Bank of England

Speeches Analysis

Final Report

B.I.U Team:

Mohammed Ahmed

Cecilia Bebeacua

Angel Casillas Joya

Francesco Giannone

Noah McQuinn Farrand

Oliver Megran

November 2025

Abstract

This study investigates whether Bank of England communication from 1998 to 2022 influenced UK financial markets and economic indicators. Using 1,209 speeches and transformer-based sentiment models, we quantify tone, classify topics via LDA, and link sentiment to equities, gilt yields, volatility, CPI, and GDP through hierarchical regression. Results show that sentiment adds meaningful explanatory power—most strongly for volatility—and is far more influential during crises, particularly the Global Financial Crisis, when policy space was greatest. Topic analysis reveals clear shifts in communication priorities across crises. Limitations include causality, measurement error, and sample constraints. Overall, findings suggest that central bank communication shapes expectations most effectively when clarity and credible policy capacity align.

Abbreviations

BoE = Bank of England

BERT = Bidirectional Encoder Representations from Transformers

ECB = European Central Bank

ESG = Environmental, Social, and Governance

FOMC = Federal Open Market Committee

GFC = Global Financial Crisis

LDA = Latent Dirichlet Allocation

NLP = Natural Language Processing

ONS = Office for National Statistics

Table of Contents

Technical Report

Bank of England Speech Sentiment Analysis and Financial Market Impact

1. Background and Context of the Business Problem

The Bank of England plays a central role in shaping expectations across UK financial markets, using communication and forward guidance as key tools to support monetary and financial stability. Yet these speeches are delivered across shifting economic environments, from the high-rate world of the late 1990s to the near-zero-rate eras of Brexit and COVID. This project investigates whether the tone of BoE speeches from 1998 to 2022 can meaningfully explain or anticipate movements in financial indicators and economic conditions, and whether communication remains effective when policy space is constrained. By analysing sentiment across major crises, we examine how tone aligns with policy action, how it shapes expectations, and ultimately, whether the Bank's words retain their influence when traditional levers lose their power. Clarifying when communication is effective enables policymakers and market participants to better anticipate how guidance will translate into real market behaviour.

---

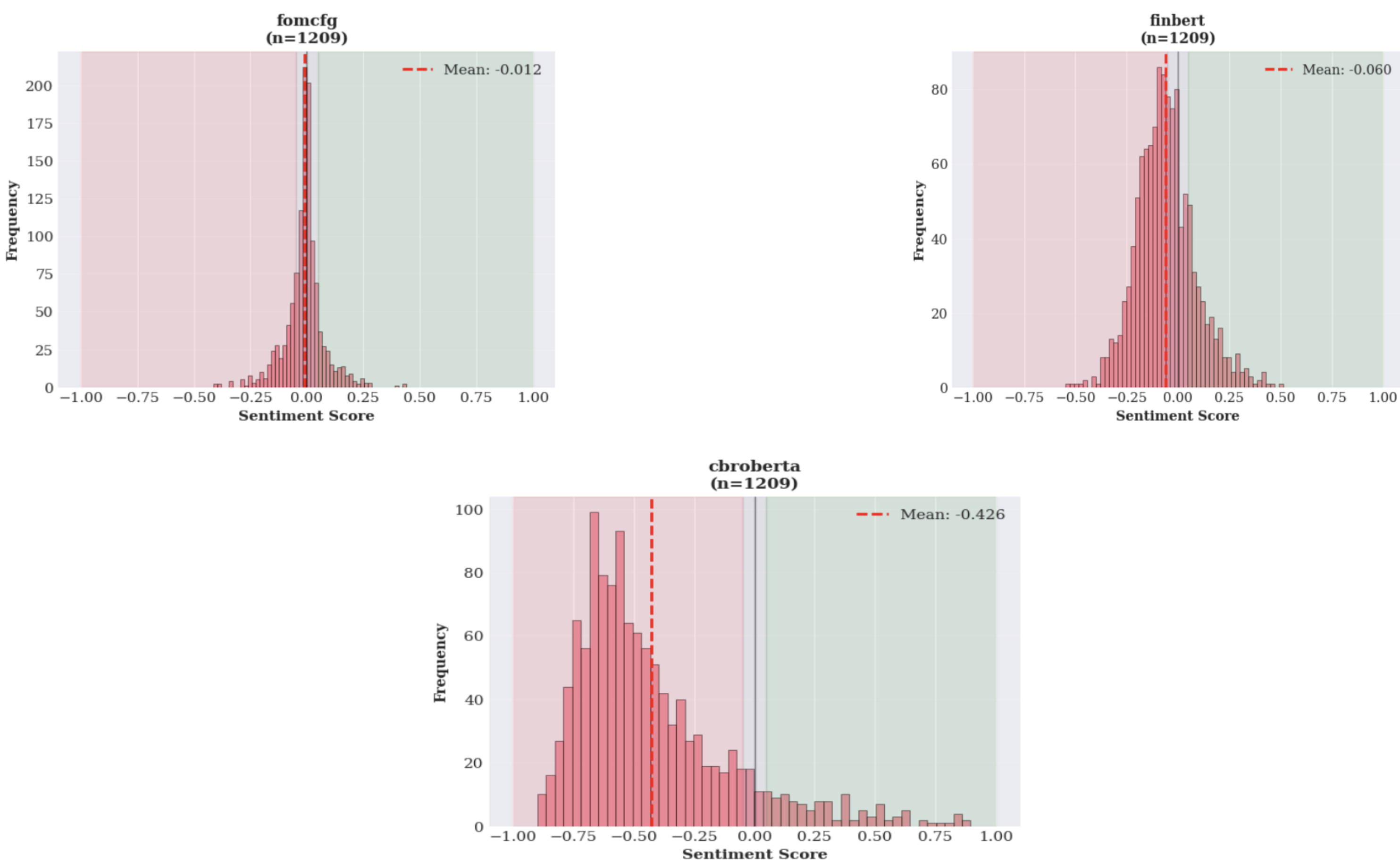
2. Project Development Process

The project began with 7,721 world central bank speeches, from which 1,209 Bank of England speeches (1998–2022) were selected. Early data exploration revealed uneven spacing between speeches and multiple speeches delivered on the same day. These were aggregated to build a clean, daily-level sentiment structure suitable for linking with financial and macroeconomic indicators.

Financial market data was retrieved via Yahoo Finance (FTSE100, gold, VIX, gilt yields), while CPI and GDP series were collected from the ONS. Cleaning and harmonisation involved standardising dates, removing duplicates, resolving missing values, and aligning datasets operating at different frequencies. As markets behave forward-looking, sentiment was matched to future windows (7–30 days, 6–18 months, 2–5 years). Speeches with insufficient future data were excluded to ensure statistical integrity.

Initial trials highlighted the limitations of lexicon approaches: extremely low token coverage and strong polarity bias. These shortcomings motivated the shift to transformer-based models. Three BERT-style architectures were tested, and FOMC-RoBERTa was selected in light of its balanced, economically coherent sentiment that aligned with recognised tightening and easing cycles. This ensured that the sentiment measure captured communication tone rather than artefacts of dictionary design.

Figure 1: Per model sentiment distribution. FOMC yields the most balanced results.



Because BoE speeches often exceed 5,000 words, a sentence-based 300-token chunking strategy was adopted heuristically. This approach preserved semantic coherence while maintaining computational efficiency. Chunks containing monetary-policy vocabulary received additional weight during aggregation, ensuring that sentiment reflected the tone of policy-relevant passages rather than peripheral remarks.

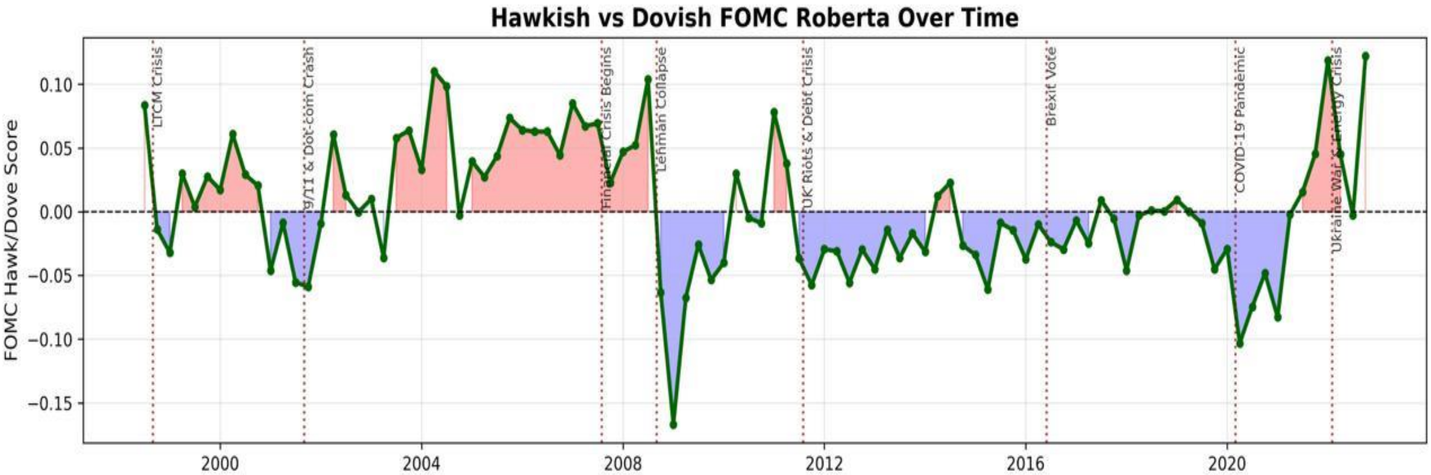
Topic modelling added a further analytical layer. A 20-topic LDA model was trained and subsequently grouped into seven macro-themes such as Monetary Policy, Household Debt, ESG, and Global Economy. This classification allowed systematic comparisons across crises, revealing not just how the Bank spoke, but what it chose to emphasise.

To evaluate the relationship between sentiment and market behaviour, hierarchical regression was used. A baseline model containing only time controls was compared with a full model incorporating sentiment; incremental R^2 represented sentiment's independent contribution. Applying this structure across crisis periods, topics, and volatility groups allowed the analysis to isolate the contexts in which communication meaningfully influenced markets.

Visualisation choices were made to emphasise clarity and accessibility. Trend lines illustrated long-term sentiment evolution, bar charts displayed differences in incremental R^2 across crises, and heatmaps showed shifts in topic prevalence. Crisis periods were colour-coded coherently and all charts used colour-blind-safe palettes. Sample sizes were displayed to ensure transparency and avoid over-interpretation of smaller subsets.

These methodological choices ensured that sentiment was measured in a way that reflects genuine policy communication rather than noise, directly addressing the uncertainty around whether communication can be treated as a reliable policy tool.

Figure 2: Hawkish vs. dovish sentiment lifecycle with crisis markers.



---

3. Technical Overview of Code Implementation

Python was selected as the primary environment due to its rich ecosystem for data analysis, NLP, and statistical modelling. Core libraries included *pandas* for data transformation, *NLTK* for preprocessing, *yfinance* for market data retrieval, *scikit-learn* for topic modelling and regressions, and the *Hugging Face Transformers* library for implementing the RoBERTa-based sentiment model.

The codebase followed a modular design, with distinct components for ingestion, preprocessing, sentiment modelling, topic modelling, and regression. This structure improved maintainability and ensured that changes in one stage did not propagate unintended effects into another.

The sentiment pipeline relied on RoBERTa, with each speech divided into 300-token chunks using the AutoTokenizer. This length was chosen after empirical evaluation showed it captured coherent semantic units while generating enough chunks to represent the full speech. Aggregation used weighted means to prioritise policy-relevant segments. This helped prevent long speeches with extended introductions or ceremonial content from diluting the underlying policy tone.

Topic modelling was implemented using *scikit-learn*'s LDA with 20 topics. This configuration achieved the best balance between thematic separation and interpretability. Topics were then grouped into seven macro-categories for downstream analysis.

The regression framework used *scikit-learn*'s LinearRegression due to its interpretability and efficiency. Hierarchical models allowed clear attribution of variance explained by sentiment beyond time trends. Custom code for incremental R^2 and associated F-tests preserved analytical transparency and avoided reliance on black-box approaches.

Lagged sentiment matching required bespoke rolling logic with "regime-aware" boundaries to prevent lookahead contamination across crisis periods. Defensive programming (skip conditions, error handling, and logging) ensured consistent behaviour even when data was sparse.

Overall, the technical implementation balanced sophistication with clarity, creating a transparent framework that is reproducible, analytically rigorous, and easily extendable to other central banks, periods, or future research questions.

---

4. Patterns, Trends, and Key Insights

Communication tone exhibited distinct patterns across crises. During the Global Financial Crisis, sentiment deteriorated sharply in line with escalating liquidity stress, and only stabilised once large-scale policy interventions were implemented.

Brexit displayed a notably steadier tone despite high market volatility, reflecting the Bank's emphasis on reassurance in a politically-driven shock.

COVID-19 showed an initial sharp drop in sentiment, followed by rapid stabilisation once fiscal and monetary support packages were announced.

Together, these patterns demonstrate that the nature of the shock, the communication strategy, and macroeconomic policy space jointly determine when sentiment meaningfully affects markets.

Figure 3: Sentiment per crisis with 10-speech mean average: chaotic in the GFC, steady in Brexit, and quick to recover in COVID.

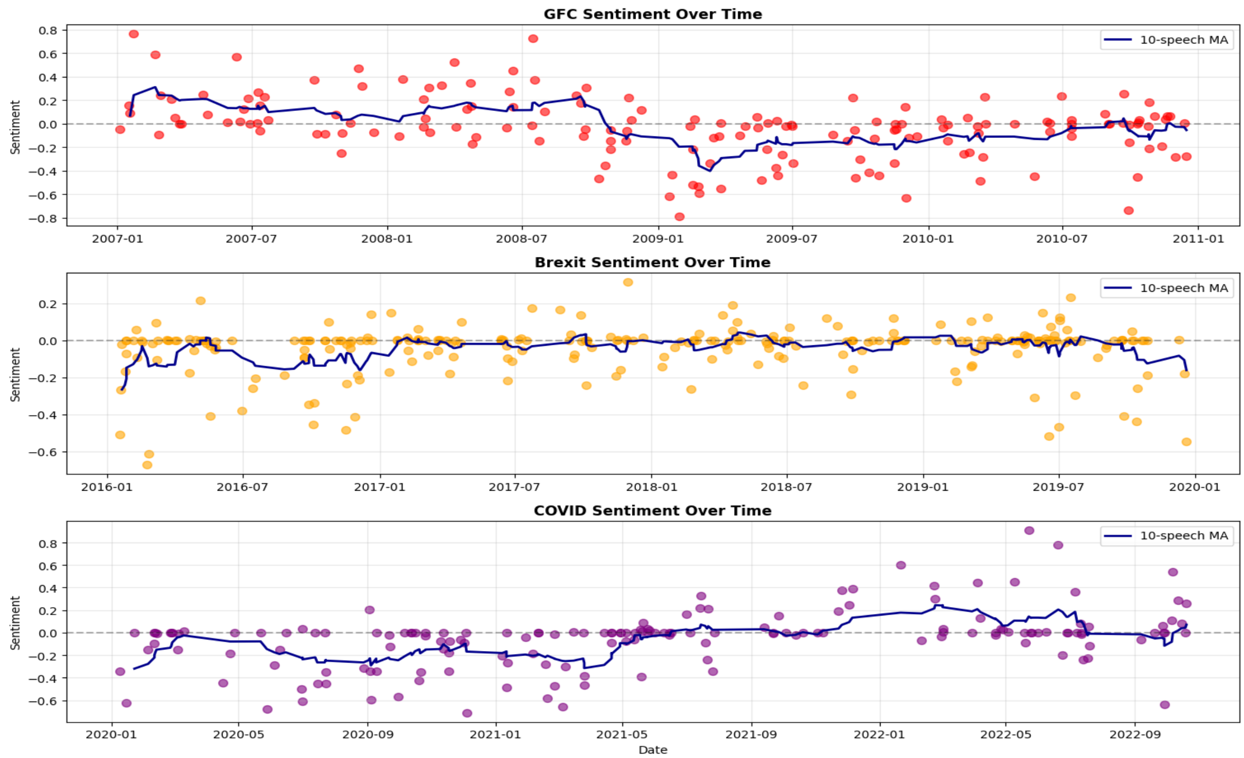


Figure 4: Sentiment effect on the market.

Controlling for time trends, sentiment consistently added explanatory power across markets, most strongly in volatility-sensitive indicators. Sentiment explained 4.76% of the VIX, contributed 6.70% to FTSE100 variation, and 1.96% to 5-year gilt yields. These effects unfolded gradually, indicating that speeches shape expectations rather than triggering immediate price adjustments.

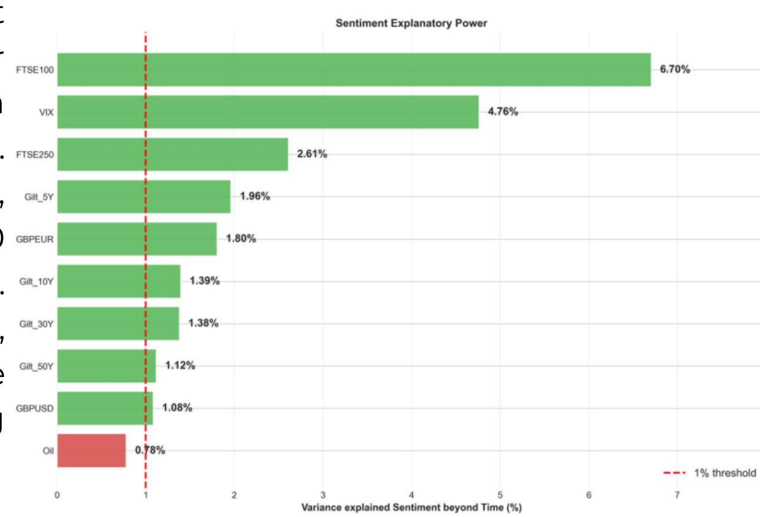
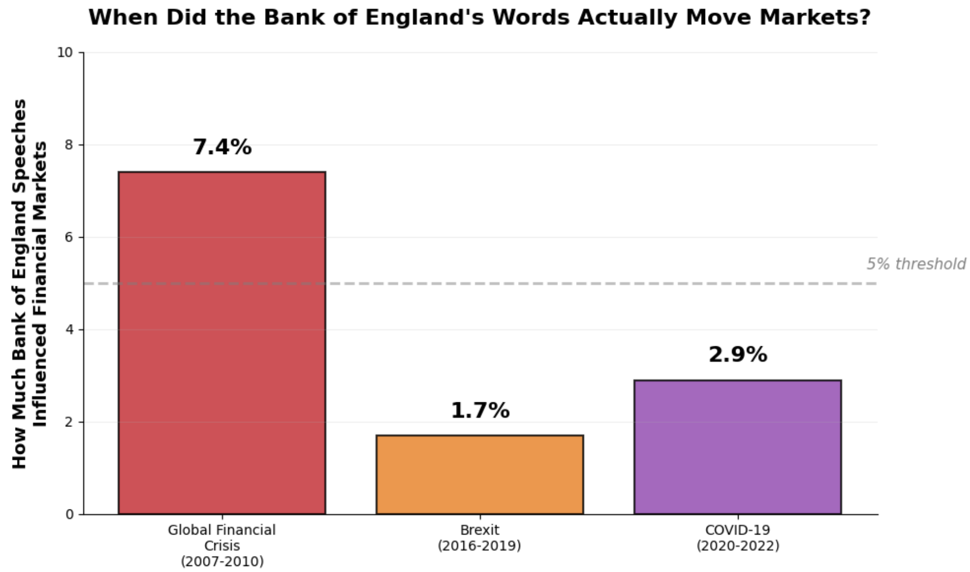
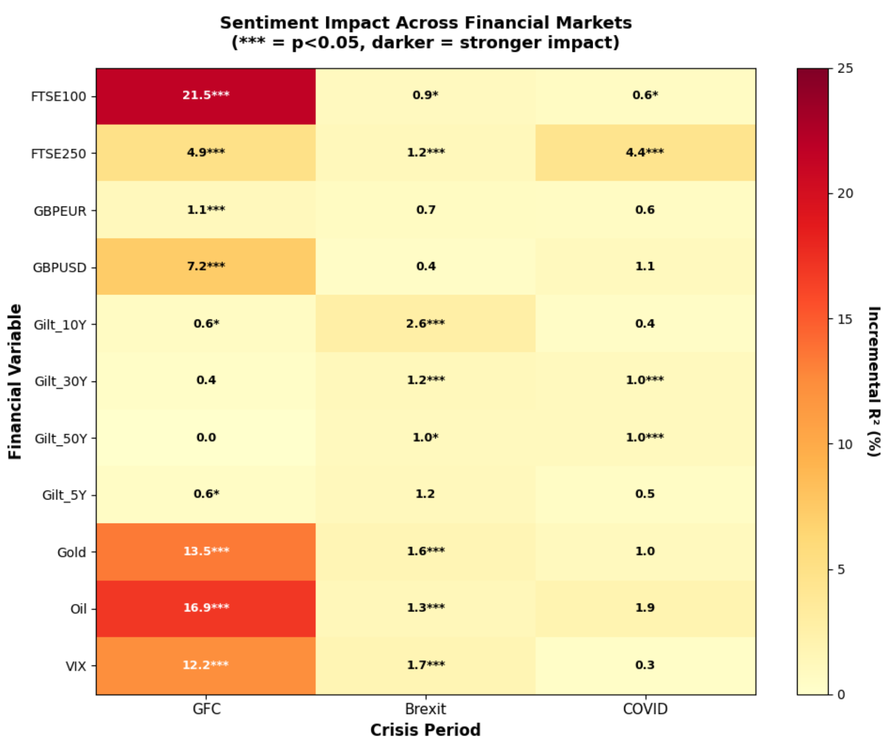


Figure 5: Aggregated sentiment impact during crisis



Crisis periods amplified communication effectiveness. During the GFC, sentiment explained 7.4% of market variation beyond time effects, i.e., four times the pre-crisis level. Monetary policy speeches in particular had a large influence (12.88%). Brexit and COVID displayed weaker effects (1.7% and 2.9%), mirroring the minimal policy space available at near-zero interest rates.

Figure 6: Sentiment impact on specific financial indicators during crisis.



Forward–reverse causality tests showed asymmetry. Equities tended to respond to sentiment, whereas gilt markets frequently led sentiment changes—evidence that the Bank often "leans against the wind" when financial stress emerges. Ultimately, markets listen most closely when clarity and credibility converge.

Figure 7: Does market influence tone, or does tone influence markets?

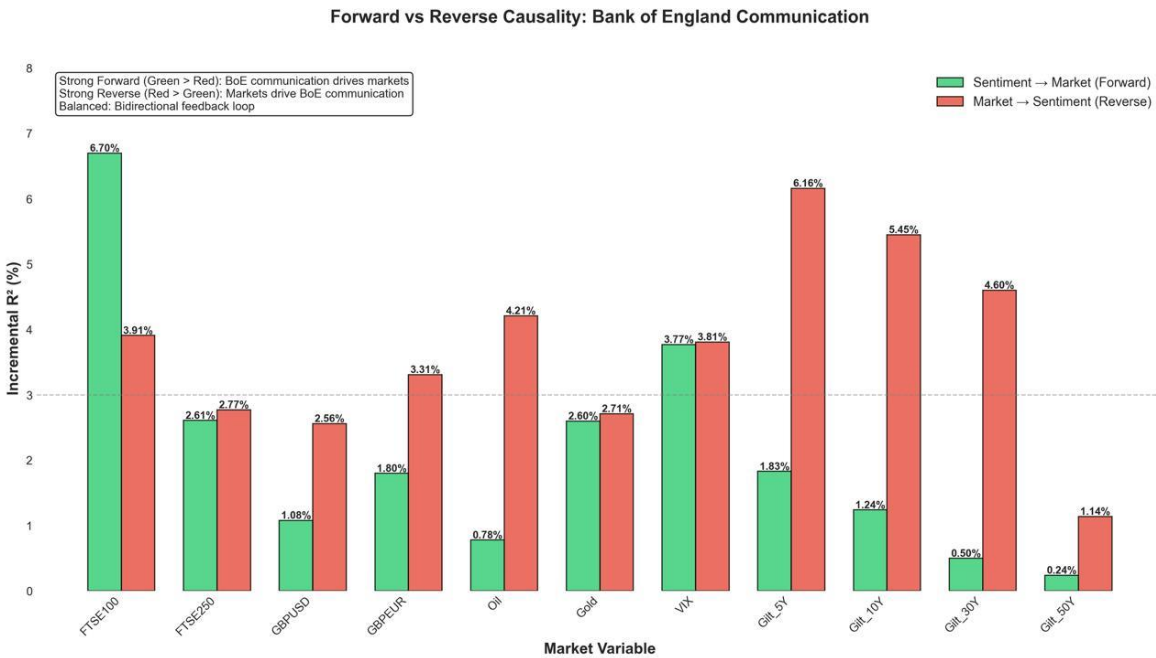
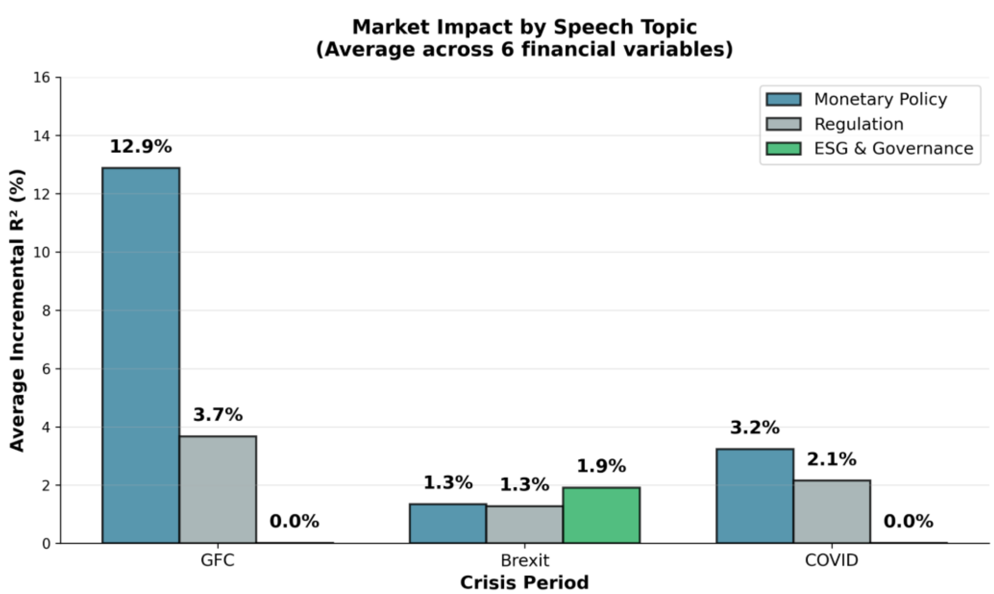


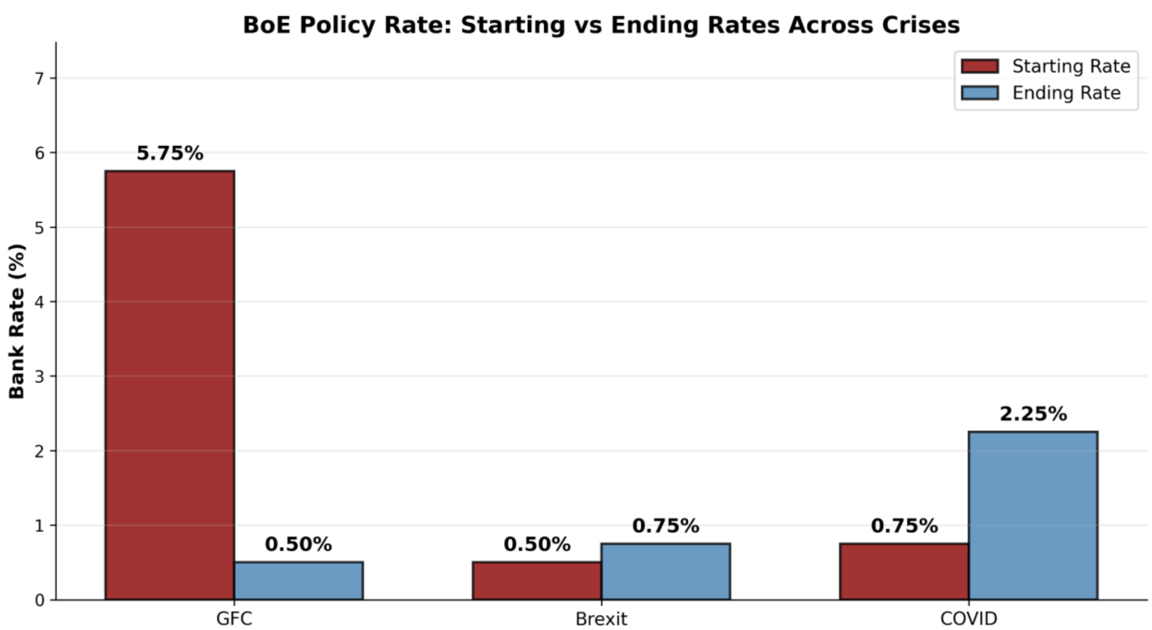
Figure 8: Topical analysis.



Topical analysis revealed structural shifts in communication priorities. The GFC was dominated by monetary policy messaging (12.9%), while Brexit saw increased emphasis on climate governance and digital innovation. COVID saw a partial return to monetary focus alongside heightened concern for household debt. Markets appeared highly responsive when communication was both clear and directly relevant to policy action.

This is due to large space for policy movement in GFC in comparison:

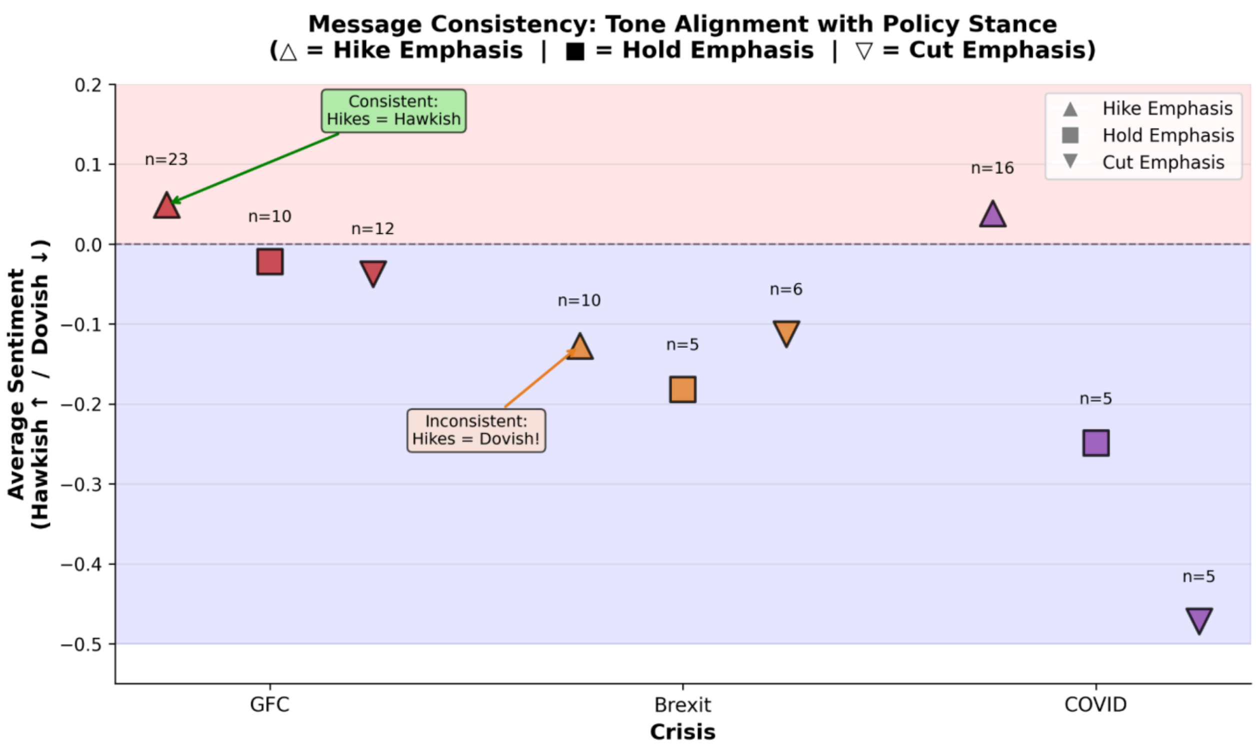
Figure 9: Space for policy movement.



Because of this difference, disconnect could be observed between policy and sentiment: Brexit was dovish overall, including Hike speeches and holding periods.

In Covid, anything but a hike emphasis was very dovish, despite no cuts ever as big as the GFC, which overall was neutral, displaying the bank's strategy to dampen major crisis effects.

Figure 10: Tone vs. policy alignment.



Finding the top 20 words pre, During and post-crisis show clear crisis-specific language patterns: Mid crisis the GFC emphasised “market”, “credit” and “liquidity”, reflecting acute financial dysfunction and large policy space; Brexit centred on “uncertainty” and “stability”, signalling reassurance without tools; COVID shifted to “risk” and “support”, matching rapid, intervention-led communication.

See figures b 1-5 in appendix

5. Limitations

These aforementioned findings were found within the context of several methodological constraints.

Firstly, correlation does not imply causation; sentiment from a speech may reflect underlying economic conditions rather than independently driving market behaviour. Reverse causality was particularly visible in gilt markets.

Secondly, omitted variable bias remains possible, as market movements reflect fiscal decisions, global shocks, and investor sentiment that are not captured in the models.

Thirdly, measurement error is inherent in transformer-based sentiment classification, and while FOMC-RoBERTa performed best, misclassifications remain.

Fourthly, some topic–crisis combinations lacked sufficient sample sizes for reliable inference. Finally, the findings reflect the unique monetary environment of 1998–2022, including two periods of near-zero rates that may not resemble future crises.

These limitations should be interpreted as boundary conditions rather than weaknesses, and they frame sentiment as one lens within a broader ecosystem of market drivers.

6. Recommendations for Further Exploration

Three recommendations emerge from the findings:

1. Maintain clarity when policy space is constrained. Markets respond far more strongly to focused, unambiguous messaging during periods of limited policy flexibility.

2. Implement real-time sentiment monitoring. A dashboard tracking tone alongside gilt yields and volatility would provide valuable insight into how communication is absorbed during unstable periods.

3. Expand the analytical framework. Future work should investigate intraday market reactions, speaker-specific effects, forward-guidance signals, and comparisons with the ECB and Federal Reserve to test generalisability.

These recommendations may help strengthen the Bank's ability to calibrate communication to market conditions, especially in periods when traditional policy tools are constrained.

APPENDIX

A. Model validation

B. wordlcouds

B1-3 = GFC ‘during’ crisis communication

B4 - BREXIT referendum language

B5 - First lockdown COVID language