Speeches in Motion:

Unravelling the Bank of England's Impact on Markets

Final Report

Team 7

The Magnificent

Craig McCracken

Emily Coughlan

Emma Roberts

Ian Munro

Joel Baksh

Will Burton

1 Contents

2 Executive Summary

2.1 Introduction to Bank of England

The Bank of England (BoE) aims to maintain price stability and support government economic policies through monetary policy tools and public speeches. These speeches aim to foster confidence and stability in financial markets.

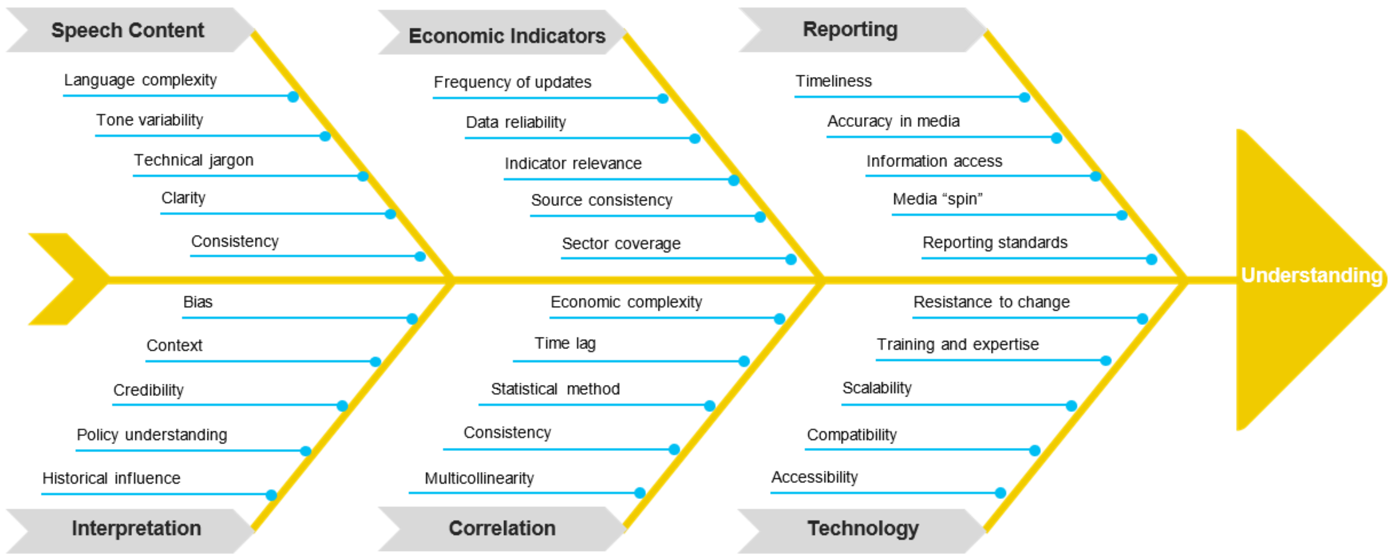
2.2 Problem Statement

This project assesses how BoE speeches shape the economy and financial markets using data-driven methods. Key questions to ask the data include:

1. Investigating changes in the sentiment of central bank speeches over time.
    a. Which NLP tool is most appropriate to use to classify BoE speeches?
2. Does the sentiment of BoE speeches correlate with key events, such as:
    a. Bank rate decisions,
    b. the Monetary Policy Report (MPR) publication,
    c. and Financial Stability Report (FSR) releases?
3. Does the sentiment of BoE speeches correlate to key economic indicators such as:
    a. Gross Domestic Product (GDP) growth,
    b. Inflation (CPI),
    c. Labour market statistics?
4. Can BoE speeches be used to forecast any market behaviour?

Initial thoughts and avenues to explore to support the business case were identified and captured in a fishbone diagram, Figure 1.

Figure 1 - Fishbone diagram illustrating key factors for comprehending the impact of Central Bank speeches.



3 Data Sources

3.1 Speech Data Provided

A summary of the data provided by the BoE in 'all_speeches.csv' can be seen in Table 1.

Table 1 – Summary of data provided by the BoE.

Column	Data	Comments
Reference	Unique reference number for the speech	
Country	Country where the speech was given	
Date	Date the speech was given	Spans 1990 - 2022
Title	Title of the speech	
Author	Person who wrote the speech	May not be the speech giver.
Is_gov	Boolean: 1 = Speech delivered by the BoE governor	
Text	Transcript of the speech	

Key decision: With only 6 weeks to complete the project, and a project brief focusing on the BoE and UK economy, the decision was made to focus solely on BoE speeches. The frequency of BoE speeches plotted over time indicated setting the timeframe from 1998 to 2022, as this is when BoE speeches occurred in the provided dataset.

3.2 Economic Data Scraping

In Table 2, the reasons for selecting economic indicators to analyse in conjunction with speech sentiment can be found. These were selected following the project briefing and early research into the UK economy.

The data was pulled in daily, monthly and quarterly intervals, depending on what was available without needing to pay.

Table 2 – Economic indicators

	Indicator	Reason for Choice	Source
Macro-Economic Indicator	Base Interest Rate	Specified in Assignment	Bank of England
	CPI	Specified in Assignment	Office for National Statistics
	GDP	Specified in Assignment	Office for National Statistics
	Unemployment	Specified in Assignment	Office for National Statistics
	Wage Growth	Specified in Assignment	Office for National Statistics
	House prices	Hypothesis: BoE base rate changes will impact	Office for National Statistics
Market	FTSE 100/250	Indicator of UK economy	Yahoo finance
	Bonds	BoE guidance in kick-off / Q&A	Investing.com
	Forex	Comparator of UK economy vs. USA / Europe	Yahoo finance
	Gold	Possible indicator of UK economy	Yahoo finance
Energy	Oil Price (Brent)	Indication of fuel costs which impacts inflation	Yahoo finance
	Natural Gas	Indication of fuel costs which impacts inflation	Unable to source (paywall)

4 Data cleaning, exploration & wrangling

4.1 Speech data

Details of cleaning the speech data can be found in Appendix A: Data Cleaning.

Multiple NLP models were tested to determine the most suitable model for the financial text use case. The chosen models were TextBlob, VADER, Loughran McDonald (lm) and, FinBERT.

The data was then prepared for Natural Language Processing (NLP), considering the text requirements of each of the chosen NLP models. A summary of all the NLP models and data preparation can be seen in Table 3.

Key decision: FinBERT was selected for its in-depth financial sentiment analysis where context and nuanced understanding are crucial.

Full details of NLP model Selection can be found in Appendix B: NLP Modelling.

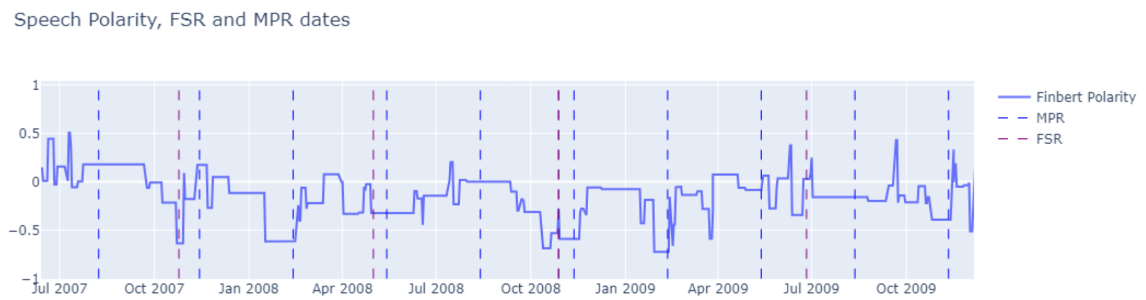
4.2 Economic Data

Economic data pulled from the sources outlined in Table 2 was cleaned and combined in one place. Details of this process can be found in Appendix A: Data Cleaning.

Key Decision: Given the varied granularity of economic data and the absence of speeches on certain days, the choice was made to standardize all economic data to daily availability. For indicators with less frequent updates, forward filling was employed, assuming constancy. Interpolation was considered, but discounted, because it could potentially introduce inaccuracies or assumptions that might not accurately reflect the underlying economic conditions.

Key Decision: It was noted in early exploration that the speeches close to when FSR and MPR reports were issued often moved in sentiment (see Figure 2). It was therefore decided to explore the FSR and MPR reports in more detail and extract the contents from them to compare their sentiment to that of the speeches. Text extraction description can be found in Appendix C: FSR and MPR data scraping.

Figure 2 – Excerpt from exploratory plotly chart showing how speech sentiment often changes when the MPR and FSR are published.



4.3 Combining speech & economic data

Like the process of amalgamating economic data in Section 4.2, the sentiment scores derived from NLP models were integrated with the economic data. Forward filling was applied, ensuring that on any given day, the sentiment reflected that of the last speech delivered.

Table 3 – Summary of the different NLP models considered and evaluated.

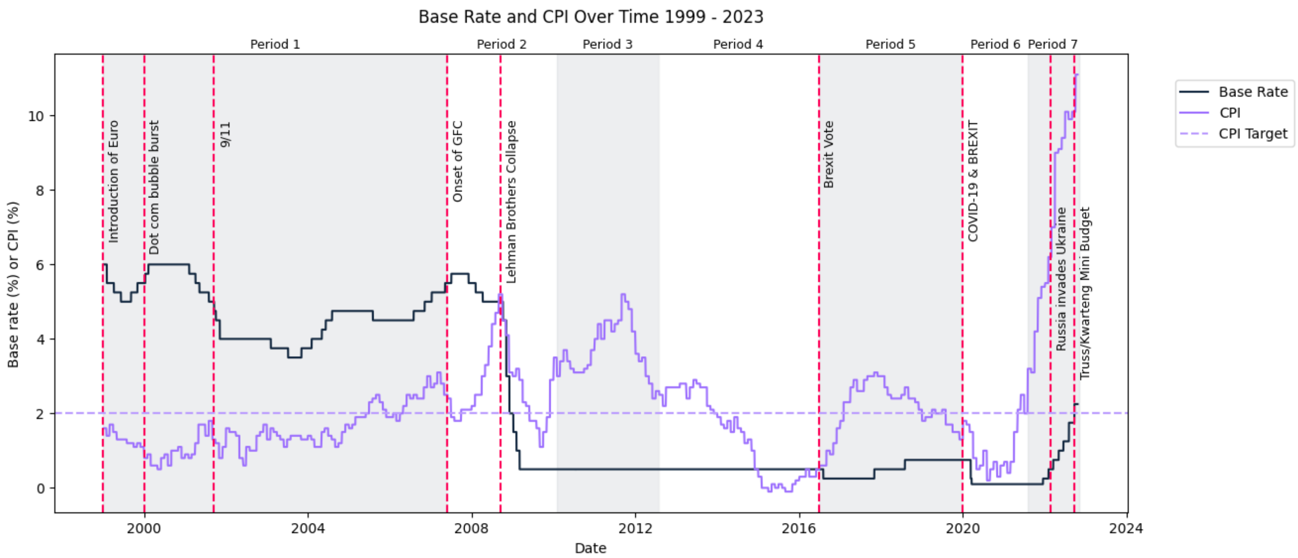
Model	Base Dictionary	Approach	Domain Specificity	Last Updated	Key Strengths	Key considerations	Text Data Preparation
TextBlob	**Base:** General English	General lexicon (Bag of Words)	General	N/A (updates vary by library version)	- Simple, transparent, and user-friendly - Versatile for general NLP tasks	- May not capture domain-specific nuances. - Less accurate for complex sentiments.	URLs, Hashtags, and mentions were identified and removed. Stopwords were removed.
VADER	**Base:** General English with social media emphasis	General lexicon and rule-based	General; focus on social media		- Understands nuances through rule-based analysis.	- Less effective outside social media contexts. - Static rules can miss specific speech language trends.	URLs, Hashtags, and mentions were identified and removed.
Loughran McDonald (lm)	**Base:** PySentiment2 **Financial:** EDGAR 10-K archive; CapIQ earnings calls	General lexicon (Bag of Words) with financial focus	Financial documents	Feb 2024	- Tailored for financial text analysis. - Transparency in sentiment scoring	- Limited in capturing complex context. - Static lexicon may miss specific speech language trends.	URLs, Hashtags, and mentions were identified and removed.
FinBERT	**Base:** Google's BERT **Financial:** Corporate Reports 10-K & 10-Q: 2.5B tokens; Earnings Call Transcripts: 1.3B tokens; Analyst Reports: 1.1B tokens	Large Language Model (LLM), deep learning with financial focus	Financial documents	2022	- Deep contextual understanding - High accuracy, especially for negative sentiments - Deep understanding of financial context.	- Requires substantial computational resources - Complexity in model understanding, less transparent.	URLs, Hashtags, and mentions were identified and removed.

5 Analysis, trends and patterns

5.1 Preliminary

The preliminary analysis aimed to identify notable periods of time to focus on. A rough timeline was developed, dividing time based on comparable base rates, inflation rates, and significant geopolitical events. Base rate and inflation were selected as key factors, as adjustments to the base rate are instrumental in the Bank of England's pursuit of its 2% inflation target. This can be seen in Figure 3.

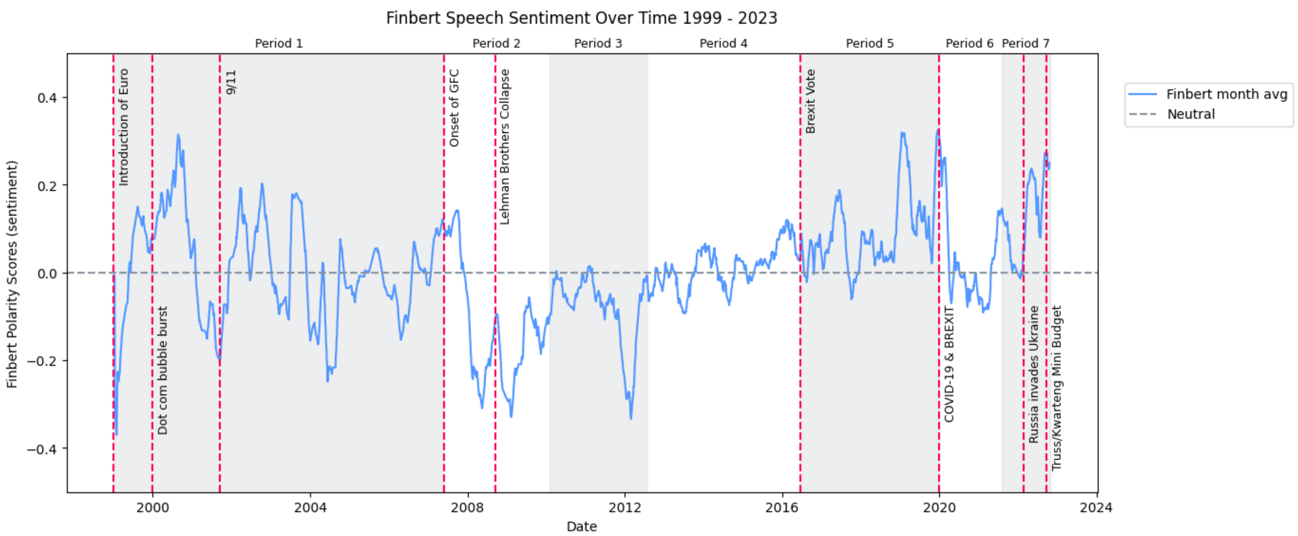
Figure 3 – Base rate and CPI between 1999 and 2023 to identify financial periods of interest.



5.2 How does sentiment of central bank speeches change over time?

As can be seen in Figure 4, over time, speech sentiment generally maintains a stable position around neutral. Significant deviations from neutrality typically coincide with key global geopolitical events. This observation is unsurprising, as the speeches often reflect prevailing global circumstances.

Figure 4 – BoE Speech Sentiment over time.

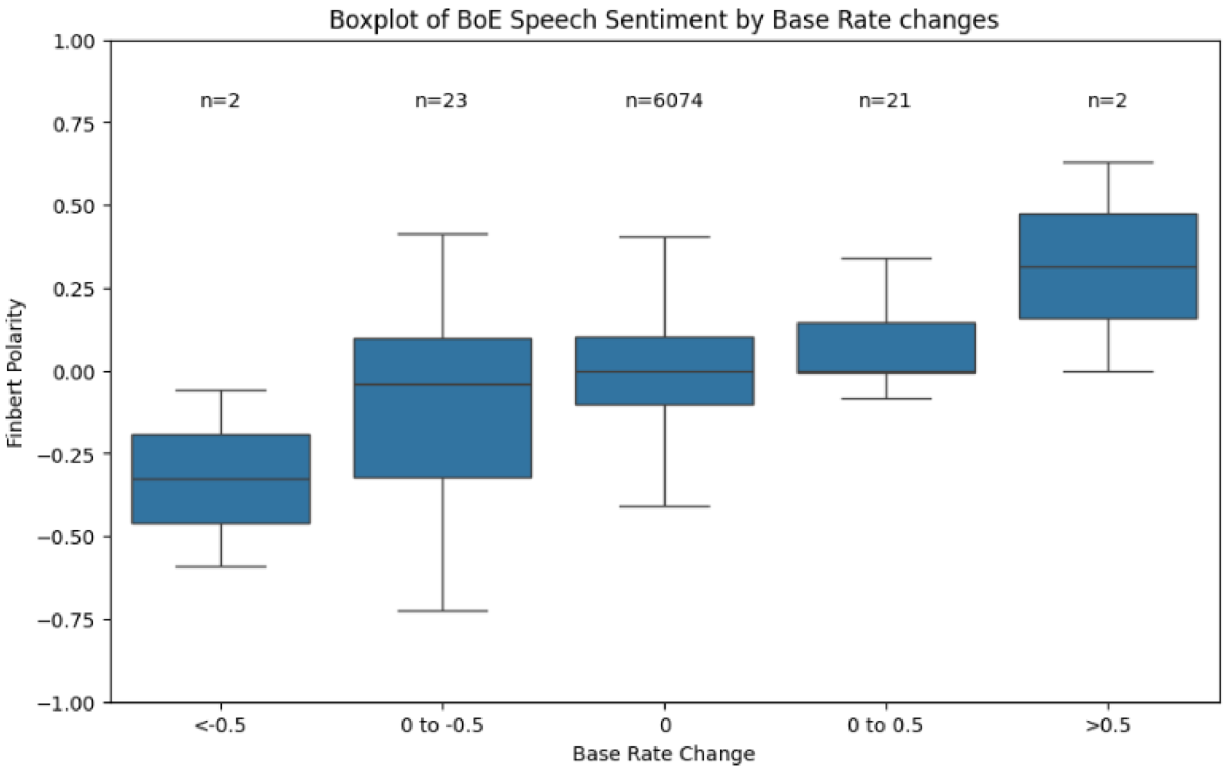


5.3 How does sentiment of BoE speeches correlate with key events?

5.3.1 Bank rate decisions

Figure 5 shows a subtle correlation between speech sentiment and BoE base rate adjustments. Furthermore, the chart depicts the spectrum of sentiment conveyed in speeches relative to both the prevailing time frame and specific base rate alterations. For instance, even amidst zero base rate changes, a range of sentiments is expressed, indicating that base rate adjustments are not the sole determinant of speech sentiment.

Figure 5 – Boxplot of BoE Speech Sentiment by Base Rate Change



Key Observation: Excluding periods of base rate change, a simple Ordinary Least Squares (OLS) linear regression was fitted to the data. This revealed a statistically significant but limited correlation between speech sentiment and BoE base rate changes (R^2 = 0.08). This suggests that speech sentiment can predict base rate changes to a degree, though the model's predictive power is low.

Key Decision: Due to the wide range of sentiment observed, the limited predictive power, and the Bank's direct control over both the base rate and its speech content, this avenue was not pursued further.

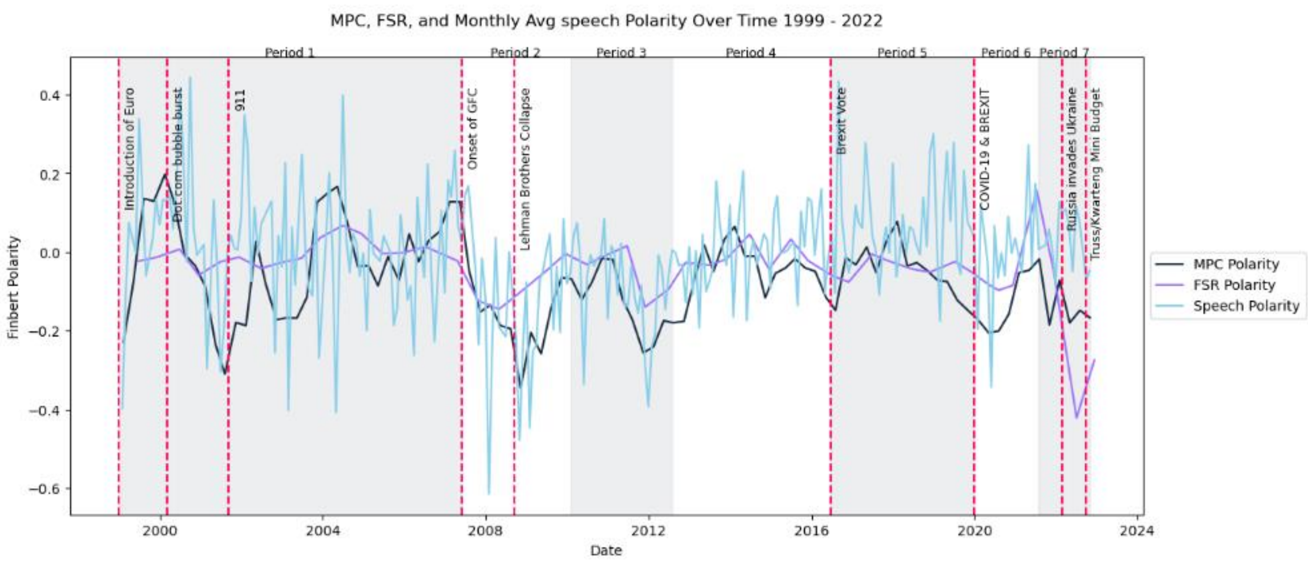
See Appendix D: Linear Regression between base rate and speech sentiment for full details.

5.3.2 MPR and FSR publication

Key Observation: Referring to Figure 6, MPR and FSR report sentiment track each other most of the time. However, the high variability in bank speeches, can be clearly seen.

Key Decision: Investigate the high variance of BoE speech sentiment during a time of relative economic stability (June 2003 to June 2004) to understand further.

Figure 6 – Sentiment of MPR, FSR and all bank speeches (monthly average due to high number of speeches)



5.3.3 Variability of BoE speech sentiment

Key Observation: Despite economic stability, polarity scores varied significantly between June 2003 to June 2004 as seen in Figure 7. Notably, Sir Edward George's banquet speech in June 2003 exhibited optimism, while Kate Barker's November 2003 lecture highlighted economic uncertainty. Barker's subsequent speech in April 2004, at a CBI dinner, emphasized economic challenges, contrasting with Mervyn King's anecdotal June 2004 address in Scotland. The diverse settings and autonomy of speakers likely influenced speech content and sentiment, underscoring the complexity of interpreting speech data during stable economic phases.

Appendix E: June 2003 to June 2004 Speech Sentiment provides an in-depth overview of the June 2003 to June 2004 speech sentiment period.

5.4 Does the sentiment of BoE speeches correlate to key economic indicators?

The economic indicators from Table 2 were explored against BoE speech sentiment. Spanning the entire period, a heatmap was generated to visualize potential correlations with speech sentiment, Figure 8.

Key Observation: Speech sentiment shows a minor positive correlation with FTSE100 and FTSE250 but exhibits a negligible to slightly negative correlation with bonds, base rate, and CPI. The bonds have strong positive correlations with each other, but strongly negative with the FTSEs. This is likely due to investors shifting between asset classes based on risk appetite or economic outlook, causing bonds and stocks to move in opposite directions.

Key Decisions: Following BoE feedback from the kick-off meeting, initial Q&A, and interim presentation:

- Further exploration of FTSE indices, house prices, oil prices, and natural gas prices was discontinued, shifting attention to alternative markets.
- Investigation of how the relationship with speech sentiment varies during a "Lower Bound Regime" (base rate below 1%) and "Hiking Cycle" (continuous base rate increase in 2022) was initiated.

Figure 7 – Variance of speeches during a period of relative financial stability (June 2003 – June 2004)

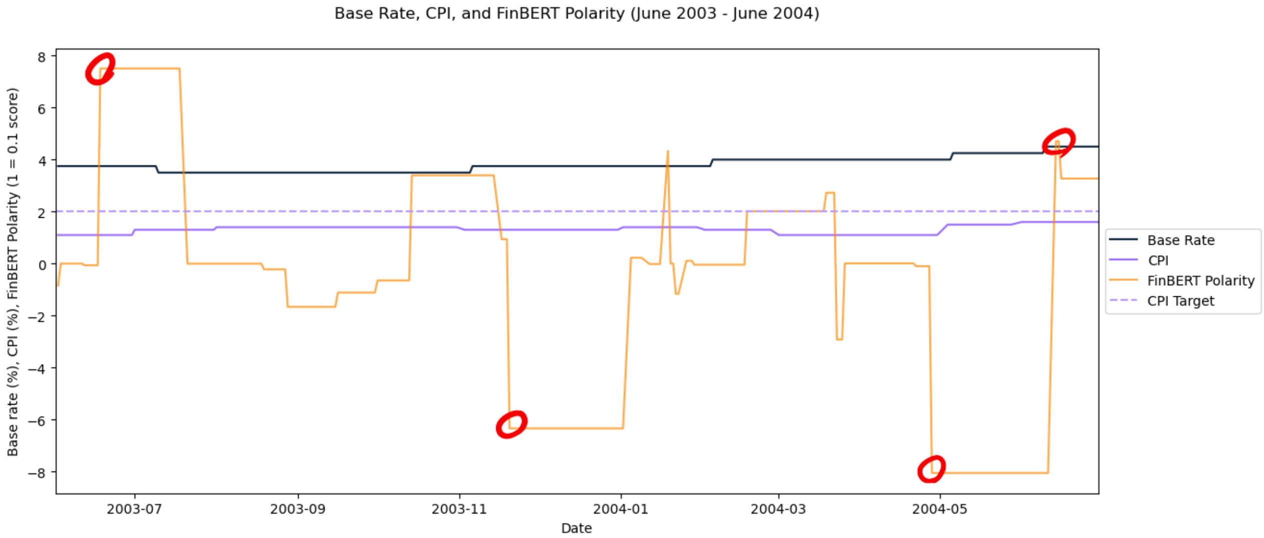
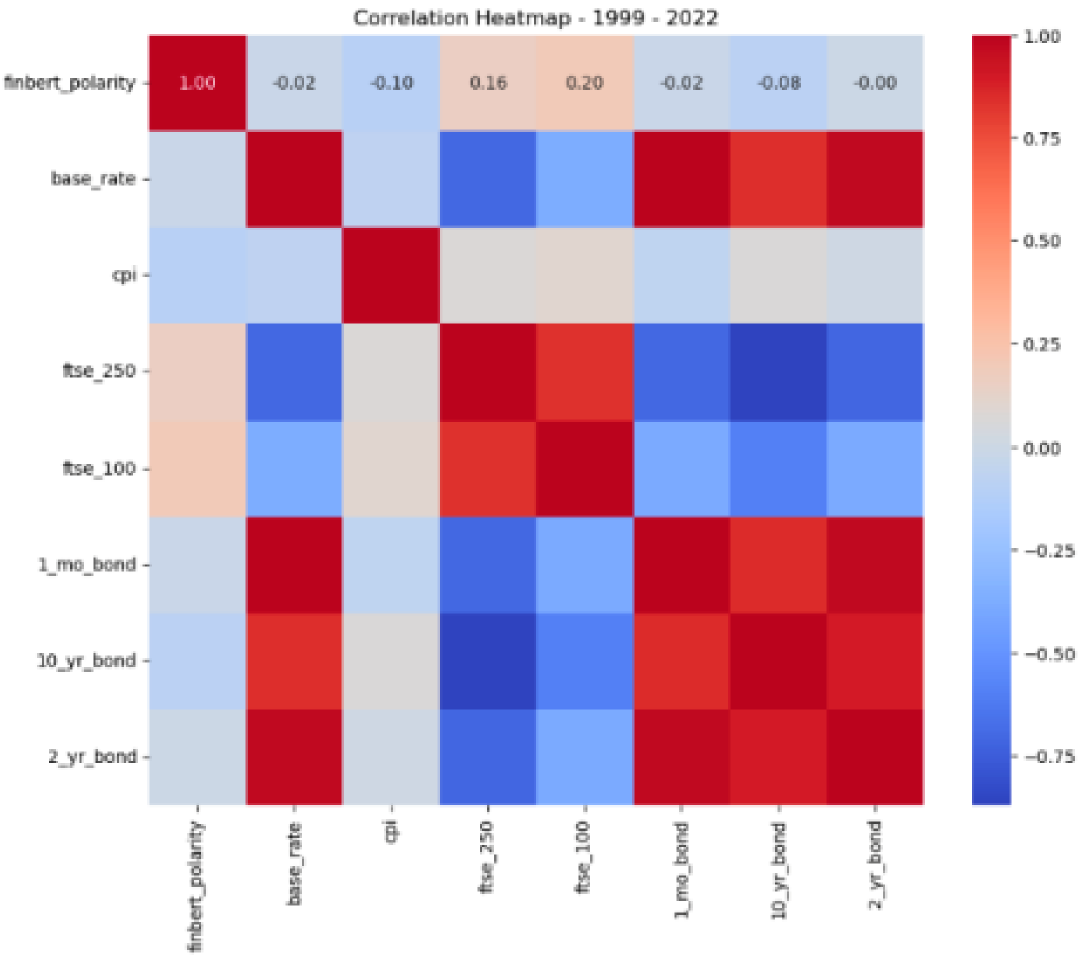


Figure 8 – Speech correlation with key economic indicators 1999 – 2022 covering 1165 speeches.



5.4.1 Lower Bound Regime and Hiking Cycle

The **key observations** from Figure 8, Figure 9, and Figure 10 can be found in Table 4.

Key recommendation: Segmenting time periods into regimes and analysing them with separate heatmaps revealed noteworthy similarities and differences in correlation. It is advisable for the BoE to consider this analytical approach further, if not already undertaken, to uncover trends that can enhance the Bank's comprehension and forecasting of market behaviour.

Full details on the work undertaken into regimes can be found in Appendix F: Monetary Policy Regimes.

Table 4 - Comparing correlations in the whole period, a lower bound regime and hiking cycle regime.

	Whole Period (control / average)	Lower Bound Regime	Hiking Regime
Speech polarity vs FTSE	Slight positive correlation	Slight positive correlation	Slight positive correlation
Speech polarity vs. bonds	Negligible correlation	Negligible to slight negative correlation	Slight negative correlation
Base rate vs. bonds	Strong positive correlation	Positive correlation	Very strong positive correlation
CPI vs bonds	Slight correlation	Slight to moderate positive correlation	Moderate positive correlation
CPI vs base rate	Slight negative correlation	Slight positive correlation	Strong positive (base rate increased to curb inflation)

Figure 9 – Speech correlation with key economic indicators during a lower bound regime.

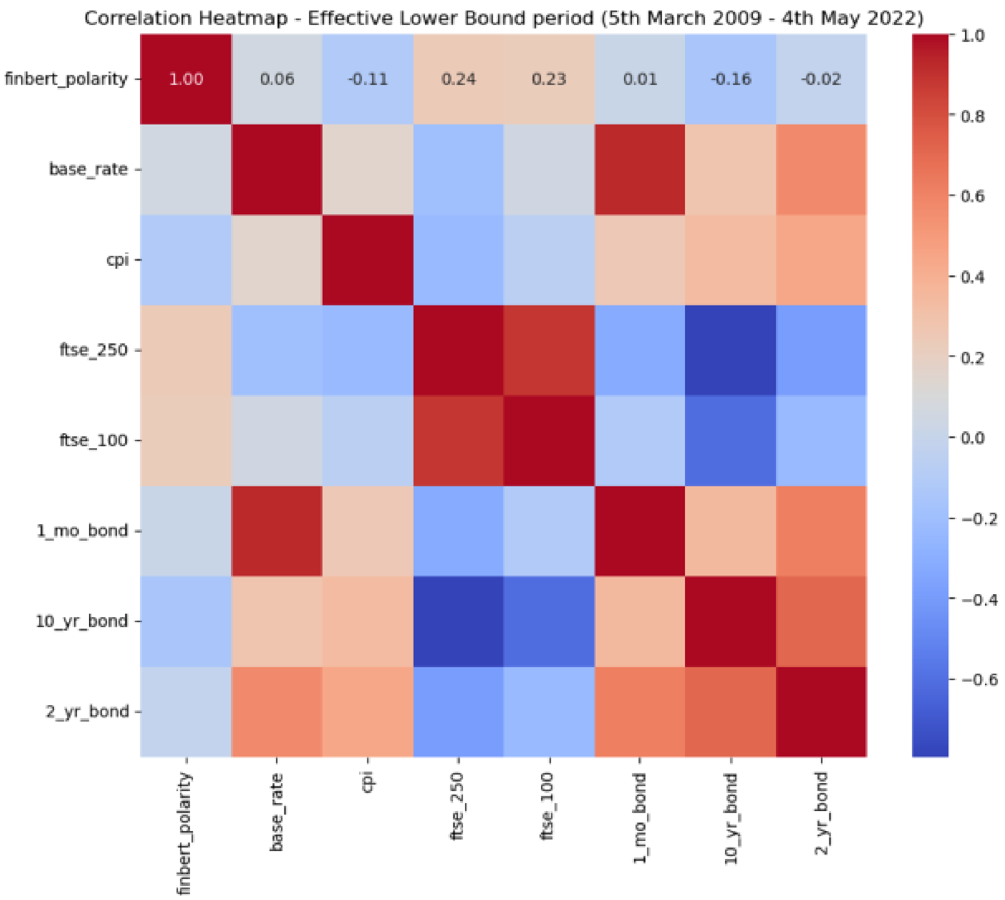


Figure 10 – Speech correlation with key economic indicators during a hiking cycle.

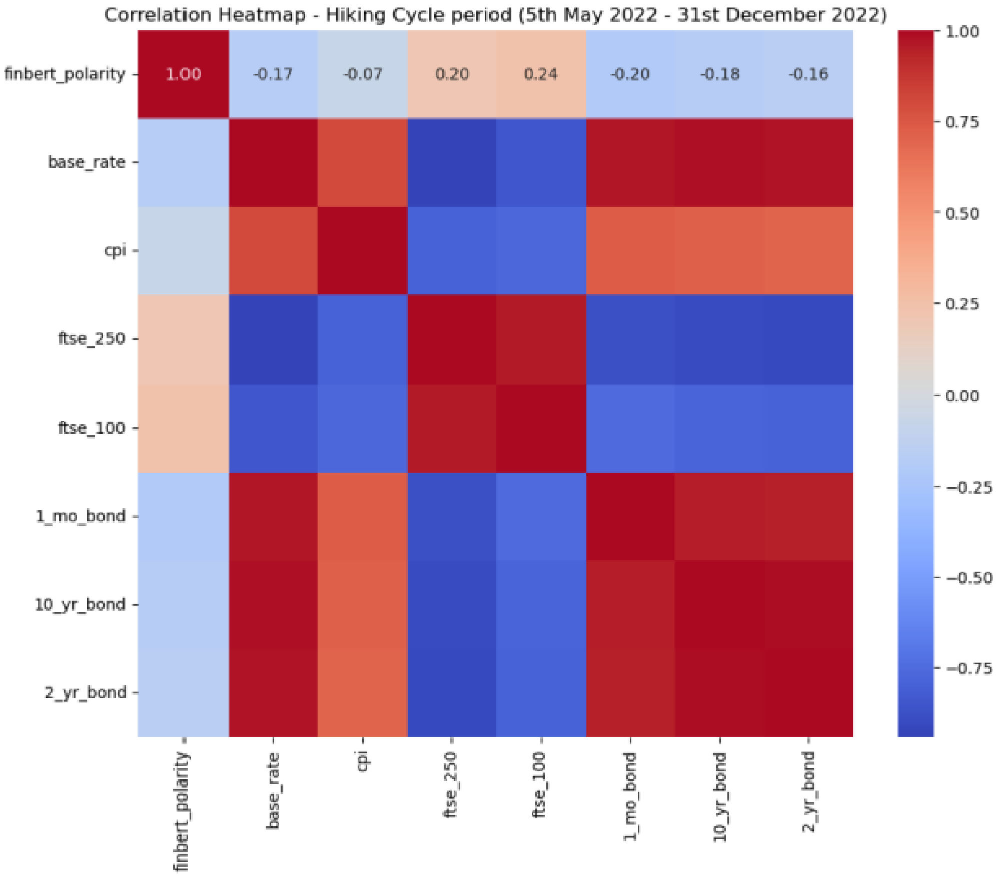
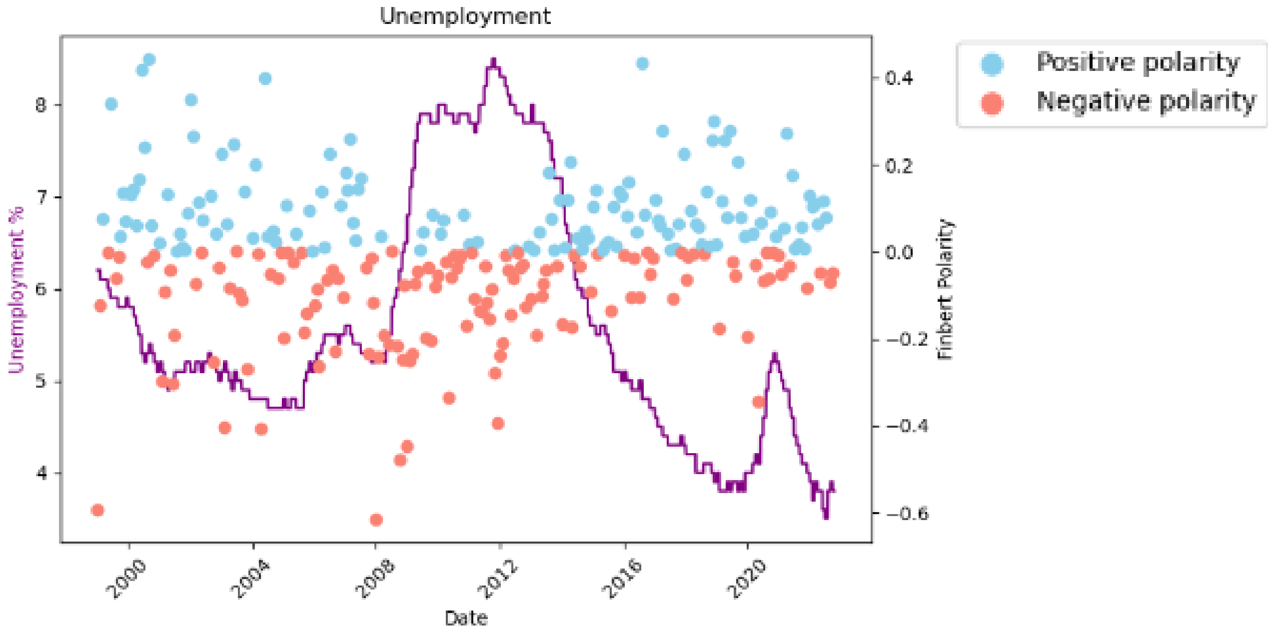


Figure 11 – Relationship between speech sentiment and unemployment.



5.4.2 Unemployment

Key Observation: Figure 11 highlights a correlation between unemployment and speech sentiment during the Global Financial Crisis (GFC), with speeches tending to be more negative as unemployment rose.

Key decision: This observation was not investigated further as per BoE guidance, which prioritized analysis of bonds. Additionally, outside of the GFC, this correlation did not persist, as seen during the sharp increase in unemployment amid the COVID-19 pandemic.

Full details on the work undertaken looking at economic indicators with speech sentiment can be found in Appendix G: Economic indicators and speech sentiment.

5.5 Possible predictive power in BoE speeches

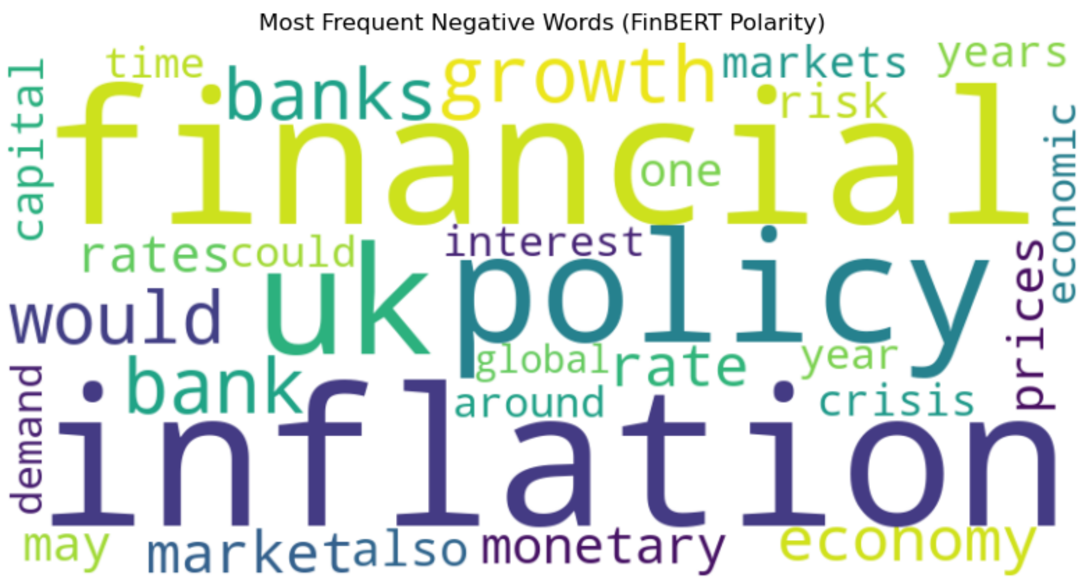
Extensive analysis into word frequencies was carried out for speeches, FSR and MPR reports, full results can be found in Appendix H: Word clouds and word frequency analysis. An example of the wordclouds produced can be seen in Figure 12.

Key Observations:

- Unique negative words such as "prices," "global," "crisis," "demand," and "around" reflect various economic challenges and uncertainties.
- Positive sentiment triggers like "risks," "system," and "data" highlight proactive management of financial systems.

Key Recommendation: Tailor word choice speeches and reports to ensure consistent sentiment across communications.

Figure 12 – Most frequent negative words in BoE speeches.



6 Analysis Tools and Visualizations

6.1 Analysis Tools

Throughout this report the rationale behind the choice of analysis tools and techniques has been summarised. A broader, high-level summary of the analysis tools used can be seen in Table 5.

For more detailed information on the NLP work see Appendix B: NLP Modelling. For more information on scraping the FSR and MPR reports for text data see Appendix C: FSR and MPR data scraping.

Table 5 – Analysis Tool Selection Overview

Tool	Selected?	Selection Basis
Excel	Yes	Excel was chosen as it accommodated the provided data sources in both excel and csv format, allowing for easy assessment of data formatting. For small datasets, it provided a quick and straightforward method for preliminary data cleaning.
Python	Yes	Python was selected for its versatility and robust data analysis capabilities. It offers extensive libraries for data manipulation, statistical analysis, and machine learning, making it suitable for in-depth analysis tasks.
R	No	R was not chosen as heavy-duty statistical methods were not required for this analysis. While R is powerful for statistical analysis, Python was deemed sufficient for the tasks at hand.
Tableau	No	Tableau was not selected as the project did not require the creation of dashboards, which is where Tableau excels. Additionally, given the exploratory nature of the project and reliance on publicly available data sources, developing a dashboard for the BoE may not align with their internal systems.

6.2 Visualisations

Table 6 summarizes the selection basis or design decisions for various visualization tools and techniques used in the project.

Table 6 – Table outlining the approach to visualisations.

Item / tool	Selection Basis / design decisions
Plotly	Utilized during data exploration for its ability to share interactive charts over the entire timeframe. This facilitated zooming, filtering, and exploration of specific project segments.
Bank of England Colour scheme	Employed BoE branding HEX codes sourced from the brand website to ensure consistency and colourblind-friendly visualizations.
Exploratory Plots	Numerous exploratory plots were internally shared and extensively discussed within the team to guide project decisions. These plots were intentionally kept minimally annotated to prioritize clarity and focus on key insights.
Time series plots	Implemented to meet the business case's objective of exploring trends over time, maintaining a consistent visual style throughout the analysis to guide the audience effectively.
Heat maps	Chosen for their effectiveness in quickly displaying potential correlations and predictive power within the dataset, aligning with the BoE's objective of exploring the predictive capability of speech sentiment.
Box Plot	Selected as a clean and clear visualization method to demonstrate the slight correlation between base rate and speech sentiment. It effectively conveyed the range of speech sentiment variability, especially compared to noisy scatterplots.
Word Cloud	Utilised to provide an easy to understand visual of word frequencies and communicate how word choice in speeches impacts speech sentiment scores.

7 Findings, Trends and Insights

A summary of the findings of this project, consistent with the business case outlined in 2.2 Problem Statement, can be found below.

1. A robust NLP model, FinBERT, has been selected for quantifying speech and financial report sentiment.
2. BoE speeches generally maintain a level of neutrality in speeches. Movements away from neutral coincide with key global geopolitical events.
3. Slight correlation between speech sentiment and:
    a. Base rate changes,
    b. unemployment,
    c. FTSE, and
    d. bonds during hiking cycle.
4. Speeches have limited predictive power over market movements on a daily timeframe:
    a. Speech sentiment exhibits significant variability during periods of relative economic stability and,
    b. Markets fluctuate independently of speeches on most days.
5. The strength of correlations changes depending on what regime the BoE base rate is in, i.e. effective lower bound or hiking cycling.

8 Recommendations

From the findings and insights above, the following recommendations are made consistent with the BoE's mission to maintain price stability and support government economic policies through monetary policy tools and public speeches:

1. Maintain alignment between report and speech sentiment to build an aligned external messaging platform,
    a. Ensure consistency between FSR, MPR and Bank Governor Speeches,
    b. Consider refining speech sentiment, particularly during periods characterized by significant sentiment fluctuations (ECB European Central Bank, 2011).
    c. Tailor word choice speeches and reports to ensure consistent sentiment across communications.
2. Implement 'Event Studies' according to the guidelines outlined in the IMF technical handbook to look at markets in the 30-minute period following a central bank speech (IMF International Monetary Fund, 2022).
3. Conduct analysis on a regime, e.g. effective lower bound or hiking cycle, basis to better assess and define speech sentiment predictive power.
4. Enhance data collection methods by categorizing speeches based on speaker, author, and target audience to gain deeper insights.
5. Monitor major news outlets to cross-check the sentiment conveyed by these outlets against the intended message of the speeches delivered by central bank officials.

9 References

Alexandria Technology. (n.d.). *Using NLP news signals to forecast volatility and generate positive returns*. Retrieved from Alexandria Technology

Arjun. (2024). *Deploy and host a financial Bert model as API in AWS sagemaker*. Retrieved from Medium.

BoE Bank of England. (n.d.). *BoE Financial Stability Reports*. Retrieved from BoE FSR.

Central Banks' communication strategy and its impact on the Financial Markets: A Comparison of the Bank of England, the US Federal Reserve and the European Central Bank. (2023). Retrieved from ESthinktank.

Cjhutto. (2014). *Vadersentiment: Vader sentiment analysis.* Retrieved from GitHub.

ECB European Central Bank. (2011, April). *ECB*. Retrieved from ECB.

Huang, A. H. (2022). *"FinBERT: A Large Language Model for Extracting Information from Financial Text."*. Retrieved from Contemporary Accounting Research: ResearchGate

IMF International Monetary Fund. (2022, Jan). *IMF International Monetary Fund*. Retrieved from IMF International Monetary Fund: IMF.

McMahon, M. a. (n.d.). *BoE*. Retrieved from Getting through: Communicating Complex Information, BoE.

NIESR National Institute of Economic and Social Research. (2023, Sept 25). *The Truss/Kwarteng Mini Budget: One Year On*. Retrieved 04 18, 2024, from The Truss/Kwarteng Mini Budget: One Year On: NIESR.

Purohit, P. (2021b). *Financial sentiment analysis using FinBert*. Retrieved from LinkedIn.

Services, W.R.D. (n.d.). *SEC filings dictionary-based sentiment analysis*. Retrieved from WRDS.

TextBlob. (n.d.). *Simplified text processing*. Retrieved from TextBlob.

University of Notre Dame. (n.d.). *Loughran-McDonald master Dictionary W/ sentiment word lists*. Retrieved from Software Repository for Accounting and Finance: SRAF.

yiyanghkust. (n.d.). *Finbert-tone*. Retrieved from HuggingFace.

10 Appendix A: Data Cleaning

10.1 Speech data

The following steps were taken to explore, clean, and pre-process the all_speeches.csv dataset ahead of analysis.

<u>1. Initial exploration and cleaning</u>

1a. Library and package import, exploration

Firstly, the packages and libraries that are envisaged to be required were imported. These include numpy, pandas, matplotlib, nltk, wordcloud, among others. The all_speeches.csv file was then imported into a dataframe and initial exploration was conducted. This involved viewing the first five rows of the dataset, reviewing the count of rows, datatypes, column names. The nature of each column was also analysed, to determine whether the corresponding datatype was appropriate for future analysis. The reliability of the use of the 'reference' column was also determined, by running a code to ensure all reference IDs within this column were unique.

1b. Cleaning of data

The following steps were taken to ensure the data was in an appropriate format to be analysed:

- The 'date' column was converted from an 'object' to 'datetime' format.
- Remove leading/trailing spaces from 'object' datapoints to determine the true count of null cells.
- Calculated total null cell count, and replaced these with an appropriate tag "NO_TITLE"
- The 'is_gov' column contained exclusively 1 or 0 as values, therefore this column was converted into a Boolean data type.
- Values in the 'country' and 'author' column were capitalised for better readability.

<u>2. Initial Summarisation of data</u>

A number of iterations were performed to obtain valuable summaries and insights of the all_speeches data. These include:

- Identifying unique values in the 'Country', 'Author', and 'is_gov' columns, in order to understand the range of data at hand.
    - This led us to identify a type in an author name which initially appeared as ']'. The related speech was successfully identified through internet search, and the author's name was correctly added back in.
- We identified that, at the end of each reference ID is a 3-4 letter acronym identifying the related central bank for that speech (i.e. 'BOA' = Bank of Australia).
    - This acronym was pulled into a separate column, and an additional column stating the full central bank's name was created for easier analysis.

<u>3. Visualise the data</u>

The team considered a number of visualisation options for the purpose of initial exploration, and to get a better understanding of the data at hand.

- The number of speeches (by Central Bank) was plotted on a line graph. This was then re-run to focus solely on the period 1992 to 2022.
    - The frequency values by year and Central bank were also visualised in a table and saved as a csv.

- The number of BoE speeches by author was plotted on a timescale graph. This allowed us to identify the author who has written the most speeches.
    - A tabular format of this summary was also created and saved as a csv file.

<u>4. Preparing data for NLP</u>

The team carefully considered the importance of pre-processing the speech data for NLP, as each model will require the text to be presented in a different way. We have concluded that we will utilise Vader, FinBERT, and Textblob in our initial sentiment analysis.

The following pre-processing steps were taken to prepare the data for NLP:

- General (for all NLP models intended to be used)
    - The language of the speeches were analysed using the langdetect package.
    - URLs, Hashtags, and mentions were identified and removed.
        - As each element was removed, the count of rows was cross checked to ensure it has reduced to the expected value.
- TextBlob pre-processing
    - Stopwords were removed after the above pre-processing tasks were completed.
- Subsetting
    - Once the above pre-processing was completed, the dataset was subset into the specific data range that we are interested in. The criteria were:
        - Central_bank = Bank of England
        - Period = 1992 – 2022 inclusive
        - Language = English

Two csv files were saved and passed through to the next analysis phase.

- 'speech_vader_1.csv' for the subset dataset suitable for Vader and FinBERT analysis
- speech_blob_1.csv for the subset dataset suitable for TextBlob analysis

<u>5. Calculate wordcount for speech</u>

After initial exploration and discussion with the team, it was decided that identifying the word count of each speech and adding this detail to the dataframe would be useful to have. This datapoint was then added to the speeches dataset.

10.2 Economic data

Economic data was pulled from the sources outlined in Table 2 in various excel spreadsheets and comma separated variable files. It was desirable to combine this data into one place to allow the whole team to work from the same data set in parallel.

This was done using Python in a Jupyter Notebook. The key steps included:

- Individual files imported as dataframes,
- Checked for missing values and correcting unusual datatypes,
- Ensuring all dates were in UK date format and stored as datetime objects,
- Merging the data on a daily basis.

11 Appendix B: NLP Modelling

In this section of the report, we assess both general-purpose and finance-specific natural language processing (NLP) tools, considering their methodological fit, domain specificity, and performance in capturing the subtleties of financial language.

Why should NLP matter to The Bank of England?

From The Bank itself, "*Central banks, like other policymakers and possibly more so, have to communicate complex messages to different audiences with different degrees of informedness. For instance, they must communicate to financial markets as well as, increasingly, to the general public. The importance of effective communication has grown in recent years, as communication has increasingly become a key tool in the central bank policy shed.* [1]"

Advantages of NLP:

1. **Enhanced Scalability**: NLP allows for the processing of large datasets, like the one provided to us, swiftly and efficiently.
2. **Consistency in Analysis**: NLP provides a standard method for text analysis, which brings uniformity to the interpretation of financial communications, removing individual variability that often affects human analysis.
3. **Quantitative Insight**: By converting qualitative data into quantitative outputs, NLP facilitates allows for statistical analysis of sentiment trends.

limitations of NLP:

1. **Contextual Nuances**: NLP systems currently struggle with complexities of language that involve context, subtext, or cultural nuances, particularly in the specific language used in finance; a problem the models will encounter frequently.
2. **Data Dependency**: The accuracy and reliability of NLP models are contingent upon the quality and diversity of their training datasets. Inherent biases in these datasets can lead to misinterpretations and analytical errors.
3. **Adaptability to New Terminologies**: The financial sector frequently introduces new concepts and terms. NLP systems require continuous updates to incorporate such changes.

Throughout the entire NLP section of the report the "expected neutrality of a central bank communication" is often referenced. While there are multiple sources that provide reasonable backing for this statement, here is a particularly useful summary of the consequences of not maintaining an expected neutrality; "*Indeed, markets and external actors might adapt and respond in advance to central bank decisions if the central bank became totally transparent, freely expressing its preferences and decision-making processes.*

[1] McMahon , M. and Naylor, M. (no date) Getting through: Communicating Complex Information, The Bank of England. Available at: https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2023/getting-through-communicating-complex-information.pdf (Accessed: 12 April 2024).

This might jeopardise the central bank's capacity to achieve its goals since market expectations can impact economic actors' behaviour and change the expected outcomes of monetary policy.[2]"

11.1 Outlining the NLP models

The Bank of England provided a file called "*BankSpeeches_StarterCode*" which provides a preliminary overview. The key takeaway from this document is that two NLP models, **Vader** and **TextBlob**, are mentioned. As such these will be our general-purpose sentiment models.

Also provided by The Bank is the **Loughran McDonald ("lm" in the Notebooks)** financial lexicon, so this will be our first financial specific model. Our second financial model, after discussion from the team, will be **FinBERT** from Hugging Face. We are excited to employ Google's BERT on the speech data as it uses a considerably more advanced approach to hopefully offer a more sophisticated contextual analysis.

These four potential models should give us a nice spread of sentiment scores to analyse and deliberate over. Useful too that a finance model Loughran McDonald and a general model TextBlob both have a dictionary-based approach allowing us to compare dictionary methods to the more advanced Vader and FinBERT.

The subsequent findings will guide the selection of an optimal model for future analytical tasks within financial sentiment analysis.

All charts and plots can be found in the accompanying Notebooks.

<u>General sentiment analysis tools:</u>

TextBlob

"*TextBlob is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, and more.*[3]"

- TextBlob uses a dictionary-based approach for sentiment analysis, relying on part-of-speech tagging and noun phrase extraction to interpret the sentiment of textual data. This method provides the user with a simple and transparent approach to sentiment analysis.
- TextBlob can be loaded into a python environment for use.

Vader

"*VADER Sentiment Analysis. VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media and works well on texts from other domains.*[4]"

- Vader combines a dictionary-based approach with a set of grammatical and syntactical rules. This approach allows VADER to capture both the lexical content and the contextual nuances of language used in social media and other text types.

---

[2] 18, O.A. et al. (2023) Central Banks' communication strategy and its impact on the Financial Markets: A Comparison of the Bank of England, the US Federal Reserve and the European Central Bank -, -. Available at: https://esthinktank.com/2023/08/18/central-banks-communication-strategy-and-its-impact-on-the-financial-markets-a-comparison-of-the-bank-of-england-the-us-federal-reserve-and-the-european-central-bank/ (Accessed: 14 April 2024).

[3] Simplified text processing¶ (no date) TextBlob. Available at: https://textblob.readthedocs.io/en/dev/ (Accessed: 04 April 2024).

[4] Cjhutto (2014) CJHUTTO/Vadersentiment: Vader sentiment analysis. GitHub. Available at: https://github.com/cjhutto/vaderSentiment (Accessed: 04 April 2024).

- With regards to the speech data, Vader has the opportunity to offer a more nuanced insight into semantic scores with its ability to better understand sentiment contexts.
- TextBlob can be loaded into a python environment for use.

Both of these models are more general sentiment analysis tools and might provide an interesting, generalised view of the speech data but they might lack domain specificity. As such, let us now explore Loughran McDonald and FinBERT.

Financial sentiment analysis tools:

Loughran McDonald

Loughran McDonald is similar to TextBlob in the approach to sentiment analysis, however, Loughran McDonald offers the financial domain specifics where TextBlob does not. *"They analysed over 40,000 10-Ks from 1994-2007. Within these financial documents, they scored individual words across different categories of sentiment* [5]"

- Loughran McDonald is provided as a large .csv file from The Bank. So, it warrants further discussion as how we implement it best. This can be found at the end of the NLP section titled "Further consideration for Loughran McDonald implementation"

FinBERT

"*FinBERT is a pre-trained NLP model to analyse sentiment of financial text. It is built by further training the BERT language model in the finance domain, using a large financial corpus and thereby fine-tuning it for financial sentiment classification.*[6]"

- In the same way that TextBlob and Loughran McDonald have a similar approach, Vader and FinBERT also share similarities that extend beyond a simple dictionary usage.
- Like Vader, FinBERT offers a more comprehensive and nuanced understanding of sentiment by considering context. Unlike Vader, FinBERT uses Google's BERT to leverage the advanced deep learning architecture of BERT to provide a more nuanced understanding of sentiment,
- FinBERT can be loaded into a python environment for use.

[5] Services, W.R.D. (no date) SEC filings dictionary-based sentiment analysis, WRDS. Available at: https://wrds-www.wharton.upenn.edu/pages/classroom/sec-filings-dictionary-based-sentiment-analysis/#:~:text=Introduction%20to%20the%20Loughran%2DMcDonald,the%20strength%20of%20the%20sentiment. (Accessed: 04 April 2024).

[6] yiyanghkust/Finbert-tone, no date. Hugging Face. Available at: https://huggingface.co/yiyanghkust/FinBERT-tone (Accessed: 01 April 2024).

11.2 Model performance.

Ideally, we would want to choose one general sentiment analysis tool and one more financial specific sentiment analysis tool. This way we might be able to see if there is a difference between general sentiment scores to financial ones.

<u>Comparison of general sentiment analysis tools:</u>

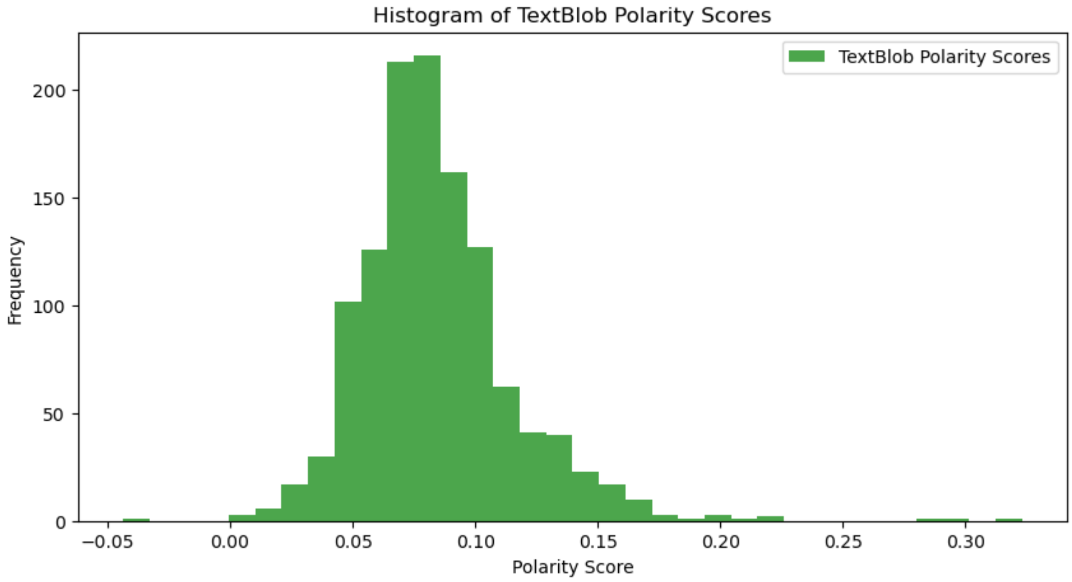
TextBlob

TextBlob had the most rigorous normalisation of all the models:

- **Ensure text is clean (Remove noise):** Remove noise such as URLs, HTML tags, hashtags, mentions (if we don't believe it would contribute to sentiment). From the above, we have decided to keep hashtags and mentioned, but removed URLs.
- **Ensure text is clean (Handle missing values):** Also ensure there are no missing values. We have completed this step above.
- **Language:** TextBlob primarily supports English out of the box. We have identified which speeches are in English.
- **Text preprocessing (Lowercase):** Processed below.
- **Text preprocessing (Remove Stopwords):** Processed below.
- **Text preprocessing (Lemmatisation):** Lemmatisation was not done at this stage (it is also optional), as we have received a sentiment labelled wordlist from BoE where the reference text has not been lemmatized.

The speech data is loaded and TextBlob run through the data to acquire sentiment.

Here is the distribution of the TextBlob sentiment across all speeches. The histogram allows for an immediate visual interpretation of the data:



- This is a healthy shaped curve but shows a surprisingly small spread across the sentiment range with the vast majority of speeches clustered around 0.1 on the sentiment scale.

- There are two ways to interpret this:
    - We could be seeing representative use of formal, cautious, and neutral language that The Bank of England might use.
    - Or this spread is a result of TextBlob's general-purpose model not grasping economic jargon and financial context. The extreme nature of the histogram suggests that this is more likely.

```
In [12]: speech_blob['textblob_polarity'].describe()

Out[12]: count    1209.000000
         mean        0.084077
         std         0.031670
         min        -0.043709
         25%         0.064897
         50%         0.079922
         75%         0.098878
         max         0.323077
         Name: textblob_polarity, dtype: float64
```

The descriptive statistics suggest:

- **Mean**: The mean sentiment score is 0.084077, suggesting a generally neutral sentiment across the dataset, with a slight positive lean.
- **Standard Deviation**: The standard deviation is 0.031670, which is relatively small, indicating that the sentiment scores do not vary widely and tend to be close to the mean.
- **Minimum**: The lowest polarity score is -0.043709, showing that the most negative sentiment is only slightly below neutral.
- **25th Percentile**: 25% of the scores are below 0.064897. Since this is above zero, it suggests that the lower quartile of data still leans towards the positive sentiment.
- **Median** (50th Percentile): The median score is 0.079922, which is close to the mean, further indicating a very slight positive sentiment trend.
- **75th Percentile**: 75% of the scores are below 0.098878, and being above the median, this suggests that the upper quartile of the scores is more strongly positive but still fairly neutral.
- **Maximum**: The highest polarity score is 0.323077, which indicates the presence of some positive sentiments but nothing extremely positive.

Vader

Vader is designed for social media text and requires less normalisation than TextBlob. Its built-in appreciation for nuances in the language, such as capitalisation and punctuation, enables it to interpret emphasis and sentiment with little preprocessing. The normalization steps we applied to Vader included:

- **Ensure text is in English:** We have tagged each speech for their language.
- **Handle missing values:** That has been completed in the previous cleaning steps.

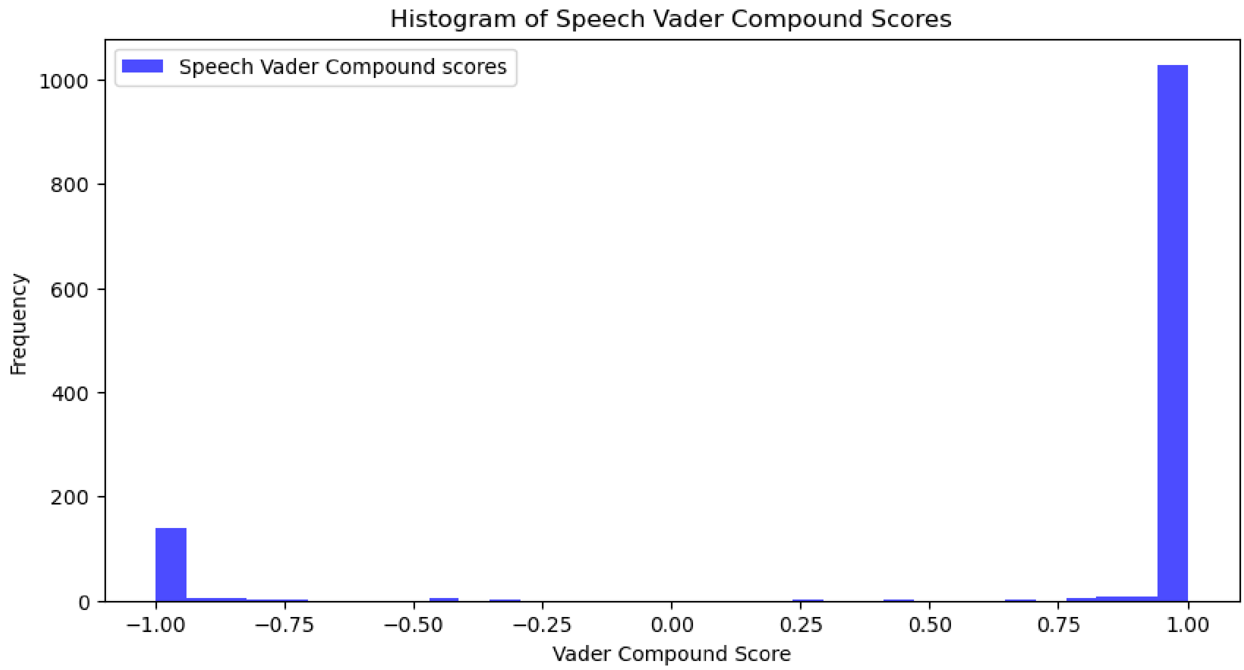
- **Standardise text data:** Vader works well if the text is standardised (i.e. ensure consistent use of quotation marks etc). We will take the assumption that the use of quotation marks etc has been applied consistently.

- **Remove unnecessary elements**: It is good practice to remove URLs and other specific mark ups. However, in the same vein, it is important not to over clean. As seen above, we have kept hashtags and mentions as they appeared to be important in providing context to the speeches, however URLs have been removed.

- **Ensure Data is in a Suitable Format:** The data is contained in a dataframe column.

The speech data is loaded, and Vader run through the data to acquire sentiment.

Here is the distribution of the Vader sentiment across all speeches. The histogram allows for an immediate visual interpretation of the data:



- Considerably different distribution of sentiment scores to TextBlob, Vader almost seems to be binary but for the occasional rise near -0.5 and near 0.8.

- This is unlikely to reflect how the Bank of England chooses the tone of its speeches, so we have to assume again that Vader's more generalised lexicon is not appropriate for more financial contexts.

- Since Vader is particularly useful at analysing language like social media, Vader's method of analysing intensity might be picking up on the presence of strong financial jargon and interpreting it as carrying more sentiment weight than intended.

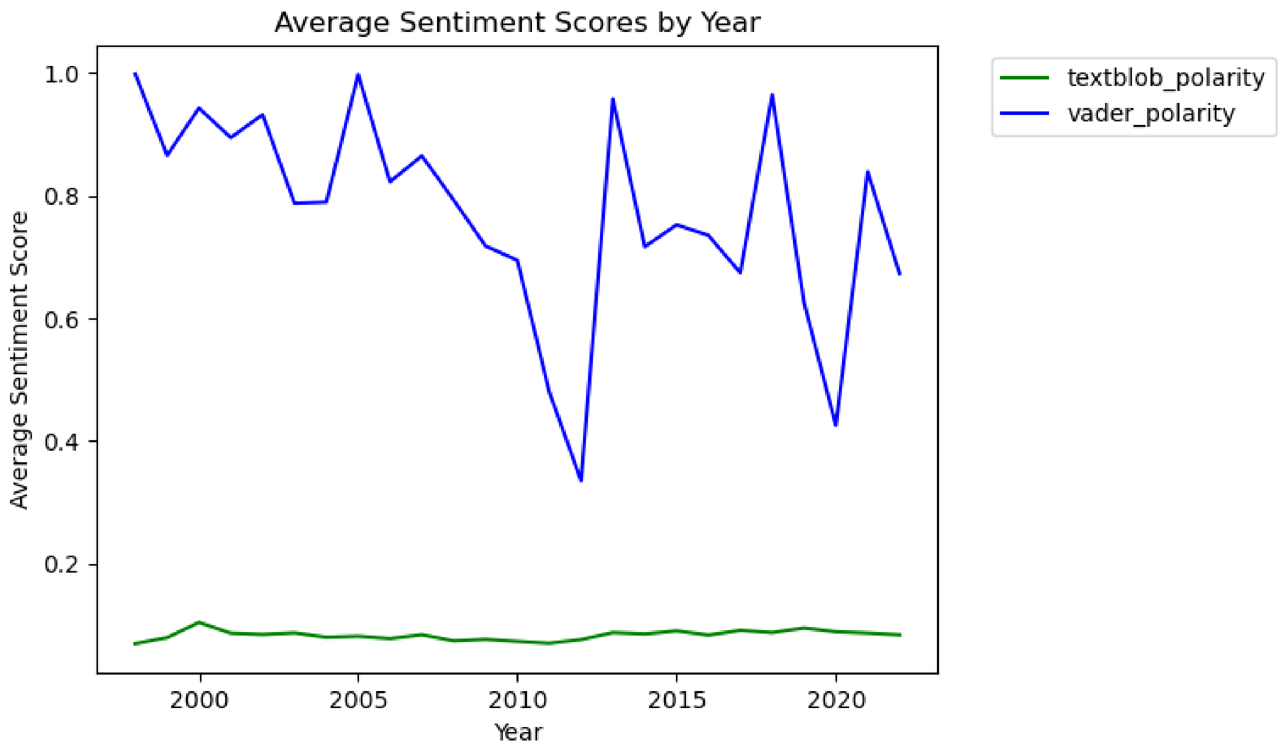
```
speech_vader['vader_compound'].describe()
```

```
count    1209.000000
mean        0.735447
std         0.664696
min        -1.000000
25%         0.996400
50%         0.999400
75%         0.999800
max         1.000000
Name: vader_compound, dtype: float64
```

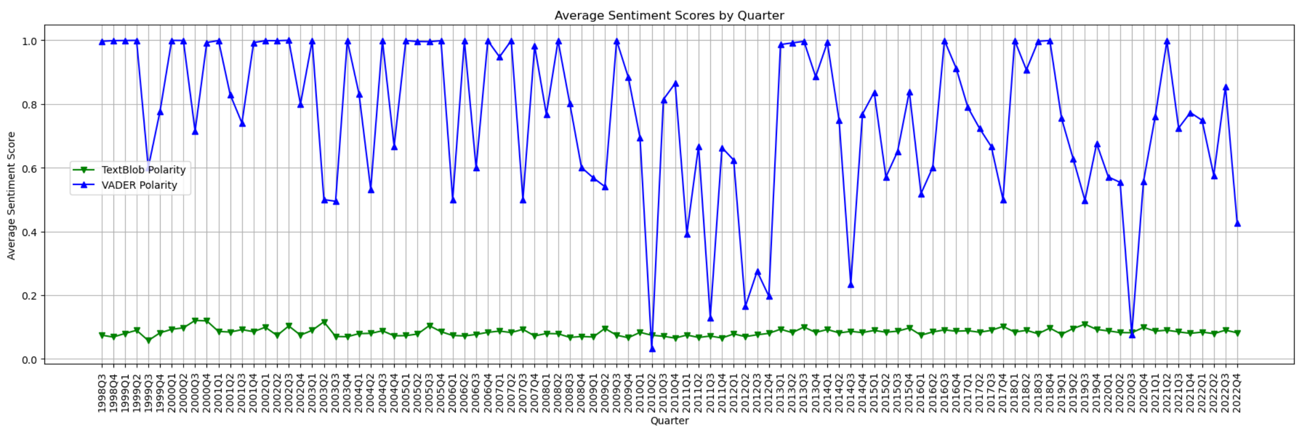
The descriptive statistics suggest:

- **Mean**: The average sentiment score is very high (0.735447), indicating an extremely positive sentiment across the dataset.
- **Standard Deviation**: The standard deviation is significant too (0.664696), suggesting a wide range of sentiment scores from highly negative to highly positive.
- **Minimum**: The lowest sentiment score is -1.000000, indicating the presence of extremely negative sentiments in some instances.
- **25th Percentile**: This value is very high (0.996400), which is unusual for sentiment scores and suggests that 25% of the data has very high positive scores, just below the maximum.
- **Median** (50th Percentile): The median is 0.999400, extremely close to the maximum, again showing a very strong positive sentiment within the dataset.
- **75th Percentile**: With 75% of scores below 0.999800, we see that the vast majority of the dataset skews towards the maximum positive score.
- **Maximum**: The highest sentiment score is 1.000000, the highest possible score, indicating extremely positive sentiments.

If we compare both tools to each other through a time-series analysis of the sentiment scores, averaged annually (mean), it neatly reveals the limited variability with TextBlob and more pronounced fluctuations with VADER.



Increasing the granularity and breaking down the data quarterly, the variance in VADER's scores becomes more evident, revealing the model's probable over-sensitivity to financial terminology and its tendency to produce extreme sentiment values. This original chart can be viewed in the Notebook for better accessibility, this chart is for illustrative purposes only.



- Vader's sharp peaks and troughs start to show the model's weakness and lack of usability here. We begin to see the earlier distribution that almost looked binary become displayed.
- Further granularity of time data of this time series chart revealed the extend of Vader's movements across the polarity scale.

With each further level of granularity Vader becomes more and more exaggerated, to the point where it is not possible to assume that The Bank of England would deliver such bipolar communications.

- Should you wish to view this data at a daily level this plot can be explored in an interactive Plotly chart in the Notebook attached.

In conclusion, it became apparent that general-purpose NLP models were not adequately aligned with the specific requirements of financial language. Given these observations, the decision was made to concentrate our analytical efforts on finance-specific models like Loughran McDonald and FinBERT. This approach was underpinned by the premise that models tailored to the financial domain would inherently offer more insightful and accurate sentiment assessments.

<u>Comparison of financial sentiment analysis tools:</u>

Loughran McDonald

We received a custom word list to help with our sentiment analysis of financial terms. This document, "*LSE_DA_BoE_Employer_project_Sentiment-labelled_wordlist.xlsx,"* was provided as part of the core dataset by The Bank of England.

Upon reviewing the document, it was found that the main the sheet is labelled "*Loughran-McDonald_MasterDiction*," pointing us towards a financial lexicon free for use in academic research.

Composition:

Loughran McDonald is formed of two parts:

- The base dictionary is release 4.0 of 2of12inf. "*This is a fairly common baseline dictionary and is oriented towards common words*.[7]"
- The financial base is specifically targeting the 10-K filings and earnings calls, specifically "*EDGAR 10-K archive and earnings calls from CapIQ* [8]."

Approach:

This model uses a "*lexicon or "bag of words" approach… that uses a dictionary of words or phrases that are labelled with sentiment*. [9]" In sentiment analysis, this method quantifies the sentiment of a text by tallying the sentiment scores of the words that appear in the text, as defined by the lexicon.

Application:

To use Loughran McDonald we installed PySentiment2, which proved a straightforward way of applying the model.

- Worth noting however that the general base dictionary for PySentiment2 is Harvard IV-4 and is different from the original 2of12inf. However, the financial specific dictionary remains the same. This also simplifies the process by loading a library, rather than finding a way to important the Loughran McDonald lexicon directly.

For the purposes of this research, PySentiment2 analysis and Loughran-McDonald (lm) amount to the same thing.

Strengths:

---

[7] Marketing Communications: Web // University of Notre Dame (no date) *Loughran-McDonald master Dictionary W/ sentiment word lists // software repository for accounting and Finance // University of Notre Dame*, *Software Repository for Accounting and Finance*. Available at: https://sraf.nd.edu/loughranmcdonald-master-dictionary/ (Accessed: 01 April 2024).

[8] Ibid.

[9] *Using NLP news signals to forecast volatility and generate positive returns* (no date) *Alexandria Technology*. Available at: https://www.alexandriatechnology.com/blog/comparing-natural-language-processing-nlp-approaches-for-earnings-calls (Accessed: 01 April 2024).

- *Simplicity and Accessibility:* Loughran McDonald's lexicon-based approach is straightforward to implement, especially through PySentiment2. It also requires less computational power.
- *Transparency:* The model offers clear insights into how sentiment scores are derived, with each word's sentiment contribution being quantifiable and directly observable within the lexicon itself.
- *Specific Financial Focus:* The lexicon is specifically tailored for financial texts with a focus on 10-K filings and earnings calls.

Weaknesses:

- *Static Analysis:* The lexicon approach does not account for context or the nuance of word usage, potentially leading to inaccuracies in sentiment analysis where the context significantly alters meaning.
- *Limited to Predefined Lexicon:* It can miss or inaccurately score sentiment for words not included in the lexicon or new financial jargon that emerges after the lexicon's last update. The financial lexicon has fewer sources than other models, for instance FinBERT.

```
speech_lm['lm_polarity'].describe()
```

```
count    1209.000000
mean       -0.245673
std         0.243331
min        -0.750000
25%        -0.421538
50%        -0.287356
75%        -0.117117
max         0.773585
Name: lm_polarity, dtype: float64
```

The descriptive statistics suggest:

- **Mean**: The mean sentiment score is -0.245673, which is somewhat negative, suggesting a general negative sentiment across the dataset.
- **Standard Deviation**: The standard deviation is relatively high at 0.243331, indicating variability in sentiment across different texts.
- **Minimum**: The most negative sentiment score is -0.750000, showing that there are texts with a strongly negative sentiment.
- **25th Percentile**: A quarter of the data falls below -0.421538, further emphasizing that a significant portion of the dataset contains fairly negative sentiments.
- **Median** (50th Percentile): The median score is -0.287356, confirming the negative trend observed in the mean.
- **75th Percentile**: Three-quarters of the data has a polarity score above -0.117117, indicating that the most negative sentiments are not shared by all texts in the dataset.
- **Maximum**: The highest sentiment score is 0.773585, revealing that there are also instances of positive sentiment, although these are less frequent given the negative mean score.

FinBERT

The second financial model is FinBERT, specifically we used FinBERTTone as it "*achieves superior performance on financial tone analysis task*.[10]" Specificall0y, it is "*fine-tuned on 10,000 manually annotated (positive, negative, neutral) sentences from analyst reports.[11]*"

Although this quote is admittedly cited from the creator of FinBERT, we were drawn to the claim that "*FinBERT substantially outperforms the Loughran and McDonald dictionary and other machine learning algorithms* [12]"

Composition:

FinBERT is formed of two parts:

- The base of the model is Google's BERT. "*BERT is a model for natural language processing developed by Google that learns bi-directional representations of text to significantly improve contextual understanding of unlabelled text across many different tasks…. BERT models have been trained on BookCorpus and English Wikipedia, which have in total more than 3.5 billion words [13]*"
- FinBERT is trained on the following three financial communication corpus. The total corpora size is 4.9B tokens. [14]
    - Corporate Reports 10-K & 10-Q: 2.5B tokens
    - Earnings Call Transcripts: 1.3B tokens
    - Analyst Reports: 1.1B tokens

Approach:

FinBERT uses BERT to present a large language model approach rather than the static lexicon approach Loughran McDonald offers. "*Unlike traditional language models that read text in one direction (either left-to-right or right-to-left), BERT considers both directions simultaneously. This bidirectional context allows BERT to capture complex relationships and nuances in language.[15]*"

Application:

We used FinBERTtone by following the official documentation on HuggingFace.

Strengths:

- *Contextual Awareness:* By using Google's BERT's deep learning architecture, FinBERT can understand the context in which words are used, allowing for more nuanced sentiment analysis that reflects the complexities of financial statements.

[10] Huang, Allen H., Hui Wang, and Yi Yang. "FinBERT: A Large Language Model for Extracting Information from Financial Text." Contemporary Accounting Research (2022).
[11] *Yiyanghkust/Finbert-tone · hugging face* (no date) *yiyanghkust/finbert-tone · Hugging Face*. Available at: https://huggingface.co/yiyanghkust/FinBERT-tone (Accessed: 01 April 2024).
[12] Huang, A. (2022) *(PDF) FinBERT: A large language model for extracting information …*, *Researchgate*. Available at: https://www.researchgate.net/publication/364070191_FinBERT_A_Large_Language_Model_for_Extracting_Information_from_Financial_Text (Accessed: 01 April 2024) .
[13] Purohit, P. (2021b) *Financial sentiment analysis using FinBert*, *LinkedIn*. Available at: https://www.linkedin.com/pulse/financial-sentiment-analysis-using-finbert-praveen-purohit (Accessed: 01 April 2024) .
[14] *Yiyanghkust/Finbert-tone · hugging face* (no date) *yiyanghkust/finbert-tone · Hugging Face*. Available at: https://huggingface.co/yiyanghkust/FinBERT-tone (Accessed: 01 April 2024).
[15] Arjun (2024) Deploy and host a financial Bert model as API in AWS sagemaker, Medium. Available at: https://medium.com/@viswanathan.arjun/deploy-and-host-a-financial-bert-model-as-api-in-aws-sagemaker-7b381dfbcd6a (Accessed: 01 April 2024).

- *Superior Performance on Financial Tone:* The additional fine-tuning on manually annotated sentences from analyst reports means that FinBERT offers even more specialised financial knowledge. This is particularly important given the nature of BOE's speeches.
- *High Accuracy in Negative Sentiment Detection:* FinBERT is particularly adept at identifying negative sentiments which, given the tumultuous period of the data provided, could be invaluable. "*Further analyses show a distinct advantage in FinBERT's detection of negative sentiments (89.7% accuracy, compared to less than 60% accuracy for non-BERT algorithms), a point worth noting given that negative sentiments have a larger impact on investors than positive ones* [16]"

Weaknesses:

- *Computational Resource Requirements:* Being a deep learning model, FinBERT requires substantial computational resources for training and inference, which proved difficult to install and time consuming to run.
- *Complexity and Opacity:* The workings of FinBERT are more complex compared to a dictionary-based approach, making it less transparent in how sentiment scores are derived.

```
speech_finbert['finbert_polarity'].describe()
```

```
count    1209.000000
mean        0.005142
std         0.234896
min        -0.805046
25%        -0.105807
50%        -0.004519
75%         0.093910
max         0.938213
Name: finbert_polarity, dtype: float64
```

The descriptive statistics suggest:

- **Mean**: The average sentiment score is very close to zero (0.005142), indicating that overall sentiment across all speeches is almost neutral.
- **Standard Deviation**: A relatively high standard deviation (0.234896) suggests there's considerable variability in sentiment scores.
- **Minimum**: The most negative sentiment score is quite low (-0.805046), showing that there have been instances of strongly negative sentiment.
- **25th Percentile**: 25% of the scores are lower than -0.105807, which could be considered mildly negative.
- **Median** (50th Percentile): The median sentiment score is slightly negative (-0.004519), reinforcing the mean's indication of a general tilt towards neutral sentiment.
- **75th Percentile**: 75% of the scores are below 0.093910, indicating that more positive sentiments are not as extreme as the negative ones.
- **Maximum**: The most positive sentiment score is very high (0.938213), showing that there have also been instances of strongly positive sentiment

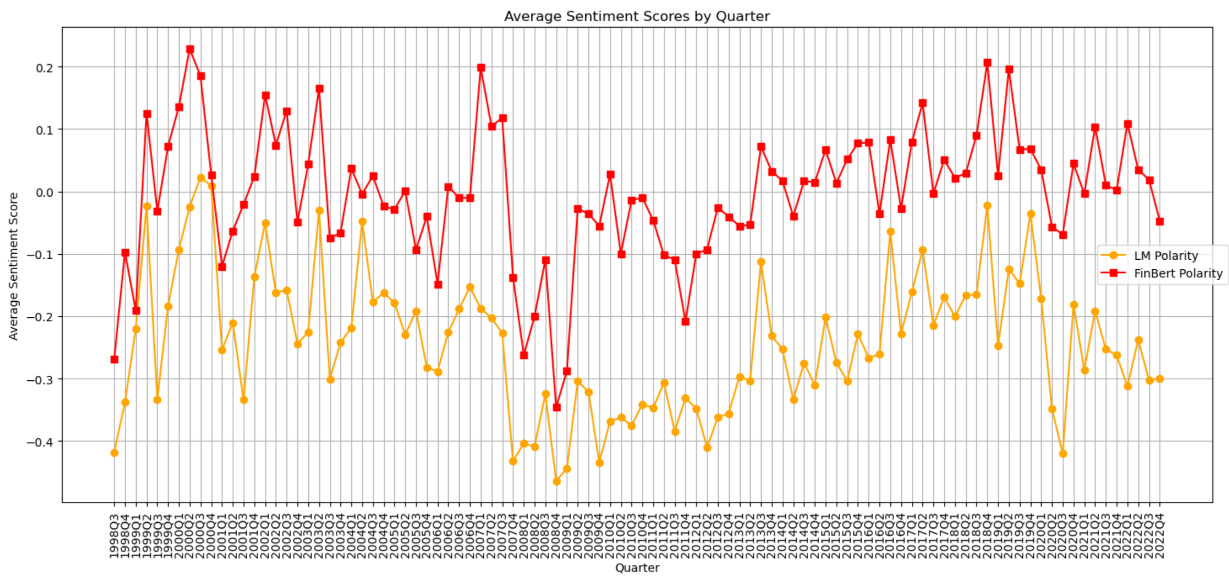
[16] Huang, Allen H., Hui Wang, and Yi Yang. "FinBERT: A Large Language Model for Extracting Information from Financial Text." Contemporary Accounting Research (2022).

Table comparison between the financial models:

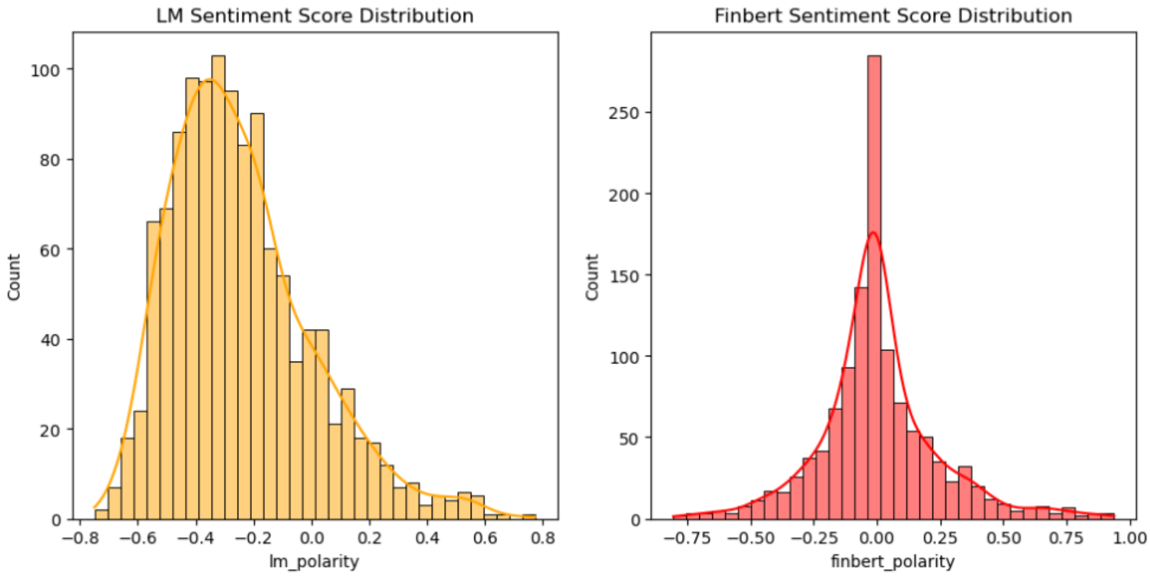
Name	Base dictionary	Financial dictionary	Approach	Last updated
Loughran-McDonald (lm)	2of12inf	- EDGAR 10-K archive - Earnings calls from CapIQ - Analyst Reports	Dictionary (BOW)	February 2024
FinBERT	Google's - BERT	- Corporate Reports 10-K & 10-Q: 2.5B tokens - Earnings Call Transcripts: 1.3B tokens - Analyst Reports: 1.1B tokens	llm (Large language model)	2022

The two models perform very similar in comparison to each other as evidenced here by the quarterly time series plot comparing models. This plot is for illustrative purposes, to show lm in orange and FinBERT in red, please view the Notebook for a full navigable image.

Both models show a comparable understanding of change of sentiment within a financial context over time. However, Loughran McDonald shows a generally more negative view and FinBERT's neutrality here perhaps reflects BERT's ability to appreciate nuanced context more successfully.



The distribution of sentiment scores goes on to reflect a similar idea, revealing how FinBERT offers a more balanced distribution.



Even though both models perform well, ultimately FinBERT proves to be the more reliable model and will be the choice moving forward. The histogram allows for an immediate visual interpretation of the data. Consider the following reasons:

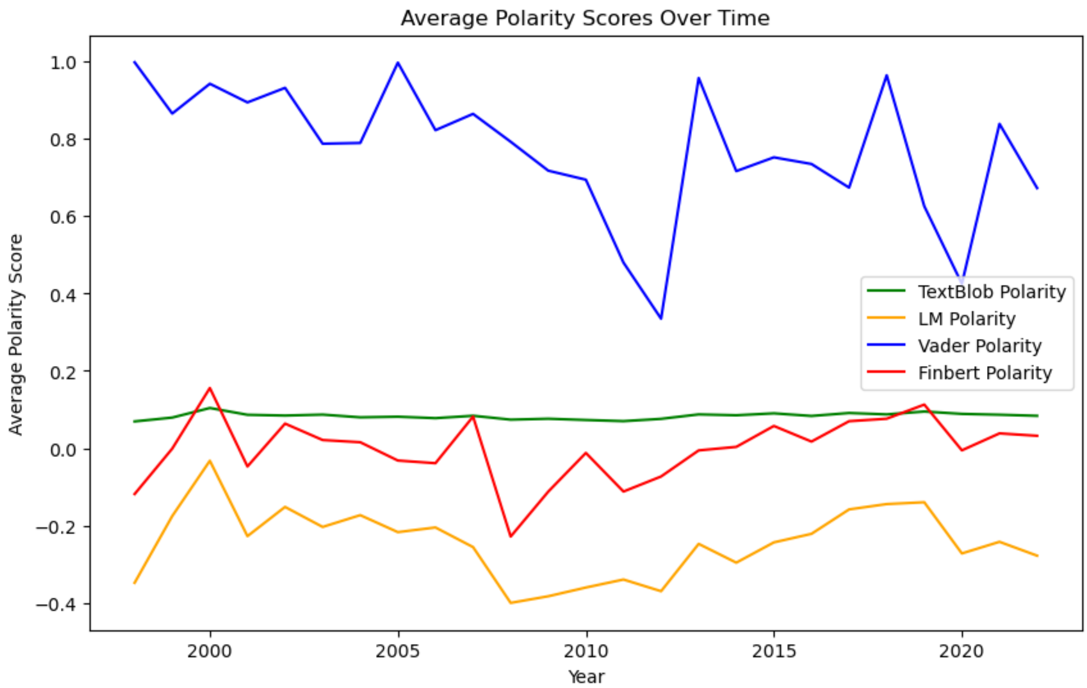
- **Deep Learning Advantage**: BERT's architecture for a nuanced understanding of financial language context is the key here, outperforming Loughran McDonald's static lexicon approach.
- **Accuracy in Sentiment Detection**: FinBERT proves superior at detecting negative sentiments, essential for Bank of England communications during several economic downturns present in the time period.
- **Balanced Sentiment Distribution**: FinBERT offers a better distributed view of sentiments, avoiding the skew towards negativity seen with Loughran McDonald, providing what feels like a fairer analysis of speeches.

Comparing all models:

In concluding our findings from the year-by-year and quarterly analyses, along with a detailed review of the sentiment score distributions, FinBERT emerges as the superior model for this speech evaluation.

Unlike the general-purpose models TextBlob and Vader, which either exhibited too narrow a sentiment range or skewed towards extremes, FinBERT demonstrated a capacity for nuanced interpretation closely aligned with the measured tone of Bank of England communications. Loughran McDonald, while closely aligned with financial terminology, still lacked the contextual depth FinBERT provided, allowing its polarity to be unusually skewed in the negative.

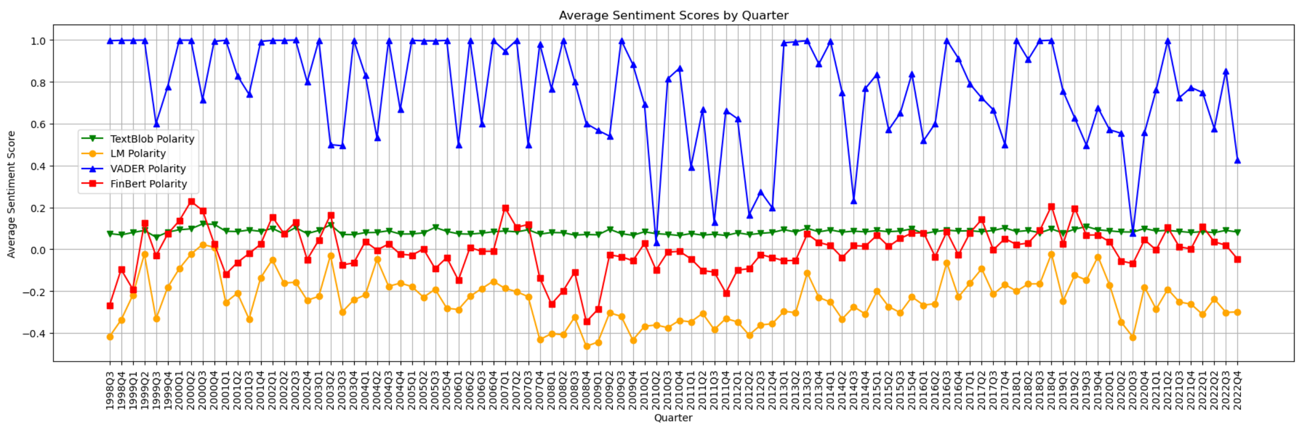
Let's compare all model's average annual sentiment now using a simple time series graph that we've used before:



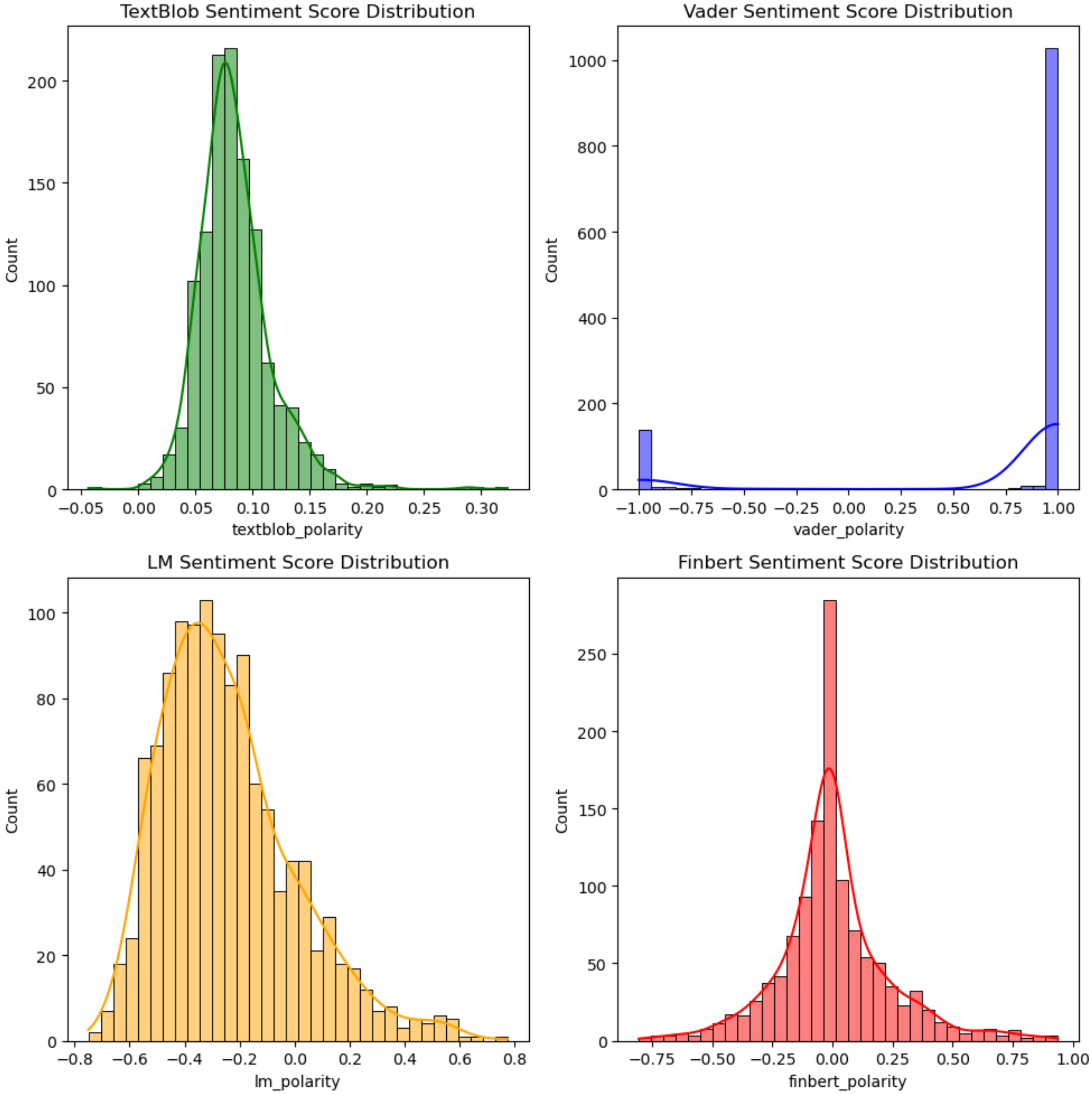
- **TextBlob** tends toward an unrepresentative neutrality, failing to capture the range of sentiments typically present in financial discourse.

- **Vader** displays a seemingly binary interpretation of sentiment, which oversimplifies the nuanced language used in financial communications.
- **Loughran McDonald** and FinBERT exhibit similar trends in sentiment analysis; however, Loughran McDonald's pronounced negative skew suggests a somewhat limited understanding of the data, in contrast to FinBERT's more balanced approach.
- **FinBERT** offers good movement around the expected neutral mark, suggesting that BERT has offered a more balanced interpretation of the speech data.

The average quarterly sentiment scores serve to show us, as above, that FinBERT offers the most balanced and consistent scores. The plot can be viewed in better detail on the Notebook provided. It is here in the report for illustrative purposes.



Reflecting on all the model's distribution with a histogram we can see similar patterns as the time series reveals the spread of the models' polarity score nicely.



- **Vader**: seems to categorise sentiment towards extremes, especially towards the positive, which seems unlikely to reflect the typically reserved language of the Bank of England.
- **Loughran McDonald**: appears to be the normally distributed, which might suggest a better fit for the type of language used in financial documents, but the significant negative skew suggests that it has struggled to make entirely accurate predictions.
- **TextBlob**: seems to be normally distributed as well but the range of sentiment scores is much too small to be considered useful.
- **FinBERT**: also appears normally distributed and the sharp peak at zero with symmetrical tails on both sides, indicates that it not only captures neutrality well but also detects variations towards both positive and negative sentiments in a balanced way. Given this, and the above analysis, FinBERT feels like the most appropriate model to continue with.

Feature	TextBlob	VADER	Loughran McDonald(lm)	FinBERT
Base Dictionary(s)	***Base*:** General English	***Base*:** General English with social media emphasis	***Base*:** - PySentiment2 ***- Financial:*** - EDGAR 10-K archive - CapIQ earnings calls	***- Base*:** - Google's BERT (BookCorpus and English Wikipedia), ***- Financial:*** - Corporate Reports 10-K & 10-Q: 2.5B tokens - Earnings Call Transcripts: 1.3B tokens - Analyst Reports: 1.1B tokens
Approach	General lexicon (Bag of Words)	General lexicon and rule-based	General lexicon (Bag of Words) with financial focus	Large Language Model (LLOUGHRAN MCDONALD), deep learning with financial focus
Domain Specificity	General	General with a focus on social media	Financial documents	- Financial documents - Financial texts
Last Updated	N/A (updates vary by library version)	N/A (updates vary by library version)	February 2024	2022
Key Strengths	- Simple, transparent and user-friendly . - Versatile for general NLP tasks.	- Understands nuances through rule-based analysis.	- Tailored for financial text analysis. - Transparency in sentiment scoring	- Deep contextual understanding - High accuracy, especially for negative sentiments - Deep understanding of financial context.
Key Considerations	- May not capture domain-specific	- Less effective outside social media contexts.	- Limited in capturing complex context.	- Requires substantial computational resources

Feature	TextBlob	VADER	Loughran McDonald(lm)	FinBERT
	nuances. - Less accurate for complex sentiments.	- Static rules can miss specific speech language trends.	- Static lexicon may miss specific speech language trends.	- Complexity in model understanding, less transparent.
Ideal Use Case	General sentiment analysis	General sentiment analysis with social media	Financial text analysis focusing on specific and static terminology	In-depth financial sentiment analysis where context and nuanced understanding are crucial

Additional LM Considerations

There were questions and complexities that arose during our consideration of Loughran McDonald, specifically how best to implement it. Do we add the words to the Vader/ TextBlob NLP model's lexicons, do we replace them, or do we use it as a standalone model?

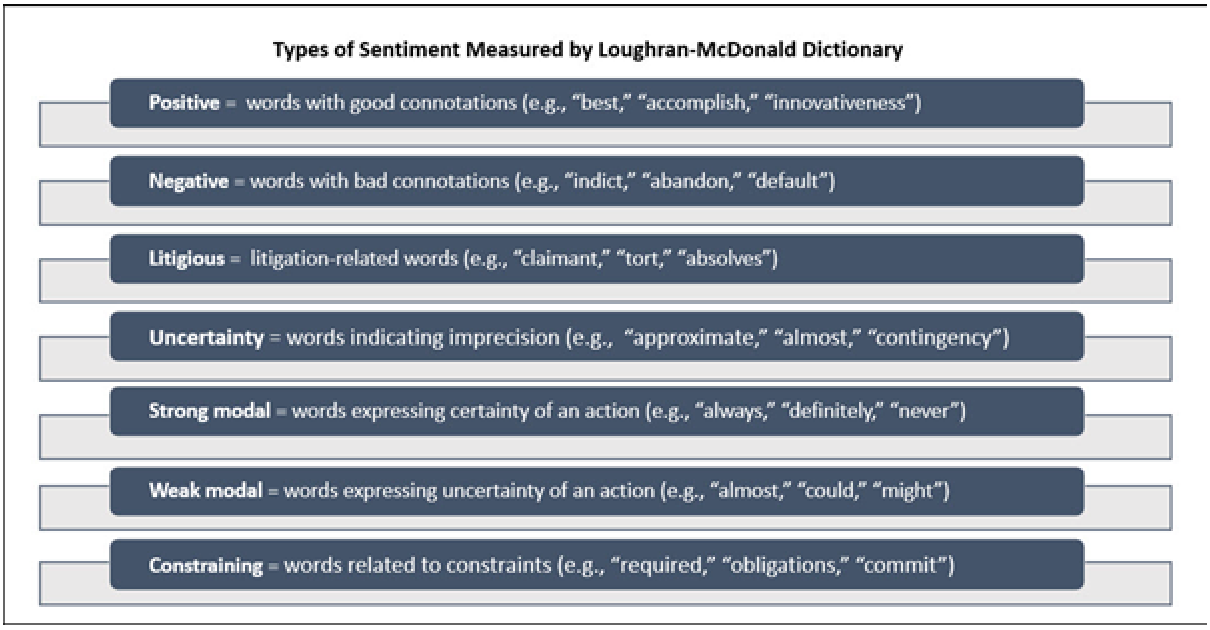
FinBERT is too complex and inaccessible a model to fuse Loughran McDonald to, and fusing Loughran McDonald to TextBlob does not offer any chance for adding dimensionality through Vader's better context appreciation. Therefore, we will be exploring pairing Loughran McDonald with Vader.

Initially, adding to VADER's existing lexicon with Loughran-McDonald feels more appropriate than replacing it because, for the most part, they could enrich each other. Loughran McDonald's domain specific lexicon should serve to increase the accuracy and relevance of VADER's already comprehensive lexicon.

However, there are several issues why introducing the Loughran McDonald lexicon to VADER 's lexicon isn't straightforward:

First:

- VADER is designed primarily for social media texts, focusing on the polarity of sentiment (positive, negative, neutral). It assigns a single score to words/phrases based on their perceived sentiment in general usage.
- Loughran-McDonald is specialised for financial texts, using multiple specific dimensions beyond the simple polarity of VADER: Negative, Positive, Uncertainty, Litigious, Strong, Weak, and Constraining. These categories reflect the nuanced way words can convey sentiment in financial contexts.



[17]

Second:

- It is technically possible to adjust these extra dimensions of Loughran McDonald to suit the VADER model, but this is a complex and time-consuming activity. Consider the following arbitrary numbers that attempt to map Loughran McDonald's polarity scale onto Vader:

- **Positive**: +1.0 (top end of the spectrum)
- **Strong**: +0.5
- **Litigious**: 0 (may not significantly affect general sentiment unless in specific contexts)
- **Constraining**: -0.25 (assuming a slight negative sentiment)
- **Uncertainty**: -0.25 (assuming a slight negative sentiment)
- **Weak**: -0.5
- **Negative**: -1.0 (bottom end of the spectrum)

  - While possible, you immediately lose out on the subtlety those dimensions created. For instance, how do you decide on the scores between Constraining and Uncertainty?
  - You also lose sight of the fact that Uncertainty, Litigious, Strong, Weak, and Constraining are all modifiers as well. They exist both as independent categories but also additional scores to the baseline of Positive or Negative in order to enhance interpretation.
  - For instance, ENCUMBER has several categories

Word	Negative	Positive	Uncertainty	Litigious	Strong	Weak	Constraining
ENCUMBER	1	0	0	1	0	0	1

[17] Services, W.R.D. (no date) SEC filings dictionary-based sentiment analysis, WRDS. Available at: https://wrds-www.wharton.upenn.edu/pages/classroom/sec-filings-dictionary-based-sentiment-analysis/#:~:text=Introduction%20to%20the%20Loughran%2DMcDonald,the%20strength%20of%20the%20sentiment. (Accessed: 04 April 2024).

Third:

- Context is important and there may be a mismatch of VADER's interpretation of words against Loughran McDonald's.
    - For instance, consider the word "Suggest"
        - From a VADER perspective that could be a simple neutral verb
        - From a Loughran McDonald perspective however, "Suggest" means Uncertainty and Weakness

Word	Negative	Positive	Uncertainty	Litigious	Strong	Weak	Constraining
SUGGEST	0	0	1	0	0	1	0

Given these considerations, operating Loughran McDonald as a standalone model emerges as the most effective approach. This method preserves the lexicon's financial domain specificity and avoids the compromises involved in integration. To facilitate this, we employ PySentiment2, allowing for a direct incorporation of Loughran McDonald's lexicon into our analysis framework.

12 Appendix C: FSR and MPR data scraping

Initially, when looking to analyse the content of the Monetary Policy and Financial Stability Repots, manual text extraction was considered, however this is time consuming, introduces unreliability and was not repeatable. Downloading these files directly was a better option and this was achieved using the Python library Requests. This was written into a function to allow specification of the file to download and the date range. This was necessary as the format, file path and name of the reports change over time.

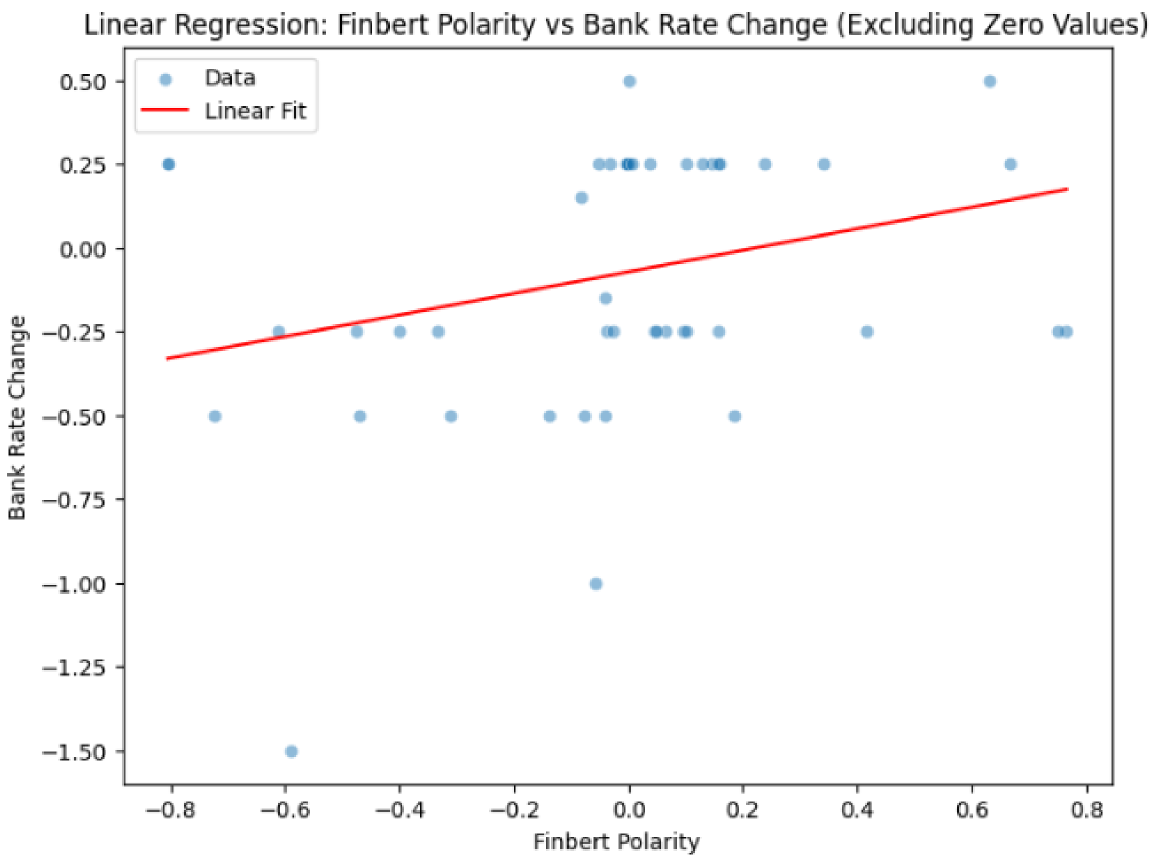
Once the files were downloaded, the intention was to extract a summary and the full text. As the files are in pdf format, PfdReader and fitz from the libraries PyPDF2 and PyMuPDF respectively were evaluated, and it was found that fitz was the most comprehensive and efficient.

Ultimately, the reliability of the summary extraction was questionable as the formatting, structure, and extent of the summary sections in the reports varied considerably over the data. The full text was cleaned, sorted and saved for to processing by FinBERT.

13 Appendix D: Linear Regression between base rate and speech sentiment

A weak correlation between base rate and speech sentiment was observed, quantified by a simple Ordinary Least Squares (OLS) linear regression. The analysis focused on periods without base rate changes under the assumption that the BoE internally anticipates these changes. Despite speech sentiment being a statistically significant predictor of base rate change (F-statistic = 4.1, Prob F-stat = 0.048), the model's explanatory power is limited (R-squared = 0.08), indicating that only 8% of the variability in base rate can be explained by speech sentiment alone. Notably, potential issues with the residuals were identified, including autocorrelation (Durbin-Watson statistic = 0.7) and departure from normality (Jarque-Bera test). If further exploration is desired, alternative regression models accounting for a broader range of variables should be considered, e.g. multiple linear regression, decision trees and/or random forests.

Visualization and results of regression are displayed on the next page.




```
                            OLS Regression Results
==============================================================================
Dep. Variable:       bank_rate_change   R-squared:                       0.082
Model:                            OLS   Adj. R-squared:                  0.062
Method:                 Least Squares   F-statistic:                     4.118
Date:                Sat, 13 Apr 2024   Prob (F-statistic):             0.0482
Time:                        16:11:30   Log-Likelihood:                -21.306
No. Observations:                  48   AIC:                             46.61
Df Residuals:                      46   BIC:                             50.35
Df Model:                           1
Covariance Type:            nonrobust
====================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const             -0.0721      0.056     -1.295      0.202      -0.184       0.040
finbert_polarity   0.3217      0.159      2.029      0.048       0.003       0.641
==============================================================================
Omnibus:                        8.088   Durbin-Watson:                   0.700
Prob(Omnibus):                  0.018   Jarque-Bera (JB):                7.100
Skew:                          -0.832   Prob(JB):                       0.0287
Kurtosis:                       3.882   Cond. No.                         2.85
==============================================================================
```

14 Appendix E: June 2003 to June 2004 Speech Sentiment

From Team 7's preliminary presentation to the Client, Bank of England, it was noted that it would be helpful for them to be able to see the movement of Base Rate, CPI, and FinBERT sentiment over time in one plot.

The team proceeded to investigate how this can be visualised in the most digestible manner, which is outlined below.

The team imported the all_data.csv file which includes both FinBERT polarity scores for the entire dataset, as well as key economic indicators during the time range of the speeches. The FinBERT polarity score, CPI, CPI target, and Base Rate was then plotted onto a line graph over several iterations until the best visualisation format was identified (see Jupyter Notebook 'BaseRate_CPI_Sentiment_Timescale' where each iteration can be seen). See

Figure 13 for the final graph created with the appropriate scaling.

From a high-level perspective, we could not see any obvious trends in FinBERT polarity and base rate, CPI, and FinBERT polarity. A further statistical analysis on correlation was performed via a heatmap in a separate Python notebook.

For the purpose of time-scale trend analysis, we identified a focussed time period where the economy in England was relatively stable, base rate did not fluctuate significantly, however FinBERT polarity was volatile. This was between June 2003 and June 2004.

The evidence of stability can be noted in the 'Overview' page of each of the Financial Stability Reports published during this period:

- June 2003: https://www.bankofengland.co.uk/-/media/boe/files/financial-stability-report/2003/june-2003.pdf
- December 2003: https://www.bankofengland.co.uk/-/media/boe/files/financial-stability-report/2003/december-2003
- June 2004: https://www.bankofengland.co.uk/-/media/boe/files/financial-stability-report/2004/june-2004.pdf

This period was delved into further, the speeches with the most extreme polarities was read through to understand their context and why they drove such a high polarity score. A line graph, showing the four key datapoints (FinBERT polarity score, CPI, CPI target, Base Rate) was plotted for this specific timeframe, see

14.1 Does the sentiment of BoE speeches correlate to key economic indicators?

The economic indicators from Table 2 were explored against BoE speech sentiment. Spanning the entire period, a heatmap was generated to visualize potential correlations with speech sentiment, Figure 8.

Key Observation: Speech sentiment shows a minor positive correlation with FTSE100 and FTSE250 but exhibits a negligible to slightly negative correlation with bonds, base rate, and CPI. The bonds have strong positive correlations with each other, but strongly negative with the FTSEs. This is likely due to investors shifting between asset classes based on risk appetite or economic outlook, causing bonds and stocks to move in opposite directions.

Key Decisions: Following BoE feedback from the kick-off meeting, initial Q&A, and interim presentation:

- Further exploration of FTSE indices, house prices, oil prices, and natural gas prices was discontinued, shifting attention to alternative markets.
- Investigation of how the relationship with speech sentiment varies during a "Lower Bound Regime" (base rate below 1%) and "Hiking Cycle" (continuous base rate increase in 2022) was initiated.

Figure 7.

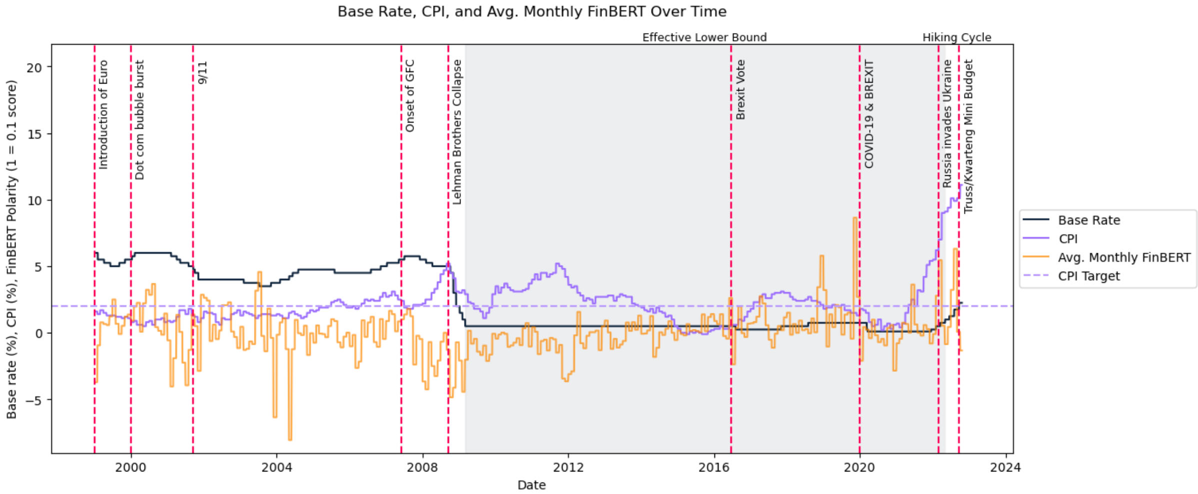
Four speeches were analysed from this specific time period:

- **18 June 2003 - Rt Hon Sir Edward George gave a speech at the At the Lord Mayor's Banquet for Bankers and Merchants of the City of London, Mansion House.**
    - This represented a polarity score of 0.75. Specific highlights from the speech include indications of optimism in the overall global environment, foreign exchange markets, expectation of a strengthening economy, ending with a cheerful toast to the guests.
- **20 November 2003 - Kate Barker, Member of the Monetary Policy Committee, Bank of England, gave a speech titled UK Monetary Policy in a Changing World. The audience was the Teesside Business School Annual Lecture.**
    - This delivered a polarity score of -0.63. Key highlights from this speech were the uncertainty driven from the increasing role of China in the world economy, offshoring of jobs to low-cost economies, the growth of the eurozone and how this will affect the UK. Uncertainty was key to this speech. The setting of the speech (at a business school) may also have had an impact on the content and tone of the speech, compared to the banquet speech mentioned previously.
- **28 April 2004 - Kate Barker, Member of the Monetary Policy Committee, Bank of England gave a speech at CBI Yorkshire and Humber Annual Dinner.**
    - This delivered a polarity score of -0.81. The contents of her speech included the decline in activity and investment in the manufacturing industry, as well as focus on the household debt-to-income ratio being the highest since comparable series started in 1987. This contrasts to the tone of the speech given at a banquet setting in June 2003.
- **14 June 2004 - Mervyn King, Governor of the Bank of England, speech at the CBI Scotland Dinner at the Glasgow Hilton Hotel.**
    - This delivered a polarity score of 0.47. This had a slightly similar theme to Kate Barker's speech in April regarding uncertainty in the industrial sector, however King's speech was more anecdotal, and touched upon what economic environment is most conductive to productivity enhancing change (i.e. innovation). Encouragement to continue the process of lowering barriers to trade.

An interesting observation is that the setting of the speeches varies widely, and will to some degree impact the tonality and quantitative/qualitative nature of the information being delivered in the speeches. The variability of tonality in speeches, and the degree of autonomy the author/speaker may have on the contents of the speech, depending on the audience may also be causing further fluctuations in sentiment, even within a narrow timeframe.

Further analysis, on which types of communication (i.e., speeches, Monetary Policy Report/Financial Stability Report etc) are being picked up and acted upon by who will be an important area of focus for the Bank of England, especially when it comes to communications that are producing materially high/low polarity scores such as these.

Figure 13 – Base Rate, CPI and monthly average speech sentiment over time.



Base Rate, CPI, and Avg. Monthly FinBERT Over Time
Effective Lower Bound
Hiking Cycle
Introduction of Euro
Dot com bubble burst
9/11
Onset of GFC
Lehman Brothers Collapse
Brexit Vote
COVID-19 & BREXIT
Russia invades Ukraine
Truss/Kwarteng Mini Budget
Base rate (%), CPI (%), FinBERT Polarity (1 = 0.1 score)
Date
Base Rate
CPI
Avg. Monthly FinBERT
CPI Target

15 Appendix F: Monetary Policy Regimes

A preliminary finding that Team 7 found from analysis was that there is only a very slight correlation between BoE speech sentiment and some key economic indicators. A comment helpfully provided by the Client (Bank of England) at this point was that it would be interesting to identify whether there are changes in the level of correlation, depending on the monetary policy regime at that time.

The two key monetary policy regimes the team focused on were:

- Base rate nearing the Effective Lower Bound (Between 5th March 2009 - 4th May 2022).
- The BoE implementing a hiking cycle (Between 5th May 2022 - 31st December 2022 (end of our dataset))

As a control, the team also decided to create a heatmap of correlations for the entire dataset period of 1999 - 2022 also to be able to compare and contrast between the three date ranges.

The data source used was the 'all_data.csv' file which included both FinBERT polarity scores for the entire dataset, as well as key economic indicators during the time range of the speeches. The data was imported and subset into the three required time periods as mentioned above.

The created heatmaps can be reviewed in Figure 8, Figure 9, and

Figure 10.

The following observations were made from reviewing the three distinct time periods via a heatmap.

- **Correlation between FinBERT polarity and the FTSE**: Slight correlation can be noted between FinBERT polarity and the FSTE 100 and 250 across all three heatmaps, with the highest scores during the Effective Lower Bound period, followed by the Hiking Cycle. It is interesting to note that the correlation scores are slightly higher during these two periods where base rates are at either side of extremes.
- **Correlation between the base rate and bond prices**: Correlation between the base rate and bond prices (1 month, 2 year and 10 year) is extremely visible, with the strongest correlations seen during the hiking cycle. This is then followed by the heatmap for the entire time period between 1999 - 2022, and lastly by the period of the Effective Lower Bound, where strong correlation can still be seen for the 1-month bond prices.
- **Correlation between CPI and bond prices**: A mild correlation between CPI and bond prices was also observed, potentially moving hand in hand with base rate movements.
- **Correlation between CPI and base rate**: Interestingly, in the heatmap during the hiking cycle period, there appears to be a relatively strong correlation between the base rate and CPI. This may be due to the fact that base rates were being pushed upwards to tame the rising inflation rate.

By splitting the time periods into regimes and running separate heatmaps, it was interesting to note both similarities as well as differences in correlation in each. It would be recommended that the Client further explores this method of analysis if not already done, in order to identify any trends that can help the Bank understand and predict market behaviour.

16 Appendix G: Economic indicators and speech sentiment

16.1 Data collection

The four economic indicators outlined by the BoE questions were collated from the ONS website. Those were CPI, GDP, Unemployment and Wage growth. CPIH which is CPI including owner occupier's housing cost was disregarded due to CPI being the flagship figure the BoE uses. RPI (Retail price index) was initially collated but dropped after further research revealed that it lost its status as a National Statistic in 2013.

As additional economic indicators the following were collated:

- Bonds: 1 month, 2-year and 10-year bonds were extracted to have a broad base of short-, mid- and long-term yields.
- FTSE250: Due to a larger representation of UK businesses.
- House prices: Thoughts were a base rate change carried out by the BoE could create variations in the market.
- The UK base interest rate: This is their direct lever for controlling CPI so clearly relevant.

The indicators were collated in various time frames due to the nature of their calculations. Most were either daily or monthly, with the exception being GDP which was quarterly.

Originally the dates 1992-2022 were gathered as that is when the speech data, we were given commenced but shortly after it was communicated that the BoE speeches were dated from late 1998 (but with only very few speeches) so it was decided that the data we would analyse would be from 1999-2022.

This meant a bit more cleaning of previous data and additional data collated by other members of the Economic team, as well as adding some additional years to some so also.

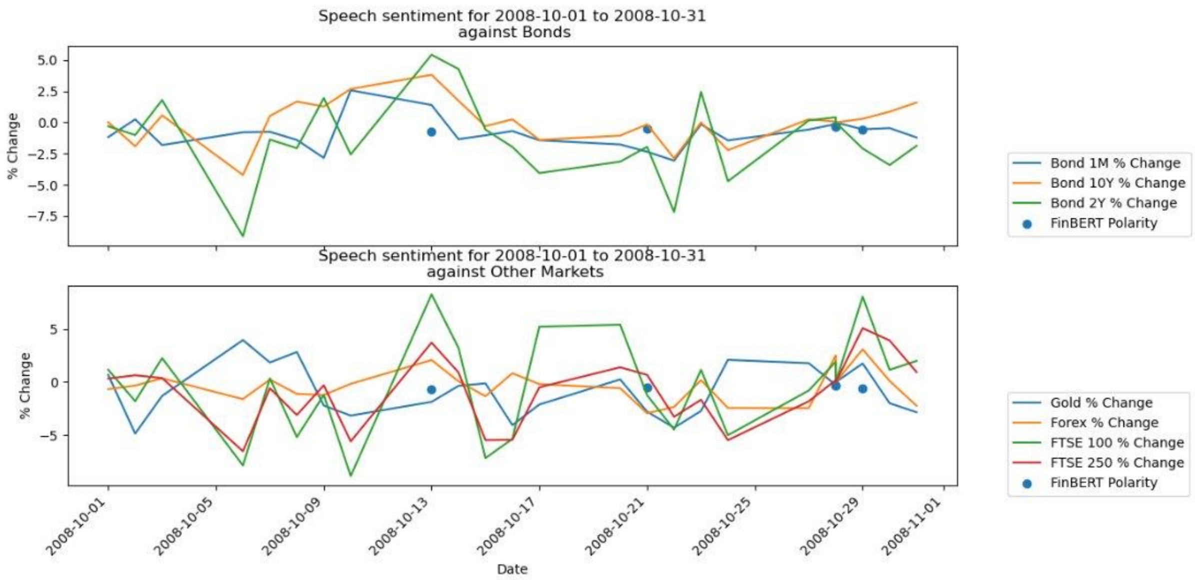
16.2 Analysis of Daily economic indicators.

Because there were differences in time frames of the collated data, the first indicators I investigated were the markets which were all in daily form. These were Gold, 1-month bonds, 2-year bonds, 10-year bonds, FTSE100, FTSE250 and USD Forex. As a better measure to display the variances in the market a new dataframe was created where the percentage change was calculated.

It was decided by the Economic team to initially to cover the GFC period from May '07 until Dec'08, subsequently this was broadened until Jan '10. With daily data and a 2.5-year time frame at the initial look, the indicators were lost in a sea of data spikes. An interesting time did stand out as having a large amount of fluctuating market changes, so it was decided to explore the dates Sep '08 until Dec '08 and view how the data was portrayed visually in a shorter period. This showed more definitive spikes but was still too large to make definitive assumptions other than the markets were moving often and dramatically at times.

A glance after this at the speeches during this time showed 15 speeches made with only one being made by the Governor of the bank. The sentiment of these speeches using the 3 NLP models was visualised showing some uncommon low Vader scores within it. The Governors speech being one of the few speeches to score low over all the NLP models in sentiment.

With the 4-month period being still too large for accurate perusal each of these months was then divided separately and visualised with the speech sentiment from each NLP model. It revealed a lot of movement around the dates of speeches but also a lot of movement in the markets on any given day. This plot of the month the Governor Mervyn King made his speech, the lowest scoring of all Governors was used in the presentation with bonds being collated in a different plot to the other markets due to the more volatile swings in bond yields this served to display the swings around the speeches but also that they swing regardless.



This visualization was chosen due to it being a month within the GFC period that contained the lowest scoring Governor speech by Mervyn King. On top of this, with Bonds and the other markets separately shown it created a good portrayal of the changes in the markets not only around the day of the speech but also when there were no speeches at all.

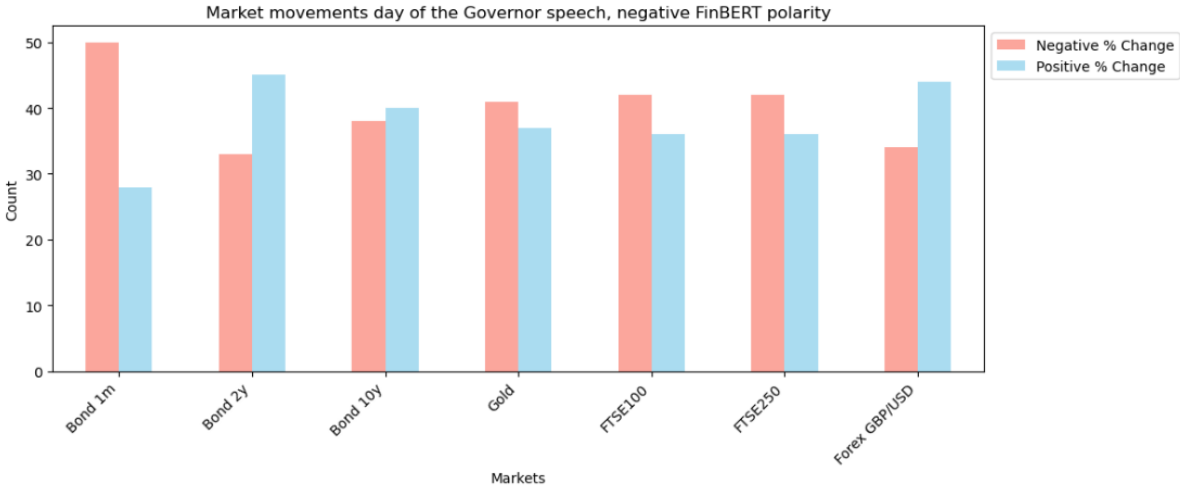
The sole speech made by the Governor was then compared to the markets that week showing large negative moves by the markets following. This led me to investigate speeches from Governors, totally 190 for our data they occur around 8 times a year. Of those only 11 scored solely negative. A product of Vader returning positive sentiment far more often, this was when I started to doubt the value of its results.

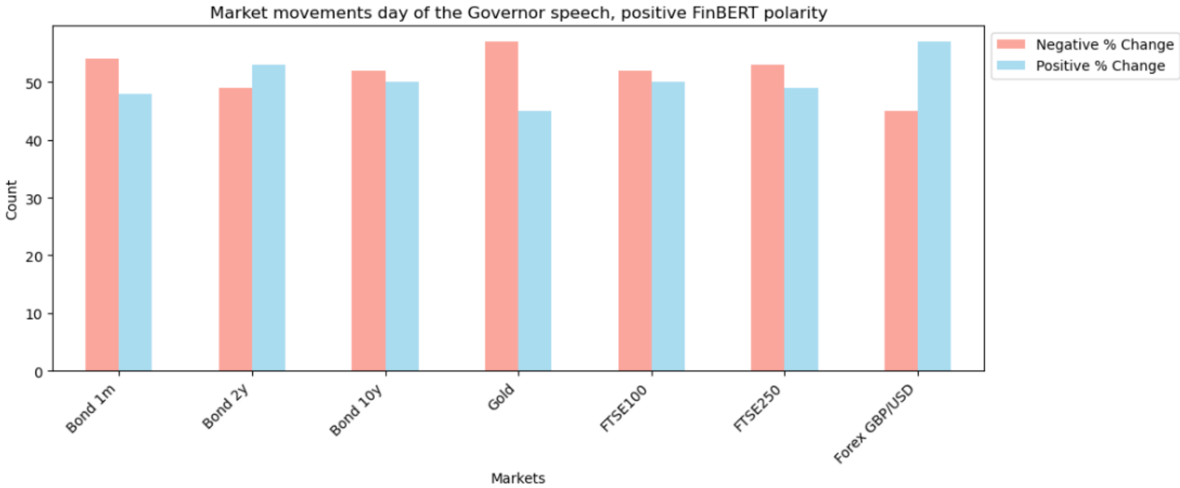
More dates were visualised in a longer time frame and 4 more months were chosen due to their negative Vader scores, again comparing the dates and sentiment of the speeches to the markets. It did seem although there is always movement in the markets at any given time but that during these negative Vader sentiment times chosen it's more likely for the markets to move in a negative direction after a speech.

Having seen some negative points I then wanted to see how the opposite looked and collated the positive speeches, of which there were 45. The markets move frequently again but there is certainly a case to be heard of the markets moving a bit more often in a positive direction than when negative speeches were made. This is certainly an area I would like to delve into further. After this, due to the greater fluctuations in the bonds market, I plotted them separately to the remaining markets to enable a clearer picture.

Data was looked at for all the Governor speeches, namely on the day of the speech and a day either side to see how the movements in the markets looked for negative and positive scoring speeches. This was done in the form of plots showing the percentage change of all the markets on the day of the speech. The top and bottom 40 speeches were looked at as well.

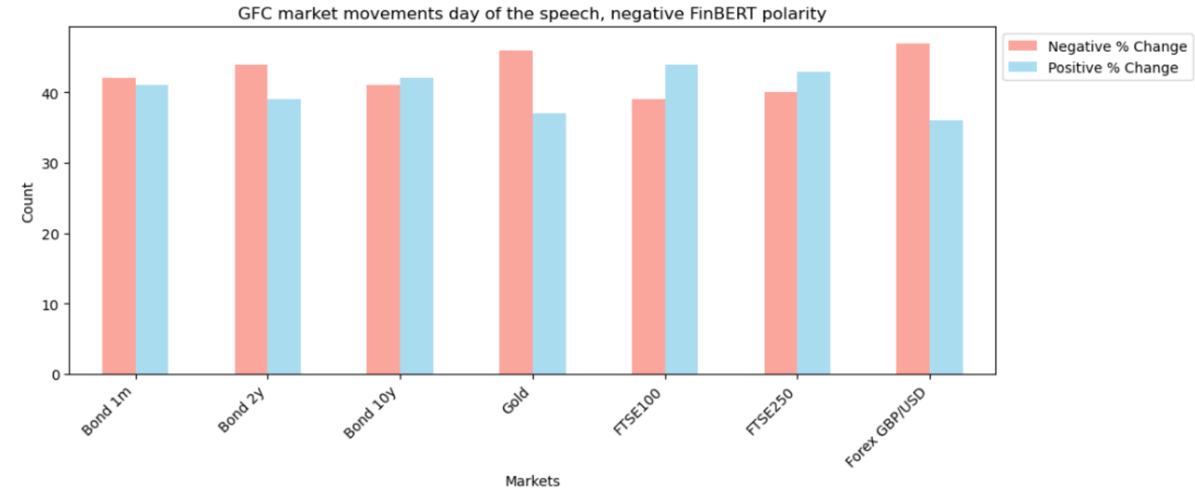
2 plots, the first for the day of Governors speeches and the second for the GFC were chosen for the presentation. Firstly, to show a correlation between bond yields moving substantially more days in the negative during negative speeches which is in line with what could be expected in times of negative sentiment when investors are searching for safe-haven assets.

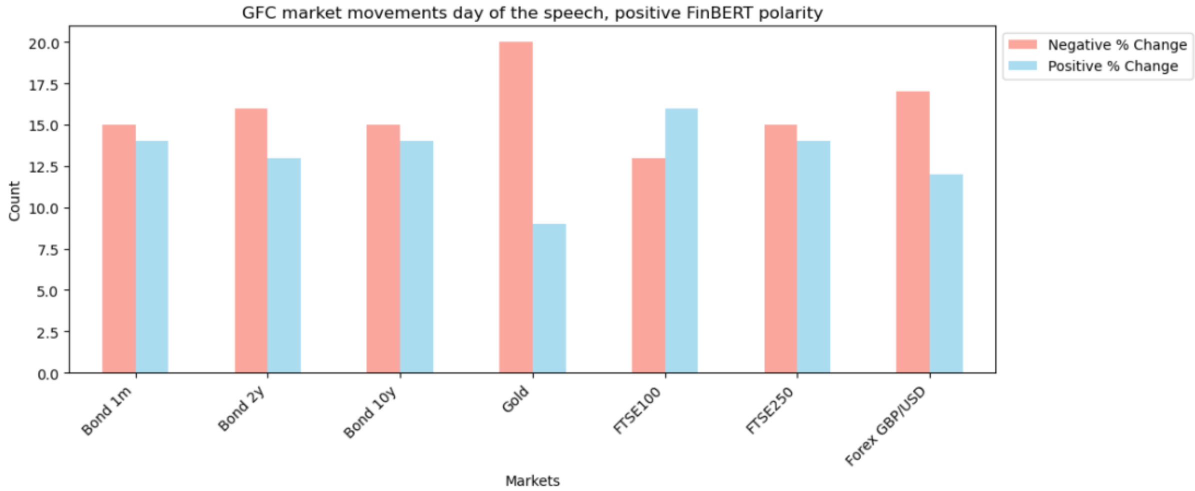




This was chosen to show the movements the day of the Governor speeches as it effectively shows that the day the speeches were made there is an even spread of movement in the positive direction other than Bond yields which is in line with what could be expected in times of negative sentiment when investors are searching for safe-haven assets.

And secondly, in the GFC showing gold prices moving more often in the negative many more days during positive speech sentiment. This movement in gold prices is also what can be expected when the economy is moving out of a negative period and into a positive era.





As with the above visualisation this was chosen as it effectively shows gold prices moving more often in the negative during positive Governor speech days, also what can be expected when the economy is moving out of a negative period and into a positive era. The 2 of these served to show that the majority of the time the speeches have not created significant movements in the markets other than an occasional correlation observed in 1month bonds and gold.

A quick exploratory look was done at individual speeches for both positive and negative sentiment for all speeches as well as the Governors alone, this shows the markets move either way and don't correlate to the speech sentiment.

Lastly, counts were made of the positive and negative changes in the markets on the day of the Governor speeches for both positive and negative polarity as well as a day on either side. Additionally, it was done for all speeches on the day and into time frames including the GFC, Golden Age, Slough of despond and Covid.

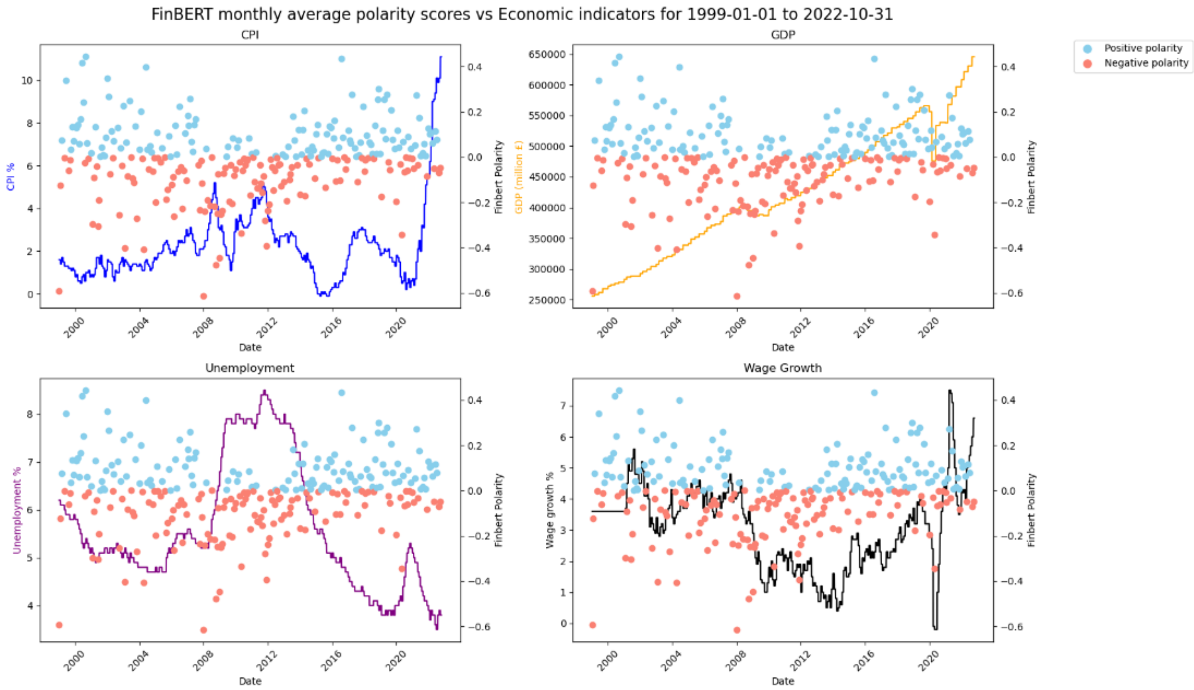
16.3 Analysis of Monthly economic indicators.

After some interesting findings from the daily economic data I then moved onto the monthly indicators, with the BoE highlighting CPI, GDP, Unemployment and Wage growth and wanting to keep things simple then these are the only indicators I chose to base my analysis on, this coincided with the team decision to use FinBERT as the NLP model additionally simplifying the analytical process, so these were my data points for exploring.

After creating a new dataframe with just those values and viewing them all over the GFC time frame we chose. No real pattern appeared at initial viewing between the speeches which varied greatly throughout the period and the economic indicators. CPI and GDP had similar looks and Wage growth and Unemployment worked inversely for each other, which to me is a logical relationship.

The speeches in the dataframe had all been forward filled so in line with accuracy a unique set of the speeches was created and then the highest and lowest 10 FinBERT polarity scores shown in a plot against CPI, which didn't really show up anything interesting. I then plotted them against all 4 indicators and again the spread of negative and positive speeches doesn't seem to change the indicator's movements.

Following this I collated the monthly average of the FinBERT polarity scores and then compared them firstly to our overall time frame in a monthly average for FinBERT from 1999-2022, this plot was used for the presentation as it nicely showed some correlation with low unemployment around the GFC.



This plot was chosen for the economic indicators as it collated the monthly FinBERT averages and displayed the negative months in red and positive in blue scatters along the time frame of our speech data. This helped to display the overall sentiment against CPI, GDP, Unemployment and Wage growth and the correlation between the negative sentiment and low unemployment and wage growth specifically around the GFC period which we had highlighted as a time of interest.

subsequently also the GFC and the 4-month period investigated in the daily markets. Additionally, 2 more periods were created for what we termed in our Economic research timeline as the Golden Age and the Slough of despond. Neither of these time frames added any weight to any conclusion that the speech sentiment had a correlation to the economic indicator's movements.

Plots were then made with the speeches at a granular daily level for the whole GFC period we chose as well as the 4-month periods in 2008 and 2019 that were investigated in the daily market analysis. The only further point of interest was that in the 2008 period when sentiment was negative there were far fewer speeches than in the 2019 timeframe when sentiment was positive. A possible further area to investigate.

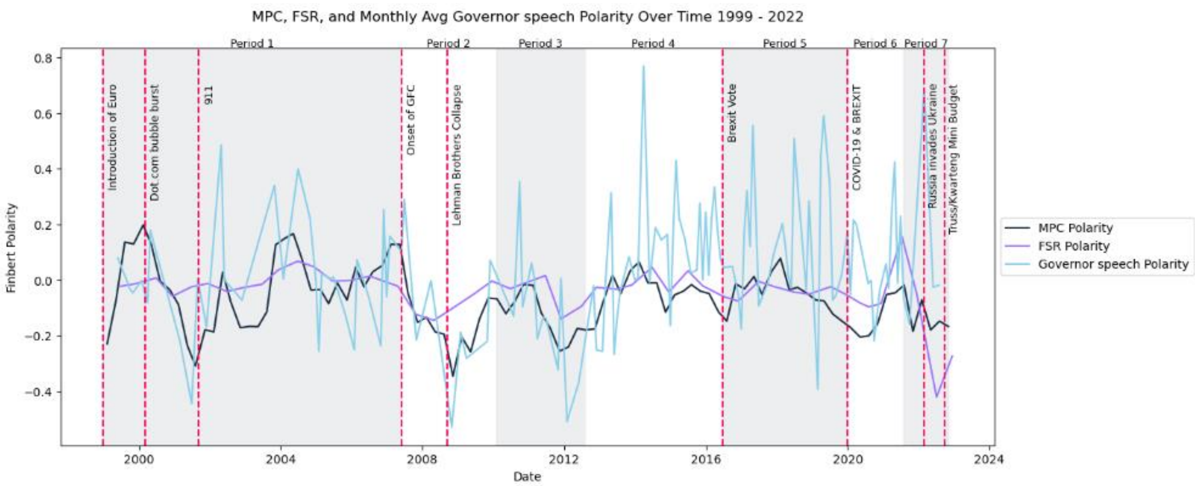
With no visible correlation from the multiple plots viewed over various time frames then my assumption is that the speech sentiment has no bearing on the economic indicators, with the data collated in monthly form and multiple speeches a month then this was not totally unexpected. For further clarification and justification of this assumption then a correlation heatmap was conducted as well as Pearson and Spearman correlation figures were calculated, and they all returned the same conclusion.

16.4 Additional Analysis post Interim Presentation

After FinBERT polarity had been calculated for the MPC and FSR reports then in line with the Governor speech analysis the movements of the markets each day were calculated for the reports either side of zero. This did not result in anything conclusive other than the reports are more often negative with both being around 80% negative.

Coupled with the small number of reports meant that the FSR for example only had 9 positive reports making it difficult to make assumptions considering the multitude of factors that can also affect markets at any given time.

The FinBERT polarity from Governor speeches was collated in a monthly average and then plotted with the MPC and FSR scores over the whole period of data and the plot below was chosen for the presentation.



The plot was chosen as it showed a general alignment between the Governor speeches and the MPC reports. A more neutral tone in the FSR reports. A positive shift in sentiment after Mark Carney started as Governor in July 2013 along with a convergence of sentiment between the reports as well also outlining that sentiment also varies greatly throughout the time shown, even at monthly averages.

Finally, heatmaps and linear regression tests were done to statistically quantify the results from the plots, helping confirm the analysis.

17 Appendix H: Word clouds and word frequency analysis

In a very simple way plotting frequencies and word clouds allows the bank to consider the business question: Do these speeches have any predictive power to assist in predicting market behaviour?

By sorting and filtering positive and negative words we get an idea of which words cause which sentiment and therefore, the opportunity to tailor communications accordingly. This is true for both speeches and the reports.

17.1 Speeches

Initially the top 20 most positive and most negative speeches were found for every model. This was done by sorting the speeches by their respective polarities and filtering by top 20.

Then sentiment word frequencies were explored for each model by setting thresholds. The positive threshold was set to > 0 and negative at < 0. The thresholds were set according to typical limits of positivity and negativity, with the exception of FinBERT. This threshold however is entirely customisable in the Notebook provided.

Following the threshold filtration, the positively and negatively associated speeches are merged into separate strings, ordered by polarity, the words extracted, and their frequencies counted. The words are ranked by frequency to determine the most prevalent in each sentiment category.

Finally, word clouds are generated from the top 30 words in each sentiment pole to visualise the most frequent positive and negative terms. These word clouds provide a graphical representation of the key terms associated with each sentiment in the Bank of England's communications.

Word clouds allow for an immediate visual interpretation of the data and offer useful insights.

Analysis of FinBERT is addressed here and in the notebook as it is our best NLP model.

First, we consider the correct positive and negative threshold to apply in order to filter the sort of the data.

```
speeches_data['finbert_polarity'].describe()
```

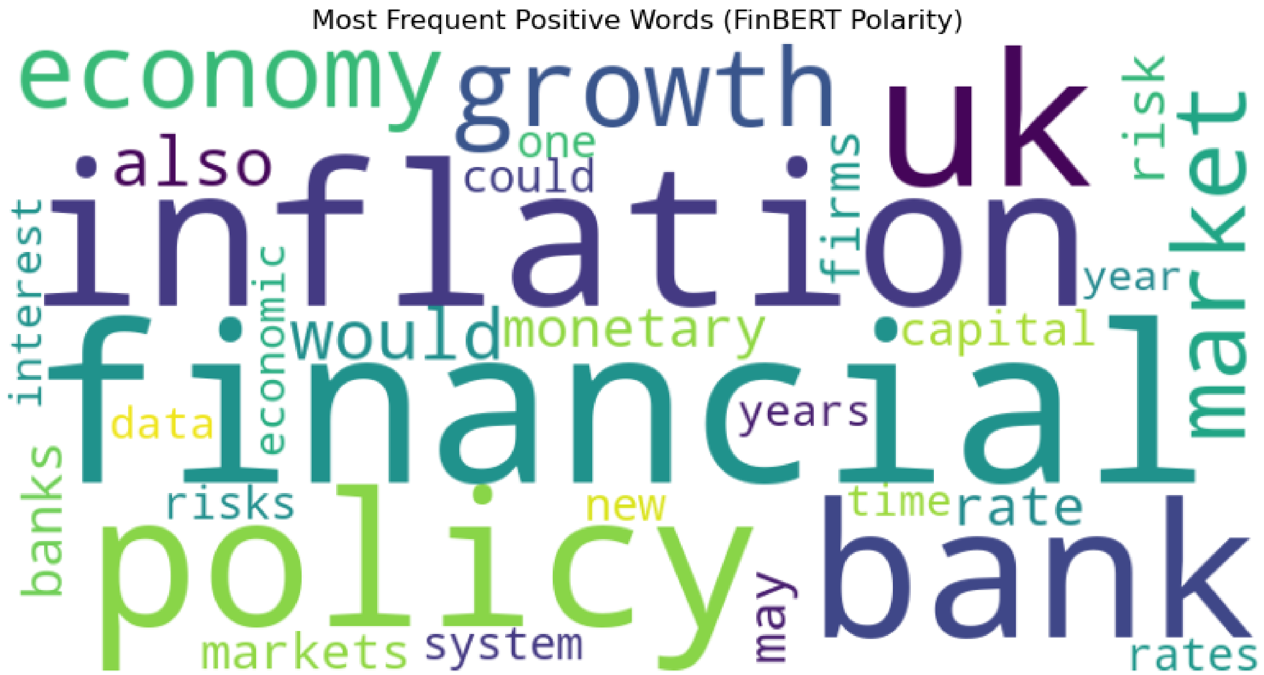
```
count    1209.000000
mean        0.085165
std         0.285261
min        -0.775184
25%         0.000168
50%         0.016790
75%         0.332369
max         0.995363
Name: finbert_polarity, dtype: float64
```

Given the descriptive statistics and the distribution above, it seems correct to set our thresholds around the interquartile range to better capture the more clearly positive or negative sentiments. For instance:

- Negative Threshold: Could be set around the 25% quartile, say at -0, to ensure that only the speeches that are clearly negative are classified as such.
- Positive Threshold: Could be set around the 75% quartile, for example, at 0.30, to classify a speech as positive only if it has a clearly positive sentiment.

Further investigation into FinBERT polarity showed us that, according to the threshold, the total positive speeches was 318 and total negative speeches was 298.

We also discovered that the average length of positive speeches was 1989 words and the average length of negative speeches was 2343 words



Positive word frequencies: [('financial', 5495), ('inflation', 4407), ('policy', 3962), ('bank', 3947), ('uk', 3879), ('growth', 3547), ('economy', 3128), ('market', 2982), ('would', 2936), ('also', 2704), ('monetary', 2567), ('rate', 2549), ('risk', 2503), ('banks', 2394), ('may', 2322), ('firms', 2299), ('markets', 2195), ('capital', 2182), ('one', 2098), ('rates', 2074), ('new', 2045), ('risks', 2034), ('could', 2020), ('system', 1995), ('interest', 1894), ('years', 1887), ('time', 1856), ('economic', 1830), ('data', 1821), ('year', 1735)]

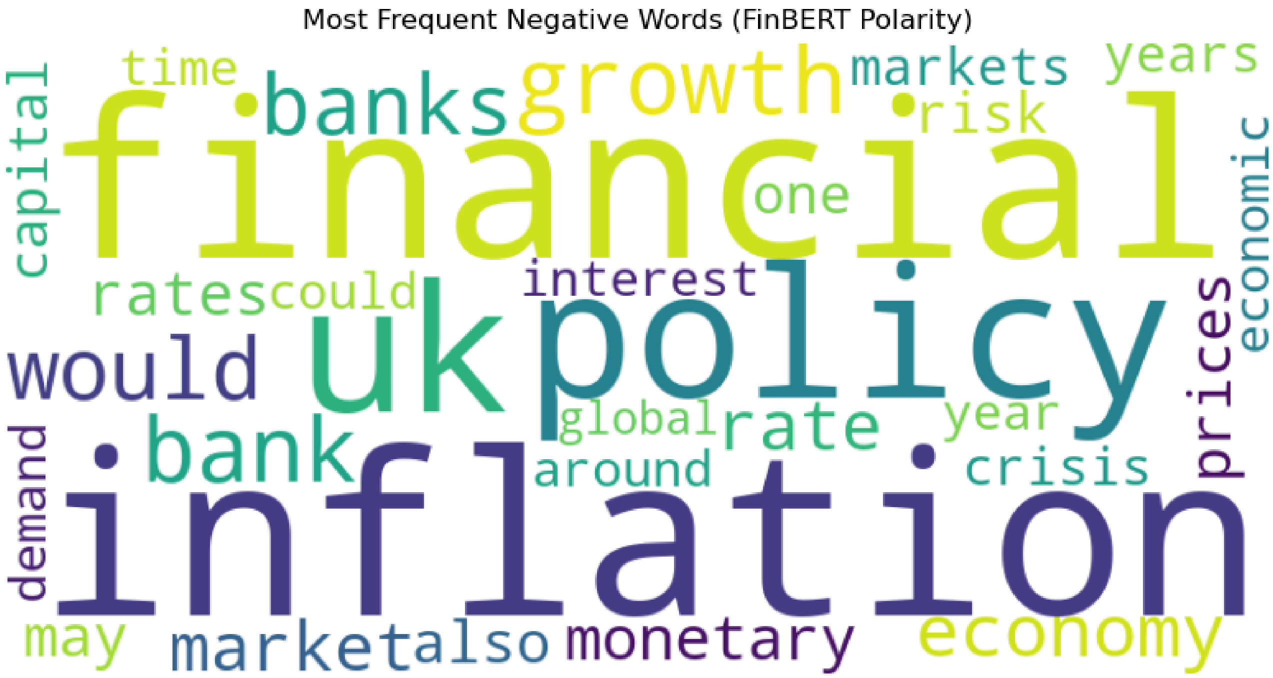
Positive Sentiment Insights

Unique Positive Words

- Risks (2034): Often discussed in a context that emphasises managing or mitigating financial risks, which might be viewed positively as proactive or cautious.
- Firms (2299): Likely refers to businesses and their activities or performance, which in a positive sentiment could relate to growth, innovation, or contribution to the economy.
- New (2045): Indicates forward-looking or innovative aspects, such as new policies, technologies, or initiatives, seen as beneficial or promising.
- System (1995): Could relate to the financial system or specific systems within banking and finance, suggesting stability, robustness, or efficiency.
- Data (1821): Points to the use of data in making informed decisions or analysis, highlighting transparency, accuracy, or advancement in banking practices.

More frequent positive words (>=500 occurrences compared to negative):

- There weren't any.



***Negative word frequencies**:* [('inflation', 5438), ('financial', 5071), ('policy', 5061), ('uk', 4884), ('growth', 4630), ('bank', 4538), ('would', 4088), ('banks', 4076), ('economy', 3975), ('rate', 3404), ('market', 3179), ('monetary', 3153), ('rates', 2967), ('may', 2801), ('prices', 2795), ('also', 2790), ('risk', 2474), ('one', 2451), ('capital', 2373), ('crisis', 2245), ('years', 2236), ('markets', 2210), ('interest', 2205), ('could', 2170), ('around', 2107), ('year', 2095), ('time', 2090), ('economic', 2074), ('demand', 2062), ('global', 2026)] ('interest', 1459), ('demand', 1454), ('years', 1451), ('credit', 1424)]

Unique Negative Words

- Prices (2795): Could be associated with inflation or rising costs, leading to negative sentiment regarding economic pressure or living costs.
- Global (2026): Refers to global challenges, crises, or economic conditions that negatively impact the UK or the global economy.
- Crisis (2245): Directly related to financial crises, uncertainties, or significant negative economic events. Which is appropriate given the tumultuous period of the data.
- Demand (2062): In a negative context, could relate to decreasing demand, economic slowdowns, or challenges in meeting demand.
- Around (2107): Usage might be in contexts discussing uncertainties or concerns surrounding various economic factors.

More frequent negative words (>=500) The numbers are the amounts by which these words appear more frequently in the negative section compared to the positive section

- Would (1152): Often used in speculating about negative outcomes or uncertainties, particularly in discussing adverse future scenarios.
- Inflation (1031): Central in discussions about the negative impacts on purchasing power, increased cost of living, and overall economic pressures.
- Policy (1099): Refers to government or monetary policies that may be viewed unfavourably or blamed for adverse economic conditions.

- UK (1005): Concerns specifically impacting the United Kingdom, such as Brexit, economic policies, or national financial challenges.
- Growth (1083): Could be discussed in terms of slowdowns, reduced forecasts, or stagnant conditions that affect economic sentiments negatively.
- Banks (1682): Discussions likely focus on challenges, failures, or issues within the banking institutions.
- Economy (847): Refers to downturns, recessions, or overall negative health of the economy.
- Rates (893): Discusses the broader implications of changes in interest rates on mortgages, loans, and economic growth.
- Rate (855): Involves interest rates, particularly increases which may suggest tightening monetary policy or concerns about inflation.
- Bank (591): Focuses on specific problems associated with banks or the banking sector as a whole.
- Monetary (586): Involves discussions on monetary policy that in a negative context could relate to criticisms or constraints of current policy measures.

Recommendations:

Positive Sentiment Triggers

- Proactivity and Management: Words like "risks," "system," and "data" are associated with proactive management and positive actions. These suggest efforts to ensure stability and thoughtful, informed decision-making.
- Innovation and Progress: Terms such as "new" and "firms" highlight forward-thinking, innovation, and positive contributions to the economy. They evoke a sense of progress and development.

Negative Sentiment Triggers

- Economic Challenges: Words like "inflation," "crisis," and "rates" often appear in contexts discussing economic pressures and challenges, which naturally tend to generate negative sentiments.
- Speculative and Uncertain Language: Use of words like "would" and "could" can introduce uncertainty and speculation, often associated with less positive or anxious sentiments.
- Banking and Policy Concerns: Terms such as "banks," "policy," and "monetary" can evoke negative sentiments when linked to problems or criticisms of banking practices and economic policies.

Final observation:

- Addressing Sentiment Imbalance: The lesser frequency of positive terms suggests a potential negative skew in communication. Increasing the use of positive words can help foster a more optimistic public perception, balancing the narrative around economic conditions.

17.2 FSR and MPC reports

We now have a justifiable analytical framework and NLP model that can be used to find correlation between sentiment scores and indicators over time.

In relation to the business questions provided, the data shows us that there are not many significant correlations between Bank of England speeches from 1999 – 2022 and economic indicators. Instead, we see a level of neutrality in sentiment scores over time consistent with a central bank’s expected communication style. When polarity does shift significantly, this is in reaction and acknowledgement to world events substantial enough to force the Bank from its intended neutral position and into a more positive or negative sentiment.

In spite of this, the analytical approach still remains useful and have applications beyond the business questions provided. For instance, you could apply the same techniques to other central banks to compare communication styles and any potential economic correlations as a result. However, as we have seen, speeches can lack specificity:

- **Addressing countries other than the UK** (19/03/2002 "*Asian Business Association Dinner*" by George
- **Focused on the future** (02/03/2018, "*The Future of Money*" by Carney
- **Or the past (03/06/2001**, "*The Beveridge Curve, Unemployment and Wages in the OECD from the 1960s to the 1990s*" by Nickell)

These are exampling all elements that could skew polarity scores of the relative present in the UK.

This led us to consider Bank of England communications other than speeches, communications like the FSR or MPC reports, which are necessarily geographically constrained to the UK and economically constrained to current relevant justifications. In particular the MPC explains The Bank's decisions of its main measure of control, the setting of interest rates.

These more localised and time-specific sources may offer a clearer reflection of the Bank's policy decisions and their impact on economic indicators. Through this, we potentially refine our understanding of the relationship between central bank discourse, sentiment, and economic health.

What are the MPC and FSR?

The Monetary Policy Report

The Monetary Policy Report sets out the economic analysis and inflation projections that the Monetary Policy Committee uses to make its interest rate decisions. Previously called the inflation report, which first published in February 1993, with the objective "to produce a wholly objective and comprehensive analysis of inflationary trends and pressures".

Typically, of the form:

1: The economic outlook, including underlying conditions and assumptions

2: Current global and domestic economic conditions, with varying items of focus, such as Wage growth and inflation, Inflation expectations

Published by the bank of England, quarterly since Nov 2019. Feb, Nov, May, Aug. and previously on the same schedule as the "Inflation Report" since February 1993.

The Financial Stability Report

Financial Stability report. First published in October 1996 as the Financial Stability Review, this was renamed in 2006 to the Financial Stability Report to reflect the change in content and aims.

The Financial Stability Report sets out the Financial Policy Committee's view on the stability of the UK financial system and what it is doing to remove or reduce any risks to it.

Form varies considerably over time – initially quite a lot of the reports appear to be qualitative, but more recently there is a great degree of transparency, with attached data files.

It is published twice a year, generally in July and December.

Exploring the reports in python: analytics and word clouds

We imported and cleaned the data to the appropriate extent that FinBERT requires, namely:

- **Ensure text is in English**: We have tagged each speech for their language.
- **Handle missing values**: There weren't any.
- **Standardise text data**: FinBERT works well if the text is standardised (i.e. ensure consistent use of quotation marks etc). We will take the assumption that the use of quotation marks etc has been applied consistently.
- **Remove unnecessary elements**: It is good practice to remove URLs and other specific mark ups. However, in the same vein, it is important not to over clean.
- **Ensure Data is in a Suitable Format**: The data is contained in a dataframe column.

Then we run the normal sentiment analysis, which can be followed as before in the notebook.

The decision was made to not merge the two dataframes in any way. By avoiding joining/concatenation, we prevent the presence of NaNs and the subsequent need to address them. By conducting NLP analysis individually for each dataset, we maintain efficiency while retaining the option to combine the scores into one DataFrame at the end should we need to.

Let us now consider the word cloud output from the models.

First, we consider the correct positive and negative threshold to apply in order to filter the sort the data. The process is the same as it was for the main body of speeches so will not be detailed here again.

	fsr_finbert_polarity
count	48.000000
mean	-0.041992
std	0.085078
min	-0.421359
25%	-0.059071
50%	-0.025743
75%	-0.005050
max	0.157170

- Average length of positive FSR reports: 63910.72 words
- Average length of negative FSR reports: 56645.51 words
- Total number of positive FSR reports: 36
- Total number of negative FSR reports: 39

Given the descriptive statistics and the distribution above, it seems correct to set our thresholds around the interquartile range to better capture the more clearly positive or negative sentiments. For instance:

- Negative Threshold: Could be set around the 25% quartile, say at -0.059, to ensure that only the speeches that are clearly negative are classified as such.
- Positive Threshold: Could be set around the 75% quartile, for example, at 0, to classify a speech as positive only if it has a clearly positive sentiment.

Most Frequent FSR Positive Words (FinBERT Polarity)

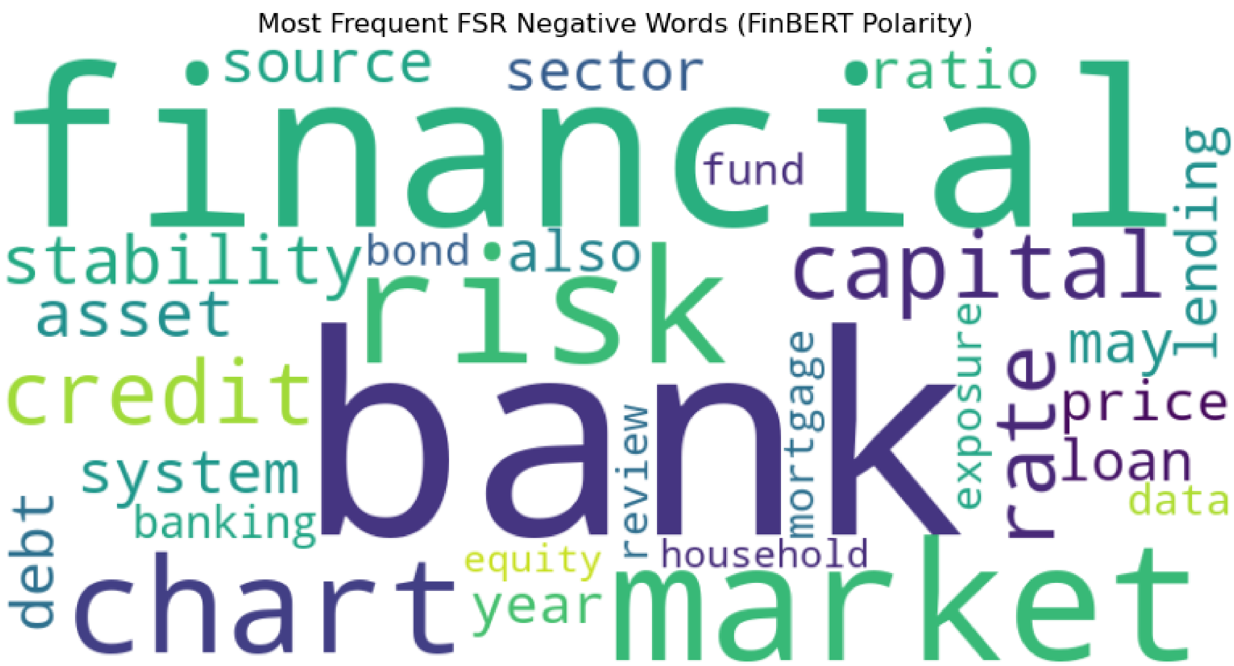
[('bank', 31499), ('financial', 19218), ('market', 14202), ('risk', 14033), ('chart', 9338), ('capital', 8500), ('stability', 8296), ('rate', 7842), ('credit', 7788), ('asset', 7059), ('system', 6791), ('source', 6140), ('may', 6045), ('debt', 5636), ('sector', 5481), ('review', 5298), ('loan', 5040), ('also', 5034), ('lending', 5009), ('year', 4957), ('banking', 4776), ('price', 4697), ('ratio', 4541), ('per', 4427), ('exposure', 4254), ('data', 4103), ('equity', 4080), ('bond', 4070), ('cent', 4018), ('mortgage', 4007)]

Unique Words:

- **Per** (4,427): Likely used in contexts discussing percentages or specific measurements, which may carry a neutral or slightly positive connotation.
- **Cent** (4,018): Similar to "per," often part of financial data presentation, implying stability or precise information.

More Frequent Words:

- **Financial** (13,998): Central to FSR, often associated with stability and positive developments.
- **Risk** (848): Surprisingly, appears more in positive contexts, possibly discussed in terms of risk management or mitigation successes.
- **Stability** (949): Directly correlates with the core aims of financial oversight and assurance.
- **System** (577): References to systemic aspects of banking, suggesting discussions on robustness and reliability.
- **May** (526): Indicates speculative or potential positive outcomes or scenarios.
- **Review** (1,210): Implicates thorough examination and control processes, reassuring stakeholders.



[('bank', 31202), ('financial', 17820), ('market', 14267), ('risk', 13185), ('chart', 8854), ('capital', 8677), ('credit', 7806), ('rate', 7526), ('stability', 7347), ('asset', 7168), ('system', 6214), ('source', 5807), ('may', 5519), ('debt', 5389), ('sector', 4996), ('lending', 4913), ('loan', 4878), ('also', 4872), ('price', 4677), ('ratio', 4605), ('year', 4529), ('banking', 4521), ('mortgage', 4168), ('data', 4153), ('fund', 4124), ('review', 4088), ('bond', 4013), ('exposure', 3961), ('household', 3890), ('equity', 3885)]

Unique Words:

- **Fund** (4,124): May relate to specific financial instruments or reserves, possibly discussed in contexts of insufficiency or stress.
- **Household** (3,890): Typically associated with vulnerabilities or stress in consumer finance segments.

Insights on Sentiment Themes

Positive Insights:

- Emphasise further analysis and discussions on financial stability, risk management strategies, and systematic strengths to maintain positive sentiments.

Negative Insights:

- Address issues related to funds management and household financial stability to alleviate negative sentiments.

The Monetary Policy Report:

First, we consider the correct positive and negative threshold to apply in order to filter the sort the data. The process is the same as it was for the main body of speeches so will not be detailed here again.

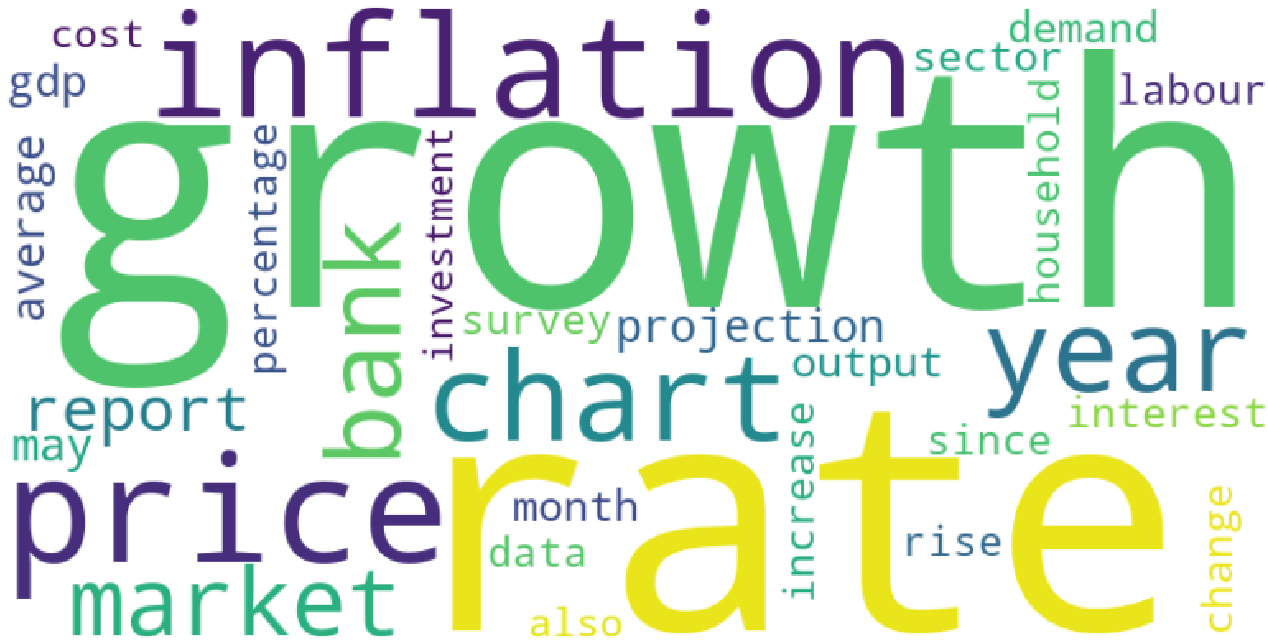
	mpc_finbert_polarity
count	96.000000
mean	-0.071532
std	0.111002
min	-0.346003
25%	-0.167058
50%	-0.069913
75%	-0.013296
max	0.197938

- Average length of positive reports: 36842.00 words
- Average length of negative reports: 33596.95 words
- Total number of positive reports: 20
- Total number of negative reports: 22

Given the descriptive statistics and the distribution above, it seems correct to set our thresholds around the interquartile range to better capture the more clearly positive or negative sentiments. For instance:

- Negative Threshold: Could be set around the 25% quartile, say at -0.17, to ensure that only the speeches that are clearly negative are classified as such.
- Positive Threshold: Could be set around the 75% quartile, for example, at 0, to classify a speech as positive only if it has a clearly positive sentiment.

Most Frequent MPC Positive Words (FinBERT Polarity)



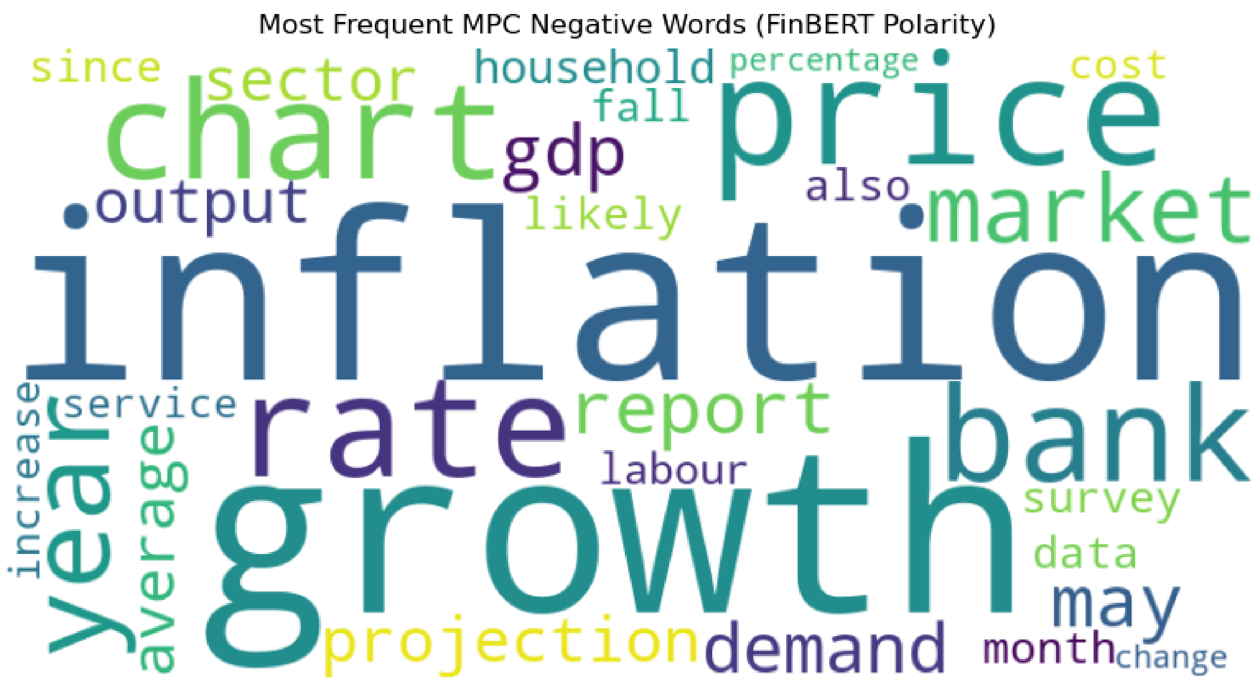
[('growth', 7173), ('rate', 6456), ('price', 6353), ('inflation', 6193), ('chart', 4401), ('year', 3777), ('bank', 3069), ('market', 2892), ('report', 2722), ('projection', 2449), ('may', 2372), ('average', 2347), ('gdp', 2227), ('demand', 2146), ('sector', 2078), ('interest', 2049), ('month', 2031), ('output', 1994), ('percentage', 1904), ('since', 1883), ('change', 1882), ('survey', 1828), ('household', 1817), ('data', 1786), ('labour', 1747), ('increase', 1744), ('rise', 1741), ('investment', 1664), ('also', 1611), ('cost', 1560)]

Unique Words

- Interest (2049): Likely associated with positive financial outcomes or rewarding investment returns.
- Investment (1664): Suggests forward-looking, growth-oriented financial activities.
- Rise (1741): Typically refers to growth in economic indicators like prices or rates, perceived positively...

More Frequent Words

- Growth (7173): Reflects economic expansion, often seen as a primary indicator of healthy economic conditions.
- Rate (6456): Can be tied to interest rates, where a favourable rate is indicative of good economic management.
- Price (6353): In a positive context, may indicate stability or reasonable inflation levels.
- Year (3777): Often used in discussions of annual progress or achievements.
- Interest (2049): Also appears as a unique word; its frequency underlines its importance in financial discussions.
- Rise (1741): As a more frequent word, it underscores positive growth trends.
- Investment (1664): Emphasizes proactive financial strategies and economic optimism.



[('inflation', 5862), ('growth', 5850), ('price', 5804), ('rate', 4978), ('chart', 4634), ('bank', 4425), ('year', 3147), ('market', 2804), ('report', 2656), ('may', 2488), ('gdp', 2431), ('demand', 2357), ('projection', 2180), ('average', 2125), ('output', 2106), ('sector', 2033), ('household', 2019), ('survey', 1928), ('data', 1904), ('also', 1873), ('likely', 1872), ('since', 1830), ('month', 1826), ('labour', 1711), ('fall', 1707), ('service', 1593), ('cost', 1582), ('increase', 1576), ('percentage', 1564), ('change', 1526)]

Negative Section

Unique Words

- Fall (1707): Often associated with declines in economic indicators, viewed negatively.
- Likely (1872): In negative contexts, may refer to the probable occurrence of undesirable events.
- Service (1593): Can imply issues in service industries or burdens caused by service sector problems.

More Frequent Words

- Bank (4425): In negative contexts, could be associated with failures, bailouts, or crises.
- Likely (1872): Reflects the anticipation or risk of negative developments.
- Fall (1707): Reinforces concerns about economic downturns or falling markets.
- Service (1593): If frequent in negative contexts, may highlight systemic issues within service-oriented sectors.

Insights on Sentiment Themes

Positive Economic Activity

- Positive sentiment in the Bank of England's MPC reports is often linked to terms that convey economic strength and forward-looking optimism. Words like "growth," "investment," and "rise" reflect a focus on economic expansion and robust financial health. These terms are not only frequent but convey a narrative of proactive financial management and the anticipation of favourable economic outcomes.

Economic Challenges and Risks

- Negative sentiment revolves around terms associated with economic setbacks or uncertainties. "Fall," "bank," and "likely" frequently appear in contexts that discuss declines, potential financial instability, or adverse economic conditions. These words highlight areas of concern within the economy, focusing on risks and the potential for negative outcomes, which could influence public and investor perceptions negatively.