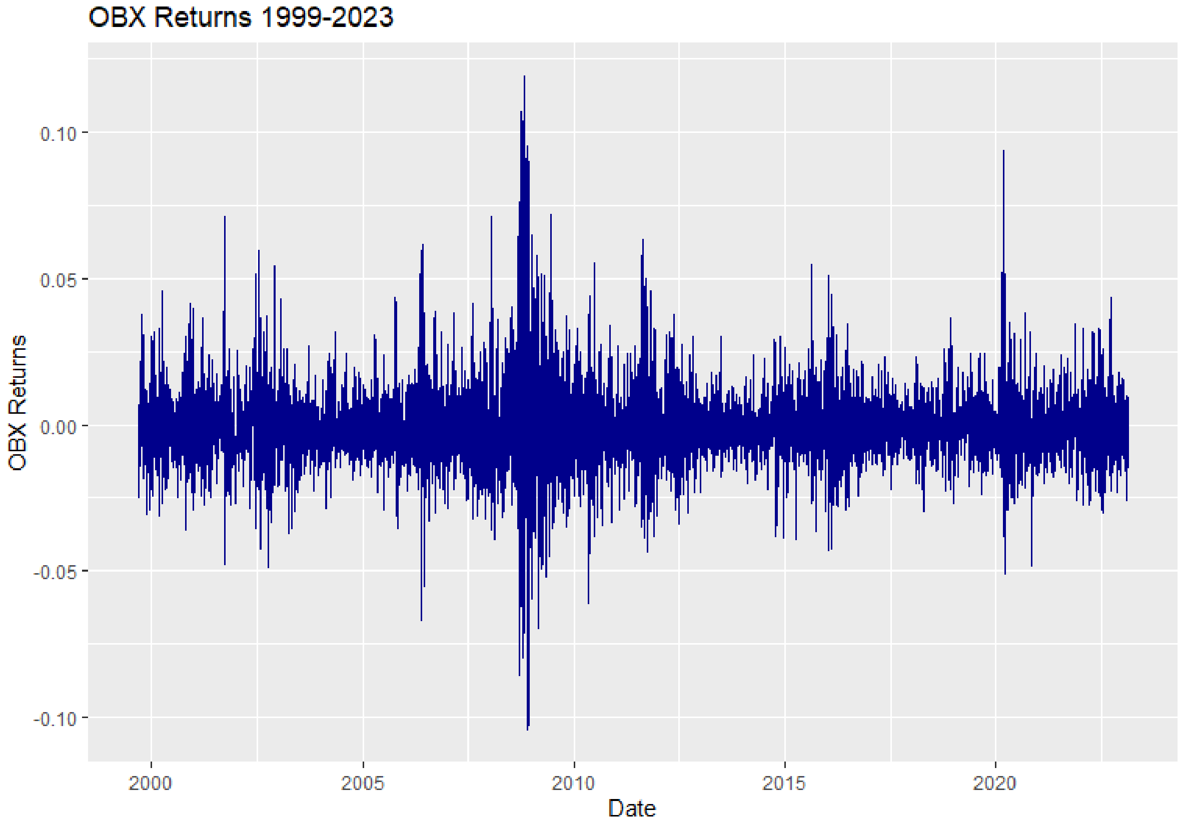

Figure description: The OBX returns are calculated as returns less returns the day before divided by returns the day before. OBX returns on the y-axis and the date on the x-axis.

8.2 B. Document Length

Figure 2 Document Length

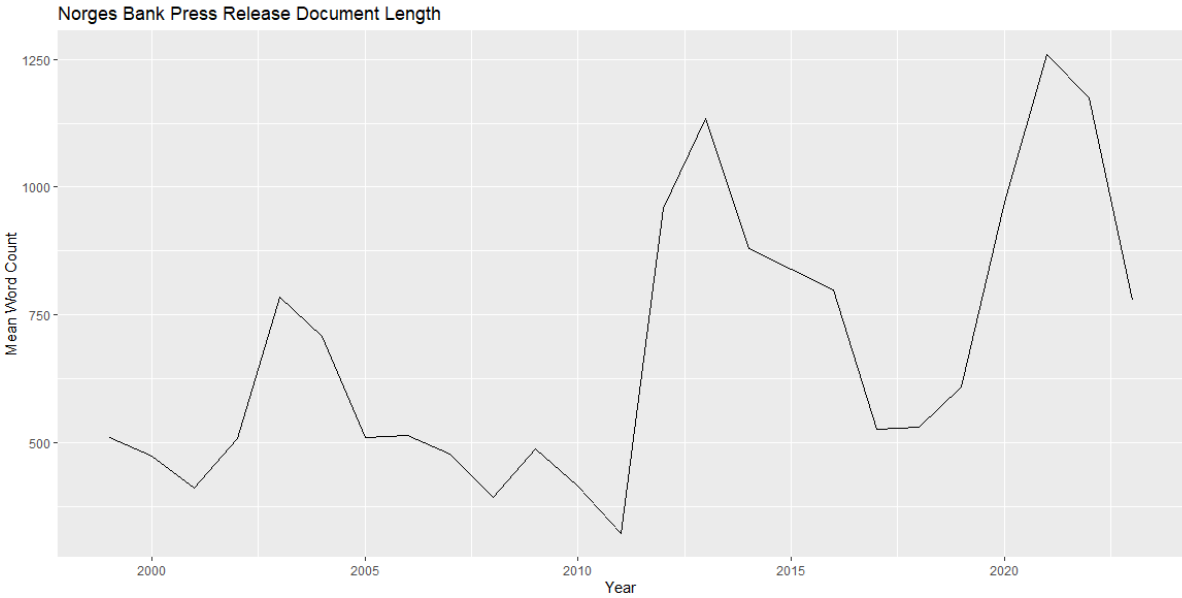

Figure description: Yearly mean word counts for our entire dataset. Word count is measured as the number of terms separated by whitespace.

8.3 C. Stock returns surrounding MPE publication

Figure 3 Event Studies Analysis

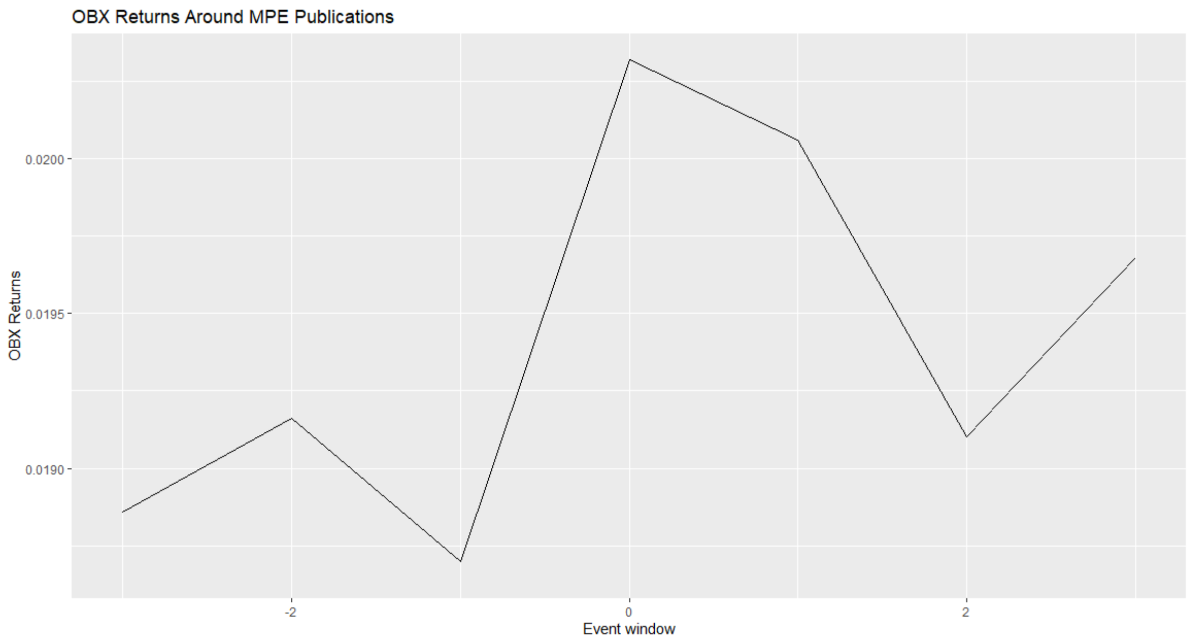

Figure description: Mean OBX returns around the event window. The event window is defined as the date of an MPE publication. We see mean returns -3 to 3 days around the event. Day 0 is the day of the MPE publication.

8.4 D. Event study for MPE publications

Table 4 Event Studies Regressions

Regression coefficients for event windows by year

	year	less_2	less_1	same_day	add_1	add_2	add_3
1	2023	-0.001	-0.017	-0.013	0.007	-0.011	-0.017
2	2022	-0.003	0.005	0.009	0.002	0.007	-0.004
3	2021	0.003	-0.002	0.002	0.0004	-0.002	0.002
4	2020	-0.003	0.007	0.008	0.007	0.015	0.004
5	2019	-0.0001	-0.001	-0.004	-0.003	0.002	-0.001
6	2018	-0.0003	0.001	0.003	0.002	0.002	0.006
7	2017	-0.005	-0.004	0.002	0.002	-0.003	-0.005
8	2016	0.006	0.002	-0.008	-0.002	-0.002	-0.008
9	2015	-0.002	-0.005	0.007	-0.006	-0.003	-0.001
10	2014	0.003	0.001	0.003	-0.006	-0.006	0.007
11	2013	0.001	0.005	-0.002	0.003	0.002	-0.0003
12	2012	-0.007	-0.003	-0.001	-0.014	-0.012	-0.0003
13	2011	-0.002	0.001	0.007	0.010	0.0001	0.018
14	2010	-0.002	0.004	0.005	0.007	-0.001	-0.009
15	2009	0.017	0.007	-0.0003	0.004	-0.0001	0.017
16	2008	0.011	0.003	0.009	0.012	-0.0002	0.007
17	2007	-0.001	-0.003	0.006	-0.001	0.008	-0.004
18	2006	-0.010	-0.015	-0.008	-0.012	-0.006	-0.004
19	2005	-0.001	-0.003	-0.006	-0.004	-0.006	-0.002
20	2004	-0.002	0.001	0.002	0.006	-0.001	0.004
21	2003	0.006	-0.003	-0.003	0.001	0.0004	-0.006
22	2002	-0.001	-0.002	-0.002	0.004	-0.002	0.0004
23	2001	0.002	0.008	0.009	0.007	0.006	0.004
24	2000	-0.0004	-0.003	0.002	0.002	0.002	-0.002
25	1999	-0.006	-0.004	-0.014	-0.004	-0.0005	-0.007

Table description: Results from running linear models for each year of MPE publications. Each value is the regression coefficient of the event, the date that the MPE was released, using the OBX returns as the dependent variable. Same_day indicates same day as the MPE publication.

Table 5 Event Studies Regressions T-Values

T-values for event windows by year

	year	less_2	less_1	same_day	add_1	add_2	add_3
1	2023						
2	2022	-0.443	0.670	1.309	0.255	0.977	-0.621
3	2021	0.706	-0.394	0.420	0.092	-0.486	0.458
4	2020	-0.373	0.749	0.899	0.741	1.665	0.447
5	2019	-0.021	-0.119	-0.803	-0.666	0.468	-0.330
6	2018	-0.064	0.205	0.692	0.325	0.467	1.179
7	2017	-1.373	-0.994	0.600	0.574	-0.715	-1.168
8	2016	0.884	0.279	-1.177	-0.265	-0.253	-1.254
9	2015	-0.423	-0.822	1.248	-1.034	-0.572	-0.148
10	2014	0.784	0.197	0.627	-1.490	-1.458	1.644
11	2013	0.144	1.058	-0.365	0.643	0.455	-0.060
12	2012	-1.157	-0.456	-0.164	-2.532	-2.070	-0.050
13	2011	-0.306	0.098	0.974	1.454	0.015	2.575
14	2010	-0.330	0.566	0.617	1.011	-0.094	-1.213
15	2009	1.782	0.746	-0.033	0.456	-0.015	1.695
16	2008	0.875	0.267	0.689	0.946	-0.013	0.558
17	2007	-0.145	-0.617	1.068	-0.165	1.527	-0.688
18	2006	-1.671	-2.462	-1.256	-2.051	-1.054	-0.648
19	2005	-0.169	-0.747	-1.250	-0.997	-1.312	-0.475
20	2004	-0.605	0.227	0.617	1.569	-0.328	1.046
21	2003	1.147	-0.682	-0.641	0.226	0.078	-1.206
22	2002	-0.106	-0.411	-0.317	0.608	-0.372	0.060
23	2001	0.347	1.105	1.303	1.053	0.892	0.545
24	2000	-0.085	-0.552	0.395	0.336	0.334	-0.499
25	1999	-0.761	-0.535	-1.936	-0.502	-0.066	-0.881

Table description: Results from running linear regressions, regressing the OBX returns on event windows. Each value is the t-statistic associated with the regression coefficients of the event variables around MPE publication dates. Same_day indicates the same day as MPE publication.

8.5 E. Empirical Model Results

Table 6 Regression Results MNIR

Empirical Model: MNIR Sentiment Score

	Dependent variable:	
	OBX Returns Normalised	
	Group a	Group b
	(1)	(2)
Sentiment Score MNIR	0.012***	-0.001
	(0.001)	(0.001)
Constant	0.020***	0.018***
	(0.001)	(0.001)
Observations	93	93
R^2	0.413	0.002
Adjusted R^2	0.406	-0.009
Residual Std. Error (df = 91)	0.014	0.014
F Statistic (df = 1; 91)	63.949***	0.207
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table description: Regression results from empirical model following the computation of sentiment scores using the MNIR dictionary and loading-adjusted z-values. Group a constitutes the training set and group b constitutes the test-set.

Table 7 Regression Results MNIR Bag-of-Words

Empirical Model: MNIR Sentiment Score using bag-of-words

	Dependent variable:	
	OBX Returns Normalised	
	Group a	Group b
	(1)	(2)
Sentiment Score MNIR bag-of-words	0.420***	-0.041
	(0.068)	(0.060)
Constant	0.027***	0.017***
	(0.002)	(0.002)
Observations	93	93
R^2	0.293	0.005
Adjusted R^2	0.285	-0.006
Residual Std. Error (df = 91)	0.016	0.014
F Statistic (df = 1; 91)	37.726***	0.470
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table description: Same notes from table 3, but for bag-of-words computation of sentiment score.

Table 8 Empirical Model Results Bag-of-Words

Empirical Models: Sentiment Score on OBX Returns

	Dependent variable:				
	OBX Returns Normalised				
	MNIR Bag Group a	MNIR Bag Group b	KL	Uncertainty	Negativity
	(1)	(2)	(3)	(4)	(5)
MNIR Sentiment Score	0.420***	-0.041			
	(0.068)	(0.060)			
KL Sentiment Score			-0.160		
			(0.186)		
Uncertainty Score				0.407	
				(0.534)	
Negativity Score					0.633**
					(0.317)
Constant	0.027***	0.017***	0.020***	0.018***	0.014***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)
Observations	93	93	186	186	186
R^2	0.293	0.005	0.004	0.003	0.021
Adjusted R^2	0.285	-0.006	-0.001	-0.002	0.016
Residual Std. Error	0.016 (df = 91)	0.014 (df = 91)	0.016 (df = 184)	0.016 (df = 184)	0.016 (df = 184)
F Statistic	37.726*** (df = 1; 91)	0.470 (df = 1; 91)	0.736 (df = 1; 184)	0.582 (df = 1; 184)	3.985** (df = 1; 184)
Note:					*p<0.1; **p<0.05; ***p<0.01

Table description: Empirical model results all sentiment scores using bag-of-words approach. KL constitutes the Kirkeby & Larsen dictionary's sentiment score regression results. The uncertainty is the uncertainty-term measure. The negativity is the self-defined list of negative words for the purpose of this study.

Table 9 Empirical Model Results Delta

Empirical Models: Delta Sentiment Score on OBX Returns

	Dependent variable:				
	OBX Returns Normalised				
	MNIR Bag Group a	MNIR Bag Group b	KL	Uncertainty	Negativity
	(1)	(2)	(3)	(4)	(5)
MNIR Sentiment Score Delta	0.393***	-0.029			
	(0.085)	(0.080)			
KL Sentiment Score Delta			-0.007		
			(0.140)		
Uncertainty Score Delta				-0.014	
				(0.410)	
Negativity Score Delta					0.246
					(0.247)
Constant	0.021***	0.018***	0.019***	0.019***	0.019***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	93	92	185	185	185
R^2	0.189	0.001	0.00001	0.00001	0.005
Adjusted R^2	0.180	-0.010	-0.005	-0.005	-0.0001
Residual Std. Error	0.017 (df = 91)	0.014 (df = 90)	0.016 (df = 183)	0.016 (df = 183)	0.016 (df = 183)
F Statistic	21.204*** (df = 1; 91)	0.134 (df = 1; 90)	0.003 (df = 1; 183)	0.001 (df = 1; 183)	0.988 (df = 1; 183)
Note:					*p<0.1; **p<0.05; ***p<0.01

Table description: Same as table 5, but using the delta-measure of the sentiment score instead as the independent variables in these regressions.

8.6 F. Sentiment scores and OBX price comparison

Figure 4 Sentiment Score

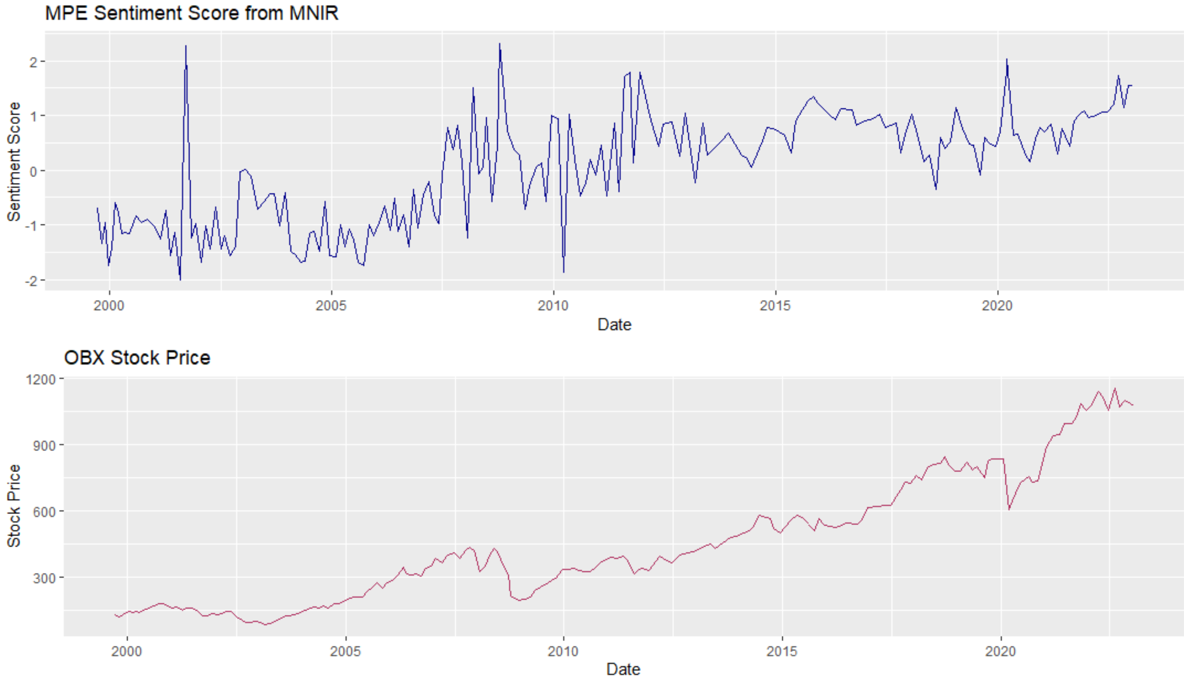

Figure description: In the top plot we see the sentiment score computed from MNIR over time when account for the loadings of individual terms. Below we see the OBX stock price for comparison.

Figure 5 Sentiment Score Bag-of-Words

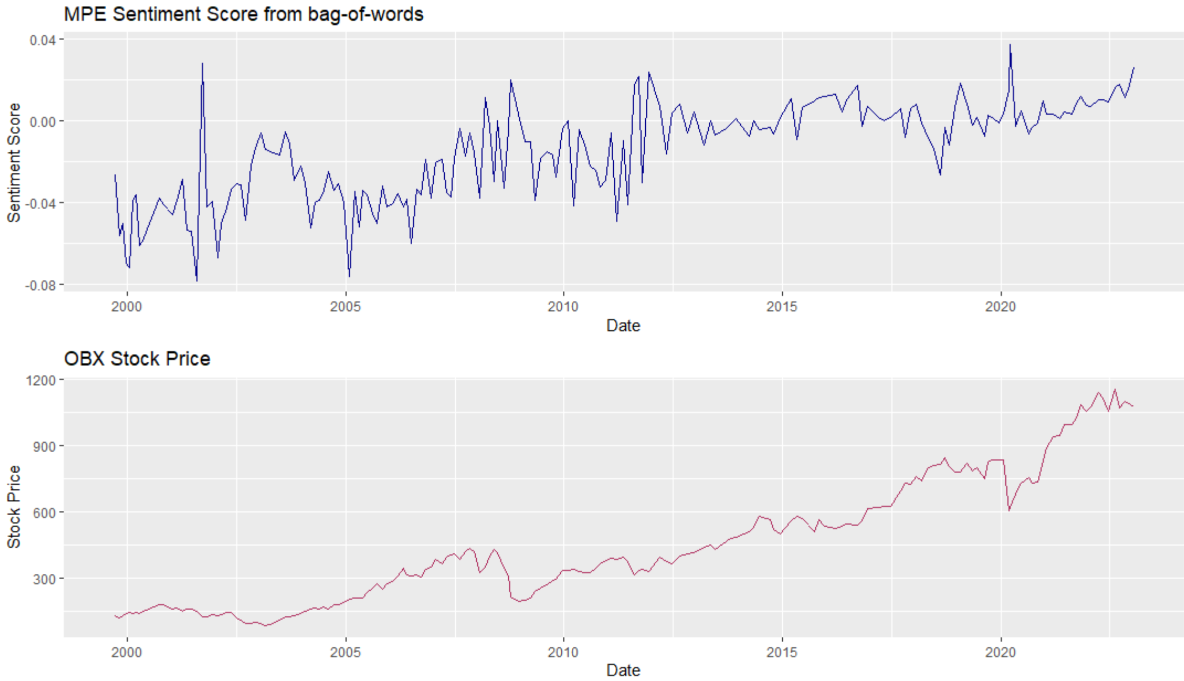

Figure description: In the top plot we see the sentiment score computed from MNIR using bag of words over time. Below we see the OBX stock price for comparison.

Figure 6 Sentiment Score KL-Dictionary

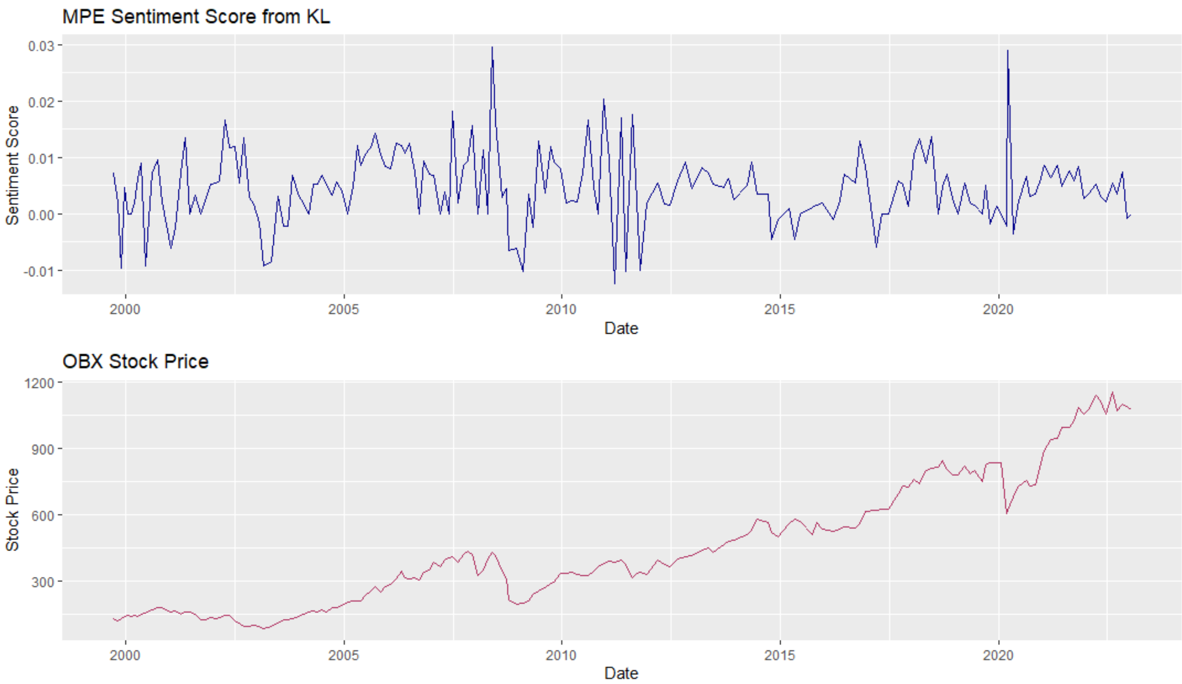

Figure description: In the top plot we see the sentiment score computed from KL-dictionary using bag of words over time. Below we see the OBX stock price for comparison.

8.7 G. Dictionary resulting from MNIR

Positive: september, desember, nærmeste, gjennom, komiteen, oppgang, prisvekst, høye, svekket, anslås, indikerer, mye, november, oktober, husholdningene, betydelig, prosentenheter, slutten, bedriftene, lenger, nettverk, handelspartnerne, finansmarkedene, pengepolitikk, holdt, virkningene, venter, usikkerheten, lavt, europeiske, euroområdet, ledigheten, pengemarkedet, større, fallet, rentesettingen, redusere, løfte, ført, olje, tiltak, arbeidskraft, rentemøtet, vekstutsiktene, gjør, prognoseperioden, uroen, komiteens, tiltatt, fleste, finansiell, husholdninger, markedet, stramt, kort, behov, storbritannia, reduserer, økonomier, avtok, kostnadsvekst, knyttet, amerikanske, verdi, anslår, svingninger, styringsrenter, komité, sentralbankene, sverige.

Negative: inflasjonsrapport, hensynet, inflasjonsrapporten, våre, utvikling, god, inflasjonsmålet, dagens, taler, rundt, tråd, tar, tilbake, synes, års, sammen, fortsetter, få, gir, forhold, figurer, grunnlag, fall, penge, forholdene, vurderingen, dermed, faren, figurene, bankenes, nedgangen, derfor, strategien, nivået, avgifter, krone, steg, vedtak, isolert, årsskiftet, retning, næringslivet, pekt, deler, vedvarende, april, kostnadsveksten, juli, enkelte, bygger, tilbakeslag, kroner, dlånsrenten, overfor, trekker, nivåer, lang, importveide, valutamarkedene, lengre, internasjonal, bringes, fattet, opprettholdes, bak, industrien, investeringene, hensyn, sist, inflasjonsrapporter, lønnsoppgjør, prosentenhet, konkurranse, kostnadsutsiktene, del, høst, presentert, hovedstyremøte, rentenivået, vokser, spesielle, dobbeltklikke, filen, henter, lagrer, powerpoint, virkemidlene, årets, tre, aksjekursene, avveiing, prosentpoeng, strømpriser, innenlandske, trekke, reduksjon, nok, trekkrettigheter, betinget, inneværende, fotnoter, økonomisk, beslutninger, valutaer, danner.

8.8 H. List of negative terms self-constructed

Inflasjon, inflasjonen, prisveksten, krig, krigen, heve, prisinflasjonen, konsumprisveksten, usikkerhet, usikkerheten, krise, krisen, dårlig, bekymrende, bekymringsverdig, bekymring, bekymringer, uvanlig, finansmarkedene, pengemarkedene, nedgang, resesjon, resesjonsfrykt.

8.9 I. Word clouds of most frequent terms

Figure 7 Most Frequent Positive Terms

Figure description: Most frequent positive terms occurring in the corpus.

Figure 8 Most Frequent Negative Terms

Figure description: Most frequent negative terms occurring in the corpus.

8.10 J. Sentiment and document length correlation plot

Figure 9 Document Length and Sentiment Score Scatterplot

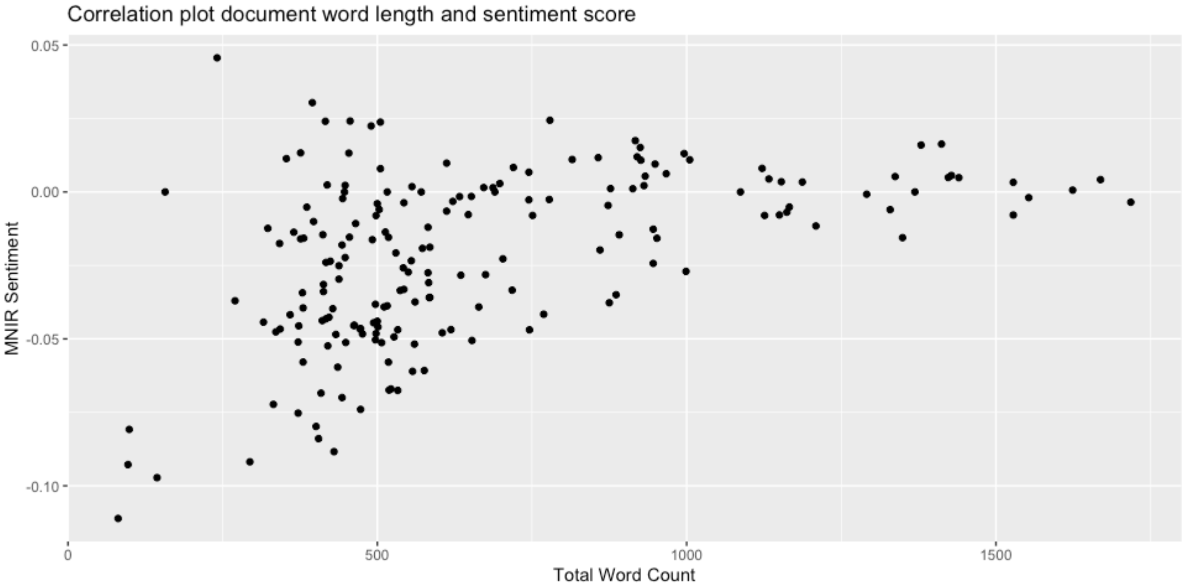

Figure description: On the x-axis we plot the total word length of the documents, and on the y-axis we plot the sentiment of the documents. Each point is an observation of a document's sentiment score computing by the MNIR dictionary and the document's word count.