MASTER

DATA ANALYTICS FOR BUSINESS

MASTER´S FINAL WORK

DISSERTATION

INFLATION FORECASTING WITH MACHINE LEARNING USING NARRATIVE FEATURES

TIAGO GONÇALVES MARTINS

FEBRUARY - 2025

MASTER

DATA ANALYTICS FOR BUSINESS

MASTER´S FINAL WORK

DISSERTATION

INFLATION FORECASTING WITH MACHINE LEARNING USING NARRATIVE FEATURES

TIAGO GONÇALVES MARTINS

SUPERVISION:
PROF. JOÃO AFONSO BASTOS
PROF. ADRIANA CORNEA-MADEIRA

FEBRUARY - 2025

Lisbon School
of Economics
& Management
Universidade de Lisboa

GLOSSARY

API - Application Programming Interface: Rules and protocols that allow different software applications to communicate and exchange data seamlessly.

AR(4) – Autoregressive Model (Order 4): A time-series forecasting model that predicts future values based on its own past four observations.

CPI – Consumer Price Index: A measure of the average change in prices paid by consumers for goods and services over time.

CPIH – Consumer Price Index including Housing Costs: An extended measure of inflation that includes housing costs associated with owning and living in a home.

DM Test – Diebold-Mariano Test: A statistical test used to compare the accuracy of two competing forecasting models.

FRED-MD – Federal Reserve Economic Data - Macroeconomic Database: A dataset of macroeconomic indicators maintained by the Federal Reserve Bank of St. Louis.

GDP – Gross Domestic Product: The total value of goods and services produced within a country over a given period.

LASSO – Least Absolute Shrinkage and Selection Operator: A regression method that performs both variable selection and regularization to enhance predictive accuracy.

LDA – Latent Dirichlet Allocation: A probabilistic model used for topic modeling in text analysis, grouping words into themes.

MFW – Master's Final Work: A dissertation or final academic research project submitted for a master's degree.

OECD – Organisation for Economic Co-operation and Development: An international organization that provides economic analysis and policy recommendations.

IMF – International Monetary Fund: global financial institution that promotes economic stability, international trade, and economic growth

RMSE – Relative Root Mean Squared Error: A metric for evaluating the predictive accuracy of a model.

rRMSE – Relative Root Mean Squared Error: A metric for evaluating the predictive accuracy of a model, relative to a benchmark.

RF - Random Forest: A machine learning ensemble model with multiple decision trees to improve predictive performance.

UK – United Kingdom

USA – United States of America

ABSTRACT

The ability to accurately forecast inflation is crucial in today's world for achieving macroeconomic stability, making informed decisions, and enabling long-term planning. Traditional models often struggle to adapt during periods of heightened economic turbulence, which increases volatility in such indicators. Recent advancements in machine learning, particularly its application to unstructured data, such as daily news articles, offer a creative and promising alternative to improve inflation forecasting accuracy. This study builds upon the hybrid methodology introduced by Hong et al. (2024), which combines both macroeconomic data and textual data using tools like Latent Dirichlet Allocation (LDA).

This research adopts a similar strategy to the referenced paper but introduces a notable difference by employing a train-test split methodology instead of a recursive approach. It systematically evaluates the impact of specific adjustments, parameters, and strategies, including algorithms such as Least Absolute Shrinkage and Selection Operator (LASSO) and Random Forest (RF), as well as a combined method leveraging multiple machine learning algorithms, all benchmarked against a baseline autoregressive model.

The results demonstrate that hybrid models significantly enhance forecasting performance, with improvements reaching up to 40% compared to models using only macroeconomic data. Furthermore, the findings highlight the importance of the topic nature depending on the forecasting horizon, with themes such as energy markets and corporate governance emerging as critical drivers. This study provides practical insights for economists to leverage advanced and alternative machine learning techniques for inflation forecasting, offering a comprehensive overview of the advantages and challenges of incorporating unstructured data into this type of research.

Keywords: Economic narratives, hybrid models, Inflation forecasting, Latent Dirichlet Allocation (LDA), machine learning, macroeconomic indicators, predictive analytics, train-test split, unstructured data.

RESUMO

A capacidade de podermos prever com assertividade a inflação é fundamental nos tempos que correm para alcançarmos estabilidade macroeconómica e podermos tomar decisões informadas, fazendo planeamentos a longo prazo.

Os modelos tradicionais tendem a ter falhas a adaptar-se a alturas com maior turbulência económica, o que gera maior volatilidade neste tipo de indicadores. Avanços recentes com *machine learning* e a sua adaptação a dados não estruturados, como a partir de um conjunto de notícias diárias, oferecem uma alternativa criativa e promissora para conseguirmos um passo em frente em prever com mais assertividade a inflação. Este trabalho tem como base o método híbrido introduzido por Hong et al. (2024) que combina tanto dados macroecómicos como dados de texto, usando ferramentas como *Latent Dirichlet Allocation*. Este estudo adota uma estratégia semelhante ao *paper* referido com uma diferença assinalável no que toca à metodologia de divisão treino-teste adotada contrariamente à abordagem recursiva. Sistematicamente avalia os impactos de algumas mudanças em pormenores, parâmetros e estratégias, incluindo algoritmos como o LASSO, *Random forest* e ainda um método combinado de vários algoritmos de *machine learning*, sempre comparando com um modelo autorregressivo de base. Resultados mostram modelos híbridos a melhorar significativamente a performance dos modelos, com melhorias a chegar a 40% comparativamente aquilo que seriam os resultados apenas com dados macroeconómicos. Ainda se realça a importância da natureza dos tópicos, mediante o horizonte em estudo, destacando-se temas como mercados energéticos e gestão empresarial em geral. Este estudo oferece ideias práticas para economistas poderem usar técnicas avançadas e alternativas de *machine learning* para previsão da inflação e proporciona uma boa *overview* das vantagens e desvantagens de usar dados não estruturados neste tipo de investigações.

Palavras-chave: Análise preditiva, dados não estruturados, divisão treino-teste, indicadores macroeconómicos, Latent Dirichlet Allocation (LDA), machine learning, modelos híbridos, narrativas económicas, previsão da inflação.

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF FIGURES

ACKNOWLEDGMENTS

I would like to express my gratitude to professors João Bastos and Adriana Cornea-Madeira for their contributions and support throughout this journey.

Also, I want to thank my friends and family for their unconditional support and willingness to help. Thank you for believing in me.

1. INTRODUCTION

Inflation is a central concern for economists, policymakers and businesses in general. It is fundamental for the stability of the economy, and it affects everything from household to central bank decisions. Inaccurate forecast can lead to wrong policy decisions, market volatility and even an economic crisis. However, forecasting inflation remains a persistent challenge particularly in the current global economy. Traditional models often fail to follow the complexity of our systems and behaviours.

So, improving inflation forecasting stays urgent with external factors like technological advancements, globalization and sudden shocks, like COVID-19 pandemic, have being introduced in the last decades. Economic stakeholders and policymakers have to navigate these uncertainties in real time making decent forecasting tools indispensable. However, the limitations of conventional econometric models, which have rigid assumptions or static relationships, prevent their ability to rapidly change conditions. This motivates the exploration of innovative data-driven approaches that can leverage the abundance of both structured and unstructured data now available.

A promising idea lies in the integration of economic narrative features from surveys and news articles into inflation forecasting. On contrast to quantitative macroeconomic indicators, these narratives capture the qualitative and daily pulse of markets, policies and public sentiment. They reflect how institutions, governments and people in general respond to economic developments or other events worldwide- information useful to understand inflation dynamics but hard to quantify through traditional data sources alone.

This research is driven by that economic narratives are more than just supplementary but critical to understand and anticipate shifts in inflation. The remain challenge relies on effectively extracting and integrating qualitative information into the models. Recent studies have demonstrated the potential of a hybrid approach combining both structured macroeconomic data with unstructured narratives but there are questions about how can they be refined, scaled and improved to meet practical forecasting need. For example, what role do narrative features play across horizons like short, medium and long terms?

The unique contribution of this work is the investigation of performance between narrative and structured data for inflation forecasting. Based on prior work, this study introduces methodological differences aiming to improve practicality and scalability of hybrid approaches. Additionally, this study explores how variations in modeling parameters affect predictive accuracy and it departs from Hong et al. (2024) by employing a train-test split methodology which reduces computational demands without sacrificing adaptability.

By demonstrating how narrative data can enhance inflation forecasts, it offers a pathway for more informed decision-making in an era of increasing data complexity. Policymakers, central banks, and financial institutions can use these insights to develop tools that better anticipate inflationary pressures, ultimately contributing to economic stability and resilience.

The remainder of this dissertation is structured as follows: Chapter 2 reviews the existing literature, focusing on traditional and modern approaches to inflation forecasting and highlighting the potential of hybrid models. Chapter 3 outlines the methodology, including data collection, preprocessing, topic modeling, and the forecasting framework. Chapter 4 presents the empirical findings, with a detailed analysis of model performance and the impact of narrative features across forecasting horizons. Finally, Chapter 5 concludes with a discussion of the research's implications, limitations, and opportunities for future exploration.

2. Literature Review

Inflation forecasting is one of the most current and talked about topics, influencing monetary policy decisions, long-term economic planning, and investment strategies. With accurate forecasts, policymakers can manage inflation expectations and move the economy toward stable growth. Traditional models, such as those based on the Phillips Curve (Atkeson & Ohanian, 2001), have served as reliable tools. However, these models and methodology fail in periods of economic turbulence as their static structures don’t capture the changing dynamics of inflation predictors. This highlights methodologies that can reach both accurate and reliability, especially in volatile economic seasons.

Advances in the machine learning world have revolutionized the field of economic forecasting by leveraging high dimensional data and handling complex and non-linear relationships. Medeiros et al. (2019) emphasized the efficacy and efficiency of machine learning methods in forecasting macroeconomic indicators. Their findings show how these models outperform traditional benchmarks in terms of predictive accuracy.

A new way in inflation forecasting involves incorporating economic narratives derived from textual data. Coibion and Gorodnichencko (2012) emphasize that misperceptions and imperfect information can lead to persistent inflationary effects, underscoring the necessity of integrating economic narratives into forecasting models. Larsen et. al (2021) highlights how economic discourse, disseminated through the media, influences business cycles and policy expectations reinforcing the predictive value of news-based inflation forecasts. Leveraging unstructured data like news articles is an alternative and innovative method to complement traditional macroeconomic indicators.

Hong et al. (2024) demonstrated this idea by employing Latent Dirichlet Allocation (LDA; Blei et al. 2003) to process a large corpus of news articles. By extracting topic distributions from the text and combining them with machine learning algorithms, they achieved more accurate and robust inflation forecasts. This new methodology represents a clear shift in inflation forecasting by introducing narrative analysis as a quantitative tool. Paranhos (2020) explored deep learning techniques, such as Recurrent Neural Networks, to detect sequential patterns in economic narratives

Hong et al. (2024) highlighted the potential of hybrid approaches that blend the predictive strengths of structured and unstructured data. For example, by combining LDA-derived topics with a comprehensive dataset of macroeconomic variables, they demonstrated how hybrid models could outperform standalone approaches. However, this integration demands careful attention to issues such as data synchronization, scaling, and the alignment of temporal structures between narrative and macroeconomic data.

Kalamara et al. (2020) and Surprenant (2022) also mentioned the importance of incorporating news text into forecasting models, demonstrating the power of narrative features about economic conditions. Their approach complements the hybrid strategies proposed by Hong et al. (2024) by reinforcing the role of narrative data as a critical source of predictive information. Additionally, the application of penalized regression methods, as outlined by Tibshirani (1996) and Zou and Hastie (2005) provide a robust mechanism for selecting relevant predictors from large datasets.

McCracken and Ng (2015) highlighted the value of structured macroeconomic datasets, such as FRED-MD, for forecasting purposes. By combining these datasets with narrative-driven features, this study seeks to improve forecast performance over a range of horizons, similar to the multi-horizon strategies employed in Boaretto and Medeiros (2023). The inclusion of models like Random Forests (Breiman, 2001) guarantee that non-linear relationships between predictors are effectively captured, while the use of evaluation metrics such as relative root mean squared error (RMSE) and the Diebold-Mariano (DM) test (Diebold, 2012) aligns with best practices in predictive accuracy assessment.

The present study builds directly on the work of Hong et al. (2024), adopting their methodological framework while introducing refinements to improve model performance and interpretability. Key similarities include the use of LDA for topic modelling, machine learning algorithms for prediction, and a hybrid approach that combines economic narratives with macroeconomic indicators. Furthermore, other decisions regarding the modelling pipeline align closely with the methodological details emphasized by Hong et al. (2024).

This study contributes to the growing body of research that seeks to enhance the accuracy and robustness of inflation forecasting in an era of increasing data availability and economic complexity.

3. Methodology

In this chapter, it is explained in depth the procedures of data collection and preparation and which methods and models were used for forecasting inflation with both narrative features and macroeconomic data using Hong et al. (2024) methodology as guidance. This process consisted of choosing a determined number of topics for LDA and, using machine learning algorithms, inspect their predictive power alone and combining them also with macroeconomic data to check their performance.

3.1 Data Collection and preprocessing steps

In preparing the dataset for predictive modelling with macroeconomic indicators the categorization and preparation of variables were organized into logical groups to facilitate clear analysis and transformation. Variables were divided into categories such as real activity, which includes indicators like production and employment; inflation metrics such as the Consumer Price Index (CPI); financial variables, which encompass interest rates, bond yields, and stock indices; and other variables, which include diverse indicators like trade indices, oil prices, and exchange rates.

Target variables, including Core CPI and Consumer Price Index including Housing Costs (CPIH), were carefully selected and checked to exclude missing values. Any records with missing data in these critical fields were excluded to maintain the dataset's quality. Transformations applied to the data included standardizing and stabilizing the dataset for statistical modelling. Techniques such as logarithmic differencing were employed to stabilize variance and achieve stationarity among variables. This method was particularly suited for data with positive values, with nulls introduced for non-applicable data points. Percentage changes were applied to financial and other variables to accurately capture relative changes over time. These transformations closely align with the methodologies used in the FRED-MD database, as outlined by McCracken and Ng (2015), providing consistency with established practices in macroeconomic forecasting. More details can be seen in Table 1 in Appendix A.

The final dataset comprised monthly data from 1999 to 2024, containing 43 categorized and transformed columns. Each variable was sourced from reputable databases to ensure the accuracy and relevance of the information. For instance, interest rates, government securities, and bonds were sourced from the International Monetary Fund (IMF) and the Organization for Economic Co-operation and Development (OECD). Some indicators were collected from FRED-MD database- a dataset originally designed for United States of America (USA) data but with several general economic indicators that were relevant for this study. The transformations applied, including log differencing and percentage changes, were also inspired by methodologies from the FRED-MD database, as utilized by Hong et al. (2024). These transformations allow the dataset to be stationarized and variance-stabilized, enabling robust statistical modeling.

The text corpus was retrieved from the digital service provided by The Guardian that gives access to their archive of published content through an Application Programming Interface (API). After an analysis of the available sections, the Business section was the most relevant among all and it was the one used consisting of all business-related articles published in The Guardian covering the period from 1989 to 2024. This journal from the United Kingdom (UK) describes daily world events, which helps to implement economic models comprehensively as documented by Kalamara et al. (2020).

To ensure the reliability and relevance of the analysis, the dataset was filtered to include only articles from 1999 onwards, due to lack of data on the first 10 years, which can easily be seen on Figure 4, Appendix A, resulting in a final corpus of 145,496 news articles.

The data preprocessing phase involved several systematic steps to clean and standardize the text data. Initially, duplicate articles and entries from non-UK production offices were removed, along with any rows containing missing values in critical columns. The publication dates were converted to a uniform datetime format, allowing for temporal alignment in subsequent analysis.

Text preprocessing followed conventional natural language processing (NLP) practices to make sure that data was suitable for machine learning. Articles were tokenized into words, with all text converted to lowercase to eliminate case sensitivity. Stop words, punctuation, and other non-informative elements were removed. Additionally, stemming and lemmatization were applied to reduce words to their base forms, thereby minimizing redundancy in the vocabulary. This process produced a clean and structured corpus, ready for further analysis.

For a more nuanced analysis, n-grams (unigrams and bigrams) were generated to capture contextual dependencies between words. This step enriched the feature set by incorporating short sequences of words, particularly valuable in identifying narrative structures and key phrases related to inflation and economic conditions.

The final news corpus, after cleansing and treatment, consists of 145,496 news articles with a vocabulary of more than 15,000 unique terms per month, covering the period from January 1999 to December 2023. On average, there are 477 articles and 222,354 terms per month, with 73% of them being unigrams and 27% bigrams.

Table 1 Descriptive analysis of the dataset

Variable	Mean	Std Dev	Min	Max
Number of articles per month	477	176	201	1,078
Number of articles per year	5,728	2,035	3,036	9,695
Number of terms per month	222,354	42,908	129,889	462,834
Number of unigrams per month	163,043	31,213	97,495	332,998
Number of bigrams per month	59,311	11,969	32,394	129,836
Number of unique terms per month	15,706	1,937	11,402	21,398
Consumer Price Index (%)	0.21	0.31	-1.77	1.38

3.2.Topic Modelling

Latent Dirichlet Allocation (LDA), a method developed by Blei, Ng, and Jordan (2003), was used to extract and articulate economic narratives from our collection of texts. LDA serves as a generative statistical model that interprets a set of observations- like articles- through unobserved topics that reveal specific underlying patterns. Each topic is defined by a distribution of words, while every document reflects a distribution across these topics. This dimensionality reduction technique streamlines high-dimensional and sparse text data into a more manageable set of topics, effectively capturing critical narrative structures.

3.2.1 Mathematical Foundations of LDA

Handling high dimensional economic narratives requires effective dimensionality reduction techniques. Huang et al. (2022) propose scalable approaches to extract meaningful features from large datasets, aligning with our use of topic modelling to identify key inflationary themes with economic news.

As a probabilistic model, LDA makes the following assumptions:

1. A Dirichlet distribution $\theta_d \sim \text{Dirichlet}(\alpha)$ parameterizes the combination of K topics in each document.

2. Each topic is a distribution over a fixed vocabulary of terms, parameterized by $\phi_k \sim$ Dirichlet(β).

The goal of this model is to maximize the marginal likelihood of the data, and this can be achieved through approximate inference techniques like Gibbs sampling or variational Bayes.

$$(1)\ p(w \mid \alpha, \beta) = \prod_{d=1}^{D} p(w_d \mid \alpha, \beta)$$

LDA assumes that each document is made from a range of topics, and each one shows a distribution over terms. To make this process easier, a Document-Term Matrix was created. The corpus underwent preprocessing steps such as tokenization, stop word removal and lemmatization- it reduces redundancy from the vocabulary.

Based on earlier research and evaluation of the model's performance metrics, 80 topics were chosen as baseline. After training, each document was converted into a vector of topic probabilities, indicating the extent to which each topic is represented in the document. Then, monthly averages of these topic distributions were computed to synchronize with macroeconomic data for forecasting aims. To identify the ideal number of topics and to address the question of whether the number of topics influence that deeply the results, a model was assessed using perplexity and coherence scores. Perplexity describes how well the model can generalize to unseen data, where lower values signify superior performance. Coherence, on the other hand, measures the interpretability of topics grounded in semantic similarities, with higher scores being more favourable.

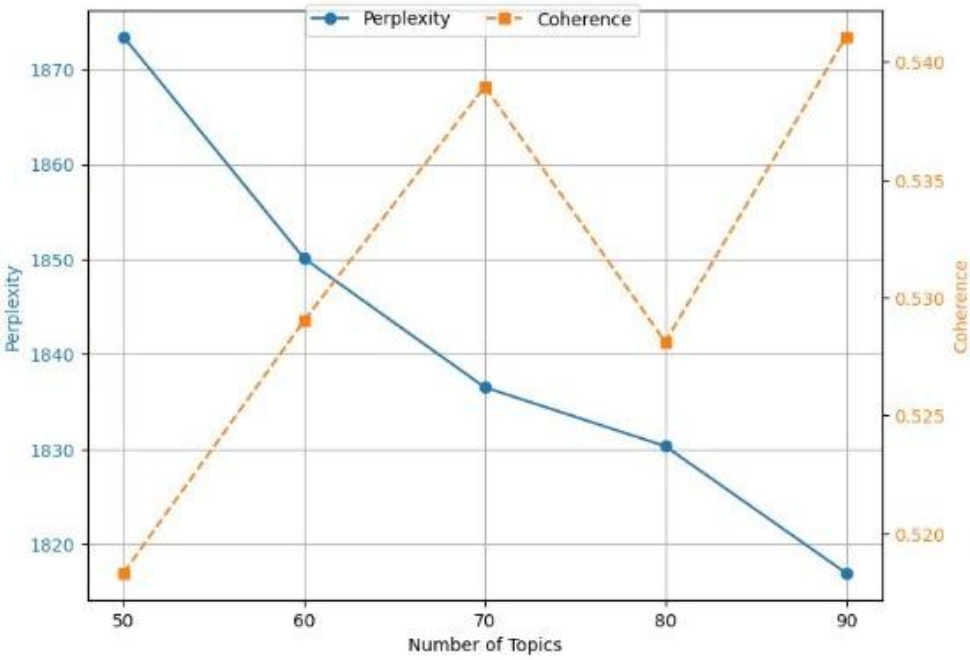

Figure 1 Perplexity and Coherence Scores vs. Number of Topics

The graph (Figure 1) demonstrates the balance between these two metrics across various topic counts. After evaluating it a new measure was taken into consideration: the ratio between coherence and perplexity (assuming equal importance for both metrics) by a specific number of topics. This measure allows us to objectively balance the trade-offs between perplexity and coherence. The ratio is calculated as follows for each topic count:

$$(2)\ Coherence - Perplexity\ Ratio = \frac{Coherence\ Score}{Perplexity\ Score}$$

Figure 2 illustrates the ratio for different topic counts. The results indicate that the ratio is maximized at 90 topics, confirming it might be the most adequate choice for this analysis. However, the analysis will be done using 70 and 80 topics to facilitate the process of labelling the set of topics.

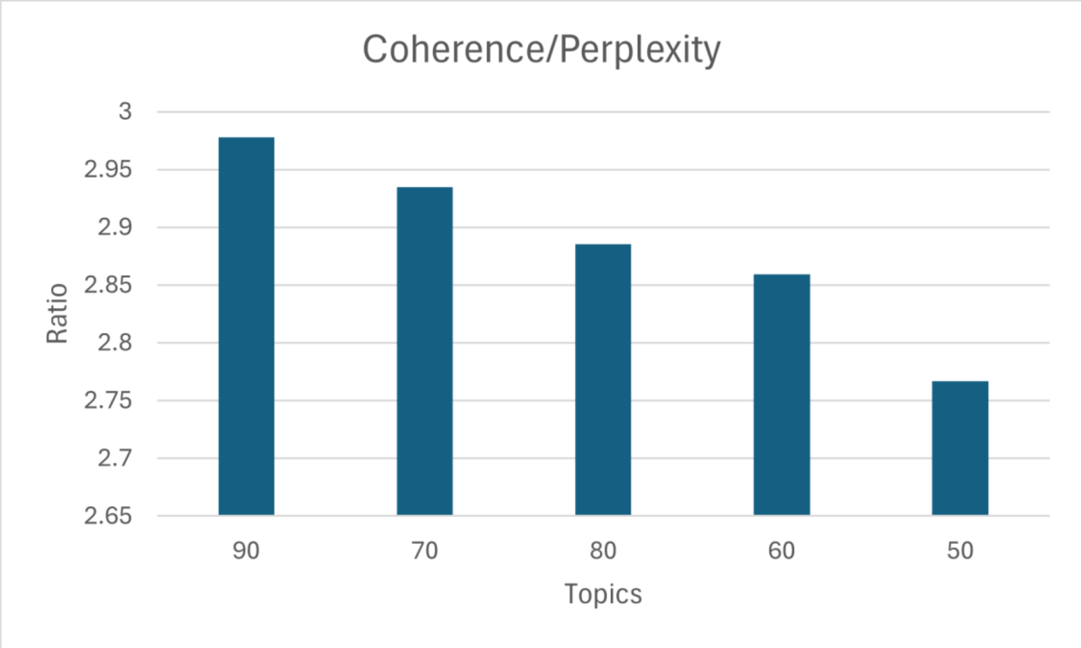

Figure 2 Coherence-to-Perplexity Ratio vs Number of Topics.

3.3 Forecasting framework and models

This study will combine structured macroeconomic indicators with economic narratives to forecast inflation both by studying it separately and together. The proposed framework shows how machine learning methods can model relationships between those aspects.

The forecasting model is expressed as:

$$(3)\ \pi_{t+1,t+h} = G\ (\Theta_t) + \varepsilon_{t+1,t+h}$$

Where:

- $\pi_{t+1,t+h}$ represents the cumulative inflation over the forecast horizon h;
- Θ_t are narrative topic distributions derived from the news corpus;
- $G\ (\cdot)$ is the predictive function learned using various forecasting models and
- $\varepsilon_{t+1,t+h}$ is the error term.

Machine learning techniques have been having a strong impact on macroeconomic forecasting with Fan, Jin and Fang (2023) demonstrating how high dimensional predictive models outperform traditional statistical methods supporting the integration of news based narrative approaches.

The machine learning models chosen align with Hong et al. (2024). Four models for the mapping were considered: least absolute shrinkage and selection operator (LASSO), elastic net (ENet), random forest (RF), and forecast combination (Comb).

LASSO, introduced by Tibshirani (1996), is a well-known method performing both variable selection and regularization with the purpose of improving prediction accuracy and interpretability of the statistical model. It minimizes the following penalized objective function:

$$(4)\ \widehat{\beta} = argmin \left(\frac{1}{T} \sum_{t=1}^{T} (y_t - X_t^T \beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right)$$

Where:

- y_t represents the response variable (e.g. inflation);
- X_t is the feature vector at time t;
- β represents the coefficients of the model;
- λ is the regularization parameter controlling sparsity.

Elastic Net extends LASSO by combining L_1 and L_2 penalties introduced by Zou and Hastie (2005). It handles various types of data in a more effective way than LASSO, particularly when dealing with highly correlated data:

$$(5)\ \widehat{\beta} = argmin \left(\frac{1}{T} \sum_{t=1}^{T} (y_t - X_t^T \beta)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2 \right)$$

Where:

- λ_1 is the regularization parameter controlling sparsity (L_1);
- λ_2 is the regularization parameter controlling shrinkage (L_2)

Random Forest (Breiman (2001)) is an ensemble learning method for classification and regression that predicts inflation using an ensemble of decision trees. Each tree partitions the feature space into regions, and predictions are averaged across all trees to reduce variance and capture nonlinear interactions. The prediction is:

$$(6)\ \hat{y} = \frac{1}{M}\sum_{t=1}^{T} f_m(\mathrm{X}_t)$$

Where:

- M represents the number of trees;
- $f_m(\mathrm{X}_t)$ is the prediction from the m-th tree trained on a bootstrap sample.

Forecast Combination is a method used by Hong et al. (2024) that averages predictions from multiple models to improve robustness:

$$(7)\ \hat{\pi}_{Comb} = \frac{1}{M}\sum_{m=1}^{M} \hat{\pi}_m$$

Where:

- M represents the number of models;
- $\hat{\pi}_m$ is the prediction from the m-th model.

Equal weights are used in the combined forecast. It is well known in the forecasting literature (e.g., Rapach et al. (2010)) that the simple equal weighting scheme performs reasonably well in practice.

The benchmark model used in this study is an autoregressive model with four lags (AR(4)), aligning again with the approach from Hong et al. (2024). The AR(4) model has been frequently used as a benchmark for forecasting inflation. It predicts inflation based on its past values and is defined as:

$$(8)\ \pi_t = \alpha + \sum_{j=1}^{4} \varphi_j \pi_{t-j} + \epsilon_t$$

Where:

- π_t represents the inflation at time t;
- φ_j represents the coefficients of the j-th lag of inflation;
- α represents the intercept term;

- ϵ_t is the error terms assumed to be independently and identically distributed (i.i.d).

To study performance, the forecasting models were evaluated using relative Root Mean Squared Error (rRMSE) and Diebold-Mariano (DM) tests across multiple horizons (3, 6, 9, and 12 months) as done by Hong et al. (2024). The benchmark AR(4) model served as a comparison point for assessing improvements in predictive accuracy.

The Root Mean Squared Error (RMSE) is calculated as:

$$(9)\ RSME_m = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\pi}_{m,i} - \pi_i\right)^2}$$

where $\hat{\pi}_{m,i}$ and π_i denote the forecasted and actual values of inflation for model m, and n is the total number of observations.

The relative RMSE (rRMSE) is used to compare each model's performance against the benchmark AR(4) and is given by:

$$(10)\ rRMSE_m = \frac{RSME_m}{RSME_{AR(4)}}$$

The Diebold-Mariano (DM) test evaluates whether the forecast errors of a given model are statistically significantly better than the benchmark. The DM statistic is calculated as

$$(11)\ DM = \frac{\bar{d}}{\sqrt{\frac{1}{T}Var(\bar{d})}}$$

Where:

- $\bar{d}$ is the mean of the loss differential $d_t = \left(e_{m,t}^2 - e_{AR(4),t}^2\right)$;
- $e_{m,t}$ and $e_{AR(4),t}$ are the forecast errors of model m and AR(4), respectively;
- T is the number of forecasts;
- $Var(\bar{d})$ incorporates autocorrelation adjustments using Newey-West estimators to account for serial dependence.

As emphasized by Diebold, F. X. (2012), the Diebold-Mariano test is highly effective in forecast accuracy but should be used cautiously in model comparison contexts. Misapplications, such as neglecting sample dependency, can undermine its validity. Fullsample approaches often yield more robust insights, particularly in settings with small sample sizes or overlapping data points. This approach minimizes the risk of overfitting while ensuring robust performance across various subsets of the data. This process was applied during the recursive forecasting steps iteratively.

To guarantee that the forecasting models generalize well to unseen data, cross validation was employed during hyperparameter tuning and grid-search to find the optimal values within an interval. Note that, despite Hong et al. (2024) recursive approach, this study uses as methodology a train-test split approach. The recursive approach provides a robust framework for inflation forecasting by adapting to dynamic economic conditions, but at the expense of significant computational demands. The modified train-test split methodology and rolling window, on the other hand, gives a balance between efficiency and adaptability, offering a practical alternative that retains some of the strengths of the recursive method while mitigating its computational challenges. It adapts to new trends and patterns in the data which contributes to improving forecast accuracy in dynamic economic environments. The data is partitioned with 60% or 70% for training sets (40% and 30% for test sets respectively) and it incorporates some adaptability seen in the recursive approach by recalibrating the models annually during testing. The null hypothesis is that AR model forecasts have a RMSE that is less than, or equal to, that of the narrative-based forecasts.

4. EMPIRICAL RESULTS

4.1 Forecasting Results

4.1.1 Narrative Features and macroeconomic forecasting results

As mentioned before, AR(4) serves as benchmark when showing results. Results of alternative models are being analysed relatively to the benchmark. The statistically significance can be seen through *(10% level), **(5% level) and ***(1% level) in DM test columns showing Diebold and Mariano t-statistic values. Different train-test splits were tested as well, and 80 topics and 70 topics were chosen for LDA. Two different target variables were tested: Core CPI and CPIH.

In this section, the approach described will be using a train-test split, a number of topics of 70 for LDA and Core CPI as the target variable. Results with variations of these parameters can be seen in Appendix A.

Table 2 Performance of forecasting the next 3-, 6- 9-, and 12- month inflation rates with narratives: Train-test split with 70/30. LDA with 70 topics. Core CPI as target variable

Horizon	LASSO rRMSE	LASSO DM	ElasticNet rRMSE	ElasticNet DM	RF rRMSE	RF DM	Combo rRMSE	Combo DM
3	0.942	-1.757**	0.943	-1.743**	0.947	-0.631	**0.927**	-1.617*
6	0.985	-0.769	0.985	-0.769	0.983	-0.144	**0.967**	-0.626
9	1.007	0.456	1.007	0.456	0.992	-0.058	**0.979**	-0.376
12	**1.033**	2.151**	**1.033**	2.151**	1.111	1.617*	1.035	1.196

Table 2 reports the out-of-sample forecasts for various forecast horizons. All the models, generally, show better improvements over AR(4) for shorter horizons- horizons 3 and 6 have all the values below one. Combo outperforms the benchmark in the first three horizons, as RF does, showing the lowest rRMSE (0.927, for horizon 3, as the lowest achieved) and also being statistically significant. In this case, Random Forest shows potential but is less robust across variations. When shifting to CPIH as the target variable, or increasing LDA topics to 80, these models still perform consistently but may exhibit slightly reduced DM significance. On the other hand, there are cases in which rRMSEs are higher with statistical significance.

In general, the results found show LASSO and Elastic Net with higher efficiency for short-term forecasting outperforming the benchmark with statistical significance. For medium horizons, no model consistently outperforms AR(4). However, their performance diminishes at medium-to-long horizons as the macroeconomic patterns become more complex and nonlinear, which limits their ability to capture evolving economic trends. For longer horizons, the benchmark remains the most reliable model. However, the variations tested in terms of the number of topics for LDA, train-test splits and the target variable can amplify or reduce statistical significance. When the alternative models outperform AR(4), their rRMSE values tend to cluster around 0.9, particularly for shorter horizons. The Diebold-Mariano (DM) test confirms the statistical significance of some of these improvements, particularly at the 5% and 10% levels, indicating that these gains are unlikely to be due to chance. This indicates that while the models offer improvements, these gains are relatively modest in magnitude. Although results primarily focus on the 70/30 split, alternative splits such as 60/40 were also tested, revealing that changes in the training proportion can amplify or reduce statistical significance depending on the forecasting horizon and feature selection. The results emphasize the importance of carefully tunning parameters and dataset configurations to maximize performance across horizons.

Moving to the results with the dataset with collected macroeconomic data, Table 3 shows results in the same conditions mentioned before. As before, Combo and RF consistently show the best results across horizons. With macroeconomic data, the results are sometimes lower than 0.9 reaching 0.754 at its minimum with Combo at horizon 3 with statistical significance at 5% level. Random Forest is the only model outperforming the benchmark across all horizons. As expected, in general, the results with macroeconomic data reveal better performance compared to AR(4) comparing to narrative features alone.

Table 3 Performance of forecasting the next 3-, 6- 9-, and 12- month inflation rates with macroeconomic data. Train-test split with 70/30. Core CPI as target variable

Horizon	LASSO rRMSE	LASSO DM	ElasticNet rRMSE	ElasticNet DM	RF rRMSE	RF DM	Combo rRMSE	Combo DM
3	0.774	-1.258	0.774	-1.260	0.885	-1.832**	**0.754**	-1.665**
6	0.889	-0.414	0.887	-0.426	0.853	-1.434*	**0.824**	-0.831
9	1.115	0.515	1.107	0.489	**0.870**	-1.173	0.977	-0.124
12	1.466	2.161**	1.453	2.132**	**0.928**	-0.888	1.251	1.940**

4.1.2 Narrative Features with Macroeconomic forecasting results

This section evaluates the performance of forecasting models across different horizons, considering the integration of macroeconomic and narrative data using two methods: Method 1 and Method 2 adopted by Hong et al. (2024). The analysis focuses on improvements in predictive accuracy, measured by relative root mean squared errors (rRMSE) and improvement percentages (IMP %), with comparisons being made against the macroeconomic rRMSE baseline for each horizon.

Method 1 directly combines macroeconomic and narrative features into a single dataset, leveraging their complementary strengths. Models trained on this integrated dataset are expected to capture both the structural trends in macroeconomic variables and the contextual insights from narratives.

Method 2 adopts a two-step residual modelling approach. First, a macroeconomic model generates forecasts, and residuals (forecast errors) are computed. These residuals are then modelled using narrative data, with the final forecast being the sum of the macroeconomic predictions and the narrative residual forecast.

Table 4 Performance of forecasting inflation with Macroeconomic data and Narrative features combined. Train-test split with 70/30. Core CPI as target variable

Horizon	Methods	ElasticNet rRMSE	IMP %	LASSO rRMSE	IMP %	RF rRMSE	IMP %	Combo rRMSE	IMP %
3	Macro (Baseline)	0.774	-	0.774	-	0.885	-	0.754	-
3	Narrative	0.943	-21.83%	0.942	-21.71%	0.947	-7.01%	0.927	-22.94%
3	Combined (Method 1)	0.682	**11.89%**	0.682	**11.89%**	0.791	10.62%	0.696	**7.69%**
3	Combined (Method 2)	0.744	3.88%	0.744	3.88%	0.735	**16.95%**	0.739	1.99%
6	Macro (Baseline)	0.889	-	0.887	-	0.853	-	0.824	-
6	Narrative	0.985	-10.80%	0.985	-11.05%	0.983	-15.24%	0.967	-17.35%
6	Combined (Method 1)	0.675	**24.07%**	0.673	**24.13%**	0.792	7.15%	0.685	**16.87%**
6	Combined (Method 2)	0.735	17.32%	0.735	17.14%	0.683	**19.93%**	0.716	13.11%
9	Macro (Baseline)	1.115	-	1.107	-	0.870	-	0.977	-
9	Narrative	1.007	9.69%	1.007	9.03%	0.992	-14.02%	0.979	-0.20%
9	Combined (Method 1)	0.877	21.35%	0.876	20.87%	0.833	4.25%	0.836	14.43%
9	Combined (Method 2)	0.76	**31.84%**	0.760	**31.35%**	0.701	**19.43%**	0.739	**24.36%**
12	Macro (Baseline)	1.466	-	1.453	-	0.928	-	1.251	-
12	Narrative	1.033	29.54%	1.033	28.91%	1.111	-19.72%	1.035	17.27%
12	Combined (Method 1)	1.167	20.40%	1.167	19.68%	0.896	3.45%	1.053	15.83%
12	Combined (Method 2)	0.821	**44.00%**	0.821	**43.50%**	0.762	**17.89%**	0.800	**36.05%**

At shorter horizons, the Macro Baseline demonstrates strong performance, often outperforming narrative-only, reflecting declines of over 21%. This suggests that narratives alone may add noise rather than value in short-term forecasts, where macroeconomic indicators dominate. However, at longer horizons, particularly beyond 9-12 months, narratives contribute meaningful information that enhances forecasts beyond macroeconomic indicators alone. This aligns with the idea that narratives help capture slow-moving economic shifts that may not yet be reflected in traditional macroeconomic variables. Method 1, however, outperforms the Macro Baseline across most models at shorter horizons. The lower improvements suggest that the residual modelling approach in Method 2 may be less effective for capturing short-term dynamics.

At Horizon 6, Method 1 continues to excel, achieving its highest improvements for ElasticNet. In contrast, Method 2 exhibits steady gains across all models, with RF achieving the largest improvement of 19.93%.

In line with Hong et al. (2024), the results indicate that narratives contribute more significantly at longer horizons, with Method 2 consistently outperforming macroeconomic baselines at 9 and 12 months. However, unlike the mentioned paper where the predictive power of narratives remained modest, this study finds stronger improvements- especially with the Combo model. This could be attributed to differences in preprocessing, where this study applies a more refined LDA filtering process, ensuring that only the most relevant topics contribute to forecasting. Additionally, differences in dataset sources (e.g., The Guardian articles) may provide a more representative measure of economic narratives, potentially explaining the stronger predictive performance observed.

Method 1 also performs well at horizon 9, particularly for Elastic Net and Combo with improvements of roughly 21% and 14%, respectively. However, its effectiveness diminishes at Horizon 12, where it underperforms Method 2 across most models. For instance, at Horizon 12, Method 1 achieves an rRMSE of 1.167 for Elastic Net Method 2's 0.821 which highlights the limitations of direct feature combination at extended horizons.

Among models, Elastic Net, LASSO and Combo consistently deliver robust performance. Elastic Net, a linear model well-suited for high-dimensional data, achieves its highest improvements in Method 2 at longer horizons, with rRMSE values of 0.760 and an improvement of 31.84% and 0.821 for horizon 9 and an improvement of 44% for horizon 12. The Combo model, which averages predictions across multiple methods, exhibits strong and consistent gains across all horizons, particularly in Method 2 where it reaches an improvement of 36.05% underscoring its ability to integrate diverse predictive signals effectively.

Random Forest (RF) shows strong performance at shorter horizons in Method 1 but excels in Method 2 at longer horizons. For instance, at Horizon 9, RF in Method 2 achieves an rRMSE of 0.701 (IMP % = 19.43%). These results demonstrate RF's ability to capture nonlinear relationships, particularly when narratives are modelled as residuals.

The approach of integrating macroeconomic and narrative data was previously explored by Hong et al. (2024), who found that residual modeling (Method 2) significantly improved long-term forecasting accuracy. Consistent with their findings, this study also demonstrates that Method 2 outperforms alternative approaches at longer horizons, particularly at 9 and 12 months. However, a key difference is that Hong et al. (2024) found limited improvements for shorter horizons, whereas in this study, the integration of macroeconomic and narrative features (Method 1) also shows gains at 3- and 6-month horizons, particularly for Elastic Net, Lasso and Combo models. This suggests that the predictive power of narratives may vary depending on dataset composition and preprocessing techniques, emphasizing the role of data-driven feature engineering.

Overall, these findings reinforce the value of integrating macroeconomic and narrative-based forecasting approaches, supporting the conclusions of Hong et al. (2024). However, this study extends their work by demonstrating that machine learning models- particularly Random Forest and Combo- can offer substantial predictive gains across both short and long horizons when properly tuned. Additionally, the results emphasize that while narrative data alone may be less effective for short-term forecasting, its integration through residual modeling (Method 2) enhances inflation predictability over extended periods. These insights contribute to the growing body of research exploring the role of economic narratives in forecasting and highlight the importance of optimizing model selection and feature engineering for practical application.

4.2 Topic Importance

After understanding some results, it is also important and interesting to identify topic importance. The value of topics varies with each specific horizon and to evaluate it a random forest regressor was used to study the impact on predictive performance by R^2

reduction when specific topics were excluded from the model as done by Gu, Kelly, and Xiu (2020) and Bali et al. (2020) to measure the relative importance of each topic.

For a specific forecast horizon, the R^2reduction is calculated by setting all values of a given topic, or group of topics, to zero while holding the remaining model estimates fixed.

The set of 70 topics was named manually and then grouped into different categories which can be seen in Appendix B.

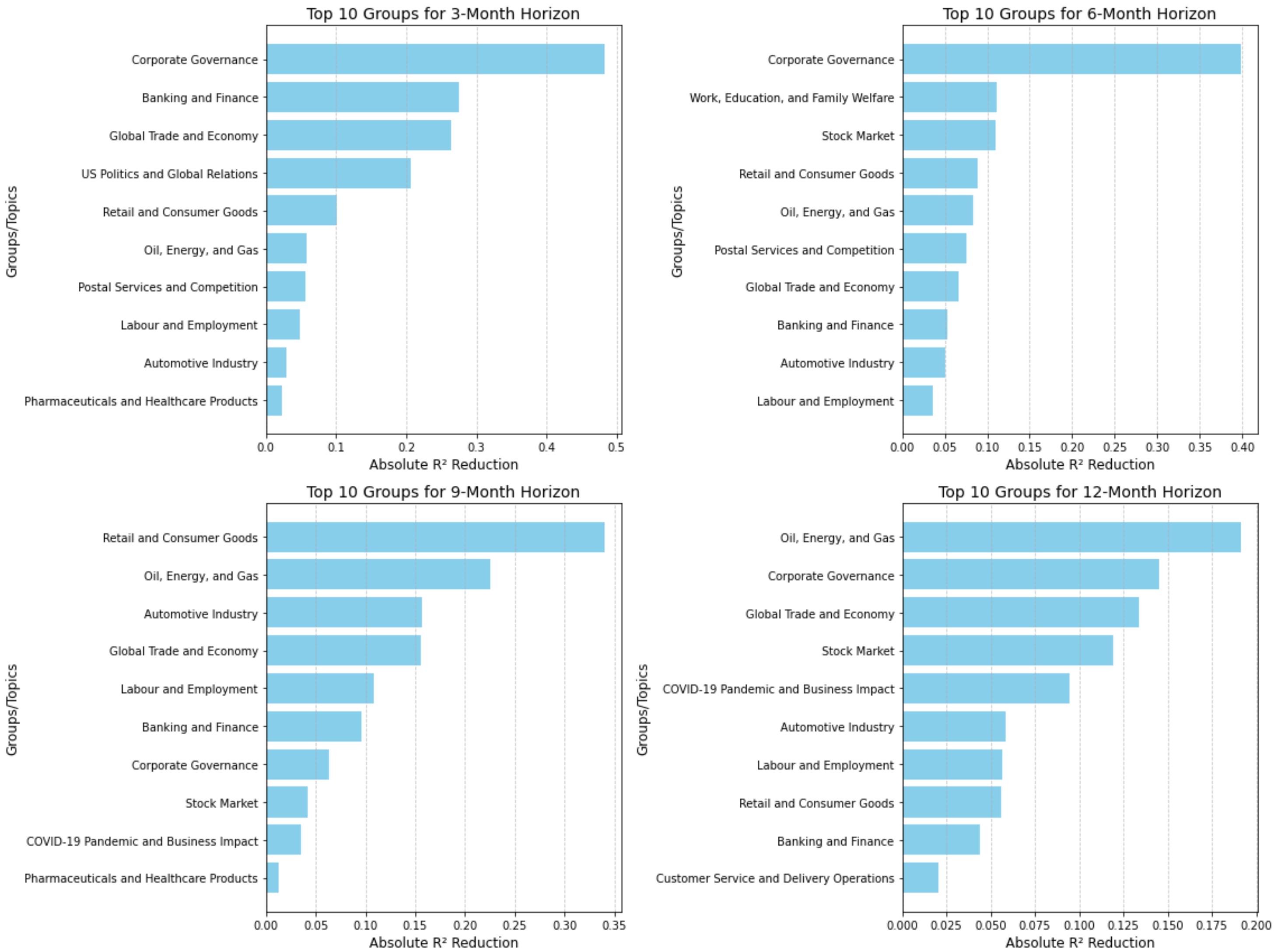

Figure 3 Topic Group's importance for different forecast horizons. This figure shows the top 10 most important categories. For each forecast horizon, it's calculated the reduction in R squared by setting all values of a given group topic to zero in each model.

For the 3-month horizon, Corporate Governance is shown as the most influential group indicating that narratives about governance and decision-making processes strongly influence in short-term inflation. The importance of the financial market and macroeconomic dynamics with categories like Banking and Finance and Global Trade Economy show also a strong impact. Short-term inflation is closely tied to market stability, which corporate governance and financial-sector news can directly impact. Disruptions in corporate practices or banking liquidity immediately affect credit markets and consumer confidence, leading to price volatility.

Moving to the 6-month horizon, Corporate Governance continues to show a strong impact, but Work, Education, and Family Welfare became the second most important group. Narratives around employment and welfare policies show a growing influence on inflation over a more medium-term period. This aligns with Larsen et al. (2021) suggest that changes in labor markets and consumer welfare directly impact medium-term inflation trends.

The 9-month horizon shows the emergence of consumer demand patterns influence on inflation with Retail and Consumer Good category. Also, groups like Oil, Energy and Gas and Automotive Industry showed preponderance in this timeframe.

Finally, over the long-term horizon, Oil, Energy and Gas takes the lead as the most dominant group. This supports the findings of Hooker (2002) and Gospodinov & Ng (2013), who argue that energy price changes have delayed, persistent effects on inflation due to their systemic role in production costs. Other consistent groups like Corporate Governance and Global Trade and Economy remain key contributors but we can see also an importance on the Covid-19 pandemic and Business Impact which is very relevant in this context.

This chapter highlights the importance of understanding the complex factors influencing inflation. Just like Hong et. al (2024) energy-related topics play an important role which is consistent with the existent literature (Hooker (2002); Stock and Watson (2003); Gospodinov and Ng (2013)) but corporate governance also has a strong impact aligning with Bybee et al. (2021), who emphasize how governance failures can influence consumer confidence and market stability, directly affecting inflation expectations. It was clear that across horizons some patterns change, and some categories gain or lose influence emphasizing the need for policymakers and economists to consider sectoral and temporal variations when addressing inflationary trends.

5. Conclusion

The main objective of this research was to continue the work introduced by Hong et al. (2024) by building a hybrid methodology with economic narratives and macroeconomic data for a different country in a distinct environment, United Kingdom. This study aimed to refine their framework by applying a different approach in the forecasting process with a train test split approach and systematically evaluating the role of modeling parameters. This research was motivated by the huge potential of narrative data being an innovative tool to complement traditional macroeconomic indicators, capturing dynamic and non-linear inflationary drivers. This study contributes to existing literature by demonstrating that hybrid models incorporating both narrative features and macroeconomic data can significantly impact positively inflation forecasting accuracy.

Despite using a smaller and differently distributed dataset (as evident from the varying number of The Guardian articles across years), the findings confirm that narrative data provide valuable context, particularly for short- and medium-term horizons. For instance, narrative topics related to energy markets and corporate governance emerged as critical drivers of inflation forecasts. However, this distribution of data throughout the years and the train-test-split may potentially contributed to less favourable comparative results when compared to Hong et al. (2024).

The empirical results reveal that hybrid models consistently outperformed standalone macroeconomic models, with forecasting improvements reaching up to 40% in specific horizons. However, the results varied across horizons, emphasizing the importance of tailoring model configurations to the forecasting timeframe. The study also highlighted the limitations of the dataset, particularly its shorter time span and unequal distribution of narrative sources, which may have constrained the generalizability of findings.

By providing evidence of the value of hybrid models, policymakers and economists can leverage these insights to enhance decision-making, particularly in environments where traditional macroeconomic models struggle to adapt to rapid economic changes and narrative features may be useful.

The limitations expressed suggest that future research should explore the use of larger and more diverse datasets, alternative narrative sources such as social media or policy reports, and advanced machine learning techniques like deep learning or transformer-based models for text analysis.

In conclusion, this study contributes to the growing field of hybrid economic forecasting by refining existing methodologies and demonstrating the value of narrative data. While the results are promising, further exploration is needed to fully unlock the potential of narrative-driven forecasting, particularly in diverse economic contexts and under varying methodological frameworks.

References

Atkeson, A., & Ohanian, L. E. (2001). Are Phillips Curves Useful for Forecasting Inflation? *Quarterly Review, 25*(1), 2–11.

Bali, T. G., Brown, S. J., Murray, S., & Tang, Y. (2020). A lottery-based explanation of the beta anomaly. *Journal of Financial and Quantitative Analysis, 55*(1), 1-37.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Boaretto, R., & Medeiros, M. C. (2023). Disaggregated Inflation Data and Machine Learning Models for Forecasting. *Journal of Econometrics, 227*(2), 130–156.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32.

Bybee, L., Kelly, B., Manela, A., & Xiu, D. (2021). The structure of economic news. *Journal of Financial Economics, 141*(3), 1119–1140.

Coibion, O., & Gorodnichenko, Y. (2012). What can survey forecasts tell us about informational rigidities? *Journal of Political Economy, 120*(1), 116-159.

Coibion, O., & Gorodnichenko, Y. (2015b). Is the Phillips curve alive and well after all? Inflation expectations and the missing disinflation. *American Economic Journal: Macroeconomics*, *7*(1), 197–232.

Diebold, F. X. (2012). Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests. *SSRN Electronic Journal.*

Fan, X., Jin, W., & Fang, K. (2023). High-dimensional forecasting in macroeconomics: A machine learning approach. *Journal of Economic Perspectives, 37*(2), 25-48.

Gospodinov, N., & Ng, S. (2013). Commodity prices, inflation expectations, and inflation forecasting. *The Review of Economics and Statistics, 95*(4), 1346-1359.

Gu, S., Kelly, B. T., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*(5), 2223-2273.

Hong, S., Park, Y., & Lee, J. (2024). Inflation Forecasting Using Economic Narratives: A Hybrid Approach. *Economic Modelling, Forthcoming.*

Hooker, M. A. (2002). Are oil shocks inflationary? Asymmetric and nonlinear specifications versus changes in regime. *Journal of Money, Credit and Banking, 34*(2), 540-561.

Huang, D., Jiang, F., Li, K., Tong, G., & Zhou, G. (2022). Extracting economic signals from high-dimensional text data: A machine learning perspective. *Quantitative Economics, 13*(4), 1225-1253.

International Monetary Fund. (n.d.). Exchange Rate Statistics and Macroeconomic Indicators. Retrieved from https://www.imf.org

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2020). Making text count: economic forecasting using newspaper text. *Bank of England.*

Larsen, V. H., Thorsrud, L. A., & Zhulanova, J. (2021). News-driven inflation expectations and macroeconomic fluctuations. *Journal of Monetary Economics*, *117*, 1-17.

McCracken, M. W., & Ng, S. (2015). FRED-MD: A Monthly Database for Macroeconomic Research. *Federal Reserve Bank of St. Louis Working Paper Series, 2015-012*. Retrieved from https://doi.org/10.20955/wp.2015.012

Medeiros, M. C., Vasconcelos, G. F., Veiga, A. F., & Zilberman, E. (2019). Forecasting Inflation Using High-Dimensional Data: A Machine Learning Approach. *International Journal of Forecasting, 35*(1), 23–42.

Organisation for Economic Co-operation and Development (OECD). (n.d.). OECD Statistics: Composite Leading Indicators and Economic Data. Retrieved from https://stats.oecd.org

Paranhos, D. (2020). Deep Learning for Inflation Forecasting Using Economic Narratives. *Computational Economics*, *56*(4), 801–822.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies, 23*(2), 821–862.

Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature, 41*(3), 788-829.

Surprenant, S. (2022). Media narratives and inflation expectations: The role of news sentiment in shaping economic outlooks. *Journal of Economic Behavior & Organization, 198*, 145-168.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*(1), 267-288.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301-320.

APPENDICES

Appendix A

Table 5 Macroeconomic Indicators preparation for forecasting. On the column "Indicators" the sources of the respective variables are represented with numbers: (1) for IMF, (2) for OECD and (3) for FRED-MD Dataset.

Variable Category	Transformation Applied	Description	Indicators
Real Activity	Log Difference	Applied to indicators like industrial production, labour hours, and unemployment rate. Converts level data to growth rates, stabilizes variance, and ensures stationarity.	(3) Industrial Production Index, (3) Industrial Production for Final Goods, (3) Industrial Production for Consumer Goods, (3) Average Weekly Hours (Manufacturing), (3) Average Weekly Overtime (Manufacturing), (2) Unemployment Rate (%), (2) Labor Compensation in Manufacturing, (2) Hourly Earnings Index, (2) Retail Trade Index
Inflation Variables	Log Difference	Applied to inflation-related metrics (e.g., CPIH, Core CPI). Ensures the data are stationary by analysing proportional changes.	(2) Consumer Price Index, (2) Core CPI, (2) CPI: All items, (3) Producer Price Index for Commodities, (3) Global Commodity Price Index, (3) Global Commodity Price Index, (2) Producer Price Index (PPI), (2) Producer Price Index (PPI) - Commodities, (2) PPI Finished Goods, (2) PPI Crude Materials, (3) Personal Consumption Expenditures Price Index
Financial Variables	Percentage Change	Applied to yields, interest rates, and stock market indices. Captures relative changes in financial metrics over time.	(3) S&P 500 Stock Market Index, (3) Volatility Index (VIX), (3) Federal Funds Rate, (3) 10-Year Treasury Yield, (3) 5-Year Treasury Yield, (3) 1-Year Treasury Yield, (2) 3-Month Treasury Securities Yield, (2) Government Bonds Interest Rate, (2) 3-Month Interbank Rate, (2) Long-Term Government Bond Yields, (3) BAA Corporate Bond Yield, (3) AAA Corporate Bond Yield, (2) 2-Year Fixed Rate Mortgage (75% LTV), (2) UK Private Final

			Consumption Expenditure, (2) Money Supply (M2) Index, (2) Monetary Aggregate (M1)
Other Variables	Percentage Change	Applied to trade indices, oil prices, and exchange rates. Highlights relative movements and trends.	(1) USD to GBP Exchange Rate, (1) Trade-Weighted U.S. Dollar Index, (1) Trade Weighted Exchange Index, (1) Exchange Rate Index, (1) Real Broad Effective Exchange Rate, (3) Oil Prices, (2) Composite Leading Indicators, (2) Population Count

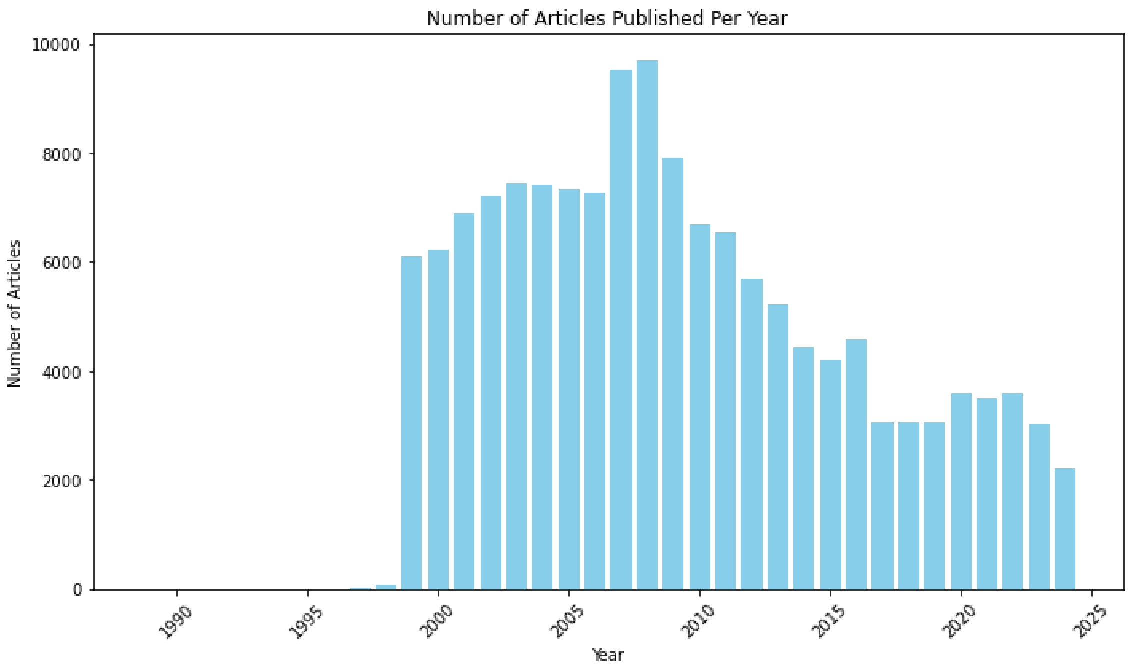

Figure 4 Number of articles published per year

Appendix B

Table 6 Performance of forecasting the next 3-, 6- 9-, and 12- month inflation rates with narratives: Train-test Split with 70/30. LDA with 80 topics. CPIH as target variable

Horizon	LASSO rRMSE	LASSO DM	ElasticNet rRMSE	ElasticNet DM	RF rRMSE	RF DM	Combo rRMSE	Combo DM
3	**0.933**	-1.109	**0.933**	-1.109	1.134	1.567	0.959	-1.317*
6	1.051	1.073	1.051	1.073	1.407	1.761*	**1.017**	0.692
9	**0.952**	-1.737**	**0.952**	-1.737**	1.529	1.860*	0.970	-2.131**
12	0.994	-0.290	0.994	-0.290	1.717	2.257*	**0.991**	-1.056

Table 7 Performance of forecasting the next 3-, 6- 9-, and 12- month inflation rates with narratives: Train-test split with 60/40. LDA with 80 topics. CPIH as target variable.

Horizon	LASSO rRMSE	LASSO DM	ElasticNet rRMSE	ElasticNet DM	RF rRMSE	RF DM	Combo rRMSE	Combo DM
3	**0.935**	-1.148	**0.935**	-1.148	1.080	0.933	0.960	-1.373*
6	1.036	1.074	1.036	1.074	1.242	1.416	1.011	0.677
9	**0.978**	-1.169	**0.978**	-1.169	1.346	1.953*	0.985	-1.515*
12	1.018	1.721*	1.018	1.721*	1.386	1.648*	**1.006**	1.044

Table 8 Performance of forecasting the next 3-, 6- 9-, and 12- month inflation rates with narratives: Train-test split with 60/40. LDA with 80 topics. Core CPI as target variable

Horizon	LASSO rRMSE	LASSO DM	ElasticNet rRMSE	ElasticNet DM	RF rRMSE	RF DM	Combo rRMSE	Combo DM
3	0.929	-2.218**	0.929	-2.211**	**0.918**	-0.933	0.960	-2.423***
6	0.982	-0.920	0.982	-0.920	**0.948**	-0.375	0.990	-1.022
9	1.004	0.254	1.004	0.254	1.033	0.284	**1.001**	0.170
12	1.025	0.962	1.025	0.962	1.165	3.8905***	**1.012**	0.922

Table 9 Performance of forecasting the next 3-, 6- 9-, and 12- month inflation rates with narratives: Train-test split of 70/30. LDA with 80 topics. Core CPI as target variable

Horizon	LASSO rRMSE	LASSO DM	ElasticNet rRMSE	ElasticNet DM	RF rRMSE	RF DM	Combo rRMSE	Combo DM
3	0.942	-1.757**	0.942	-1.749**	**0.930**	-0.728	0.967	-1.933**
6	0.985	-0.769	0.985	-0.769	**0.929**	-0.481	0.992	-0.851
9	1.007	0.456	1.007	0.456	**1.000**	-0.001	1.003	0.393
12	1.033	2.151**	1.033	2.151**	1.121	3.513***	**1.016**	2.111**

Table 10 Performance of forecasting the next 3-, 6- 9-, and 12- month inflation rates with narratives: Train-test split with 60/40. LDA with 70 topics. CPIH as target variable

Horizon	LASSO rRMSE	LASSO DM	ElasticNet rRMSE	ElasticNet DM	RF rRMSE	RF DM	Combo rRMSE	Combo DM
3	**0.935**	-1.148	**0.935**	-1.148	1.188	1.337*	0.960	-1.373*
6	1.036	1.074	1.036	1.074	1.175	1.048	**1.011**	0.677
9	**0.978**	-1.169	**0.978**	-1.169	1.195	1.266	0.985	-1.515*
12	1.018	1.721**	1.018	1.721**	1.332	1.273	**1.006**	1.044

Table 11 Performance of forecasting the next 3-, 6- 9-, and 12- month inflation rates with narratives: Train-test split with 70/30. LDA with 70 topics. CPIH as target variable.

Horizon	LASSO rRMSE	LASSO DM	ElasticNet rRMSE	ElasticNet DM	RF rRMSE	RF DM	Combo rRMSE	Combo DM
3	**0.933**	-1.109	**0.933**	-1.109	1.262	1.434*	0.959	-1.317*
6	1.051	1.073	1.051	1.073	1.453	4.119***	**1.017**	0.692
9	**0.952**	-1.737**	**0.952**	-1.737**	1.504	4.209***	0.970	-2.131**
12	0.994	-0.290	0.994	-0.290	1.713	4.762***	**0.991**	-1.056

Table 12 Performance of forecasting the next 3-, 6- 9-, and 12- month inflation rates with narratives: Train-test slit with 60/40. LDA with 70 topics. Core CPI as target variable

Horizon	LASSO rRMSE	LASSO DM	ElasticNet rRMSE	ElasticNet DM	RF rRMSE	RF DM	Combo rRMSE	Combo DM
3	0.929	-2.218**	0.929	-2.205**	0.948	-0.819	**0.920**	-2.109***
6	0.982	-0.9200	0.982	-0.9200	0.962	-0.499	**0.961**	-0.9794
9	1.004	0.2536	1.004	0.2536	0.976	-0.227	**0.975**	-0.536
12	1.025	0.9623	1.025	0.9623	1.070	1.253	**1.024**	1.269

Table 13 Performance of forecasting the next 3-, 6- 9-, and 12- month inflation rates with macroeconomic data: Train-test split with 60/40; Core CPI as target variable

Horizon	LASSO rRMSE	LASSO DM	ElasticNet rRMSE	ElasticNet DM	RF rRMSE	RF DM	Combo rRMSE	Combo DM
3	0.788	-1.246	0.787	-1.250	0.902	-1.131	**0.785**	-1.481*
6	0.906	-0.363	0.903	-0.376	0.950	-0.398	**0.866**	-0.610
9	1.113	0.495	1.105	0.469	**1.011**	0.0906	1.017	0.094
12	1.425	1.297**	1.412	1.281	**1.138**	2.709***	1.272	1.208

Table 14 Relative Root mean squared errors using Macroeconomic Data and Narrative Features simultaneously. Train-test split with 70/30. Core CPI as target variable

	Method 1				Method 2			
Horizon	ElasticNet	LASSO	RF	Combo	ElasticNet	LASSO	RF	Combo
3	**0.682**	0.682	0.791	0.696	0.744	0.744	**0.735**	0.739
6	0.675	**0.673**	0.792	0.685	0.735	0.735	**0.683**	0.716
9	0.877	0.876	**0.833**	0.836	0.760	0.760	**0.701**	0.739
12	1.167	1.167	**0.896**	1.053	0.821	0.821	**0.762**	0.800

Table 15 Relative Root mean squared errors using Macroeconomic Data and Narrative Features simultaneously. Train-test split with 60/40. Core CPI as target variable

	Method 1				Method 2			
Horizon	**ElasticNet**	**LASSO**	**RF**	**Combo**	**ElasticNet**	**LASSO**	**RF**	**Combo**
3	0.740	**0.736**	0.841	0.743	0.772	0.772	0.**731**	0.766
6	0.744	0.737	0.807	**0.730**	0.772	0.772	**0.690**	0.759
9	0.927	0.921	**0.855**	0.864	0.801	0.801	**0.772**	0.781
12	1.202	1.198	**0.887**	1.060	0.842	0.842	**0.756**	0.826

Appendix C

Table 16 Topic Key Terms. A label was assigned to each topic based on the reading of its content.

Topic Number	**Topic Label**	**Key Terms**
Topic 0	Oil and Gas Industry	oil, shell, company, north, sea, gas, fuel, production, reserve
Topic 1	Fashion Retail Industry	mamps, fashion, spencer, mark, store, retailer, amp, rise, clothing
Topic 2	Employment and Unemployment Trends	job, people, year, unemployment, number, rise, figure, work
Topic 3	COVID-19 Pandemic and Business Impact	pandemic, uk, high, covid19, lockdown, rise, month, business
Topic 4	Corporate Finance and Market Performance	per, cent, share, year, company, market, last, price, analyst, million
Topic 5	Energy and Power Sector	energy, gas, power, price, electricity, nuclear, company, year, bill
Topic 6	Legal and Court Cases	court, claim, case, investigation, legal, action, allegation, charge, law
Topic 7	Greek Economic Crisis	greece, greek, minister, government, debt, bailout, european, finance, athens

Topic 8	Labour Party and UK Economic Policy	labour, brown, government, britain, chancellor, minister, economic, economy
Topic 9	Global Economic Development and Debt	country, world, international, global, imf, develop, poor, debt, development
Topic 10	Inflation, Interest Rates, and Banking Policy	rate, inflation, bank, interest, rise, price, year, cost, high
Topic 11	Stock Market Analysis	share, price, group, market, target, company, analyst, low
Topic 12	UK Budget, Tax, and Public Spending	year, tax, budget, government, spending, public, deficit, cut, chancellor
Topic 13	Banking and Financial Institutions	bank, Barclays, banking, financial, banker, committee, rb, city
Topic 14	Investment Funds and Debt Management	fund, share, company, investor, debt, investment, equity, market, sell
Topic 15	Business Strategy and Public Sector Reforms	business, government, change, need, industry, make, company, new, public
Topic 16	Enron Scandal and Corporate Governance Issues	group, enron, uk, final, schedule, amp, sunday, andersen, agms
Topic 17	Environmental Audits and Pension Reforms	green, say, company, coop, committee, audit, report, bh, pension, railtrack
Topic 18	Mining and Commodity Prices	mine, mining, rio, company, bhp, anglo, commodity, price, miner
Topic 19	Defense Contracts and Aerospace Industry	defence, contract, bae, airbus, company, government, order, uk, system
Topic 20	Supermarkets and Food Retail Industry	tesco, supermarket, store, price, food, year, chain, customer, retailer

Topic 21	Transportation and Rail Services	rail, train, london, transport, city, network, service, year, passenger
Topic 22	Corporate Leadership and Resignations	mr, yesterday, former, chief, executive, company, diageo, chairman
Topic 23	Automotive Industry Trends	car, vehicle, new, uk, year, sale, industry, electric, manufacturer
Topic 24	Shareholder Governance and Executive Decisions	shareholder, company, investor, board, executive, vote, chairman, director, meeting, share
Topic 25	Corporate Profits and Market Performance	year, profit, sale, company, group, share, market, business, rise
Topic 26	Automotive Manufacturing	car, rover, ford, company, plant, year, gm, motor
Topic 27	Media and Journalism	newspaper, black, paper, coffee, magazine, daily, medium, editor, news, mirror
Topic 28	Corporate Fraud and Investigations	company, fraud, tax, account, money, firm, year, former, investigation
Topic 29	Executive Leadership and Governance	executive, chief, director, board, company, chairman, year, group, business
Topic 30	EU Regulations and Corporate Taxation	tax, eu, uk, european, would, commission, rule, country, company
Topic 31	Market Volatility and Brexit Impact	market, brexit, stock, investor, today, pound, uk, point, dollar
Topic 32	Water Utilities and Customer Services	water, company, thames, customer, year, industry, safety, waste, utility
Topic 33	Steel Industry and Global Trade	steel, russian, russia, port, say, tata, uk, company, british, corus

Topic 34	Sports, Clubs, and Beer Sponsorships	club, football, beer, sport, league, brewer, deal, right, united
Topic 35	General Observations and Public Opinions	get, go, one, time, like, think, would, even, look, make
Topic 36	Oil Prices and Market Dynamics	price, oil, market, rise, house, year, high, barrel, increase
Topic 37	BP and Russian Energy Partnerships	oil, bp, company, gas, energy, bps, russian, russia
Topic 38	Wages and Minimum Wage Policies	worker, pay, wage, drink, wine, direct, sport, work, minimum
Topic 39	Mergers, Acquisitions, and Corporate Branding	company, deal, business, group, market, firm, bet, year, brand
Topic 40	Financial Regulation and Investment Banks	financial, firm, goldman, regulator, market, fsa, investment, morgan, wall
Topic 41	Pharmaceuticals and Healthcare Products	drug, food, product, company, year, pharmaceutical, use, treatment, farmer
Topic 42	European Economy	germany, france, eurozone, spain, european, german, ireland, europe, italy, country
Topic 43	China and Global Trade	trade, china, chinese, export, year, tariff, import, japan, global
Topic 44	Internet Services and Digital Business	service, internet, customer, bt, company, online, new, business, broadband
Topic 45	Postal Services and Competition	competition, mail, royal, commission, government, deal, service, company
Topic 46	Labor Unions and Worker Strikes	union, strike, worker, action, member, pay, dispute, staff, unite

Topic 47	Government Pension and Tax Schemes	government, pension, scheme, tax, business, minister, plan, fund
Topic 48	General Corporate and Business Discussions	year, one, business, make, go, company, people, new, like
Topic 49	Music Industry and Entertainment	music, morrison, emi, company, warner, wpp, group, safeway, buffett
Topic 50	Corporate Takeovers and Mergers	bid, offer, takeover, share, deal, group, shareholder
Topic 51	UK Economic Growth and Quarterly Trends	growth, economy, month, year, uk, quarter, sector, show, rise
Topic 52	Customer Service and Delivery Operations	customer, work, office, service, day, staff, people, driver, delivery
Topic 53	Real Estate and Urban Development	property, london, home, pub, restaurant, new, centre, build, year
Topic 54	Travel and Hospitality	travel, holiday, book, hotel, film, thomas, cook, year, uk
Topic 55	Airline Industry	airline, flight, airport, passenger, ba, air, ryanair, heathrow, carrier
Topic 56	Retail Chains and Consumer Goods	store, chain, retailer, business, group, company, retail, year, sport
Topic 57	Executive Compensation and Bonuses	pay, year, bonus, executive, company, share, salary, receive, last, chief
Topic 58	European Central Banking and Financial Policy	bank, central, euro, ecb, european, policy, eurozone, financial
Topic 59	Job Cuts and Downsizing	job, cut, business, staff, company, uk, would, loss, plan
Topic 60	Supermarkets and Grocery Chains	sainsburys, food, asda, king, group, walmart, sainsbury, supermarket, sale

Topic 61	Work, Education, and Family Welfare	work, people, child, home, school, family, health, university, help
Topic 62	US Politics and Global Relations	president, american, state, world, bush, america, war, washington, political
Topic 63	Stock Markets and FTSE Index	share, company, market, stock, ftse, group, close, yesterday, point, 100
Topic 64	Bonds, Investors, and Financial Markets	market, point, bond, investor, stock, rate, low, index, fall, bank
Topic 65	Mobile Technology and Telecommunications	mobile, phone, technology, company, use, computer, vodafone, software, network, microsoft
Topic 66	Global Economy and Investments	economy, market, world, economic, year, growth, investment, financial, global, one
Topic 67	Hong Kong and Global Business	hong, kong, hsbc, hotel, world, canary, note, wharf, whitbread, game
Topic 68	Retail Sales and Consumer Trends	sale, retailer, retail, christmas, year, consumer, store, high, last
Topic 69	Mortgage and Banking Services	bank, loan, credit, mortgage, banking, financial, lender, market, lloyds

Table 17 Topic Groups. Based on the previous name, categories were created gathering all the topics specific to a single theme

Topic Group	**Associated LDA Topics**
Oil, Energy, and Gas	Topic_0, Topic_5, Topic_36, Topic_37, Topic_18
Retail and Consumer Goods	Topic_1, Topic_20, Topic_56, Topic_68, Topic_59, Topic_67
Corporate Governance	Topic_22, Topic_28, Topic_29, Topic_57, Topic_16, Topic_17, Topic_24, Topic_39, Topic_49

Banking and Finance	Topic_13, Topic_14, Topic_40, Topic_58, Topic_69, Topic_7, Topic_30, Topic_47
Automotive Industry	Topic_23, Topic_25, Topic_26
Stock Market	Topic_4, Topic_11, Topic_63, Topic_64, Topic_31
Global Trade and Economy	Topic_9, Topic_42, Topic_43, Topic_66, Topic_52, Topic_33, Topic_50
Labour and Employment	Topic_2, Topic_38, Topic_46, Topic_58
COVID-19 Pandemic and Business Impact	Topic_3
Work, Education, and Family Welfare	Topic_60
US Politics and Global Relations	Topic_62
Customer Service and Delivery Operations	Topic_51
Pharmaceuticals and Healthcare Products	Topic_41
General Observations and Public Opinions	Topic_35
Postal Services and Competition	Topic_45
Travel and Hospitality	Topic_53
Internet Services and Digital Business	Topic_44
Technology and Telecommunications	Topic_65