High-Frequency Inflation Expectations from Big Data: A Natural Language Approach

Firmin Ayivodji*

Universite de Montreal and CIREQ

[Click here for the latest version]

February 1, 2024

Abstract

In this study, I leverage large language models (LLMs) in natural language processing to scrutinize a comprehensive dataset of more than 2 million newspaper articles and 40 million tweets across Canadian provinces. This method is employed to develop novel high-frequency and real-time indicators of consumer inflation expectations at both national and subnational levels. I first identify news articles and tweets related to inflation or prices. Additionally, I apply deep learning methods, particularly LLMs to extract information specifically related to future price dynamics. Then, I construct daily measures of text-based inflation expectations as the difference between the number of news articles or tweets about inflation and the number of news articles or tweets about deflation. The results indicate a high correlation between the resulting text-based inflation expectations indices with consumers' survey-based inflation expectations and realized inflation. Subsequently, I use a mixed-frequency machine learning approach to generate nowcasts/forecasts of quarterly inflation expectations and actual inflation based on large sets of text indicators and Google Trends search volume data for inflation-related terms. The analysis demonstrates that news and social media data contain valuable information regarding inflation dynamics and my newly developed indicators effectively anticipate consumer inflation expectations and actual inflation. The paper further explores the application of Shapley additive explanations (SHAP) values to enhance the interpretability of complex, nonlinear models. The findings suggest that newspaper and social media data can serve as a timely source for market participants and policymakers to elicit beliefs on inflation or future price dynamics.

JEL Classification: E31, C53, C55, D84, E58.

Keywords: inflation expectations, LLMs, news articles, Twitter, big data, mixed-frequency data, machine learning, random forests, neural networks, shrinkage.

*Département de Sciences Economiques and CIREQ, Université de Montréal. Email : firmin.ayivodji@umontreal.ca.
I am deeply indebted to my dissertation committee: Benoit Perron (**chair**), Marine Carrasco, and Christopher Rauh for their invaluable advice and generous help. I would like to thank Maciej Augustyniak, Philippe Coulombe, Tatjana Dahlhaus, Prosper Dovonon, Bruno Fenou, Eric Ghysels, Juste Djabakou, Aristide Houndetoungan, Felicien Goudou, Fansa Kone, Mathieu Marcoux, Jean-François Rouillard, Christian Montes Schütte, Rodrigo Sekkel, Guillaume Sublet, Nicolas Vincent, Tao Wang, and Solo Zerbo, as well as conference and seminar participants at NBER-NSF Time Series Conference, Bank of England, IMF, 1st CIREQ Interdisciplinary Conference on Big Data and Artificial Intelligence, Concordia University, UdeM Department of Mathematics and Statistics Brownbag Seminar, 18th CIREQ PhD Students Conference, HEC Montreal, SCSE, UdeM Macroeconomic Brown Bag, and 3^{rd} GREDI/CREATE/CIREQ PhD Student Research Workshop for extremely constructive feedback and discussions. The work is partly financed by the Fonds de Recherche du Quebec - Societe et Culture (FRQSC) grant.

"[E]xpected inflation feeds directly into the general level of interest rates. Well-anchored inflation expectations are critical for giving the Fed the latitude to support employment when necessary without destabilizing inflation." **Jerome Powell**, Chair of the Federal Reserve, August 2020.

"*Inflation expectations are terribly important. We spend a lot of time watching them.*"
Jerome Powell, Chair of the Federal Reserve, September 2021.

1 Introduction

Inflation expectations play a crucial role in monetary policy as they have a significant influence on the decisions made by households, firms, and financial markets regarding their spending, saving, pricing, wages, and investment (Bernanke et al., 2007). These decisions, in turn, have a bearing on the economy as a whole and hence on actual inflation. The recent surge in realized inflation in Canada and other developed economies raises new questions about the anchoring of inflation expectations and the consequences of shifting expectations for economic activity. Incorporating inflation expectations into macroeconomic models and central banks' formulation of their inflation expectations is essential for developing monetary policy (Stock and Watson, 2007).

However, accounting for the different ways inflation expectations are formed across various economic agents is a complex exercise (Coibion et al., 2018a). For households, expectations are closely tied to their perception of inflation (Weber et al., 2022). According to survey evidence, households gather information about current or future price levels from personal shopping experiences, as well as through social media, television/radio news, and newspapers (Kumar et al., 2015; Cavallo et al., 2017; Adämmer and Schüssler, 2020; D'Acunto et al., 2021). Existing measures of consumer inflation expectations are inherently low-frequency indicators and often come with significant delays. Therefore, there is an urgent need to develop high-frequency and real-time indicators of inflation expectations to address these challenges and discrepancies. This need is particularly pronounced during severe economic downturns, such as the COVID-19 pandemic and the 2008 financial crisis (Lewis et al., 2022). Additionally, as highlighted by Weber et al. (2022), the nuances of survey methodologies from the wording of questions to unintentional priming effects and limited sample sizes can introduce biases and distort results.

In this paper, I develop novel high-frequency and real-time consumer text-based inflation expectations indices at both subnational and national levels in Canada. To achieve this, I use a comprehensive dataset comprising over 2 million newspaper articles and 40 million tweets from various Canadian provinces. My methodology involves five key steps. First, I compile all news articles from a database that mentions specific regions in their titles and texts, spanning from January 1977 to June 2023, following the methodology used by Mueller and Rauh (2018). Second, I collect English and French tweets posted between January 1, 2008, and June 30, 2023, containing selected keywords related to inflation or expected price dynamics for each major city or province in Canada, utilizing Twitter's geo-localization information (Jiao et al., 2020; Angelico et al., 2022).

Subsequently, I deploy large-scale pre-trained neural network language models, specifically the Robustly Optimized BERT Pre-training Approach (RoBERTa), to ensure the dataset includes only relevant articles and tweets that address inflation or prices while excluding off-topic material (Liu et al., 2019; Ash and Hansen, 2023). I then apply advanced deep learning methods, particularly large language models (LLMs) to isolate information specifically related to future price dynamics. I further combine RoBERTa and dictionary-based algorithms to determine the direction of inflation or prices mentioned in the text, distinguishing between rising and falling trends. Using these insights, I construct daily measures of text-based inflation expectations as the difference between the number of news articles or tweets about inflation and the number of news articles or tweets about deflation.

To validate my methodology, I evaluate the correlation between my novel text-based inflation expectations, consumer inflation expectations, and actual inflation. I also propose an out-of-sample prediction approach that combines unrestricted mixed-data sampling (U-MIDAS) with machine learning techniques to handle mixed-frequency and high-dimensional problems. This approach enables me to test the forecasting power of text-based inflation expectations. My forecast target variables include quarterly consumer inflation expectations and actual inflation, with predictions extending up to four quarters ahead. I employ an autoregressive (AR) model as a benchmark model for comparison against a spectrum of machine learning models, ranging from linear techniques such as LASSO, ridge, and elastic net, as well as nonlinear models based on random forests, extreme gradient boosting, and feedforward neural networks. These models are well-equipped to handle large sets of predictors or features. Furthermore, I enhance prediction accuracy by combining linear and nonlinear models and supplement my analysis with economic narratives and comprehensive monthly Google Trends search data for inflation-related terms, ensuring the reliability of my forecasts.

Economic narratives, which encompass the accounts recorded and tales recounted by individuals to interpret and articulate their understanding of the world, are potent tools in shaping economic expectations (Andre et al., 2021; Kendall and Charles, 2022; Flynn and Sastry, 2022). In the economic realm, news generated by both traditional and social media captures pivotal economic events and offers valuable insights into the current economic state (Mullainathan and Shleifer, 2005; Bybee et al., 2021; Ayivodji and Rauh, 2023). Furthermore, these economic narratives significantly influence expectations, thereby impacting macroeconomic outcomes and economic fluctuations (Shiller, 2017, 2020; Larsen et al., 2021; Bertsch et al., 2021; Macaulay and Song, 2022). However, quantifying these narratives poses a considerable challenge. Addressing this challenge, I compile and decompose economic narratives into quantitative and time-varying topics via Latent Dirichlet Allocation (LDA) in the field of natural language processing. Topic models can transfer the high-dimensional unstructured news or tweets into a relatively low-dimensional set of common "topics" (Blei et al., 2003; Calomiris and Mamaysky, 2019). The higher the media's attention to a certain topic, the more likely it is that this topic transfers something for the economic state at that given point in time (Adämmer and Schüssler, 2020). Therefore, the time series of topics provide a quantitative description and measurement of the changing economic narratives.

The findings of the study present several notable insights. Initially, the analysis indicates a strong alignment between the trends captured in both news articles and Twitter-based text indicators of inflation expectations and the existing measures of consumer inflation expectations and actual inflation. To facilitate a direct comparison, the high-frequency indices derived from these unconventional data sources were aggregated into lower-frequency intervals. Over the period from October 2014 to June 2023, the correlation coefficients are telling. The news-based indicator correlates highly with actual inflation, yielding a coefficient of approximately 0.89, and with the household expectations survey at approximately 0.90. In contrast, the Twitter-based indicator shows a marginally higher correlation with the household expectations survey, at around 0.92, for the period from the third quarter of 2014 to the first quarter of 2023. However, it correlates slightly lower with actual inflation, at about 0.87. These results underscore the substantial relevance of the extracted information from news and Twitter feeds, which mirrors the prevailing survey-based expectations and actual inflation figures, thereby affirming their utility as a consistent proxy for the dynamics of inflation or future prices.

The study further discloses that the incorporation of high-dimensional text indicators and Google Trends search volume data significantly enhances the accuracy of forecasting consumer inflation expectations and actual inflation. Models enriched with text-based inflation expectations consistently surpass the AR benchmark model in predictive performance, achieving considerable reductions in the root mean squared error (RMSE). Throughout the out-of-sample period assessed, predictions anchored in text-based indicators invariably yielded superior results when measured against the AR model, with RMSE reductions reaching up to 65 percent. Additionally, the study underscores the robustness of machine learning techniques, with random forests and neural networks demonstrating consistent outperformance over all the prediction horizons tested.

Moreover, the study also assesses the out-of-sample forecasting capability of economic narratives, which closely aligns with real-time forecasting practices. The results indicate that forecasting based on economic narratives yields smaller forecast errors compared to the AR benchmark model in an out-of-sample setting. The above result highlights the non-negligible economic significance of narrative-based forecasts, as policymakers are in great need of an accurate inflation forecast to develop monetary policies in times of economic turmoil.

To better understand how the mixed-frequency machine learning approach improves out-of-sample prediction, I compute variable-importance measures (Molnar, 2020; Buckmann et al., 2022; Medeiros et al., 2022; Borup et al., 2023) for the individual predictors in the fitted models. The variable-importance measures allow us to see which predictors are the most relevant in the fitted models. My analysis reveals that certain economic narratives closely linked to inflation, such as the housing market, monetary policy, interest rates, transportation costs, energy, and cost of living, rank among the most pivotal topics. Interestingly, other significant narratives such as certain political topics, indirectly exert an impact on inflation, eluding capture by conventional methods. This diversity of influential media topics underscores the unique advantage of narrative-based forecasting and partly elucidates the superior performance of economic narratives. Additionally, Google Trends search data for terms associated with inflation or prices (such as home prices, high gas prices, higher inflation, rent, rising house prices, groceries prices, interest rates rise, rise credit, expensive bills) gain marked importance. As anticipated, my novel text-based inflation expectations indices and the lagged data on target variables emerge as key factors.

Furthermore, text-based inflation expectations can be influenced by a myriad of factors, including public sentiment, economic forecasts, and market dynamics. To delve deeper into the nature of these expectations, I differentiate between the perspectives of experts and non-experts. Experts are defined as individuals or entities with specialized knowledge or expertise in economic matters, such as officials from the central bank, economists, traders, real estate agents, and analysts, among others. Non-experts encompass the broader public whose views are shaped by general news consumption and personal experience, rather than formal economic training or professional experience. The remarkably high correlation between the two sub-indices suggests that the delineation between experts and non-experts may not be as stark as traditionally perceived. This convergence could imply a widespread consensus or a shared sentiment about inflation trends among both experts and non-experts.

Finally, I observe significant regional disparities in text-based inflation expectations. Specifically, Alberta, Ontario, British Columbia, and Quebec experienced an initial peak in early 2021, driven by rising prices in essential goods such as groceries and housing-related expenses (rent, mortgage costs, insurance, and so on). In contrast, the Atlantic provinces (New Brunswick, Newfoundland and Labrador, Nova Scotia, and Prince Edward Island) exhibited their first peak toward the end of the second quarter in 2020. These findings align with the research conducted by Miller and Sabourin (2023) and Jain et al. (2022), who analyzed the Bank of Canada's Canadian Survey of Consumer Expectations data and reported disparities in consumers' survey inflation expectations across different provinces.

Overall, my findings have important implications for policymakers and researchers alike, as they offer timely and accurate insights into the formation of inflation expectations across various economic agents and regions. With my news articles and Twitter's high-frequency and real-time indicators, policymakers can make informed decisions and implement effective strategies to manage inflation and stabilize the economy. My results also show the importance of considering regional text-based inflation expectations. The intuition for this result is that it pays to consider a broader set of information than solely that resulting from survey-based expectations that account only for experts' opinions. Furthermore, researchers can use my methodology as a foundation for further investigations into inflation expectations and their impact on economic behavior and outcomes.

Contributions to literature. The present paper contributes to the economic literature in several ways. My primary contribution lies in introducing a new data source for eliciting subnational and national inflation expectations, along with the development of a robust methodology for classifying inflation-related news and Twitter content, and proposing a robust unrestricted mixed-data sampling approach combined with machine learning techniques. This approach offers distinct advantages over conventional measures. First, my approach incorporates a diverse range of individuals and news articles, involving a large sample size that surpasses the limitations of survey-based measures reliant on small samples of agents. Second, the high frequency of the data allows for the construction of daily indicators, providing more timely information compared to the typically available monthly or quarterly polls. Additionally, a notable aspect of my work is the provision of high-frequency and real-time inflation expectations for each Canadian province, offering valuable insights for researchers and policymakers. These contributions are made possible through the use of innovative data sources.

This paper also contributes to developing a methodology for classifying news articles or tweets as including or excluding content on inflation or prices and also for deriving the direction of inflation/prices (rising, falling, or stable) that may be mentioned in the article or tweet. I deploy large-scale pre-trained language models and adjust them both to account for the specific language structure of inflation news and, more importantly, to predict five thousand human-labeled sequences. The RoBERTa algorithm substantially outperforms existing methods in terms of out-of-sample classification accuracy, including the language models that underlie SVM, GPT-4, and ALBERT.

Furthermore, my research aligns with the emerging literature on the use of alternative data sources such as news articles, social media, credit card or online purchases, and internet search volume data to predict macroeconomic variables (Ayivodji, 2021; Medeiros et al., 2021; Chahrour et al., 2021; Barbaglia et al., 2022; Marcellino and Stevanovic, 2022; Ayivodji and Rauh, 2023; Borup et al., 2023). The mixed-frequency machine learning approach complements the work of Babii et al. (2021), who employ the sparse-group LASSO (Simon et al., 2013) to estimate MIDAS models with a large number of predictors. In contrast, I adopt the U-MIDAS approach and consider feed-forward neural networks (FNNs) as well as regression trees (random forests and extreme gradient boosting) to accommodate more general nonlinearities. Moreover, I explore ensemble predictions that combine multiple mixed-frequency machine learning models. These advantages highlight the potential power of news articles and Twitter as sources for extracting agents' expectations.

While machine learning (ML) methods show great capacity for both approximating highly complicated nonlinear functions and forecasting, they are routinely criticized for lacking interpretability and are considered a "black box," in the sense that they do not offer simple summaries of relationships in the data. Recently though, a number of studies (e.g., Wager and Athey (2018), Buckmann et al. (2022), Medeiros et al. (2022), Borup et al. (2023)) have tried to make ML output interpretable. In this article, I also try to understand, which variables have more of an impact on the forecasting performance of the mixed-frequency machine learning approach. To this end, I use Shapley Additive Explanation (SHAP) values, as proposed by Lundberg and Lee (2017), which have started to become a standard tool for interpretability in ML methods.

My work also contributes to the field by demonstrating how machine learning techniques and text analysis, coupled with a semantic approach, can extract meaningful information on macroeconomic variables and expectations from noisy and textual big data. Finally, my research adds to the existing literature exploring the usefulness of social media data in predicting real-world outcomes (Antenucci et al., 2014; Jiao et al., 2020; Gorodnichenko et al., 2021; Angelico et al., 2022; Bianchi et al., 2023) by focusing specifically on subnational and national inflation expectations.

Layout. The paper is organized as follows. In Section 2, I describe the data sources. The methodology for text-based inflation expectations is discussed in Section 3. I compare news, Twitter, and survey-based inflation expectations in Section 4. Section 5 explores how economic narratives are measured using topic models. The forecasting methodology and inference are outlined in Section 6. The main results of my research are presented in Sections 7 and 8. Finally, Section 9 concludes.

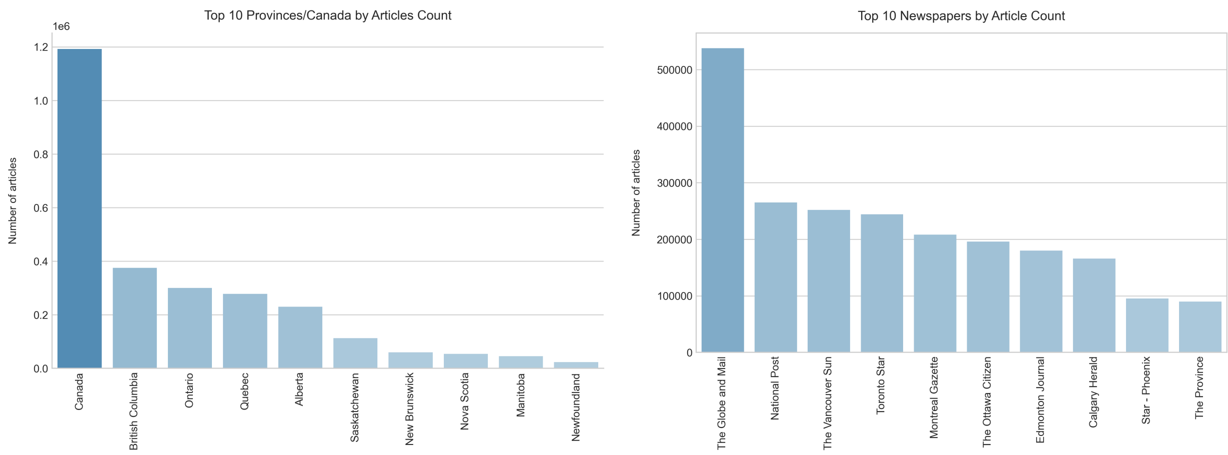
2 Data sources

In this section, I provide a detailed description of the data employed in my study. My analysis uses four primary sources of data: (1) a comprehensive database of Canadian newspapers at a granular level; (2) Twitter data, which includes tweets from the public (consumers, firms, and so on) on price developments; (3) Google search volume data related to prices or inflation; and (4) data from Consumer Survey Inflation Expectations.

2.1 Newspaper text

To collect the news data, I adapt the methodology used by Mueller and Rauh (2018) to the regional level. I download newspaper articles containing one of the province or territory names, or abbreviations thereof, in the title and assign each article to the respective region. The articles spanning from 1977 until June 2023 are downloaded from a database of a wide range of Canadian news sources including the *National Post, Calgary Herald, Edmonton Journal, Montreal Gazette, Ottawa Citizen, Regina Leader-Post, The Globe and Mail, Vancouver Sun, and the Victoria Times-Colonist*.

Another crucial aspect of my data is that it covers local newspapers, and, as a result, it provides an ideal setup to measure inflation expectations at a granular level (for example, city or province).



(a) **Articles distribution by province** (b) **Articles distribution by media source**

Figure 1: **Distribution of news articles by province and media source**

Figure 1a provides an illustration of articles' distribution based on Canadian provinces and the unique category for national news labeled "Canada". The latter, representing articles with a specific mention of "Canada" in their title or body, stands out prominently with a contribution of 1,192,318 articles. On a provincial breakdown, British Columbia reigns supreme at 375,457 articles, closely followed by Ontario and Quebec with counts of 300,303 and 278,422 respectively. In stark contrast, territories such as Yukon, Nunavut, and the Northwest Territories are characterized by minimal article counts, indicating their peripheral coverage in the media landscape. The figure thus serves as an illustrative representation of both regional and overarching national coverage.

Figure 1b sheds light on the distribution of articles by specific news sources. "The Globe and Mail" unmistakably dominates, boasting 537,792 articles. Close competitors "National Post" and "The Vancouver Sun" log 265,246 and 251,981 articles, respectively. Among the enumerated top 10 sources, "The Province" records the smallest number, with a count of 89,780 articles. This illustration provides an overview, revealing the varying degrees of influence each news source holds within the comprehensive Canadian newspaper archive.

2.2 Twitter data: keywords and initial dataset

I use messages, or "tweets," from the social media platform Twitter. A tweet is a brief message limited to 140 characters or less. Once a user writes and sends a tweet, it is distributed to the other users with whom they are linked, known as followers, who may share the message with their followers through a process called "re-tweeting." Tweets typically include news, links, opinions, advertisements, or personal information that the user wishes to disseminate to the public. They are publicly accessible on social networks and can be searched using the platform's search engine.

To build my database of tweets, I select several keywords in English/French related to inflation, prices, and price dynamics and collect all the tweets that contain at least one of them following Angelico et al. (2022). The dictionary of selected keywords is organized into the following categories:

- price, prices, cost of living, tariff: capture tweets about prices in general that do not provide information on price dynamics unless further analyzed;
- expensive bills, inflation, expensive, high prices, high gasoline prices, high bill, high rents, high gasoline price, high petrol prices, high gas prices: reflect certain price dynamics and capture tweets about increasing price(s);
- deflation, disinflation, sales, sale, less expensive, less expensive bills: select tweets about decreasing prices.

The initial dataset is a corpus of about 40 million tweets sent between 01/01/2010 and 06/30/2023, containing at least one of the aforementioned keywords, posted by 3 million unique users. The sample contains the body of the posted tweet and a broad set of metadata related to the tweet and the author: date, time, retweet identifier, hashtags, and detailed information on the user, including geographic location.

The tweets are composed in both English and French. While this diversity is advantageous for distinguishing differences among Canadian provinces, it complicates textual analysis and inter-provincial comparisons. To address this challenge, I convert all French tweets to English through automated translation before further text processing.[1]

Although this set of tweets contains words that are important in the context of inflation/Prices, these keywords could also be used in another context. For example, the word price could likely be related to any form of advertisement or another context that could be considered as noise for my purposes. Therefore, I need to reduce the amount of noise in my data set, as I describe below.

2.3 Google search volume data

GT data is sourced directly from the Google Trends website. As far back as 2006, Google began offering public access to a portion of its search data, dating from 2004, which offers insights into the number of searches conducted for a specific keyword. Essentially, Google Trends operates as an index showcasing the relative popularity of search terms based on geographical region. The number of searches for a keyword (e.g., "inflation") is divided by the total number of searches conducted in Canada over a monthly period. This data is subsequently normalized on a scale ranging from 0 to 100, which represents the proportion of the chosen keyword in relation to all search queries within a designated time frame.

[1] I use the GoogleTranslator package in Python to translate the text.

In my study, I have formulated a high-dimensional set of predictors hinging on monthly GT data for an extensive array of search terms related to inflation and its facets. As a starting point, employing terms like "inflation", "energy prices", and "commodity prices", I turned to the Google Keyword Planner tool. This tool is tailored to provide pivotal terms that can bolster web traffic when integrated into web content. For instance, using the term "inflation" led me to the identification of associated top 15 terms like (1) inflation, (2) energy prices, (3) commodity prices, (4) grocery prices, (5) cost of living, (6) home prices, (7) real interest rate, (8) retail price, (9) rent prices, (10) rising inflation, (11) high gas bills, (12) fuel price, (13) price increasing, (14) price rise, and (15) credit interest rates rising. I have termed these as “primitive terms” since they naturally resonate with the broader public’s perceptions and concerns about inflation.

Building upon this foundation, I further expanded my keyword repertoire by employing GT’s feature, which offers a list of 25 top-related terms for every primitive term. This process not only added specificity pertaining to nuances within the Canadian context but also incorporated a wide spectrum of inflation-related themes. After filtering out duplicate entries, we were left with a rich tapestry of 52 unique terms, all poised to provide a deeper understanding of inflation sentiment in Canada.

Given the constraints of GT in terms of data retrieval for extended periods, I adopted a strategy of accessing data in overlapping blocks spanning seven months each. This segmented approach necessitated precision in data alignment across the blocks. Each GT data segment was adjusted for congruency using the ratio of average search volume from overlapping months. Furthermore, recognizing that each GT data retrieval is a random sample (approximately 1% of total search queries), I sought to minimize sample variability. This was achieved by repetitively requesting the same monthly data ten times and subsequently deriving an average from the retrieved sets, ensuring a more consistent and accurate representation of search trends.

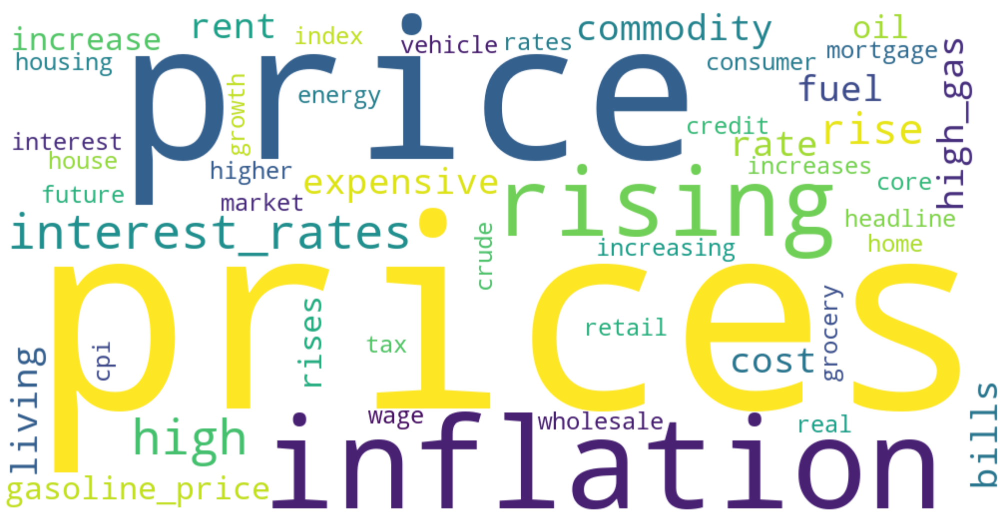


Figure 2: **Google Trends inflation related keywords WordCloud**

Figure 2 presents a WordCloud, visually emphasizing the Google Trends keywords tied to inflation.

A cursory glance immediately reveals that "price (s)" and "inflation" are the most prominent terms, underscoring their significance in public discourse and searches. The prevalence of these words suggests that the general populace is keenly interested in or concerned about price movements and inflationary trends. Other notable keywords include "energy prices," "home prices," "interest rates," and "fuel prices," pointing toward specific sectors or areas of the economy that people associate with inflationary pressures. For instance, the prominence of "energy prices" and "fuel prices" likely reflects concerns about the cost of energy and its impact on daily expenses. Similarly, "home prices" and "rent prices" highlight the housing market's relevance in the inflation narrative. The presence of terms like "rising interest rates" and "consumer price index" (cpi) suggests that individuals are not only searching for the effects of inflation but also the underlying causes and official metrics. Overall, WordCloud offers a comprehensive overview of the public's most pressing concerns and areas of interest regarding inflation.

In Figure 3, The GT Inflation Index for Canada provides a comprehensive visualization of public interest in inflation over the years. Notably, the pronounced peak in the recent period underscores an intensified public concern and attention towards inflation. This surge in interest is a reflection of the real-world situation, where Canada is currently grappling with heightened inflationary pressures. Such an uptick in online searches is indicative of the populace's attempts to understand the implications of this economic challenge, be it in terms of rising commodity prices, increased living costs, or potential policy responses from the government. The magnitude of this recent spike, in comparison to previous years, accentuates the depth of concern and suggests that inflation has become a predominant topic of discourse among Canadians. In the context of the study, this elevated search interest further emphasizes the importance of real-time alternative data sources, like Google Trends, in gauging and responding to rapidly evolving economic conditions.

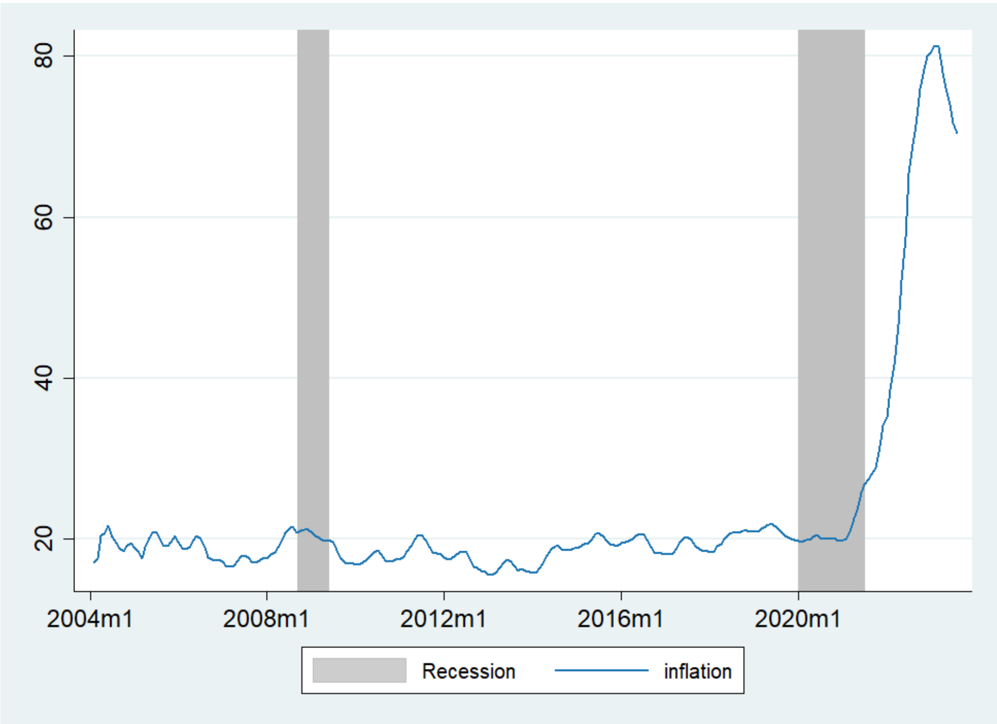


Figure 3: **Google Trends inflation index**

This finding is in line with the analysis in Bracha and Tang (2022), which finds that consumers' attention to inflation is higher when inflation is high, and in Coibion et al. (2018b), which documents that firms spend few resources on collecting and processing information about inflation when they think it is not so relevant to their decisions.

2.4 Consumer Survey Inflation Expectations Data

The data I used for inflation expectations was obtained through an information experiment conducted as part of an online survey. The Canadian Survey of Consumer Expectations is a nationally representative, internet-based quarterly survey of a rotating panel of approximately 2,000 heads of households. It is administered by a large polling firm on behalf of the Bank of Canada. Respondents participate in the panel for up to a year, with a roughly equal number joining and leaving the panel each quarter. This reduces variability caused by changes in composition, allowing for greater stability and precision in the estimates. The survey's target population is adult residents of Canada aged 18 or older. The survey is conducted in February, May, August, and November and is offered in both English and French. Respondents answer questions about inflation, the labor market, household finances, and demographic questions about themselves and their households.

The survey included specific questions related to households' expectations regarding inflation. **One-year-ahead inflation expectations** are based on the following questions:

1. "**Over the next 12 months, do you think that there will be inflation or deflation? (Note: deflation is the opposite of inflation.)**"

   Please choose one.

   - Inflation
   - Deflation (the opposite of inflation)

2. "**What do you expect the rate of [inflation/deflation] to be over the next 12 months? Please give your best guess.**"

   - Over the next 12 months, I expect the rate of [inflation/deflation] by $-\%$.

Between 2021 and 2023, Canadian consumer expectations have been deeply influenced by various factors ranging from the effects of the COVID-19 pandemic to geopolitical events like Russia's invasion of Ukraine. Key takeaways include heightened concerns about inflation, the impact of supply disruptions, and expectations related to wage growth and job markets.

In particular, as highlighted in Figure 4, 2022 marked a significant year where one-year inflation expectations surged to 7.18% by the end of the year. This period also witnessed current inflation perceptions reaching a high of 7.96% in the third quarter. Although there was a decline in these metrics in 2023, they continued to remain at concerning levels with expectations at 5.09% and perceptions at 7.02% by the second quarter. It's essential to emphasize that during the latter part of 2022, the inflation perceptions notably exceeded the one-year expectations, reflecting that consumers were experiencing the pinch of inflation more intensely than they had forecasted.

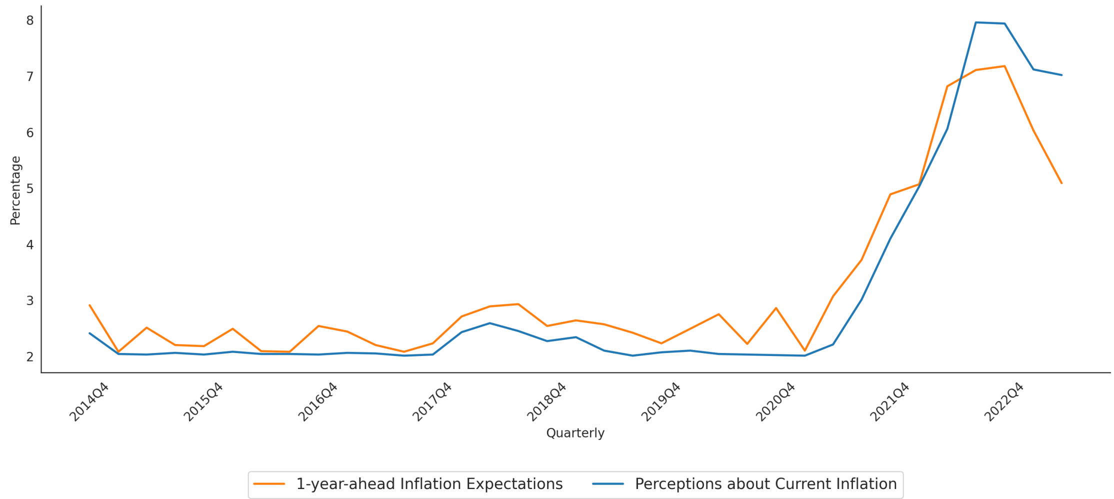


Figure 4: **Consumer's survey inflation expectations**

This sentiment of elevated inflation perceptions has resonated across multiple quarters of the Canadian Survey of Consumer Expectations. For instance, the surveys pointed out factors like the higher cost of living, rising interest rates, and concerns about food prices as prevalent issues among consumers. These concerns, coupled with the uncertainties surrounding global events and their potential impacts on the Canadian economy, led to a cautious and sometimes pessimistic outlook among consumers.

The figure highlights the dynamic shifts in Canadian consumers' inflation expectations and perceptions from the fourth quarter of 2014 to the second quarter of 2023. The sharp rise in one-year inflation expectations during 2022 at 7.18%, was paralleled by an increase in current inflation perceptions, which peaked at 7.96% in the third quarter of the same year. However, as 2023 unfolded, a decline was observed in these metrics, yet they persisted at high levels. A crucial observation from the latter half of 2022 indicates that inflation perceptions substantially exceeded one-year expectations, suggesting a deeper felt impact of inflation by the public than previously forecasted. This alignment, and at times a disparity between perceived inflation and short-term expectations, offers insights into public sentiment and potential future economic behaviors. Ultimately, the sustained high levels of both inflation perceptions and expectations throughout this period highlight the significant inflationary concerns and their potential ramifications on the broader Canadian economy.

3 Methodology: text-based inflation expectations

To derive inflation expectations from large textual data, I implement a carefully constructed four-step process. Initially, I process the raw text to weed out redundancies, ensuring the data adheres to a consistent format. In my second step, I harness the power of Large Language Models, particularly *Robustly Optimized BERT Pre-training Approach (RoBERTa)*, to scrutinize the refined data, and earmarking segments related to inflation or prices. For my third phase, I pivot my attention to identifying text indicative of future price trends, enhancing my analysis with a temporal perspective. Using RoBERTa once more, I categorize pertinent texts into two distinct classifications: "Up" and "Down". This enable me to formulate two distinct indexes based on the daily text frequency. Finally, I aggregate these indexes in directional indicators that increase (decrease) with expectations of increasing (decreasing) inflation.

3.1 Raw text preprocessing

Preprocessing is of utmost importance before textual analysis, especially for researchers working with data gathered from social media and newspaper articles. For instance, people don't speak on the web as they would in normal life and the language used for tweets and posts is noticeably different from the traditional one used in books or newspapers. Users often include hashtags and emojis in their posts to gain visibility and express feelings and they use specific acronyms and abbreviations which are typical of social network language; typos are also quite common. I use a filter to keep tweets in English/French language and exclude any duplicates or retweets.

Before feeding the corpus into my pipeline I conduct standard cleaning and preprocessing procedures. First, each news article is divided into individual sentences to facilitate granular analysis. I then eliminate stopwords, common words like "and" or "but," which add little value to my analysis. Additionally, I apply stemming to words, reducing them to their root form (e.g., "production" becomes "produc") to standardize my text data. Punctuation and digits are also removed to refine the dataset further, ensuring that the focus remains on textual content. Certain tags such as GEP and LOC are excluded from the dataset. To capture the significance of terms in the corpus, I implement Term Frequency-Inverse Document Frequency (TFIDF) weighting. These data cleaning and preprocessing steps collectively enable me to prepare the dataset for in-depth analysis and modeling, ensuring that my findings are both accurate and insightful (see Gentzkow, Kelly and Taddy, 2019 for review).

Next, I use advanced Large Language Models (LLMs), specifically the RoBERTa, to sift through the processed data, pinpointing passages that pertain to inflation or prices. My third step involves isolating text that signals future price movements, thereby infusing my analysis with a forward-looking dimension. I once again deploy RoBERTa to segregate relevant texts into two categories: “Up” and “Down”. This categorization facilitates the creation of two separate indices based on daily text frequencies. Finally, I synthesize these indices into directional indicators that rise (fall) with anticipations of increasing (decreasing) inflation.

3.2 Inflation text binary classification using Large Language Models

In this study, I tackle the challenging task of classifying a large volume of narrative data from news articles and tweets into relevant and irrelevant categories. To achieve this, I employ Natural Language Processing (NLP) methods for classification, with a particular focus on the RoBERTa model. RoBERTa is a self-supervised NLP system based on BERT's language masking strategy. By predicting hidden sections of unannotated text, RoBERTa has proven to achieve extraordinary performance on various NLP tasks (Liu et al., 2019).

The classification process begins with humans labeling a subset of data, which trains the model to understand and replicate the labeling on a larger scale autonomously.

RoBERTa's acclaim in the NLP community stems from its deep learning capabilities, enabling it to comprehend human language nuances. It has been pre-trained on a vast corpus exceeding

160GB, encompassing diverse sources such as Wikipedia, CC-News, OpenWebText, and Stories. This extensive training endows RoBERTa with a comprehensive grasp of English language semantics and syntax. When discerning price change information in media narratives, RoBERTa can interpret even previously unseen vocabulary thanks to this embedded knowledge.

The flexibility of deep learning models like RoBERTa lies in their rule-agnostic nature. Instead of rigidly defining rules, the model learns language patterns from human-labeled sentences, adapting to detect pricing changes without explicit instruction. The bidirectional design of RoBERTa allows it to process entire sentences in unison rather than sequentially, a significant leap from unidirectional models. This holistic view enables RoBERTa to understand multiple meanings of phrases based on context, a feat unachievable by earlier models such as Word2Vec (Mikolov et al., 2013) and Glove (Pennington et al., 2014).

In my study, I aim to harness RoBERTa to classify texts by their content. I curated a dataset of 5000 randomly selected articles and tweets, each annotated with a binary label indicating relevance to inflation or prices. This dataset was then split into a training set (80%) and a test set (20%). The training set was utilized to fine-tune RoBERTa, enabling it to learn from the patterns and relationships within the data. The test set, comprising unseen data, served to assess the model's classification efficacy.

RoBERTa's transformer-based architecture is adept at interpreting semantic meanings and contextual nuances within texts. By fine-tuning the pre-trained model with my dataset, I tailored RoBERTa's parameters to my specific classification task, leveraging its advanced representation learning capabilities for accurate predictions on new text samples.

The results of my RoBERTa experiments were remarkable, demonstrating high accuracy in classifying texts as relevant or irrelevant to inflation or prices. These findings affirm RoBERTa's effectiveness in text classification tasks within large language models, attributed to its sophisticated pattern recognition and contextual understanding.

To validate RoBERTa's classification accuracy, I conducted manual proofreading of 200 news articles. The results corroborated the model's high accuracy, which in some cases, may even surpass that of a well-trained research assistant. The model's proficiency is reflected in the F1 scores presented in Table 1, with an overall classification accuracy of 94.2 percent. [2]

	Precision	Recall	F1
1 - Relevant News	0.95717	0.92739	**0.94294**
0 - Irrelevant News	0.92637	0.95655	**0.94122**
Accuracy			0.94164
Macro Average	0.94177	0.94197	0.94163
Weight Average	0.94212	0.94163	0.94163

Table 1: **RoBERTa Text-Binary Classification Results**

Additionally, as depicted in Figure 5, an analysis of the data highlights disparities in media coverage of inflation across Canadian provinces. Provinces with denser populations and higher economic activity, such as Ontario and British Columbia, exhibit more extensive coverage, indicated by darker

[2] F1 Score $= 2 \times \text{Precision} \times \text{Recall}/(\text{Precision} + \text{Recall})$. See Appendix H for more details on F1 scores.

shades. In contrast, sparsely populated regions like the Northwest Territories and Nunavut receive minimal attention. This geographical analysis of media focus offers crucial insights into the distribution of inflation-related narratives.

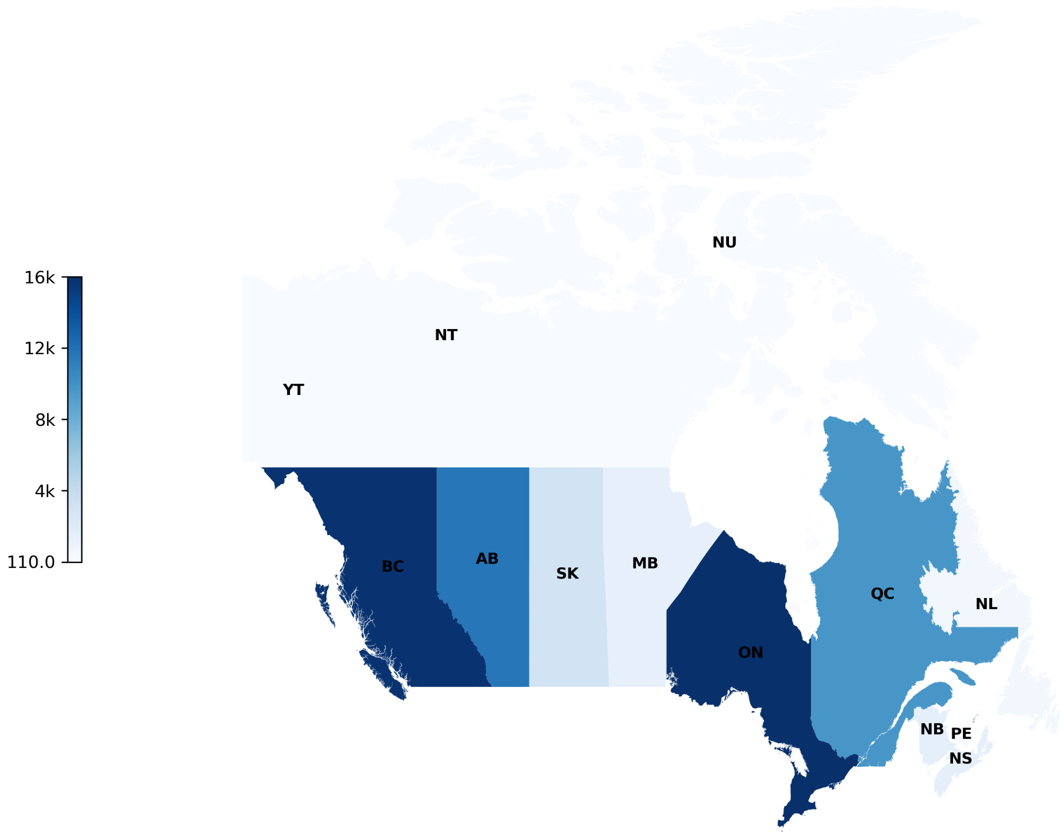


Figure 5: **Inflation Newspapers Coverage Across Provinces**

3.3 Detecting future price dynamics sentences

In this section, I outline my cutting-edge strategy for extracting information relevant to future price dynamics from news articles and tweets. My approach involves dissecting the text into sentences and applying a multidimensional temporal tagging methodology to isolate those that convey forward-looking or future-oriented insights.

3.3.1 Detecting tenses

In this subsection, I detail the methodology employed for tense detection in my textual analysis. The process begins by analyzing the grammatical structure of sentences, identifying the relationships between words, and discerning their temporal context. This is achieved through an advanced algorithm that interprets the grammatical structure as a network of interconnected words. [3]

[3] I deploy spaCy's dependency parsing algorithm to process the grammatical structure of a sentence (see https://spacy.io/usage/linguistic-features).

The core of this analysis hinges on identifying the main verb, or 'root node,' of a sentence. The tense of this main verb often dictates the overall tense of the sentence. For instance, if the main verb is in the past tense, or if it's accompanied by a past-tense helper verb, the sentence is categorized as past-tense. Conversely, a sentence is deemed present tense if the main verb is in the present tense without any past tense or modal auxiliary verbs.

Consider the sentence, "Price hikes are everywhere." Here, the main verb "are" is in the present tense, classifying the sentence as present tense. In contrast, the sentence "Experts predict that inflation will rise further next year." is future-tense because the helper verb "will" is modal, indicating a future action.

However, identifying future tenses can be more complex, as English often expresses future actions using the present tense, supplemented by modal verbs like "will," "shall," or "might." I consider a sentence future tense if the main verb is in the present tense and is accompanied by a modal auxiliary verb.

In the realm of price and inflation communication, certain phrases such as "we expect," "we forecast," or "we predict" commonly occur. Despite being grammatically present-tense, these phrases inherently convey future-oriented meanings. To accurately reflect this, I reclassify such phrases from present to future tense within my analysis. This nuanced understanding of language ensures that my temporal classification aligns with the intended meaning conveyed by the text.

3.3.2 Detecting forward-looking keywords

To create a comprehensive dictionary of forward-looking keywords, I drew inspiration from a blend of authoritative sources, ranging from SEC filings' safe harbor disclosures to the prognostic communications of leading central banks such as the Bank of Canada, the Bank of England, the European Central Bank (ECB), and the Federal Reserve (Fed). [4] These institutions often use specific language to signal expectations or predictions about future events. Common terms like "anticipates," "believes," "estimates," "expects," "intends," "plans," "predicts," "will," and similar expressions are standard in their communications, serving as indicators of forward-looking statements.

The identification of forward-looking statements is crucial, as these entities are meticulous in their language to avoid potential liabilities should their predictions not come to fruition. The legal framework, particularly the Private Securities Litigation Reform Act of 1995, provides a safe harbor for such statements, offering protection as long as they are not misleading and are accompanied by adequate cautionary language.

In enhancing my methodology for detecting forward-looking keywords, I employed state-of-the-art large language models (LLMs), specifically ChatGPT, developed by OpenAI. Large language models, like ChatGPT, represent the pinnacle of contemporary natural language processing capabilities. LLMs, like GPT-4, have become increasingly popular and extremely attractive for a wide range of NLP applications because of their state-of-the-art performance, zero-shot learning, scalability, general-purpose design, and ease of use. Leveraging billions of parameters, these models are trained on diverse datasets, encapsulating a vast range of linguistic nuances and patterns. By incorporating ChatGPT, I not only benefit from its exceptional proficiency in understanding and

[4] see https://www.sec.gov/Archives/edgar/data/846926/000095014403003122/g81231exv99w1.htm.

generating human-like text but also tap into its unparalleled capability to discern subtle contexts and forward-looking linguistic cues that might escape traditional models. This adds a layer of robustness and sophistication to my approach, ensuring that my detection of forward-looking keywords is both comprehensive and precise.

3.3.3 Detecting forward-looking sentences using Deep Learning

The FinBERT-FLS method (Huang et al., 2020) is a cutting-edge text classification model designed specifically to detect forward-looking statements (FLS) at the sentence or paragraph level. Forward-looking statements refer to beliefs and opinions expressed about a firm's future events or results, typically found in financial reports. The ability to accurately identify forward-looking statements from corporate documents can greatly assist investors in conducting thorough financial analysis.

The methodology behind the FinBERT-FLS method involves a sophisticated approach based on the BERT (Bidirectional Encoder Representations from Transformers) family of models. Initially, the model was trained using a dataset consisting of 3,500 manually annotated sentences extracted from the Management Discussion and Analysis section of annual reports of Russell 3000 firms. This dataset served as the foundation for teaching the model to recognize and understand the context and language used in forward-looking statements.

To ensure optimal performance, the FinBERT-FLS model underwent a fine-tuning process carried out in-house by JSL (name of the organization/department). During this phase, particular attention was given to addressing low-performing examples, allowing the model to refine its understanding of complex financial language and nuances specific to forward-looking statements.

By leveraging the power of BERT and its large language model capabilities, the FinBERT-FLS method demonstrates remarkable accuracy in identifying forward-looking statements within financial reports. This breakthrough model has the potential to revolutionize the way investors approach financial analysis, providing them with valuable insights into a company's future prospects and aiding in informed decision-making.

To illustrate the practical application of the FinBERT-FLS method, consider its use in analyzing statements related to inflation. For instance, when the sentence "We expect the inflation rate to increase in the upcoming quarter" is processed through the FinBERT-FLS model, it is adeptly categorized as a 'Specific Forward-Looking Statement'. This exemplifies the model's proficiency in identifying and interpreting anticipatory inflation narratives. For a detailed breakdown of the operational process in Python, refer to the appendix I.

3.4 Tense Aggregation

Having detailed the three distinct methods to identify future price sentence-oriented, it becomes crucial to aggregate these findings into a cohesive understanding of the text's temporal orientation, especially when discussing future price dynamics.

To address the variance in results produced by the trio of methods, I adopt a streamlined aggregation technique. Specifically, for each sentence under scrutiny: if any of the methods flag the sentence as future-oriented (i.e., award it a value of 1), it is treated as a future reference, irrespective of the results from the other methods. This implies that a sentence is categorized as referencing future price dynamics even if only one of the three methods discerns a future orientation.

This method of aggregation capitalizes on the unique strengths of each individual technique, enhancing the probability of detecting genuine future references while minimizing potential oversights. This unified approach provides a comprehensive view of the manner in which future price dynamics are broached, whether in news articles or tweets.

3.5 Computation of aggregate directional indicators

A second supervised machine learning model using RoBERTa Large Language Models Classification presented above is trained to determine the direction of the price trends mentioned in the articles and tweets sentences. I then generate directional indicators, which point towards increasing or decreasing inflation expectations. As there is no straightforward way to do this, I propose here two indicators and check how each of these performs:

1. **Indicator 1**: I compute the difference between the two indexes, $\pi_0^e = (\# \text{ Up} - \# \text{ Down})$. The resulting index is smoothed using a (backward-looking) moving average (MA) of 30 and 60 days. This smoothing aims to capture the idea that most likely it is not just the single piece of information received in a day that is important for inflation expectations, but also information obtained in the recent past.

2. **Indicator 2**: I compute the following indicator: $\pi_{ln}^e = (\ln(\#\text{Up} + 1) - \ln(\#\text{Down} + 1))$. In this case, as I am taking the log, extreme values affect the indicator less. Nevertheless, the resulting indicator is smoothed using a (backward-looking) MA of 30 and 60 days.

All these indicators are based on the difference between Index Up and Index Down and reflect the intuitive idea that when there are more (fewer) news and tweets about higher inflation than there are about lower inflation, then the overall signal should be of increasing (decreasing) inflation expectations.

The methodology described above is summarized in Figure 6. This figure highlights the multi-level filtering process, leveraging a cutting-edge natural language approach to develop text-based inflation expectations. The workflow is adaptable and can be applied to other countries and extended to both housing and labor market expectations, for instance.

4 Benchmarking granular textual expectations against hard data

To ascertain whether the suggested News/Twitter-based inflation expectation indicators are capturing inflation expectations, I compare them with both survey-based and CPI inflation measures. While survey-based measures offer greater precision, their availability is restricted to a lower frequency, specifically on a quarterly basis. Conversely, CPI inflation data is accessible at a monthly frequency. To facilitate a meaningful comparison, I convert my high-frequency indicators into lower-frequency data points, aligning them with the periodicity of the traditional inflation measures.

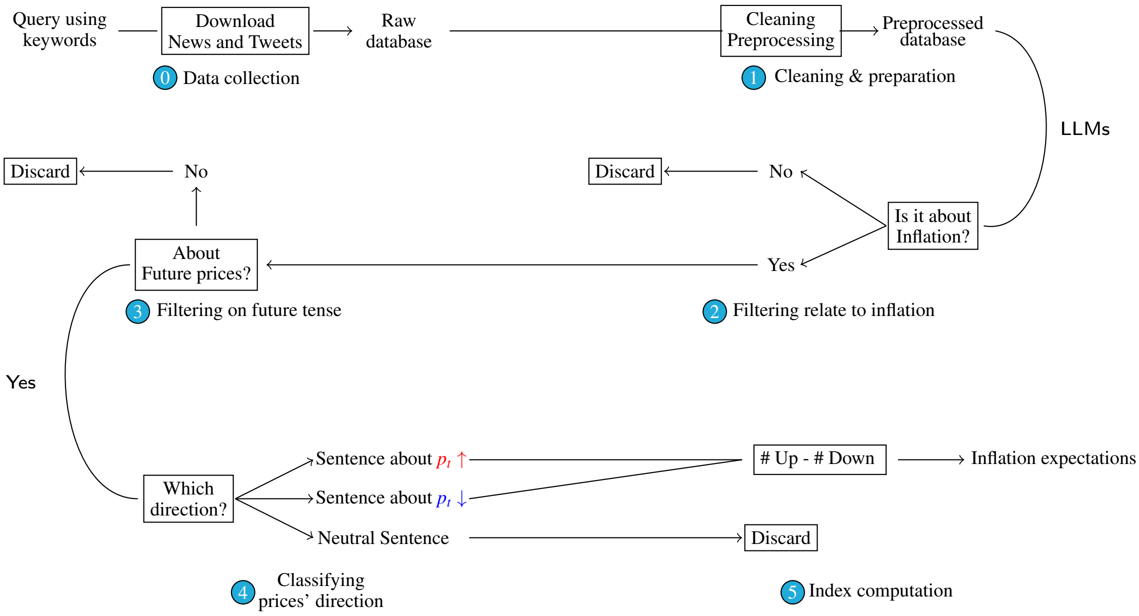
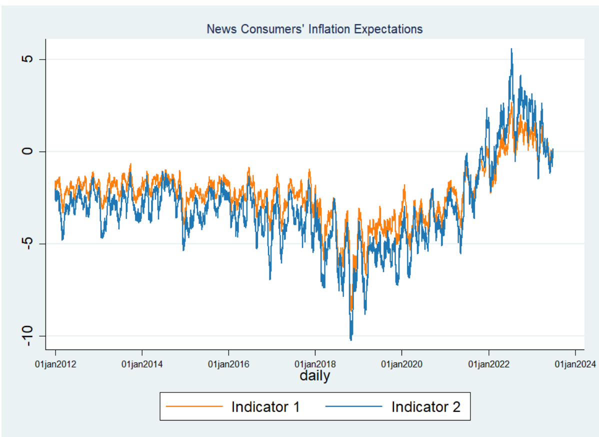


Figure 6: **Overview of the text-based inflation expectations pipeline**

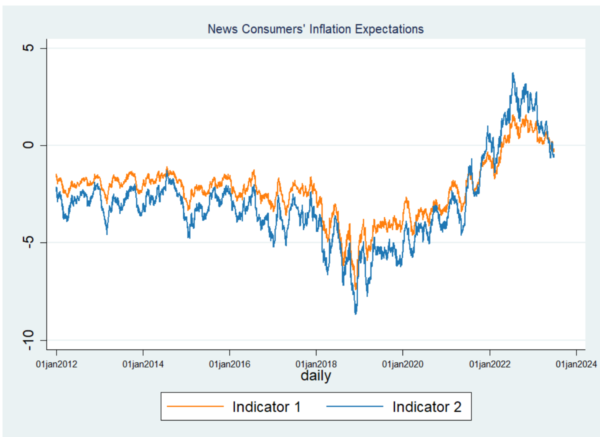
Notes: The figure describes the six major steps used, from data collection to the construction of the indicator. Details of each step are provided in the subsections above.

4.1 News-based inflation expectations

Figures 7a and 7b illustrate the two standardized inflation expectation indicators harnessed from news sources, using both the 30-day and 60-day moving averages.[5] The exceedingly high correlation (approximately 0.97) between these two indicators affirms that the method of aggregation, whether it be 30 days or 60 days, yields consistent and aligned results.



(a) **News inflation expectations MA(30)**



(b) **News inflation expectations MA(60)**

Figure 7: **Daily News-based inflation expectations over time**

A closer examination of the figures reveals that the indicators based on news effectively mirror inflation trends. This is particularly noticeable during periods of high inflation, showcasing their

[5]All measures are normalized to have mean equal to zero and standard deviation equal to one.

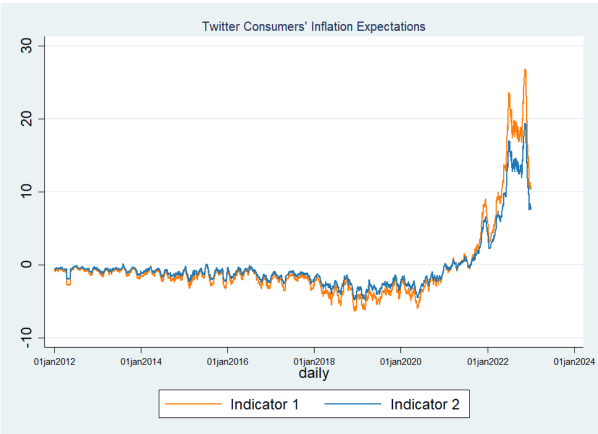
effectiveness in capturing economic developments. The consistent trends observed in both indicators, regardless of the moving average period, underscore their reliability and robustness in tracking price movements.

Furthermore, their alignment with actual inflationary patterns underlines the enduring influence of traditional news outlets in shaping economic anticipations. It reiterates the continued relevance of these news sources and their adaptability amidst evolving data sourcing techniques.

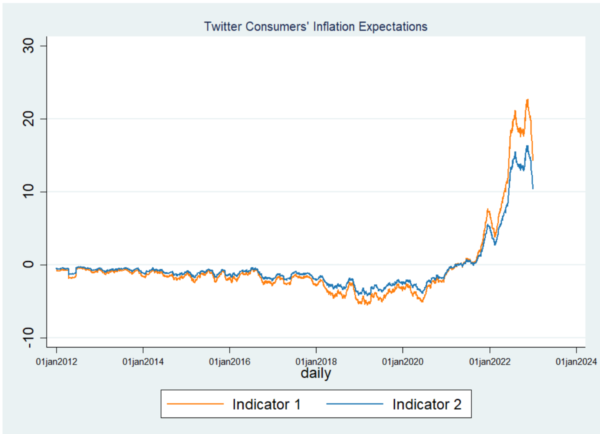
To sum up, these news-based inflation expectation indicators amplify the merit of traditional news sources in economic forecasting. Their ability to timely and precisely delineate inflationary patterns renders them invaluable for economists, market experts, and policymakers striving for a thorough comprehension of inflationary tendencies.

4.2 Twitter-based inflation expectations

Figures 8a and 8b depict two standardized inflation expectation indicators derived from Twitter data, using 30-day and 60-day moving averages respectively.[6] The strikingly high correlation between them (approximately 0.98) insinuates that the chosen aggregation method exerts minimal influence on these indicators.



(a) **Twitter inflation expectations MA(30)**



(b) **Twitter inflation expectations MA(60)**

Figure 8: **Daily Twitter-based inflation expectations over time**

A closer inspection of these charts reveals that the Twitter-based inflation expectation indicators adeptly capture inflationary fluctuations, particularly evident during periods of inflationary spikes. This observation remains consistent, irrespective of whether a 30-day or 60-day moving average is considered, thereby attesting to the robustness and reliability of these indicators.

Moreover, the overarching trend of both indicators remains largely congruent regardless of the moving average period employed. This further underscores the notion that these Twitter-based metrics are in sync with recent price dynamics in Canada, presenting a pivotal tool for anticipating inflationary trends leveraging real-time social media data.

In conclusion, by faithfully mirroring inflationary movements, these indicators substantiate the growing significance of social media data sources like Twitter in contemporary economic analysis.

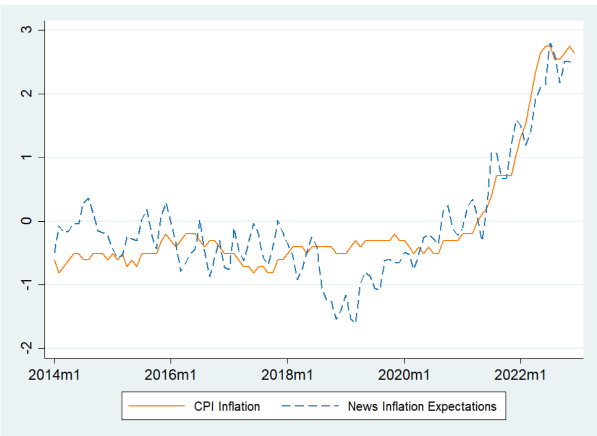
[6] All measures are normalized to have a mean equal to zero and a standard deviation equal to one.

For analysts, policymakers, and market participants, these findings emphasize the immense value and potential of alternative data in effectively gauging inflationary pressures.

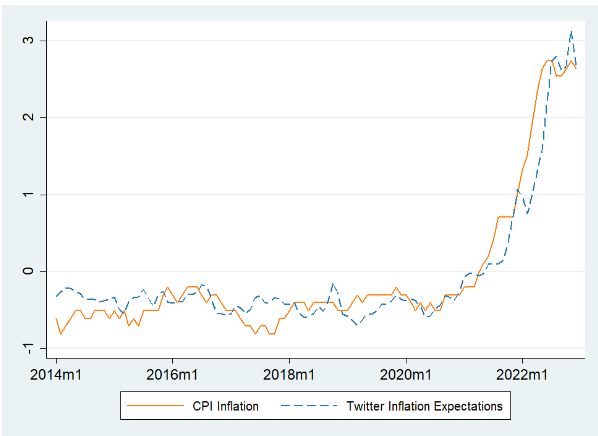
4.3 Comovement between actual inflation and media expectations

In assessing the comovement between the actual Consumer Price Index (CPI) inflation and inflation expectations inferred from media sources, I observe a multifaceted understanding of inflationary dynamics in today's information-rich environment. Figure 9a showcases a discernible alignment between the CPI trajectory and the "News Inflation Expectations". This observable correlation underscores the integral role that journalistic narratives play in both influencing and reflecting the prevailing macroeconomic sentiments.

Turning my attention to Figure 9b, which contrasts CPI with "Twitter Inflation Expectations", the congruence between these data sets is palpable. The parallel movements between the structured CPI metrics and the more fluid Twitter-derived sentiments emphasize the prompt responsiveness of social media platforms in capturing the zeitgeist of real-time economic shifts. Nevertheless, it's pertinent to note periodic deviations between these metrics. A nuanced examination of tweets, especially during these outlier periods, might shed light on exogenous events or narratives that might momentarily skew the correlation.



(a) **News Inflation Expectations and CPI inflation**



(b) **Twitter Inflation Expectations and CPI inflation**

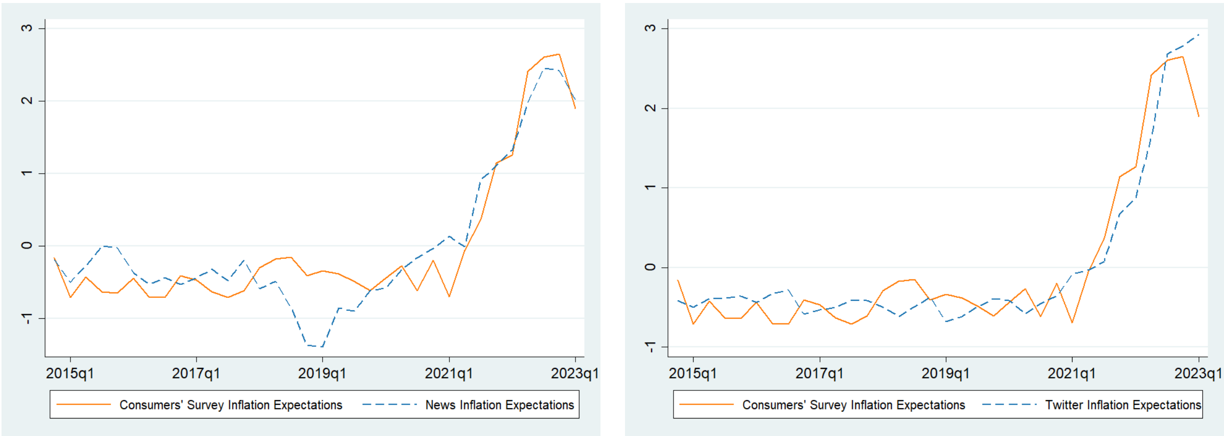
Figure 9: **Comovement between actual inflation and media expectations**

Despite the intermittent disparities, the overarching alignment between the CPI and both sets of media-driven expectations reinforces the significance of integrating high-frequency, real-time data into contemporary economic frameworks. As global economies grapple with fluctuating inflationary pressures, the immediacy and granularity offered by these alternative data sources are paramount. This confluence of traditional economic measures with burgeoning digital narratives affirms that both conventional news mediums and emergent social platforms are pivotal in gauging public economic sentiment. For policymakers and economic scholars, these insights suggest that embracing a diverse data-driven approach, which incorporates both traditional and digital narratives, can yield a more rounded, timely, and anticipatory economic analysis.

4.4 Comovement between consumer inflation expectations and media expectations

Figures 10a and 10b provide an insightful exploration into the dynamic interplay between consumers' survey inflation expectations and text-based inflation expectations.[7] The notable correlation, manifesting in coefficients of 0.91 and 0.94, accentuates the reliability and significance of these innovative indicators. Their real-time, high-frequency nature makes them especially relevant for capturing swift changes in price dynamics and inflationary directions.

Figure 10a demonstrates the enduring impact of traditional news media in influencing consumers' inflation expectations. It highlights that, even amidst a rapidly evolving media environment, classic news outlets continue to play a critical role in molding public perceptions about inflation. On the other hand, Figure 10b showcases the rising significance of social media platforms like Twitter, illustrating their effectiveness in reflecting immediate changes in inflation expectations and the broader economic conversation surrounding inflationary movements.



(a) **News and Consumer's Survey Expectations** (b) **Twitter and Consumer's Survey Expectations**

Figure 10: **Comovement between consumer inflation expectations and media expectation**

For market participants and policymakers, the implications of these findings are profound. The real-time and high-frequency nature of these text-based indicators ensures that they're consistently ahead of the curve, capturing nuances in inflation expectations as they unfold. By integrating these indicators into their analytical arsenal, stakeholders can harness a more immediate pulse on inflationary trends, facilitating timely and strategic responses to market shifts. In essence, the figures underscore the transformative potential of these text-based indicators, positioning them as invaluable tools for anyone keen on decoding the intricate ebbs and flows of price expectations in today's fast-paced economic environment.

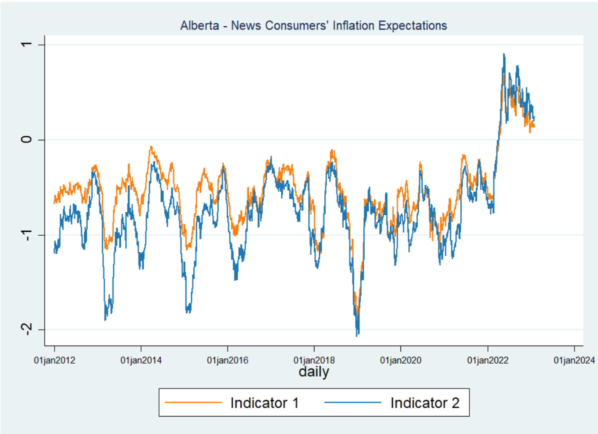
4.5 News-based inflation expectations across provinces

Leveraging local newspaper coverage, my data uniquely positions us to gauge inflation expectations at an unparalleled granular level, spanning cities to provinces. My expansive sample encompasses local newspapers from all 13 Canadian provinces and territories. These sources, deeply ingrained in their respective communities, provide rich and relevant insights into regional economic landscapes.

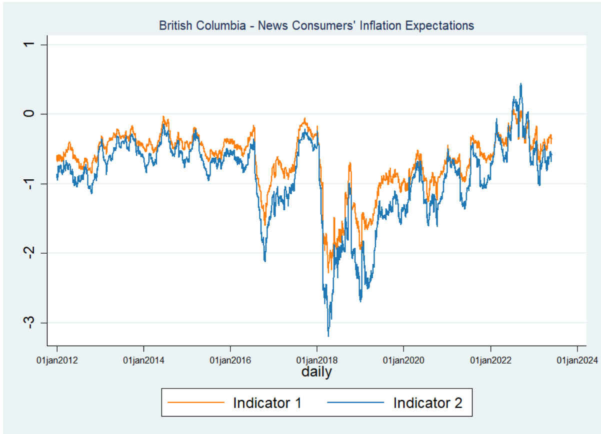
[7] All measures are normalized to have a mean equal to zero and a standard deviation equal to one.

By harnessing this data, I have crafted finely tuned inflation expectation indicators at the provincial level, offering a deeper dive into the intricacies of regional inflation sentiments.

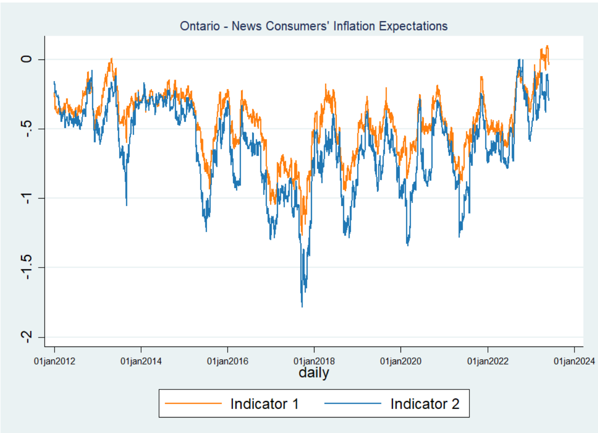
Figures 11a to 11d depict the inflation expectations for four of Canada's most populous provinces: Alberta, British Columbia, Ontario, and Quebec. A striking similarity in inflationary trends is evident across the provinces, especially during Canada's recent surge in inflation. This pronounced trend is particularly evident in Alberta, British Columbia, and Quebec, mirroring the overarching economic dynamics currently shaping the country.



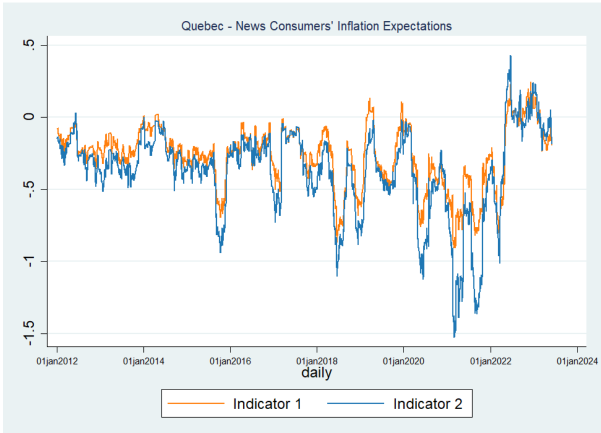
(a) Alberta



(b) British Columbia



(c) Ontario



(d) Quebec

Figure 11: **Regional Inflation Expectations from Daily News Data**

Notes: These indices are developed using the same approach for four major Canadian regions, showcasing the potential of region-specific insights.

The capacity to discern these province-specific inflation expectations underscores the invaluable potential of local news data. Such granularity aids in a more nuanced understanding of region-specific economic sentiments, proving indispensable for analysts, policymakers, and market participants eager to decode the economic intricacies of Canada's diverse landscape.

4.6 Major Cities Twitter-based inflation expectations

Upon examining the inflation expectation trends presented in Figures A.2a, A.2b, A.2c, and A.2d, a clear uniformity emerges across the major Canadian cities of Toronto, Montreal, Vancouver, and Calgary. This comovement in the inflationary trends of these cities, as captured through Twitter data, underscores a shared sentiment in urban economic expectations. It suggests that despite the individual nuances of each city, the macroeconomic factors impacting inflation perceptions are broadly similar. These figures, capturing real-time sentiments from a platform as dynamic as Twitter, offer invaluable insights into the urban economic dynamics. Analysts and policymakers can greatly benefit from such granular data, which provides a consistent pulse on the economic sentiments across Canada's primary urban centers. The uniformity in these trends emphasizes the robustness of the methodology employed and the potential of social media data in understanding city-specific economic dynamics.

4.7 Extensions to 'Experts' and 'Non-Experts'

Inflation expectations can be influenced by a myriad of factors, including public sentiment, economic forecasts, and market dynamics. To delve deeper into the nature of these expectations, I differentiate between the perspectives of experts and non-experts. Experts are defined as individuals or entities with specialized knowledge or expertise in economic matters, such as officials from the Bank of Canada, economists, traders, real estate agents, and analysts, among others.[8] Non-experts encompass the broader public whose views are shaped by general news consumption and personal experience, rather than formal economic training or professional experience.

Figure 12 presents a comparative analysis of inflation expectations as voiced by these two distinct groups. The remarkably high correlation of 0.94 between the two indices suggests that the delineation between experts and non-experts may not be as stark as traditionally perceived. This convergence could imply a widespread consensus or a shared sentiment about inflation trends among both experts and non-experts. It may also reflect the efficacy of communication strategies employed by central banks and other economic institutions in disseminating information, thereby aligning public understanding with expert analysis.

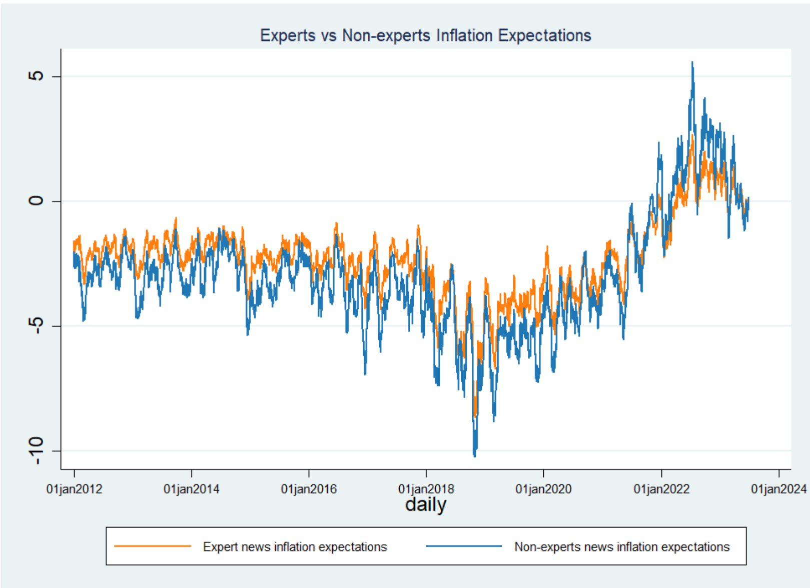


Figure 12: **Experts vs Non-experts news inflation expectations**

[8] Experts: Bank of Canada, Governor, Deputy Governor, Competition Bureau Canada, Statistics Canada, Economist, Professional, Real Estate Agent, General Manager, Purchasing Manager, Trader, Manufacturer, Price Comparator, Economist, Analyst, etc.

To discern the nature of the messages contributing to these indices, I employed a keyword-based approach to categorize articles into expert and non-expert groups. This classification not only included professionals with direct expertise in pricing, such as purchasing managers and real estate agents but also extended to broader economic analysts and investors.

The alignment observed in Figure 12 raises intriguing questions about the drivers of inflation expectations. It suggests that the informational divide between experts and the general public may be narrowing, potentially due to increased accessibility to economic data and analysis. This trend carries significant implications for policymakers and market participants, as it underscores the need to consider a wider range of voices in economic forecasting and decision-making.

In summary, the high correlation between expert and non-expert inflation expectations highlights a synchronized understanding of inflationary trends, suggesting a cohesive narrative permeating through diverse segments of society. This finding enriches my comprehension of how inflation expectations are formed and propagated, offering valuable insights for both economic research and policy formulation.

4.8 Can the text-based indices explain actual CPI and inflation expectations?

In the pursuit of identifying robust predictors for inflation expectations and actual CPI data, my analysis delves into the predictive power of both news and Twitter-based inflation indices. Table 2 presents the regression results examining the relationship between media-based inflation expectations and actual consumer price index (CPI) inflation as well as 1-year-ahead consumer inflation expectations ($E_t^{BoC}\pi_{t,t+12}$). The table is divided into four columns, each representing a different regression model with distinct dependent variables.

In the first column, I regress actual CPI inflation (CPI_t) on its lagged value (CPI_{t-1}) and the news-based inflation index. The coefficient for CPI_{t-1} suggests strong persistence in CPI inflation. The news-based inflation index also demonstrates a positive association with CPI inflation, indicating its relevance in explaining actual inflation dynamics.

In the second column, I replaced the news-based inflation index with the Twitter-based inflation index. The lagged CPI inflation remains significant, and the Twitter-based inflation index shows a positive relationship with CPI inflation, reinforcing its role as an informative predictor.

In the third column, the focus shifts to 1-year-ahead consumer inflation expectations ($E_t^{BoC}\pi_{t,t+12}$), regressing it on its lagged value and the news-based inflation index. Both past inflation expectations and the news-based inflation index are found to be strong predictors of future inflation expectations.

Finally, in the fourth column, the same dependent variable is considered, but with the Twitter-based inflation index replacing the news-based index. The Twitter-based inflation index shows a notable association with future inflation expectations, suggesting its superior predictive power compared to the news-based index.

The robust adjusted R^2 values across all models, particularly those with actual CPI inflation as the dependent variable, underscore the substantial explanatory power of the models. These findings highlight the significant explanatory capabilities of my text-based inflation expectations in forecasting both actual CPI inflation and future inflation expectations. Their efficacy underscores their potential as indispensable tools for policymakers and market participants alike.

Table 2: Can Media Inflation Expectations Explain Inflation Expectations and Actual CPI?

	CPI_t	CPI_t	$E_t^{BoC}\pi_{t,t+12}$	$E_t^{BoC}\pi_{t,t+12}$
CPI_{t-1}	0.982***	0.873***		
	(0.000)	(0.000)		
News inflation index$_t$	0.0354**		0.418***	
	(0.034)		(0.004)	
Twitter inflation index$_t$		0.0573*		0.650***
		(0.064)		(0.008)
$E_t^{BoC}\pi_{t-1,t+11}$			0.600***	0.320
			(0.000)	(0.135)
Constant	0.0246**	0.0301**	0.0370	0.0150
	(0.038)	(0.013)	(0.519)	(0.800)
N	110	109	33	33
$Adj.R^2$	0.985	0.985	0.899	0.882

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable CPI_t represents the monthly Consumer Price Index (CPI) for Canada, obtained from Statistics Canada. The dependent variable $E_t^{BoC}\pi_{t,t+12}$ represents 1-year-ahead quarterly consumer inflation expectations, as reported by the Bank of Canada. My Media-based Inflation Indices, including the News and Twitter inflation indices, are computed as described in the methodology section. All variables have been standardized, with means set to zero and standard deviations to one. P-values are reported in parentheses. Significance levels are denoted by *** for $p < 0.01$, ** for $p < 0.05$, and * for $p < 0.1$.

The extended analysis delves into the role of news-driven inflation expectations in shaping sector-specific historical monthly CPI inflation rates. As illustrated in Table 3, findings across key CPI categories reveal a significant impact of news inflation expectations on all examined disaggregated CPI components, including shelter, goods, services, core CPI, and clothing and footwear. Notably, the influence on CPI for shelter and clothing and footwear shows a parallel trend, whereas a similar pattern emerges among CPI for goods, services, and core CPI. The consistency of these results holds when incorporating the Twitter-based inflation expectations index, indicating the robustness of these relationships across different indicators.

These insights are pivotal for understanding the broad influence of news-driven inflation expectations on total CPI inflation, especially during economic crises or periods of market turbulence, where swift and accurate insights into inflation trends are essential. The ability to quickly interpret and respond to inflation trends informed by real-time public sentiment can be critical for economic stabilization and guiding recovery efforts. Moreover, the findings offer valuable direction for policymakers, particularly central banks, in fine-tuning their communication strategies and adjusting their monetary policy actions to achieve their desired objectives effectively. Embracing a comprehensive approach that includes both traditional and alternative data sources allows policymakers to better navigate and respond to dynamic economic conditions.

Table 3: News-inflation expectations driven sectoral inflation

	Shelter	Goods	Services	Core CPI	Clothing and Footwear
Lagged CPI Shelter	0.258***				
	(0.001)				
News inflation index	0.119***	0.0974*	0.0901**	0.0890**	0.123***
	(0.002)	(0.051)	(0.027)	(0.034)	(0.002)
Lagged CPI Goods		-0.0254			
		(0.530)			
Lagged CPI Services			0.270***		
			(0.000)		
Lagged Core CPI				0.311***	
				(0.000)	
Lagged CPI Clothing and Footwear					0.0262
					(0.599)
Constant	-0.00152	-0.00190	-0.000163	0.000614	0.0000942
	(0.971)	(0.965)	(0.997)	(0.988)	(0.998)
N	522	522	522	522	522
$Adj.R^2$	0.0893	0.00582	0.0826	0.108	0.0126

p-values in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable represents the sectoral or disaggregated monthly Consumer Price Index (CPI) inflation for Canada, obtained from Statistics Canada. All variables have been standardized, with means set to zero and standard deviations to one. P-values are reported in parentheses. Significance levels are denoted by *** for $p < 0.01$, ** for $p < 0.05$, and * for $p < 0.1$.

5 Measuring economic narratives with topic models

Shiller (2017, 2020) introduces narrative economics which states that popular stories weaved into daily conversation, such as how much to save and consume, as well as when and where to invest, can eventually affect individual and collective human behaviors and thus drive economic outcomes. According to Shiller, studying these economic narratives can enhance my ability to predict or prepare for future major economic events. Shiller (2020) points out nine major economic narratives that have been mutating over the past two centuries, namely, panic versus confidence, frugality versus conspicuous consumption, gold standard versus bimetallism, unemployment due to labor-saving machines, automation and artificial intelligence replacing human jobs, real estate booms and busts, stock market bubbles, boycotts and evil business, and wage-price spiral and evil labor unions. Although narrative economics is intuitive, simple to understand, and important, the theory has received little attention from most economists because of its lack of a well-developed framework.

In this study, I use Blei et al. (2003) Latent Dirichlet Allocation (LDA) approach to quantify the economics narratives from inflation news and tweets sample separately.[9] LDA is a Bayesian factor model that uncovers topics in the articles and represents each article in terms of these topics.

It reduces the dimensionality of the text from the entire corpus of articles to just K "topics", or groupings of words that tend to appear together. LDA is essentially a very flexible clustering algorithm for words that groups words into topics on the basis of repeated co-occurrence across paragraphs. There are two inputs to the algorithm. The first input that the user must supply is a corpus of the documents of text to be analyzed; in this paper, the corpus is the full history of inflation newspapers and tweets where I group words at the level of an individual sentence in a statement. However, before using the words in the LDA analysis, I first remove stop words (such as 'the', 'a' and 'and') and also stem the remaining words which reduces them to a common linguistic root ('economy' and 'economic' both become 'economi'). The second input is a number of topics that the algorithm should form; I use a 50-topic model.

There are two broadly defined outputs. The algorithm will form, in my case, 50 topics which are probability distributions over words, and tell the user the words that tend to go together. The algorithm also forms document distributions which contain probabilities that capture the fraction of words policymakers devote to the different topics in their communications.

To get more precise, topic models estimate K topics each of which is a distribution $\beta_k \in \Delta^V$ over the V unique tokens (words) in the corpus vocabulary. LDA is flexible enough to allow unique tokens to belong to more than one topic. LDA will also generate a predictive distribution over topics $\widehat{\boldsymbol{\theta}}_d \in \Delta^K$ for each document, where Δ^K is the K-simplex. However, given that I estimate the topic model at the sentence level, rather than use the predictive distribution, I prefer to work with the word-to-topic allocations directly (this is an intermediate step in the LDA algorithm to generate $\widehat{\boldsymbol{\theta}}_d$. In particular, let $\phi_{p,k,d} = n_{p,d}(k)/n_{p,d}$ be the fraction of sentence p words allocated to topic k, where $n_{p,d}(k)$ is the number of sentence p words allocated to topic k, and $n_{p,d}$ is the total number of words in the paragraph. I will define a sentence as being about topic k when this estimated topic allocation fraction $\phi_{p,k,d}$ is greater than some critical proportion (α). In fact, I estimate the LDA model using a collapsed Gibbs sampling algorithm. As such, I get measures of topic allocation for every iteration of the chain. The data that I work with has been extracted from the best-performing (in an information-matching sense) chain but I draw 20 samples from points in the chain that are thinned using a thinning interval of 50. I then take an average of over 20 samples.

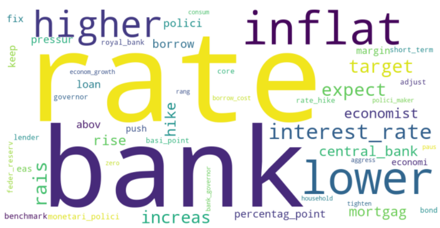
I estimate my 50-topic LDA separately on the full corpus of inflation newspapers and tweets, split into sentences. The LDA-estimated topics cover different aspects of the inflation newspapers and tweets. I select four key topics which relate to the discussion of the economic situation and monetary policy decisions. The key words (tokens) in the economic topics are presented as word clouds in Figure 13.

Building on the understanding that economic narratives significantly influence economic expectations and consequently affect macroeconomic outcomes and fluctuations, as highlighted in studies like Shiller (2020), Larsen et al. (2021), Bertsch et al. (2021), and Ayivodji and Rauh (2023), I have integrated these narratives into the forecasting of actual inflation. This integration is done in conjunction with my News and Twitter-based inflation expectations indices. In the subsequent

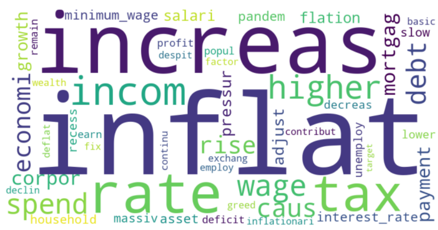
[9] Also see Hansen et al. (2018) for a discussion on LDA and its application in macroeconomics.

sections, I will elaborate on how this integration of economic narratives with data-driven indices enhances our ability to predict actual inflation, providing a more holistic view of inflation dynamics.

Figure 13: WORDCLOUDS OF KEY TOPICS NARRATIVES



(a) Interest rates



(b) Inflation/prices

(c) Market expectations

(d) Housing

Notes: These are the top 50 words of 4 out of 50 topics computed using LDA with $\alpha = 3.33$ and $\beta = 0.01$ for the entire news articles sample. The size of a term represents its probability within a given topic. The position conveys no information. A list of the 50 topics is exhibited in Appendix E.

For a detailed visualization and analysis of the Twitter and newspaper narratives, please refer to Appendices C and D, respectively; and for the comprehensive word cloud representations of these topics, delve into Appendices E and F.

6 Forecasting methodology and inference

I now explore whether my News and Twitter-based inflation indices help in forecasting inflation expectations and actual inflation. To fix ideas - and in line with my application suppose the target is sampled at a quarterly frequency, while the predictors are sampled at a monthly frequency. I start by considering the following Unrestricted Mixed Data Sampling (U-MIDAS) model developed by Foroni et al. (2015), which also includes the lags of the dependent variable, written as

$$y^h_{t+h} = \beta_0 + \sum_{i=0}^{p-1} \lambda_i y_{t-i} + \sum_{i=1}^{N} \boldsymbol{B}_i \left(L^{1/m}\right) x^{(m)}_{i,t} + \varepsilon_{t+h}$$

where

$$\boldsymbol{B}_i\left(L^{1/m}\right) = \sum_{k=0}^{K-1} \boldsymbol{b}_{i,k} L^{k/m}$$

and I denote the low-frequency variables by y_t that represents my quarterly consumers' survey inflation expectations or realized inflation, whereas I represent the high-frequency variables by $x_t^{(m)}$, which is sampled m times faster than y_t (e.g. with quarterly/monthly sampling rates $m = 3$). Furthermore, the low-frequency time period is indexed with discrete times $t = 1, \ldots, T$, while the high-frequency observations are represented by its time fraction j/m. Thus, $x_{t-j/m}^{(m)}$ denotes the j^{th} (past) high-frequency period of quarter t with $j = 0, \ldots, m-1$. For a quarterly/monthly mixture, for example, $x_t^{(3)}, x_{t-1/3}^{(3)}, x_{t-2/3}^{(3)}$ denote the last, second, and first month of the quarter t, respectively.

Note that in the U-MIDAS regression, each high-frequency observation $x_{i,t-k/m}^{(m)}$ within a low-frequency sample period t can be interpreted as a different low-frequency series that has a separate effect on y_{t+h} (i.e. quarterly consumers' survey inflation expectations for h-step ahead).

That is, each month within a quarter can be treated as a separate quarterly series and hence I can stack them together to construct a $(NK \times 1)$ vector of a new regressor set X_t^{HF} for the high-frequency data lags as

$$X_t^{HF} = \left[x_{1,t}, x_{1,t-1/3}, \ldots, x_{1,t-(K-1)/3}, x_{2,t}, x_{2,t-1/3}, \ldots, x_{2,t-(K-1)/3}, \ldots, x_{N,t}, \ldots, x_{N,t-(K-1)/3}\right]'$$

so that I can pretend to be working within the common lowest frequency. Ghysels et al. (2016) call this operation as the frequency alignment. In addition, if I denote the set of low-frequency regressors, i.e. lagged values of the dependent variable y_t, by X_t^{LF} such that

$$X_t^{LF} = [y_t, y_{t-1}, \ldots, y_{t-p-1}]'$$

then I can rewrite the U-MIDAS regression in a more compact way as:

$$y_{t+h}^h = X_t'\Theta + \varepsilon_{t+h} \tag{1}$$

where

$$\Theta = [\beta_0, \lambda_0, \ldots, \lambda_p, \boldsymbol{b}_{1,0}, \ldots, \boldsymbol{b}_{1,K-1}, \boldsymbol{b}_{2,0}, \ldots, \boldsymbol{b}_{2,K-1}, \ldots, \boldsymbol{b}_{N,0}, \ldots, \boldsymbol{b}_{N,K-1}]'$$

and

$$X_t = \left[1, X_t^{LF'}, X_t^{HF'}\right]'$$

are K_d-dimensional vectors with $K_d = 1 + p + NK$. Foroni et al. (2015) show that if the lag orders p and K are large enough to make the error term ε_{t+h} uncorrelated, all the parameters in Equation 1 can be estimated via OLS.

The U-MIDAS model, as specified in Equation 1, offers the advantage of not imposing a lag-polynomial structure. This flexibility allows each higher-frequency feature to have its own coefficient, enabling a separate estimation of its effect on the target variable. While this approach enhances interpretability, it also entails estimating a larger number of coefficients. This can pose challenges for traditional estimation methods, especially when dealing with a substantial number of higher-frequency features. To address this issue of high-dimensional predictors and mitigate overfitting, I employ machine learning techniques. These methods offer the flexibility to estimate coefficients without arbitrarily imposing a lag-polynomial structure, while still effectively guarding against overfitting, even when the number of predictors K_d is large. The large set of predictors or features in my analysis includes News-based inflation expectations, Twitter-based inflation expectations, 50 narratives from News, 50 narratives from Twitter, a lag of actual inflation, and 52 Google search indices. Under the UMIDAS framework, this leads to a total of $3 \times (1 + 1 + 50 + 50 + 52) + 1 = 463$ predictors.

I can rewrite Equation 1 like:

$$y_{t+h}^{h} = G_{h,t}\left(\boldsymbol{X}_t\right) + \varepsilon_{t+h} \tag{2}$$

where $G_{h,t}$ is the mapping between the predictors and consumers' survey inflation expectations. The mapping relation $G(.)$ is estimated by ten models from the machine learning literature that can handle many predictors: least absolute and selection operator (LASSO), Ridge, Elastic Net (EN), Extreme Gradient Boosting (XGB), Random Forest (RF), Shallow Neural Networks (NN1), Deep Neural Networks (NN3), and forecast combination (ensemble linear, ensemble nonlinear and ensemble all).

The LASSO, Ridge models with the Elastic Net penalty of Zou and Hastie (2005), which has demonstrated success in high-dimensional inflation forecasting (Medeiros et al., 2021; Garcia et al., 2017; Medeiros and Mendes, 2016). Neural Networks, including both shallow and deep architectures, have shown promise in prediction tasks, including inflation forecasting (Coulombe et al., 2022). The Random Forest model, inspired by previous studies, outperforms other machine learning methods in data-rich environments when forecasting US or Global inflation (Medeiros et al., 2022, 2021; Coulombe et al., 2022). The Neural Network, Extreme Gradient Boosting, and Random Forest models allow for the estimation of nonparametric nonlinear relations between predictors and the target variable. All models are implemented in Python. In the upcoming sections, I will delve into various machine learning methods, ranging from linear to nonlinear approaches, particularly suited for scenarios with a large number of predictors.

6.1 Penalized regression

The LASSO (Tibshirani, 1996) is a machine-learning device based on penalized regression. It alleviates overfitting by augmenting the objective function for estimating $\boldsymbol{\theta}$ in Eq. (1) with an ℓ_1 penalty term:

$$\underset{\boldsymbol{\theta} \in \mathbb{R}^{K_d}}{\arg\min} \frac{1}{2T}\left[\sum_{t=1}^{T}\left(y_{t+h} - \boldsymbol{\theta}'\boldsymbol{X}_t\right)^2\right] + \lambda\left\|\boldsymbol{\theta}\right\|_1 \tag{3}$$

T is the number of quarterly y_t observations used to fit the model, $\|\cdot\|_1$ is the ℓ_1 norm, and $\lambda \geq 0$ is a regularization hyperparameter for controlling the degree of shrinkage. Unlike the ℓ_2 penalty in ridge regression (Hoerl and Kennard, 1970), the ℓ_1 penalty in Eq. (3) permits shrinkage to zero (for sufficiently large λ) so that the LASSO performs variable selection. Although the LASSO effectively selects relevant predictors in certain environments (e.g., Huang et al., 2008; Bickel et al., 2009; Meinshausen and Yu, 2009), it tends to arbitrarily select one predictor from a group of highly correlated predictors. The EN (Zou and Hastie, 2005) is a refinement of the LASSO, which helps to mitigate this tendency by including both ℓ_1 (LASSO) and ℓ_2 (Ridge) components in the penalty term for the objective function:

$$\underset{\boldsymbol{\theta}\in\mathbb{R}^{K_d}}{\arg\min}\frac{1}{2T}\left[\sum_{t=1}^{T}\left(y_{t+h}-\boldsymbol{\theta}'\boldsymbol{X}_t\right)^2\right]+\lambda P_\delta\left(\boldsymbol{\theta}\right), \tag{4}$$

where

$$P_\delta\left(\boldsymbol{\theta}\right)=0.5(1-\delta)\left\|\boldsymbol{\theta}\right\|_2^2+\delta\left\|\boldsymbol{\theta}\right\|_1 \tag{5}$$

$\|\cdot\|_2$ is the ℓ_2 norm, and $0\leq\delta\leq1$ is a blending hyperparameter for the ℓ_1 and ℓ_2 components of the penalty term. When $\delta=1, P_\delta=\left\|\boldsymbol{\theta}\right\|_1$ in Eq. (5), so that the EN reduces to the LASSO or $\delta=0$ the EN reduces to the Ridge. I follow the recommendation of Hastie and Qian (2016) and set $\delta=0.5$. [10]

6.2 Artificial neural networks

For modeling $G(.)$ in Eq. (2), I also consider feedforward Neural Networks (NN), which permit nonlinearities in the conditional mean and have proven useful for prediction in numerous domains. A feedforward NN architecture is comprised of multiple layers. The first, the input layer, is the set of predictors, which I denote by $x_1,\ldots,x_{P_0}$. One or more hidden layers follow. Each hidden layer l contains P_l neurons, each of which takes signals from the neurons in the previously hidden layer to generate a subsequent signal:

$$h_m^{(l)}=g\left(w_{m,0}^{(l)}+\sum_{k=1}^{P_{l-1}}w_{m,k}^{(l)}h_k^{(l-1)}\right)\qquad\text{for } m=1,\ldots,P_l;l=1,\ldots,L, \tag{6}$$

where $h_m^{(l)}$ is the signal corresponding to the m th neuron in the lth hidden layer, $w_{m,0}^{(l)},w_{m,1}^{(l)},\ldots,w_{m,P_{l-1}}^{(l)}$ are weights; and $g(\cdot)$ is a (nonlinear) activation function. The final layer is the output layer, which translates the signals from the last hidden layer into a prediction:

$$\hat{y}_{t+h}^h=w_{m,0}^{(L)}+\sum_{k=1}^{P_L}w_k^{(L)}h_k^{(L-1)} \tag{7}$$

where $\hat{y}_{t+h}^h$ denotes the h quarterly ahead prediction of the target variable. For the activation function, I use the popular rectified linear unit (ReLU) function:

[10] To better guard against overfitting, I tune the regularization hyperparameter λ for the LASSO and EN in Eqs. (3), (4), respectively, via the Bayesian information criterion (BIC), as suggested by Zou et al. (2007). I obtain similar results when I tune λ via the extended regularization information criterion (Hui et al., 2015), which is a modification of the BIC.

$$g(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{otherwise} \end{cases} \tag{8}$$

In response to a sufficiently strong signal, Eq. (8) activates a neuronal connection and relays the signal forward through the network. To illustrate the basic structure of a feedforward NN, the following diagram in Figure 14 portrays a feedforward NN consisting of five inputs and three hidden layers with four, three, and two neurons, respectively:

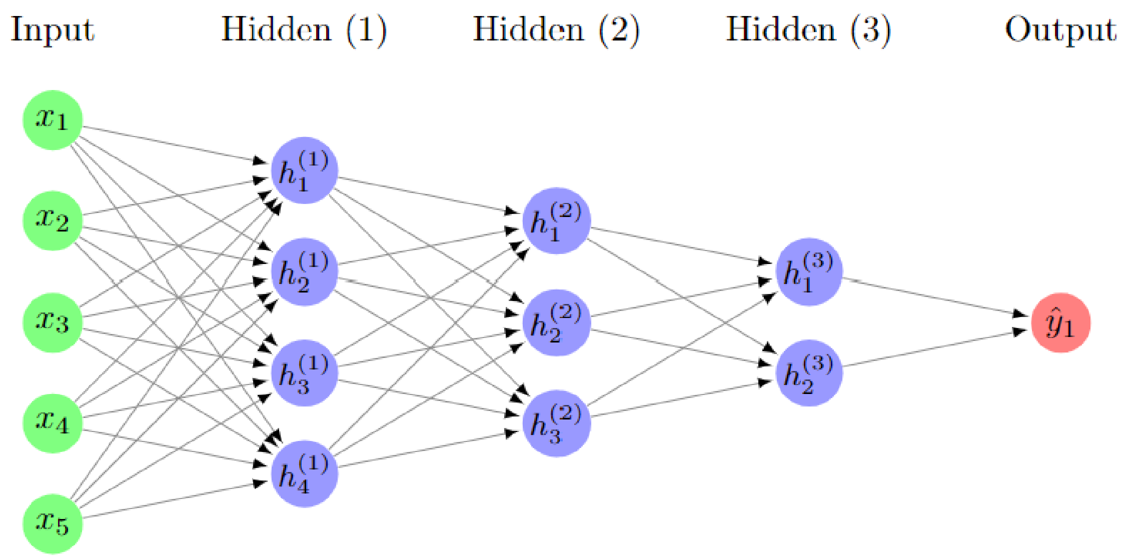


Figure 14: Graphical representation of feedforward Neural Network

The final layer is the output layer and, thus, the model's prediction. The myriad of interactions among the inputs and neurons in the network and the nonlinear activation function allow for complex nonlinear predictive relationships.

Fitting (or training) a NN requires estimating the weights. I train the NNs by minimizing an objective function based on the MSE between the predicted and observed values. I augment the objective function with an ℓ_1 penalty term to help guard against overfitting, which I set equal to 0.0001, the sklearn default value. I use the recently developed ADAM SGD algorithm (Kingma and Ba, 2014) to train the NN, which I implement in Python using the sklearn package with a constant learning rate of 0.01. I use three layers in the network and follow a conventional geometric pyramid rule (Masters, 1993) in setting the number of neurons in the hidden layers, so the shallow neural network has $\lceil\sqrt{P}\rceil$ neurons in its hidden payer, while the deep neural network has $\lceil P^{3/4}\rceil$, $\lceil P^{2/4}\rceil$, and $\lceil P^{1/4}\rceil$ in its first, second, and third hidden layers, respectively. I set the number of epochs in the algorithm to 700 and the batch size to 32. To reduce the influence of the random-number generator in the initialization of the NN estimation, I fit each model 100 times with a different seed each time and use the median of the predictions. [11]

[11] Although the ADAM stochastic gradient descent algorithm is a very powerful optimizer, it is my experience that Neural Networks tend to get stuck in local minima. Using the median of 100 networks greatly minimizes the chance of this happening.

6.3 Random Forest

The Random Forest methodology is due to Breiman (2001) who proposed to create ensembles of regression trees by bagging randomly constructed regression trees with the intent to reduce the overall variance of the model (Breiman, 1996). Regression trees are highly flexible nonparametric models that approximate nonlinear functions using local predictions from a recursive partitioning of the predictor space. More specifically, a regression tree splits the predictor space into a set of regions subject to the chosen splitting predictors and the corresponding split-points and models the response by a region-specific constant value. So, regression trees split the observations into sub-groups with similar predictor values. The objective is to choose the splitting predictors and split points such that the loss function is minimized. [12] Given my target variable, y_{t+h}^h, a vector of predictor values at time $t, \boldsymbol{q}_t$, and a partition into J regions, $R_1, R_2, \ldots, R_J$, the regression tree becomes:

$$y_{t+h}^h = \sum_{j=1}^{J} c_j \mathbf{I}\left(\boldsymbol{q}_t \in R_j\right)$$

where $\mathbf{I}$ is an indicator function and c_j is the prediction of the tree if the observation falls into region R_j. I use the sum of squared error criterion for which $\widehat{c}_j$ is simply the average of past realizations of y_{t+h}^h such that $\boldsymbol{q}_t$ belongs to R_j. I find all J regions by recursive binary splitting, which is a top-down approach meaning that all data is considered for the first split. Consider a splitting variable, m, and a split-point, s, that partitions the predictor space, Q, into two regions $R_1(m,s) = \{Q \mid Q_m \leq s\}$ and $R_2(m,s) = \{Q \mid Q_m > s\}$. I choose m and s to solve:

$$\min_{m,s}\left[\min_{c_1} \sum_{t=1}^{T} \mathbf{I}\left(\boldsymbol{q}_t \in R_1(m,s)\right)\left(y_{t+h}^h - c_1\right)^2 + \min_{c_2} \sum_{t=1}^{T} \mathbf{I}\left(\boldsymbol{q}_t \in R_2(m,s)\right)\left(y_{t+h}^h - c_2\right)^2\right] \tag{9}$$

This procedure is typically continued iteratively on each of the resulting regions, subject to a minimum requirement for the number of observations in each region. A large regression tree is typically able to explain complex nonlinear relations in the data but tends to overfit (high variance). The RF methodology reduces the variance by averaging many of such regression trees where each tree is fitted on a bootstrap sub-sample of the original data using a randomly selected subset of the predictors, indexed by $b = 1, \ldots, B$ where B is the total number of individual trees. The final forecast is a simple average of the B individual forecasts:

$$\widehat{y}_{t+h}^h = \frac{1}{B} \sum_{b=1}^{B}\left[\sum_{j=1}^{J_b} \widehat{c}_{j,b} \mathbf{I}\left(\boldsymbol{q}_t \in R_{j,b}\right)\right] \tag{10}$$

The proportion of predictors randomly selected for each split is chosen by cross-validation. I set $B = 500$ and the minimum leaf size to five. [13]

[12] See Hastie et al. (2009) for an intuitive example of a regression tree with two predictors.

[13] I also tried cross-validating the minimum leaf size as well as letting the trees grow fully deep (which is consistent with Breiman (2001)), but both alternatives produced slightly weaker results.

6.4 Extreme Gradient Boosting

Extreme Gradient Boosting model is a boosting tree algorithm that is an enhancement over tree bagging methodologies, such as Random Forests (Breiman, 2001), which have gained significant ground and are frequently used in many machine learning applications across various academic fields. Gradient boosting trees model has been proposed by Breiman (2001) and has the advantage of reducing both variance and bias. It reduces variance because multiple models are used (as in Random Forests), and it additionally reduces bias in training the subsequent model by informing it what errors the previous models made. In particular, in gradient boosting, each subsequent model is trained using the residuals (the difference between the predicted and true values) of previous models. XGB is an advanced implementation of gradient boosting algorithm, offering increased efficiency, accuracy, and scalability over simple bagging algorithms. It supports the fitting of various types of objective functions, including regression, classification, and ranking whereas it offers increased flexibility since optimization is performed on an extended set of hyperparameters.

6.5 Ensembles

I also consider ensemble predictions, which are popular in machine learning. In recognition of model uncertainty, instead of relying on a prediction based on a single model, I employ a stacking method to integrate the predictions generated by multiple models. I construct three ensembles using this approach. The first ensemble (Ensemble-Linear) uses a meta-learner to combine predictions from the linear models fitted via the LASSO, ridge, and elastic net. The second ensemble (Ensemble-Nonlinear) employs a different meta-learner to integrate predictions from the random forest, extreme gradient boosting, and neural network models NN1 and NN3. The final ensemble (Ensemble-All) combines predictions from both linear models and nonlinear models using a comprehensive meta-learner. In this stacking framework, each ensemble's meta-learner is trained to optimally weigh the input predictions from constituent models, aiming for a potentially better performance than simple averaging of predictions (Hastie et al., 2009; Gne and Tan, 2017; Luedtke et al., 2021).

6.6 Evaluation measures and interpretation tools

I evaluate the quality of the forecasts using the root mean square error (RMSE) and the median absolute deviation from the median (MAD). For some model m and horizon h, I compute RMSE and MAD as:

$$\text{RMSE}_{h,m} = \sqrt{\frac{1}{\#OOS_h} \sum_{t \in OOS} \widehat{e}^2_{t,h,m}},$$

$$\text{MAD}_{h,m} = \text{median}\left[\left|\widehat{e}_{t,h,m} - \text{median}\left(\widehat{e}_{t,h,m}\right)\right|\right],$$

where $\widehat{e}_{t,h,m} = \pi^h_t - \widehat{\pi}^h_{t,m}$ and $\widehat{\pi}^h_{t,m}$ is the forecast of the h month ahead inflation growth at time t using model m and $\#OOS_h$ is the number of out-of-sample predictions made for horizon h. I use MAD, in addition to RMSE, to ensure that outliers do not drive the results.

To analyze the stability of predictive accuracy through time, I plot the cumulative sum of squared error difference (CSSED) between the model of interest and the benchmark model (Welch and

Goyal, 2008). I compute the CSSED for model m and horizon h at time t as:

$$\text{CSSED}_{m,t}^{h} = \sum_{i=t_0}^{t} \left[\left(y_t^h - \tilde{y}_t^h \right)^2 - \left(y_t^h - \hat{y}_{m,t}^h \right)^2 \right]$$

where t_0 and t are the beginning and the end of the evaluation period and $\tilde{\pi}_t^h$ is the forecast produced by the benchmark model. A positive $\text{CSSED}_{m,t}^{h}$ at time t implies that the model of interest is outperforming the benchmark up to time t, and vice versa for a negative value.

Machine learning models, such as the RF and NN models, can be black boxes for which it is difficult to calculate the marginal effect of a variable and thus obtain a measure that allows interpreting the model. To surmount this challenge, I use SHapley Additive exPlanations (SHAP) values (Lundberg and Lee, 2017). SHAP values are currently considered the state-of-the-art method for interpreting machine learning models, as they have many desirable properties. [14] They stem from game theory as a general solution to the problem of attributing a payoff obtained in a cooperative game to the individual players based on their marginal contribution to the game (Shapley et al., 1953). In this setting, I can think of the variables as the players and the prediction as the payoff. I compute SHAP values in Python using the shap package.

I denote the SHAP value of model m corresponding to the k th predictor for the t th observation in the training sample by $\phi_{k,t}^{(m)}$. This SHAP value gives us the contribution in model m of the k th predictor to the prediction of the target variable for the t th observation (measured in terms of the deviation from the mean of the target over the training sample).

An important property of SHAP values is efficiency:

$$\hat{y}_t^{(m)} = \sum_{i=1}^{n} \phi_{k,t}^{(m)},$$

meaning that the prediction of model m at time t, $\hat{y}_t^{(m)}$, can be decomposed into the sum of SHAP values for all the predictors of the model. From this result, I have that a variable importance measure for predictor k in model m over the training sample $t = 1, \ldots, T$ can be constructed as:

$$VI_k^{(m)} = \frac{1}{T} \sum_{t=1}^{T} |\phi_{k,t}^{(m)}|.$$

$VI_k^{(m)}$ provides a measure of the overall importance of predictor k to the fitted model. Since I use a rolling training window estimation, I have a sequence of variable importance measures for each predictor (one for each training window). Therefore, to obtain an overall measure of importance across the whole sample period, I calculate the $VI_k^{(m)}$ for all predictors in each estimation window and then take an average of these.

7 Out-of-sample results

In this section, I examine the out-of-sample properties of my new indices—news-based and Twitter-based inflation expectations—using a range of linear, nonlinear, and ensemble models across four

[14] See Molnar (2020) for a detailed exposition of SHAP values.

distinct horizons.

7.1 Inflation expectations forecasting with news-based expectations

Table 4 reveals the relative RMSE of different models in predicting consumer inflation expectations for four forecasting horizons (h=1, 2, 3, and 4). The RMSE provides a measure of the model's prediction error; a smaller RMSE value indicates better predictive accuracy, with the relative RMSE providing a comparison to my benchmark. In this context, values less than one indicate that the competing model outperforms the benchmark, while values greater than one suggest it performs worse.

For short-term forecasting (h=1), nearly all models outperform the benchmark, with the shallow neural network (NN1) model providing the most significant reduction in RMSE, achieving a 22.32% improvement relative to the benchmark. LASSO, ridge, and elastic net, which are linear models, demonstrate mixed results across horizons. Notably, elastic net is especially proficient in the short term (h=1), exhibiting the best performance among linear models. The nonlinear models, which include extreme gradient boosting, random forest, and shallow and deep neural networks, generally fare better across all horizons, confirming the inherent complexities in predicting inflation expectations and the advantages of nonlinear modeling. In particular, Random Forest consistently outperforms other models in mid-term horizons (h=2 and h=3), suggesting its aptness in capturing intricate patterns within the given dataset. As I move to a longer-term prediction at h=4, no model consistently excels, with all Relative RMSE values exceeding one, except for Random Forests. This indicates the challenge posed by long-term forecasting, with most models underperforming the benchmark.

The ensemble models, which combine predictions from various models to enhance robustness, offer promising results. Ens-All, which aggregates insights from both linear and nonlinear models, consistently produces results below the 1 benchmark, highlighting its general applicability across all horizons. Interestingly, the Ens-Nonlinear model, despite harnessing the power of nonlinear models, does not always outshine its linear counterpart, underscoring the complexities of the underlying data and the unpredictable nature of inflation expectations. In summation, the table accentuates the value of nonlinear models and ensemble approaches in forecasting inflation expectations, with the latter especially viable given its balanced performance across different horizons. Moreover, the table shows that including my news-based index improves the forecasting performance of consumer inflation expectations. Therefore, I conclude that my granular news-based inflation index is well-suited to capture consumer expectations about inflation in the near future.

7.2 Inflation expectations forecasting with Twitter-based expectations

Table 5 offers a comparative analysis of various forecasting models leveraging Twitter-based inflation expectations across four horizons, with a particular emphasis on relative RMSE as an accuracy benchmark.

Short-term insights (h=1) reveal that LASSO and Elastic Net models excel, underscoring their effectiveness in assimilating Twitter data to predict immediate inflation trends. Elastic net, in

Table 4: News Relative RMSE Forecasting Comparison

	Relative RMSE			
Models	**h=1**	**h=2**	**h=3**	**h=4**
LASSO	0.9246	**0.8412**	0.8546	1.2163
Ridge	0.8385	1.2076	1.0044	1.7375
EN	**0.8235**	**0.8402**	0.8546	1.2155
XGB	0.8381	1.0530	1.1237	1.4754
RF	**0.8462**	**0.8570**	**0.8449**	1.1709
NN1	**0.7768**	1.0349	0.8945	1.4692
NN3	**0.8505**	**0.9453**	1.0636	1.7496
Ens-Linear	0.8622	0.9630	0.9046	1.3898
Ens-Nonlinear	**0.8279**	0.9726	0.9817	1.4663
Ens-All	**0.8426**	0.9685	**0.9487**	1.4336

Note: The table displays Root Mean Squared Errors (RMSE) relative to an $AR(1)$ benchmark for forecasting accuracy across models. Evaluations are based on an out-of-sample window from 2021Q3 to 2023Q1 for horizons $h = 1, 2, 3, 4$. Forecast accuracy significance is gauged using Diebold and Mariano (2002) test statistics, adjusted according to Harvey. The top-performing models for each horizon, as per the DM test, are highlighted in bold.

particular, displays robust performance across all horizons, consistently harnessing Twitter data to forecast inflation expectations.

As the forecasting horizon extends, the random forest model emerges as a frontrunner, especially for the second and third horizons (h=2 and h=3), adeptly navigating the complexity of Twitter data.

Neural networks show mixed results. The basic NN1 model fluctuates in performance, whereas the more intricate NN3 demonstrates sustained accuracy, suggesting that deeper models might be more proficient in decoding the rich information embedded in Twitter data.

Ensemble models, which amalgamate diverse forecasting approaches, exhibit notable success. Specifically, Ens-All, which integrates a mix of models, consistently delivers reliable predictions across all time horizons, validating the merit of employing a multifaceted forecasting approach.

Furthermore, the table clearly illustrates that integrating my Twitter-based index enhances the forecasting accuracy of consumer inflation expectations. This improvement corroborates the effectiveness of my Twitter-based inflation index in accurately capturing near-term consumer expectations about inflation. The robust performance of ensemble methods and models adept at unraveling complex patterns further emphasizes the significance of incorporating real-time digital data into the economic forecasting arsenal for both policymakers and market analysts.

8 Actual inflation forecasting exercise with text indices

Following my investigation into the predictive capacity of text-based indices for consumer inflation expectations, this section shifts focus to their predictive power for actual inflation. Specifically, I

Table 5: Twitter Relative RMSE Forecasting Comparison

	Relative RMSE			
Models	**h=1**	**h=2**	**h=3**	**h=4**
LASSO	0.8746	0.8646	0.8546	1.0978
Ridge	1.0676	0.9767	1.0100	1.0246
EN	0.8646	**0.8546**	**0.8546**	1.0978
XGB	1.2618	1.2701	1.1258	1.1652
RF	**0.9964**	**0.9458**	**0.9548**	1.2985
NN1	0.9629	**0.9176**	0.9730	1.0673
NN3	**0.9064**	**0.8552**	**0.8558**	1.0919
Ens-Linear	0.9257	0.8954	0.9065	1.0734
Ens-Nonlinear	1.0319	0.9972	0.9774	1.1558
Ens-All	0.9864	**0.9536**	**0.9470**	1.1205

Note: The table displays Root Mean Squared Errors (RMSE) relative to an $AR(1)$ benchmark for forecasting accuracy across models. Evaluations are based on an out-of-sample window from 2021Q3 to 2023Q1 for horizons $h = 1, 2, 3, 4$. Forecast accuracy significance is gauged using Diebold and Mariano (2002) test statistics, adjusted according to Harvey. The top-performing models for each horizon, as per the DM test, are highlighted in bold.

evaluate three distinct sets of predictors: news-based inflation expectations, Twitter-based inflation expectations, and a comprehensive set that combines both Google trends search volume data and inflation narratives derived from Twitter and news sources. This step is crucial in determining the practical utility of these text-based indices as straightforward, yet powerful tools for actual inflation forecasting, offering valuable insights for a broad spectrum of economic stakeholders.

8.1 Actual inflation forecasting with News-based expectations

In Table 6, I present the out-of-sample forecasting performance for actual inflation using news-based inflation expectations. The displayed RMSE is benchmarked against a standard AR(1) model over horizons $h = 1$ to $h = 4$, corresponding to one to four quarters ahead.

The table reveals a nuanced landscape of forecasting efficacy across different models and horizons. Notably, the LASSO and Elastic Net models demonstrate consistent outperformance relative to the AR(1) benchmark across the first three quarters, as evidenced by their Relative RMSEs falling below 1. This indicates their superior predictive capabilities in the short to medium term, likely due to their robustness in handling high-dimensional data and avoiding overfitting. The reduction in RMSE compared to the benchmark is particularly pronounced, highlighting the substantial predictive power of news-based inflation expectations, which are derived from economic institutions and media sources.

However, an interesting reversal occurs in the fourth quarter, where these models' performance dips, suggesting a possible limitation in capturing long-term inflation dynamics with the given predictors.

The Random Forest and Neural Network models, particularly deep neural networks, also exhibit commendable forecasting accuracy, especially in the second and third quarters. Their ability to model complex non-linear relationships might be attributed to their relative success in these horizons.

Ensemble models, which aggregate predictions from multiple models, generally show improved performance, underscoring the value of combining diverse predictive perspectives. Yet, they do not uniformly outperform simpler models like LASSO and Elastic Net, indicating that complexity does not always equate to superior forecasting.

This table underscores the importance of news-based inflation expectations as a predictive tool for actual inflation, particularly in the short to medium term. While no single model consistently outperforms across all horizons, the integration of news-based inflation expectations with advanced machine learning techniques offers promising avenues for improving inflation forecasts.

Table 6: News Relative RMSE Forecasting Comparison

	Relative RMSE			
Models	**h=1**	**h=2**	**h=3**	**h=4**
LASSO	**0.8556**	**0.8546**	**0.8569**	**0.8568**
Ridge	1.3577	1.3807	1.4346	1.2033
Enet	**0.8570**	**0.8562**	**0.8585**	**0.8579**
XGB	**0.7420**	**0.7973**	0.9376	0.9960
RF	**0.7593**	**0.7762**	**0.8673**	**0.8879**
NN1	**0.6586**	0.8925	**0.7602**	**0.7808**
NN3	**0.8150**	**0.8178**	**0.8144**	**0.8073**
Ens-Linear	**0.8798**	**0.8761**	**0.8830**	**0.8864**
Ens-NonLinear	**0.6824**	**0.7451**	**0.8131**	**0.8406**
Ens-All	**0.7598**	**0.7916**	**0.8378**	**0.8556**

Notes: The table displays Root Mean Squared Errors (RMSE) relative to an $AR(1)$ benchmark for forecasting accuracy across models. The forecast period spans from 2019Q3 to 2023Q1 with the training data extending from 2011Q1 to 2019Q2. The horizons considered are $h = 1, 2, 3, 4$. Forecast accuracy significance is gauged using Diebold and Mariano (2002) test statistics, adjusted according to Harvey. The top-performing models for each horizon, as per the DM test, are highlighted in bold. The predictor set contains only news inflation expectations and the lag of the target variable. The predictors used in these models are exclusively news-based inflation expectations and the lagged values of the target variable.

8.2 Actual inflation forecasting with Twitter-based expectations

In Table 7, I explore the out-of-sample forecasting performance for actual inflation using Twitter-based inflation expectations.

The LASSO and Elastic Net models consistently surpass the AR(1) benchmark across all horizons. Their Relative RMSEs, persistently below 1, highlight the robustness of Twitter data in capturing inflationary trends, especially in the short to medium term. Their ability to effectively manage high-dimensional data and prevent overfitting likely contributes to their predictive success.

The Extreme Gradient Boosting and Random Forests models exhibit exceptional performance, with Relative RMSEs significantly below 1 for all forecasting horizons. Known for their capacity to model complex non-linear relationships, these models adeptly utilize the rich and intricate information inherent in Twitter data, leading to a substantial reduction in forecasting error.

Neural Network models, represented by the shallow neural network (NN1) and the deep neural networks (NN3), show varied performance. While NN1 demonstrates improvements in specific quarters, NN3 maintains strong performance across all horizons, suggesting that network depth may be crucial in extracting more nuanced predictive signals from Twitter-based data.

The ensemble models, especially Ens-All, which combines various predictive models, showcase the most impressive results. Their consistently low Relative RMSEs across all horizons underscore the advantages of integrating diverse predictive methodologies, positioning them as the top performers in this forecasting endeavor.

These insights underscore the significant predictive power of Twitter-based inflation expectations for actual inflation forecasting, which is of particular relevance to policymakers and market participants. The exceptional performance of ensemble methods and models capable of capturing non-linear dynamics underscores the value of incorporating real-time, crowd-sourced economic indicators into models that inform monetary policy and central bank decisions. The standout results from this exercise reaffirm the complex and dynamic nature of Twitter data as a valuable resource for future inflation forecasting.

Table 7: Twitter Relative RMSE Forecasting Comparison

	Relative RMSE			
Models	**h=1**	**h=2**	**h=3**	**h=4**
LASSO	**0.8622**	**0.8759**	**0.8672**	**0.8677**
Ridge	1.1101	1.1678	1.2500	1.1237
Enet	**0.8638**	**0.8773**	**0.8686**	**0.8691**
XGB	**0.7452**	**0.7423**	**0.7453**	**0.7566**
RF	**0.7307**	**0.7319**	**0.7546**	**0.7506**
NN1	1.2346	**0.8629**	1.4167	**0.8365**
NN3	**0.8197**	**0.8149**	**0.8160**	**0.8125**
Ens-Linear	**0.8009**	**0.8200**	**0.8283**	**0.8554**
Ens-NonLinear	**0.7091**	**0.7553**	0.8475	**0.7518**
Ens-All	**0.7258**	**0.7776**	0.8329	**0.7929**

Notes: The table displays Root Mean Squared Errors (RMSE) relative to an $AR(1)$ benchmark for forecasting accuracy across models. The forecast period spans from 2019Q3 to 2023Q1 with the training data extending from 2011Q1 to 2019Q2. The horizons considered are $h = 1, 2, 3, 4$. Forecast accuracy significance is gauged using Diebold and Mariano (2002) test statistics, adjusted according to Harvey. The top-performing models for each horizon, as per the DM test, are highlighted in bold. The predictor set contains only Twitter inflation expectations and the lag of the target variable. The predictors used in these models are exclusively Twitter-based inflation expectations and the lagged values of the target variable.

8.3 Actual inflation forecasting with all predictors

In Table 8, I delve into the out-of-sample forecasting performance for actual inflation using a large set of predictors. This set encompasses News-based and Twitter-based inflation expectations, inflation narratives from both Twitter and news sources, and Google Trends search data, alongside the lagged values of the target variable.

This table showcases the synergistic potential of a multifactorial predictor set. Both LASSO and Elastic Net models exhibit exceptional performance, consistently surpassing the AR(1) benchmark across all horizons. Their adeptness in managing a complex predictor landscape is evident.

The XGB model, with its gradient boosting framework, shines particularly in the initial quarter, maintaining strong performance thereafter. This underscores its capability to iteratively refine predictions by capturing subtle patterns embedded within the extensive predictor set.

While the Random Forests model demonstrates commendable predictive abilities, the Neural Network models present a mixed bag. NN1, a shallow neural network, appears challenged by the predictor complexity, yielding less optimal forecasts. Conversely, NN3, a deep neural network, shows promise in the intermediate horizons, likely leveraging its deeper structure to unravel more intricate relationships.

The ensemble models, especially Ens-NonLinear and Ens-All, emerge as the frontrunners in forecasting accuracy. Their unmatched performance across all quarters, with Ens-All recording an impressive Relative RMSE of 0.0285 in the fourth quarter, testifies to the collective strength of diverse modeling techniques.

These insights highlight the significant predictive value drawn from a rich array of predictors. They affirm the importance of text-based inflation indices, social media sentiments, and search behavior trends while underscoring the efficacy of sophisticated machine learning techniques in actual inflation forecasting. For policymakers and market participants, these findings underscore the value of integrating a broad spectrum of real-time and heterogeneous data sources for more nuanced decision-making in monetary policy and economic analysis.

While machine learning (ML) techniques have showcased unparalleled prowess in approximating intricate non-linear functions and forecasting, their interpretability often remains a challenge. Often labeled as a "black box", they tend to obscure the underlying relationships within the data (Shapley et al., 1953; Lundberg and Lee, 2017). In my research, I endeavor to shed light on these relationships in a semi-structured manner, aiming to decipher which economic indicators predominantly influence the forecasting capabilities of mixed-frequency machine learning. I employ SHAP values, as introduced by Lundberg and Lee (2017), which have rapidly become an industry-standard interpretability tool for ML techniques. For a comprehensive understanding, I recommend Molnar (2020) for a broad overview and Borup et al. (2022) for an in-depth exploration of its relevance to financial and macroeconomic forecasting.

Figure 15 enumerates the ten most significant predictors for Elastic Net, Random Forest, Extreme Gradient Boosting, 'and' Deep Neural Network for the horizon $h = 2$. A careful examination of the feature importance across these models unveils several pivotal insights. Google Trends keywords, such as 'home prices', 'higher inflation', 'high gasoline price', 'rent', 'headline inflation', 'rise credit', 'crude oil price', 'consumer price index', 'rising house prices', 'price increasing', 'inflation rises',

Table 8: Relative RMSE Forecasting Comparison

	Relative RMSE			
Models	**h=1**	**h=2**	**h=3**	**h=4**
LASSO	**0.7153**	**0.7568**	**0.6814**	**0.7643**
Ridge	0.7906	0.7323	**0.7144**	**0.6862**
Enet	**0.7162**	**0.7565**	**0.6822**	**0.7642**
XGB	**0.6369**	**0.6832**	**0.7010**	**0.6565**
RF	**0.7200**	**0.7052**	**0.7047**	**0.7275**
NN1	2.2657	2.8710	2.5415	2.2022
NN3	1.5790	**0.8039**	**0.8020**	1.1839
Ens-Linear	**0.6507**	**0.6558**	**0.5838**	**0.6463**
Ens-NonLinear	**0.6591**	0.8430	**0.7731**	**0.6904**
Ens-All	**0.6311**	**0.6590**	**0.5994**	**0.5913**

Notes: The table displays Root Mean Squared Errors (RMSE) relative to an $AR(1)$ benchmark for forecasting accuracy across models. The forecast period spans from 2019Q3 to 2023Q1 with the training data extending from 2011Q1 to 2019Q2. The horizons considered are $h = 1, 2, 3, 4$. Forecast accuracy significance is gauged using Diebold and Mariano (2002) test statistics, adjusted according to Harvey. The top-performing models for each horizon, as per the DM test, are highlighted in bold. The predictor set contains only news inflation expectations and the lag of the target variable. The predictors used in these models include text-based inflation indices, Google Trends, media narratives, and the lagged values of the target variable.

’vehicles prices’, among others, consistently emerge as salient features across all models. These keywords, emblematic of everyday consumer apprehensions, underscore the significant role of real-time search trends in capturing public sentiment toward inflationary dynamics.

Moreover, media narratives emerge as critical drivers. Narratives revolving around themes like ’online shopping’, ’housing market’, ’monetary policy’, ’interest rates’, ’energy’, ’transport costs’, ’inflation’, ’and’ ’cost of living’ provide a glimpse into the overarching economic discourse shaping public inflation experiences. The pronounced importance of these narratives, alongside the lagged data on consumer inflation expectations, indicates a vibrant interplay between historical data, prevailing public discourse, ’and’ real-time sentiments. Collectively, these models adeptly capture a holistic spectrum, blending past data, contemporary narratives, ’and’ real-time search trends to provide a comprehensive view of future inflation.

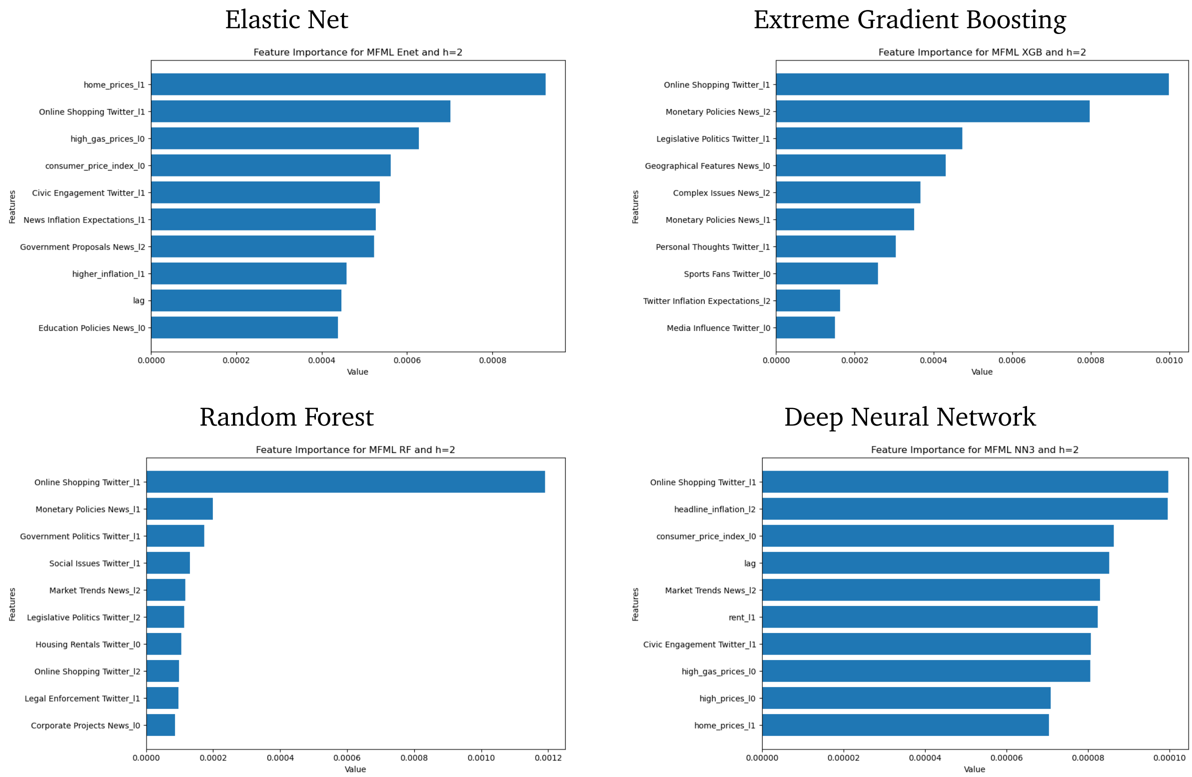


Figure 15: Inflation Forecasting Variables Importance for $h = 2$

Notes: This figure presents the top 10 important variables in the second quarter. For each selected model, SHAP values are calculated, providing an additive and easily interpretable decomposition of the model’s performance. These values attribute the impact of each variable to the prediction, allowing for an assessment of their marginal contributions to the overall forecasting accuracy.

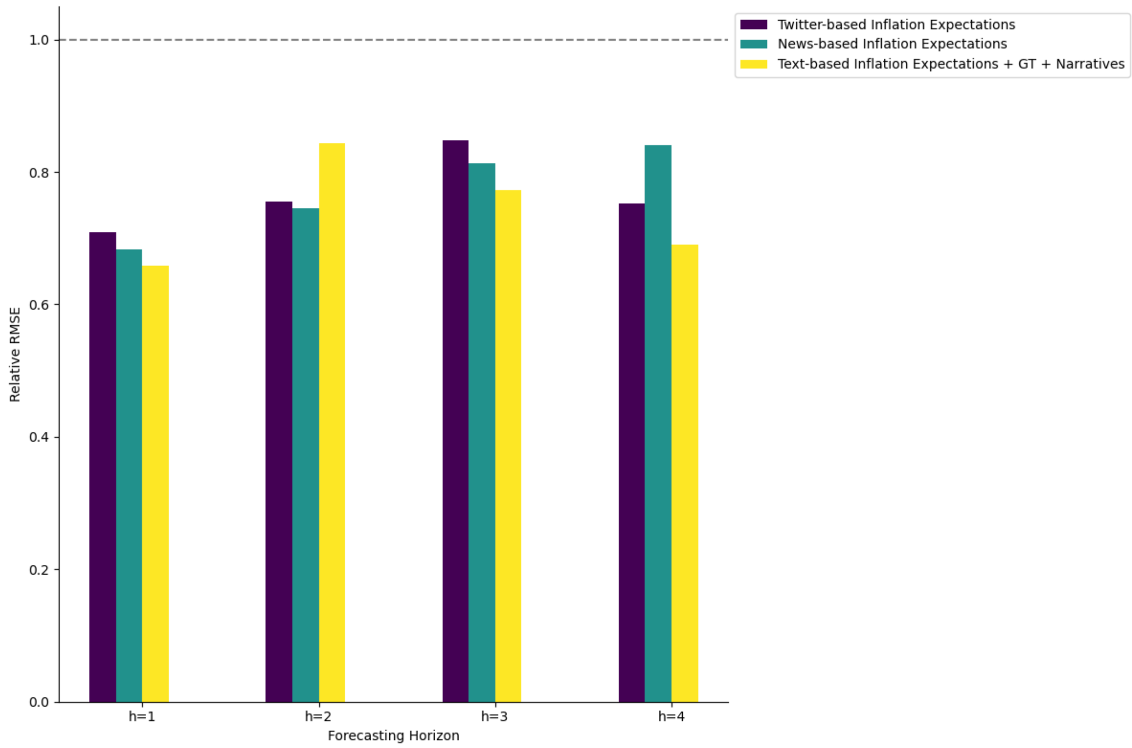
The recurrent appearance of text-based inflation expectations across different models emphasizes their robustness as a forecasting tool. This consistency is not just a pattern but a validation of their importance in understanding future inflation. It confirms the findings of previous studies, which have recognized the predictive value of consumer inflation expectations in future inflation (Verbrugge and Zaman, 2021; Rudd, 2022; D’Acunto et al., 2023).

Furthermore, the comprehensive analysis in Appendix J reinforces this result. The resilience of text-based inflation expectations as top predictors, irrespective of model variations, attests to their fundamental role in forecasting. This convergence of traditional economic measures with emerging digital narratives represents a significant advancement in the field of economic prediction.

The findings from Figure 15 and Appendix J affirm the substantial predictive capacity of my novel text-based inflation expectations. They are not mere supplements but key contributors to the understanding and forecasting of inflation. This aligns with existing research and further establishes the relevance of media narratives in the realm of economic forecasting (Bybee et al., 2021; Ayivodji and Rauh, 2023).

In Figure 16, I show the results of the ensemble nonlinear models that include various inflation expectations based on news, Twitter, and a combination of text-based expectations, Google Trends search indices, and media narratives, across different forecasting horizons. I see that for all nonlinear ensemble models and all horizons, the model provides considerable predictive power, in particular for News-based inflation expectations and full predictors. In Appendix K, I present the results for each of the prediction algorithms.

Figure 16: Comparing the performance of the ensemble nonlinear models with different sets of predictors



9 Conclusion

Inflation expectations are pivotal in shaping consumer behavior and the broader economic landscape. Traditional methods of capturing these expectations are often burdened with substantial costs, delays, and timeliness issues. Addressing this gap, my research introduces an innovative approach to assess consumer inflation expectations by leveraging a vast of unconventional data sources, notably tweets and news articles. Employing the power of state-of-the-art LLMs in NLP to develop daily text-based inflation expectations indices at both national and regional levels. These indices not only reflect the current or future sentiment toward inflation but also offer a timely and lag-free alternative to conventional measures.

My empirical analysis reveals a strong correlation between these text-based indices and existing measures of inflation expectations, such as survey data and actual inflation figures. Moreover, these indices provide a forward-looking perspective, offering additional insights that surpass those provided by professional forecasts and survey-based expectations.

At the heart of my approach is mixed-frequency machine learning methodology, an innovative fusion of U-MIDAS and machine learning. This methodology has demonstrated exceptional predictive power, indicating a significant shift in the way I can forecast consumer inflation expectations and actual inflation. By exploring the interpretability of these models, I highlight the relevance of real-time data, such as my novel text-based inflation expectations indices, Google Trends keywords, and media narratives. Terms like 'home prices', 'high gasoline price', 'rise credit', 'rent', 'high inflation', and 'rising house prices' underscore the importance of real-time data in reflecting public sentiment towards inflation.

Furthermore, media narratives play a crucial role in shaping public inflation expectations. Themes such as the housing market, monetary policy, interest rates, transport costs, energy, and cost of living provide insight into the economic narratives that influence public expectations of inflation. These narratives, coupled with historical data on consumer inflation expectations, reveal a dynamic interplay between past data, current discourse, and real-time sentiments, which my models adeptly capture.

Looking ahead, the success of my methodology opens the door to numerous applications across different economic indicators and regions. This presents a valuable opportunity for central banks, policymakers, and economic strategists to benefit from a more nuanced and comprehensive understanding of economic indicators. The convergence of traditional economic principles with modern data analytics heralds a new era in economic research, with the potential to revolutionize my approach to understanding and forecasting economic trends.

My study lays the groundwork for a range of research extensions. Adapting this methodology to various global economies, differentiating signals from household and firm sources, and delving into sector-specific inflation expectations are some of the promising directions for future research. Additionally, exploring the subtleties of different media outlets and sector-specific inflation narratives could yield deeper insights. My research not only highlights the impact of media on shaping economic expectations but also opens the door for using natural language preprocessing techniques to extract nuanced signals from diverse media sources.

In summary, the high-frequency nature of my text-based inflation expectations indices offers a contemporary lens into the public's inflation perceptions, adapting swiftly to events such as monetary policy shocks or announcements. This attribute is particularly pertinent for central banks, whose mandate includes steering inflation expectations. The capacity to evaluate the repercussions of monetary policy actions in real time could be a strategic asset for policymakers endeavoring to maintain inflation stability. This aspect of my research not only highlights the potential utility of my indices for real-time analysis but also suggests a promising avenue for future research in monetary policy effectiveness and communication.

References

Adämmer, Philipp and Rainer A Schüssler, "Forecasting the equity premium: mind the news!," *Review of Finance*, 2020, *24* (6), 1313–1355.

Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart, "Narratives about the Macroeconomy," 2021.

Angelico, Cristina, Juri Marcucci, Marcello Miccoli, and Filippo Quarta, "Can we measure inflation expectations using Twitter?," *Journal of Econometrics*, 2022, *228* (2), 259–277.

Antenucci, Dolan, Michael Cafarella, Margaret Levenstein, Christopher Ré, and Matthew D Shapiro, "Using social media to measure labor market flows," Technical Report, National Bureau of Economic Research 2014.

Ash, Elliott and Stephen Hansen, "Text algorithms in economics," *Unpublished manuscript*, 2023.

Ayivodji, Firmin, "Regional and Sectoral News-Based Indicators for Macroeconomic Forecasting," *Available at SSRN 4145663*, 2021.

— **and Christopher Rauh**, "Can Media Narratives Predict House Price Movements?," 2023.

Babii, Andrii, Eric Ghysels, and Jonas Striaukas, "Machine learning time series regressions with an application to nowcasting Online Appendix," 2021.

Barbaglia, Luca, Lorenzo Frattarolo, Luca Onorante, Filippo Maria Pericoli, Marco Ratto, and Luca Tiozzo Pezzoli, "Testing big data in a big crisis: Nowcasting under COVID-19," *International Journal of Forecasting*, 2022.

Bernanke, Ben S et al., "Inflation expectations and inflation forecasting," in "Speech at the Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute, Cambridge, Massachusetts," Vol. 10 2007, p. 11.

Bertsch, Christoph, Isaiah Hull, and Xin Zhang, "Narrative fragmentation and the business cycle," *Economics Letters*, 2021, *201*, 109783.

Bianchi, Francesco, Roberto Gómez-Cram, Thilo Kind, and Howard Kung, "Threats to central bank independence: High-frequency identification with Twitter," *Journal of Monetary Economics*, 2023.

Bickel, Peter J, Ya'acov Ritov, and Alexandre B Tsybakov, "Simultaneous analysis of Lasso and Dantzig selector," 2009.

Blei, David M, Andrew Y Ng, and Michael I Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, 2003, *3* (Jan), 993–1022.

Borup, Daniel, David E Rapach, and Erik Christian Montes Schütte, "Mixed-frequency machine learning: Nowcasting and backcasting weekly initial claims with daily internet search volume data," *International Journal of Forecasting*, 2023, *39* (3), 1122–1144.

— **, Philippe Goulet Coulombe, David Rapach, Erik Christian Montes Schütte, and Sander Schwenk-Nebbe**, "The anatomy of out-of-sample forecasting accuracy," 2022.

Bracha, Anat and Jenny Tang, “Inflation levels and (in) attention,” 2022.

Breiman, Leo, “Bagging predictors,” *Machine learning*, 1996, *24*, 123–140.

— , “Random forests,” *Machine learning*, 2001, *45*, 5–32.

Buckmann, Marcus, Andreas Joseph, and Helena Robertson, “An interpretable machine learning workflow with an application to economic forecasting,” Technical Report, Bank of England 2022.

Bybee, Leland, Bryan T Kelly, Asaf Manela, and Dacheng Xiu, “Business news and business cycles,” Technical Report, National Bureau of Economic Research 2021.

Calomiris, Charles W and Harry Mamaysky, “How news and its context drive risk and returns around the world,” *Journal of Financial Economics*, 2019, *133* (2), 299–336.

Cavallo, Alberto, Guillermo Cruces, and Ricardo Perez-Truglia, “Inflation expectations, learning, and supermarket prices: Evidence from survey experiments,” *American Economic Journal: Macroeconomics*, 2017, *9* (3), 1–35.

Chahrour, Ryan, Kristoffer Nimark, and Stefan Pitschner, “Sectoral media focus and aggregate fluctuations,” *American Economic Review*, 2021, *111* (12), 3872–3922.

Coibion, Olivier, Yuriy Gorodnichenko, and Rupal Kamdar, “The formation of expectations, inflation, and the phillips curve,” *Journal of Economic Literature*, 2018, *56* (4), 1447–1491.

— , — , and Saten Kumar, “How do firms form their expectations? new survey evidence,” *American Economic Review*, 2018, *108* (9), 2671–2713.

Coulombe, Philippe, Maxime Leroux, Dalibor Stevanovic, and Stéphane Surprenant, “How is machine learning useful for macroeconomic forecasting?,” *Journal of Applied Econometrics*, 2022, *37* (5), 920–964.

D’Acunto, Francesco, Ulrike Malmendier, and Michael Weber, “What do the data tell us about inflation expectations?,” in “Handbook of economic expectations,” Elsevier, 2023, pp. 133–161.

Diebold, Francis X and Robert S Mariano, “Comparing predictive accuracy,” *Journal of Business & economic statistics*, 2002, *20* (1), 134–144.

D’Acunto, Francesco, Ulrike Malmendier, Juan Ospina, and Michael Weber, “Exposure to grocery prices and inflation expectations,” *Journal of Political Economy*, 2021, *129* (5), 1615–1639.

Flynn, Joel P and Karthik Sastry, “The Macroeconomics of Narratives,” *Available at SSRN 4140751*, 2022.

Foroni, Claudia, Massimiliano Marcellino, and Christian Schumacher, “Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials,” *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 2015, pp. 57–82.

Garcia, Márcio GP, Marcelo C Medeiros, and Gabriel FR Vasconcelos, “Real-time inflation forecasting with high-dimensional models: The case of Brazil,” *International Journal of Forecasting*, 2017, *33* (3), 679–693.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text as data," *Journal of Economic Literature*, 2019, *57* (3), 535–74.

Ghysels, Eric, Virmantas Kvedaras, and Vaidotas Zemlys, "Mixed frequency data sampling regression models: the R package midasr," *Journal of statistical software*, 2016, *72*, 1–35.

Gne, RW Funda and PY Tan, "Stacked ensemble models for improved prediction accuracy," *SAS0437-2017. pdf*, 2017.

Gorodnichenko, Yuriy, Tho Pham, and Oleksandr Talavera, "Social media, sentiment and public opinions: Evidence from# Brexit and# USElection," *European Economic Review*, 2021, *136*, 103772.

Hansen, Stephen, Michael McMahon, and Andrea Prat, "Transparency and deliberation within the FOMC: a computational linguistics approach," *The Quarterly Journal of Economics*, 2018, *133* (2), 801–870.

Hastie, Trevor and Junyang Qian, "Glmnet vignette," *Retrieved June*, 2016, *9* (2016), 1–30.

_ , Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Vol. 2, Springer, 2009.

Hoerl, Arthur E and Robert W Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, 1970, *12* (1), 55–67.

Huang, Allen, Hui Wang, and Yi Yang, "Finbert—a large language model approach to extracting information from financial text," *Available at SSRN 3910214*, 2020.

Huang, Jian, Shuangge Ma, and Cun-Hui Zhang, "Adaptive Lasso for sparse high-dimensional regression models," *Statistica Sinica*, 2008, pp. 1603–1618.

Hui, Francis KC, David I Warton, and Scott D Foster, "Tuning parameter selection for the adaptive lasso using ERIC," *Journal of the American Statistical Association*, 2015, *110* (509), 262–269.

Jain, Monica, Olena Kostyshyna, and Xu Zhang, "How do people view price and wage inflation?," Technical Report, Bank of Canada 2022.

Jiao, Peiran, Andre Veiga, and Ansgar Walther, "Social media, news media and the stock market," *Journal of Economic Behavior & Organization*, 2020, *176*, 63–90.

Kendall, Chad W and Constantin Charles, "Causal narratives," Technical Report, National Bureau of Economic Research 2022.

Kingma, Diederik P and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko, "Inflation targeting does not anchor inflation expectations: Evidence from firms in New Zealand," Technical Report, National Bureau of Economic Research 2015.

Larsen, Vegard H, Leif Anders Thorsrud, and Julia Zhulanova, "News-driven inflation expectations and information rigidities," *Journal of Monetary Economics*, 2021, *117*, 507–520.

Lewis, Daniel J, Karel Mertens, James H Stock, and Mihir Trivedi, "Measuring real activity using a weekly economic index," *Journal of Applied Econometrics*, 2022, *37* (4), 667–687.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

Luedtke, Alex, Incheoul Chung, and Oleg Sofrygin, "Adversarial Monte Carlo meta-learning of optimal prediction procedures," *The Journal of Machine Learning Research*, 2021, *22* (1), 11534–11600.

Lundberg, Scott M and Su-In Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, 2017, *30.*

Macaulay, Alistair and Wenting Song, "Narrative-driven fluctuations in sentiment: Evidence linking traditional and social media," *Available at SSRN 4150087*, 2022.

Marcellino, Massimiliano and Dalibor Stevanovic, *The demand and supply of information about inflation*, CIRANO, 2022.

Masters, Timothy, *Practical neural network recipes in C++*, Academic Press Professional, Inc., 1993.

Medeiros, Marcelo C and Eduardo F Mendes, "L1-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors," *Journal of Econometrics*, 2016, *191* (1), 255–271.

_ **, Erik Christian Montes Schütte, and Tobias Skipper Soussi**, "Global inflation forecasting: Benefits from machine learning methods," *Available at SSRN 4145665*, 2022.

_ **, Gabriel FR Vasconcelos, Álvaro Veiga, and Eduardo Zilberman**, "Forecasting inflation in a data-rich environment: the benefits of machine learning methods," *Journal of Business & Economic Statistics*, 2021, *39* (1), 98–119.

Meinshausen, Nicolai and Bin Yu, "Lasso-type recovery of sparse representations for high-dimensional data," 2009.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, 2013, *26.*

Miller, Sarah and Patrick Sabourin, "What consistent responses on future inflation by consumers can reveal," Technical Report, Bank of Canada 2023.

Molnar, Christoph, *Interpretable machine learning*, Lulu. com, 2020.

Mueller, Hannes and Christopher Rauh, "Reading between the lines: Prediction of political violence using newspaper text," *American Political Science Review*, 2018, *112* (2), 358–375.

Mullainathan, Sendhil and Andrei Shleifer, "The market for news," *American economic review*, 2005, *95* (4), 1031–1053.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning, "Glove: Global vectors for word representation," in "Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)" 2014, pp. 1532–1543.

Rudd, Jeremy B, "Why do we think that inflation expectations matter for inflation?(and should we?)," *Review of Keynesian Economics*, 2022, *10* (1), 25–45.

Shapley, Lloyd S et al., "A value for n-person games," 1953.

Shiller, Robert J, "Narrative economics," *American economic review*, 2017, *107* (4), 967–1004.

_ , *Narrative economics: How stories go viral and drive major economic events*, Princeton University Press, 2020.

Simon, Noah, Jerome Friedman, Trevor Hastie, and Robert Tibshirani, "A sparse-group lasso," *Journal of computational and graphical statistics*, 2013, *22* (2), 231–245.

Stock, James H and Mark W Watson, "Why has US inflation become harder to forecast?," *Journal of Money, Credit and banking*, 2007, *39*, 3–33.

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

Verbrugge, Randal J and Saeed Zaman, "Whose inflation expectations best predict inflation?," *Economic Commentary*, 2021, (2021-19).

Wager, Stefan and Susan Athey, "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 2018, *113* (523), 1228–1242.

Weber, Michael, Francesco D'Acunto, Yuriy Gorodnichenko, and Olivier Coibion, "The subjective inflation expectations of households and firms: Measurement, determinants, and implications," *Journal of Economic Perspectives*, 2022, *36* (3), 157–184.

Welch, Ivo and Amit Goyal, "A comprehensive look at the empirical performance of equity premium prediction," *The Review of Financial Studies*, 2008, *21* (4), 1455–1508.

Zou, Hui and Trevor Hastie, "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, 2005, *67*, 301–320.

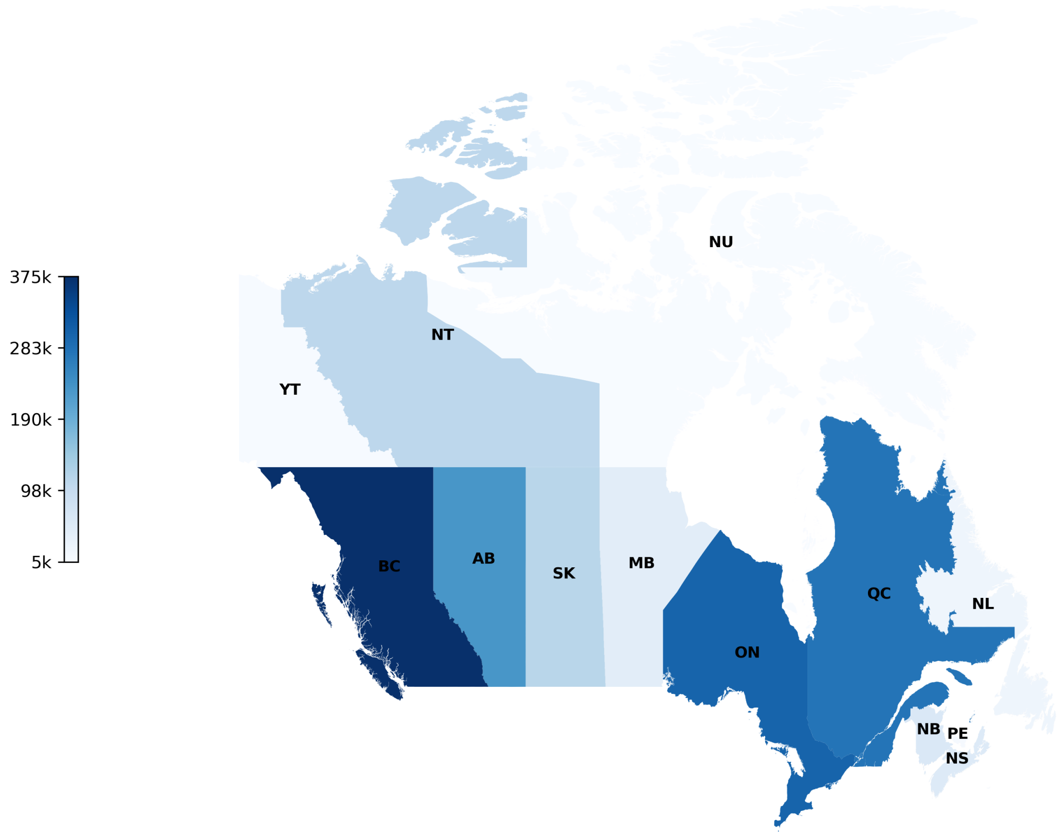
_ , _ , and Robert Tibshirani, "On the "degrees of freedom" of the lasso," 2007.

Appendices: Additional Information

A Newspaper Coverage Across Provinces

Figure A.1 presents a comprehensive visual representation of newspaper coverage across various Canadian provinces. The visualization aims to depict the geographical distribution and intensity of newspaper coverage related to inflation sentiments, thereby offering a spatial context to our data source.

Figure A.1: **Newspaper Coverage Across Provinces**



B Major Cities Twitter-based inflation expectations

Upon examining the inflation expectation trends presented in Figures A.2a, A.2b, A.2c, and A.2d, a clear uniformity emerges across the major Canadian cities of Toronto, Montreal, Vancouver, and Calgary.

(a) Toronto (b) Montreal

(c) Vancouver (d) Calgary

Figure A.2: **Major Cities Inflation Expectations from Daily Tweets Data**

Notes: These indices are developed using the same approach for four major Canadian cities, showcasing the potential of city-specific insights.

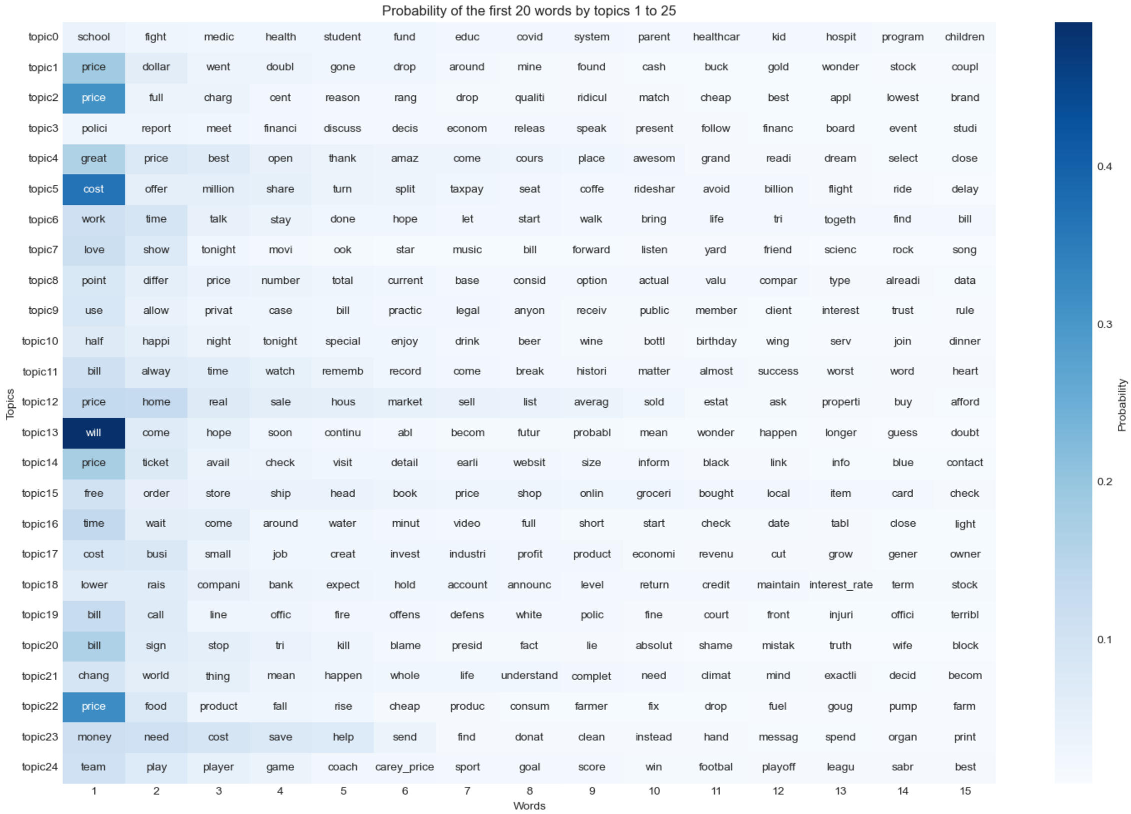
C Twitter narratives topics heatmap

Figures A.3 and A.4 visualize the LDA output surrounding keywords in all categories. Figure A.3 shows the heatmap of the first twenty-five topics. Each row represents a topic clustered by LDA, and the darkness of the cell within a topic represents the likelihood of a word appearing in the topic. Although LDA output does not label topics, it is natural to characterize some of the topics. Topic 1 relates to Education; Topic 2 relates to dollar prices; Topic 13 relates to house prices; Topic 23 relates to grocery prices; Topic 27 relates to inflation; Topic 28 relates to the cost of living, etc.

D Newspaper narratives topics heatmap

This section illuminates the LDA output for newspaper articles discussing inflation. The presented heatmaps, derived from a corpus of newspaper articles, showcase the probability distribution of terms within various topics. The graphical representation aims to capture the essence of newspaper narratives, revealing common themes and patterns prevalent in traditional media discussions surrounding inflation. Topic 2 relates to Investment; Topic 9 relates to trading; Topic 21 relates to

Figure A.3: **LDA output: Terms within Topics Ranked by Probability**

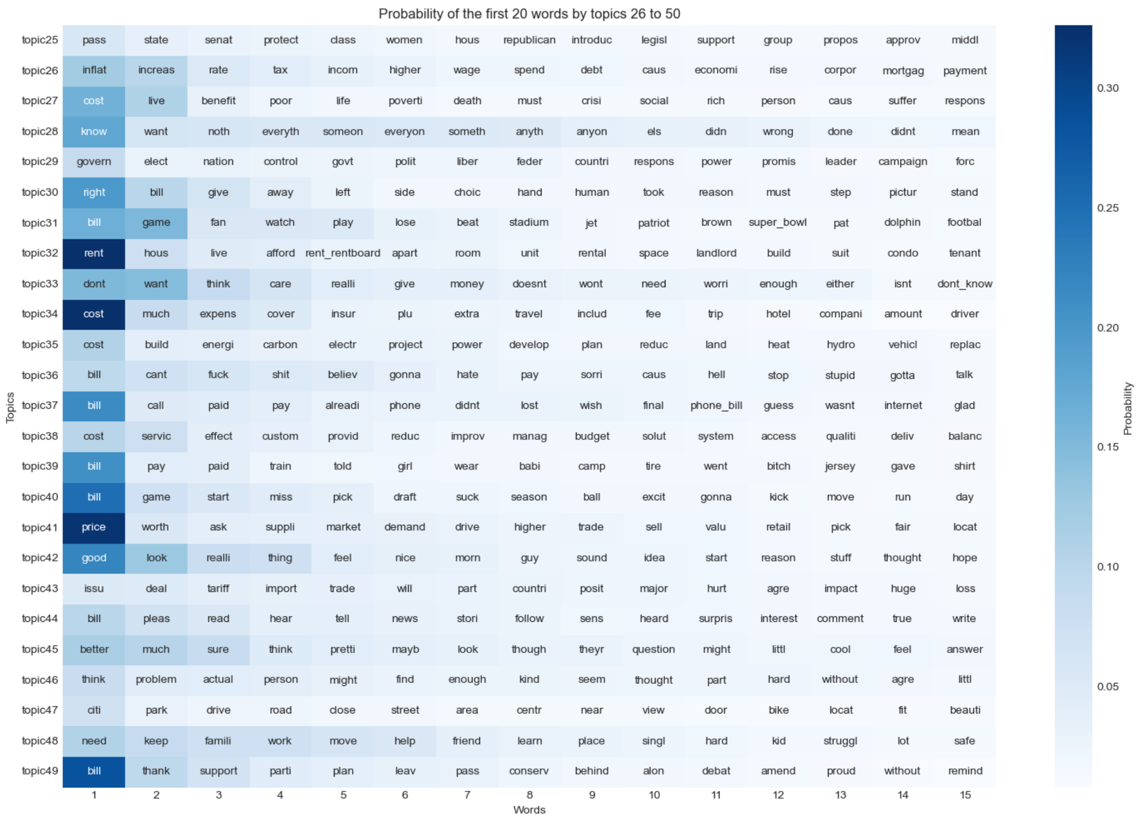


market expectations; Topic 28 relates to the housing market; Topic 42 relates to inflation; Topic 33 relates to the bank interest rates, etc.

E Newspaper Narratives LDA output

Diving deeper into the newspaper narrative analysis, Figures A.7 and A.8 present WordClouds for the 50 topics identified in the preceding heatmaps (D). Each WordCloud visually emphasizes the most dominant keywords within a topic, offering a succinct and engaging summary. The clouds provide insights into the salient themes in newspaper narratives, underscoring the elements that frequently catch public attention.

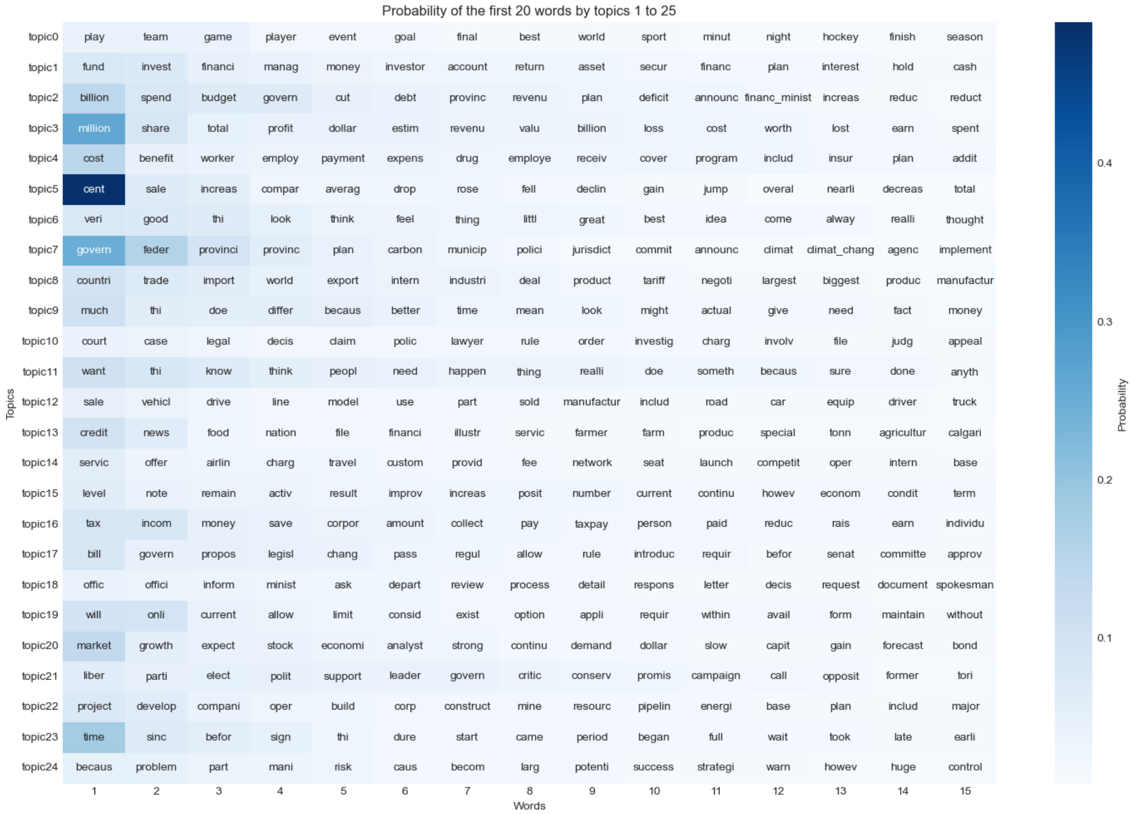
Figure A.4: **LDA output: Terms within Topics Ranked by Probability**



F Twitter Narratives LDA output

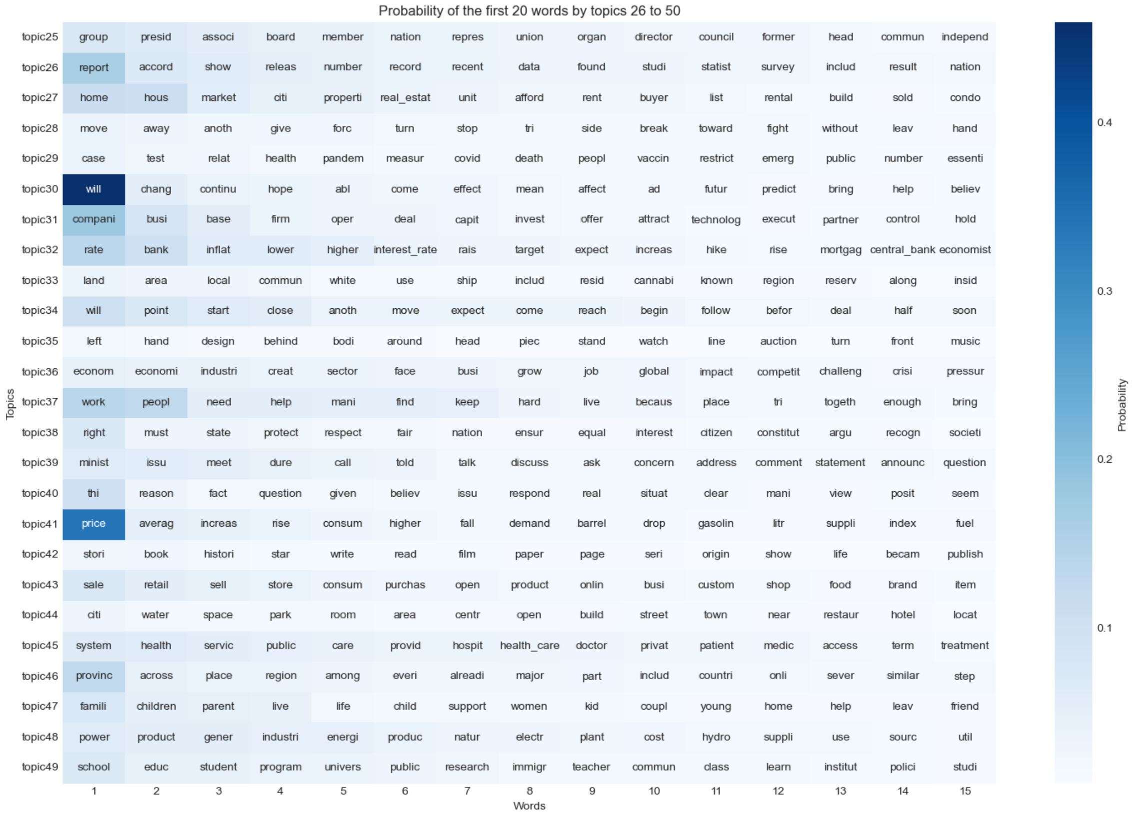
Supplementing our analysis of Twitter narratives, Figures A.9 and A.10 spotlight WordClouds for the 50 topics as identified in the heatmap section (C). These graphical depictions serve to emphasize the most prevalent keywords in Twitter discussions about inflation. By visualizing dominant themes, the word clouds offer a bird's-eye view of the prevailing sentiments and concerns voiced on the platform, further enhancing our understanding of the public's inflationary perceptions.

Figure A.5: **LDA output: Terms within Topics Ranked by Probability**



G Twitter Narratives LDA output

Figure A.6: **LDA output: Terms within Topics Ranked by Probability**



H Precision, Recall, F1 score, and Confusion Matrix.

Precision tells us that out of the results classified as positive by our model, how many were actually positive. The equation that represents precision is:

$$\text{precision} \;=\; \frac{\text{True Positives}}{\text{True Positives}+\text{False Positives}}$$

Recall tells us how many of the positive cases the classifier correctly predicted, over all the positive cases in the data. The equation that represents recall is:

$$\text{Recall} \;=\; \frac{\text{True Positives}}{\text{True Positives}+\text{False Negatives}}$$

F1 Score is the weighted average of Precision and Recall.The equation that represents F1 score is:

$$F1 = 2 \times \frac{\text{Precision} \;\times\; \text{Recall}}{\text{Precision} \;+\; \text{Recall}}$$

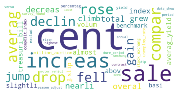
(a) Topic 1

(b) Topic 2

(c) Topic 3

(d) Topic 4

(e) Topic 5



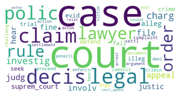
(f) Topic 6

(g) Topic 7

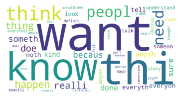
(h) Topic 8

(i) Topic 9

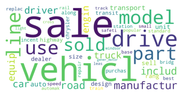
(j) Topic 10



(k) Topic 11



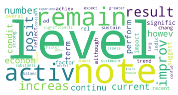
(l) Topic 12



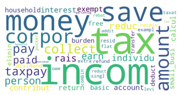
(m) Topic 13

(n) Topic 14

(o) Topic 15



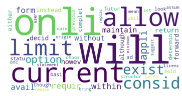
(p) Topic 16



(q) Topic 17

(r) Topic 18

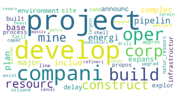
(s) Topic 19



(t) Topic 20

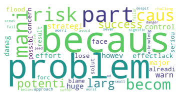
(u) Topic 21

(v) Topic 22



(w) Topic 23

(x) Topic 24

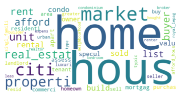


(y) Topic 25

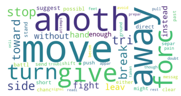
Figure A.7: Newspaper LDA topics wordcloud

(a) Topic 26

(b) Topic 27

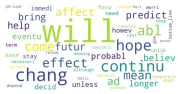


(c) Topic 28

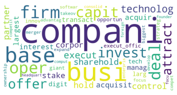


(d) Topic 29

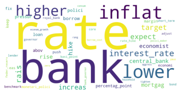
(e) Topic 30



(f) Topic 31

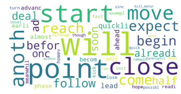


(g) Topic 32



(h) Topic 33

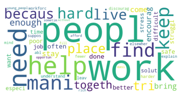
(i) Topic 34



(j) Topic 35

(k) Topic 36

(l) Topic 37

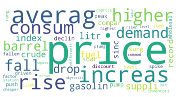


(m) Topic 38

(n) Topic 39

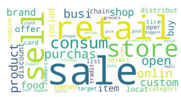
(o) Topic 40

(p) Topic 41



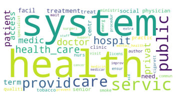
(q) Topic 42

(r) Topic 43



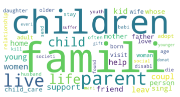
(s) Topic 44

(t) Topic 45

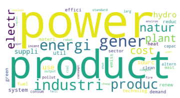


(u) Topic 46

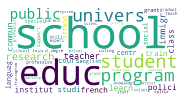
(v) Topic 47



(w) Topic 48



(x) Topic 49



(y) Topic 50

Figure A.8: Newspaper LDA topics wordcloud

(a) Topic 1 (b) Topic 2 (c) Topic 3 (d) Topic 4 (e) Topic 5

(f) Topic 6 (g) Topic 7 (h) Topic 8 (i) Topic 9 (j) Topic 10

(k) Topic 11 (l) Topic 12 (m) Topic 13 (n) Topic 14 (o) Topic 15

(p) Topic 16 (q) Topic 17 (r) Topic 18 (s) Topic 19 (t) Topic 20

(u) Topic 21 (v) Topic 22 (w) Topic 23 (x) Topic 24 (y) Topic 25

Figure A.9: Twitter LDA topics wordcloud

(a) Topic 26 (b) Topic 27 (c) Topic 28 (d) Topic 29 (e) Topic 30

(f) Topic 31 (g) Topic 32 (h) Topic 33 (i) Topic 34 (j) Topic 35

(k) Topic 36 (l) Topic 37 (m) Topic 38 (n) Topic 39 (o) Topic 40

(p) Topic 41 (q) Topic 42 (r) Topic 43 (s) Topic 44 (t) Topic 45

(u) Topic 46 (v) Topic 47 (w) Topic 48 (x) Topic 49 (y) Topic 50

Figure A.10: Inflation/Prices GT Keyword

Table A.1: Inflation/Prices GT Keywords

gasoline price	future inflation
consumer price index	home prices
grocery prices	mortgage rates
prices rise	price rises
core inflation	cost of living
prices increase	interest rates rise
high oil prices	price increases
energy prices	crude oil price
fuel price	interest rates rising
fuel prices	high rents
commodity prices	more expensive
inflation	high interest rates
inflation rises	rise credit
housing price	living cost
rent prices	price increasing
high gasoline price	wholesale price
retail price	high gas prices
vehicle prices	price
inflation rate	rising house prices
rent	tax
wage growth	expensive bills
rising inflation	high gas bills
rising interest rates	higher inflation
real interest rate	headline inflation
price increase	commodity market
cpi	high prices

Note: The table lists the 52 monthly Google Trends search terms used to construct nowcasts and forecasts of quarterly inflation expectations.

Table A.2: Confusion Matrix of Classification

	Predicted Positives	**Predicted Negatives**
Actual Positives	True Positives (TP)	False Negatives (FN)
Actual Negatives	False Positives (FP)	True Negatives (TN)

I Extracting Forward-Looking Information with FinBert-FLS

The following Python code demonstrates the use of the FinBert-FLS model to classify sentences related to forward-looking inflation or prices:

```
# Tested in transformers==4.18.0
from transformers import BertTokenizer, BertForSequenceClassification, pipeline

# Load the FinBERT-FLS model and tokenizer
```

```python
finbert = BertForSequenceClassification.from_pretrained('yiyanghkust/finbert-fls',
    num_labels=3)
tokenizer = BertTokenizer.from_pretrained('yiyanghkust/finbert-fls')

# Create a pipeline for text classification
nlp = pipeline("text-classification", model=finbert, tokenizer=tokenizer)

# List of sentences to classify
sentences = [
    'We expect inflation rates to climb significantly by the end of the year.',
    'Analysts predict a steady increase in commodity prices over the next few
   months.',
    'There is a consensus that inflationary pressures will persist into the next
   quarter.'
]
# Classify each sentence
for sentence in sentences:
    results = nlp(sentence)
    print(f"Sentence: {sentence}")
    print(f"Classification: {results}\n")
```

J Inflation Forecasting Variables Importance

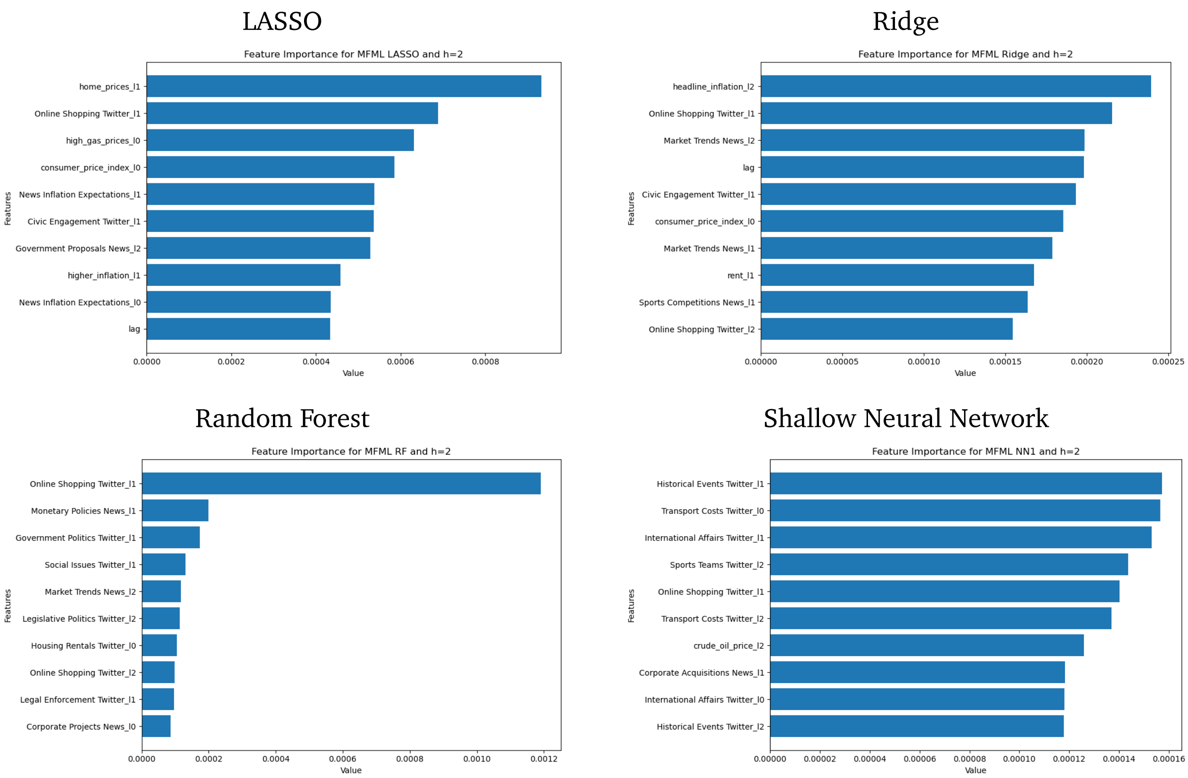


Figure A.11: Inflation Forecasting Variables Importance for $h = 2$

Notes: This figure presents the top 10 important variables at the second quarter. For each selected model, SHAP values are calculated, providing an additive and easily interpretable decomposition of the model's performance. These values attribute the impact of each variable to the prediction, allowing for an assessment of their marginal contributions to the overall forecasting accuracy.

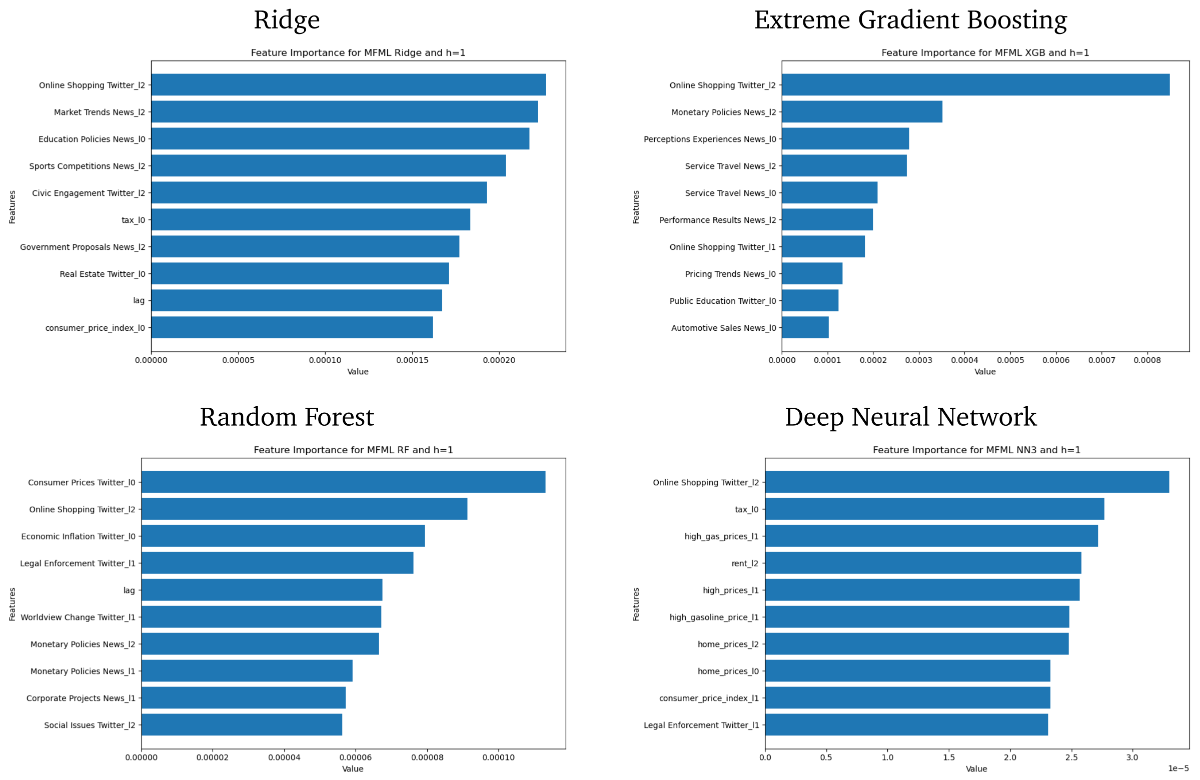


Figure A.12: Inflation Forecasting Variables Importance for $h = 1$

Notes: This figure presents the top 10 important variables at the second quarter. For each selected model, SHAP values are calculated, providing an additive and easily interpretable decomposition of the model's performance. These values attribute the impact of each variable to the prediction, allowing for an assessment of their marginal contributions to the overall forecasting accuracy.

K Comparing RMSE Across Models and Features

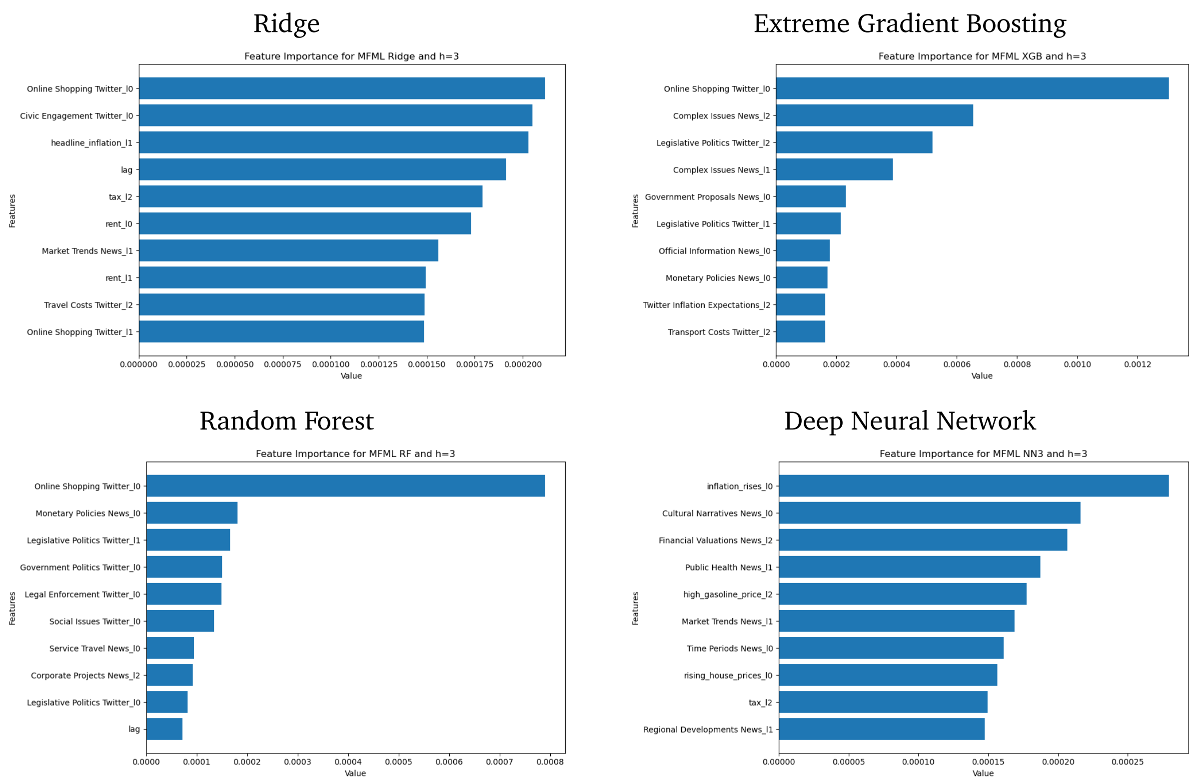


Figure A.13: Inflation Forecasting Variables Importance for $h = 3$

Notes: This figure presents the top 10 important variables at the second quarter. For each selected model, SHAP values are calculated, providing an additive and easily interpretable decomposition of the model's performance. These values attribute the impact of each variable to the prediction, allowing for an assessment of their marginal contributions to the overall forecasting accuracy.

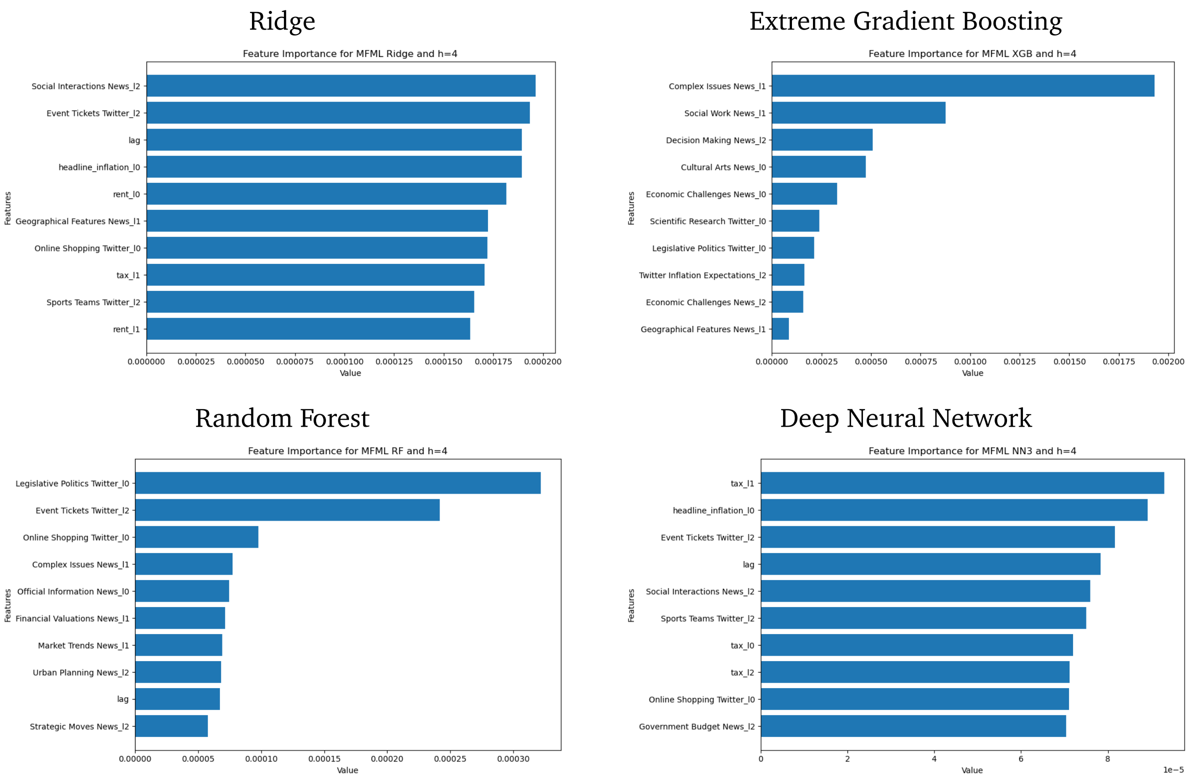


Figure A.14: Inflation Forecasting Variables Importance for $h = 4$

Notes: This figure presents the top 10 important variables at the second quarter. For each selected model, SHAP values are calculated, providing an additive and easily interpretable decomposition of the model's performance. These values attribute the impact of each variable to the prediction, allowing for an assessment of their marginal contributions to the overall forecasting accuracy.

Figure A.15: Comparing Ensemble Linear models with different sets of predictors

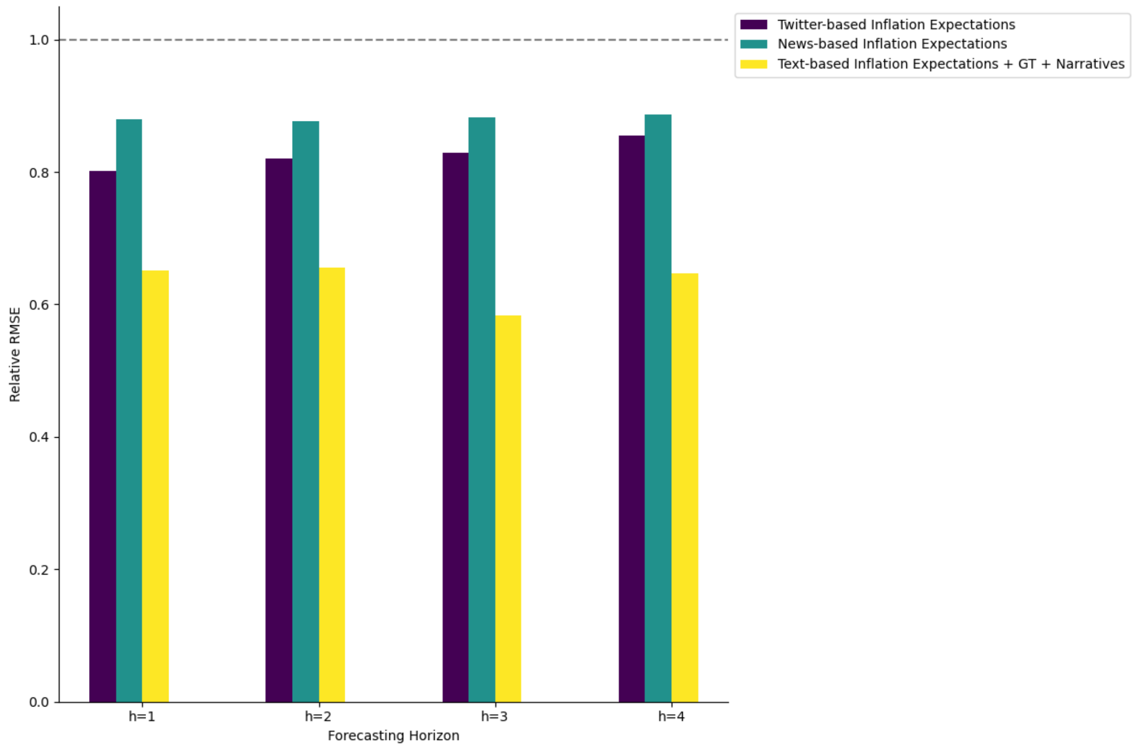


Figure A.16: Comparing Elastic Net model with different sets of predictors

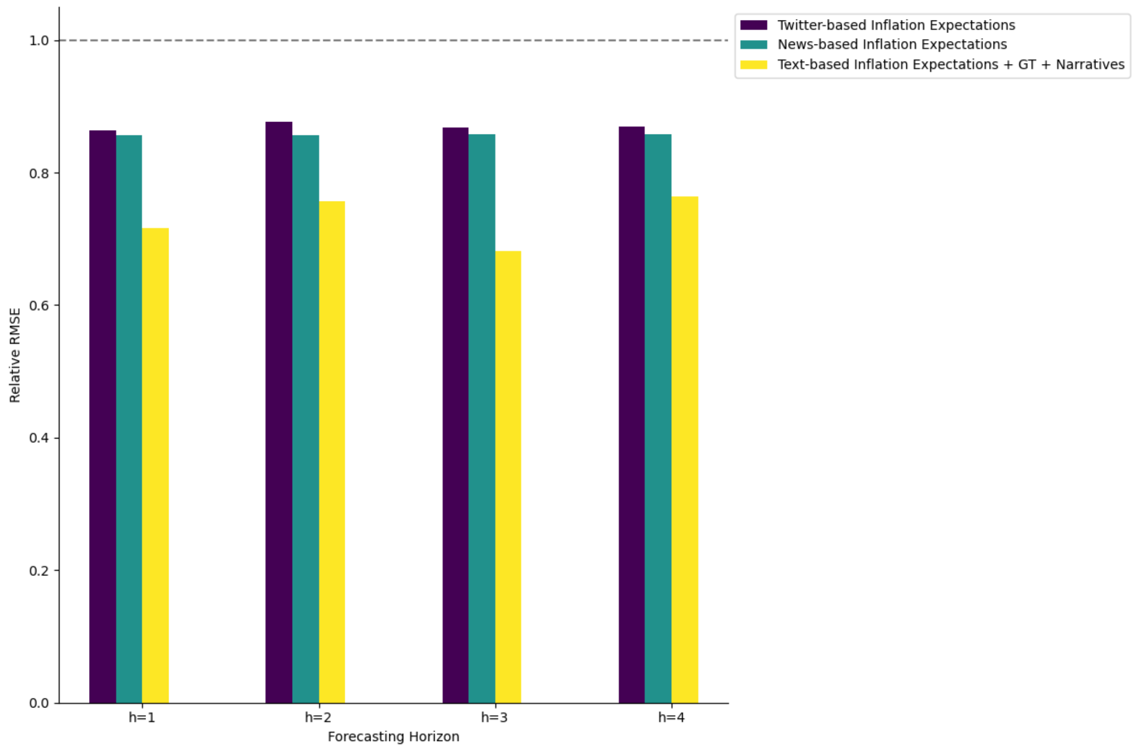


Figure A.17: Comparing LASSO model with different sets of predictors

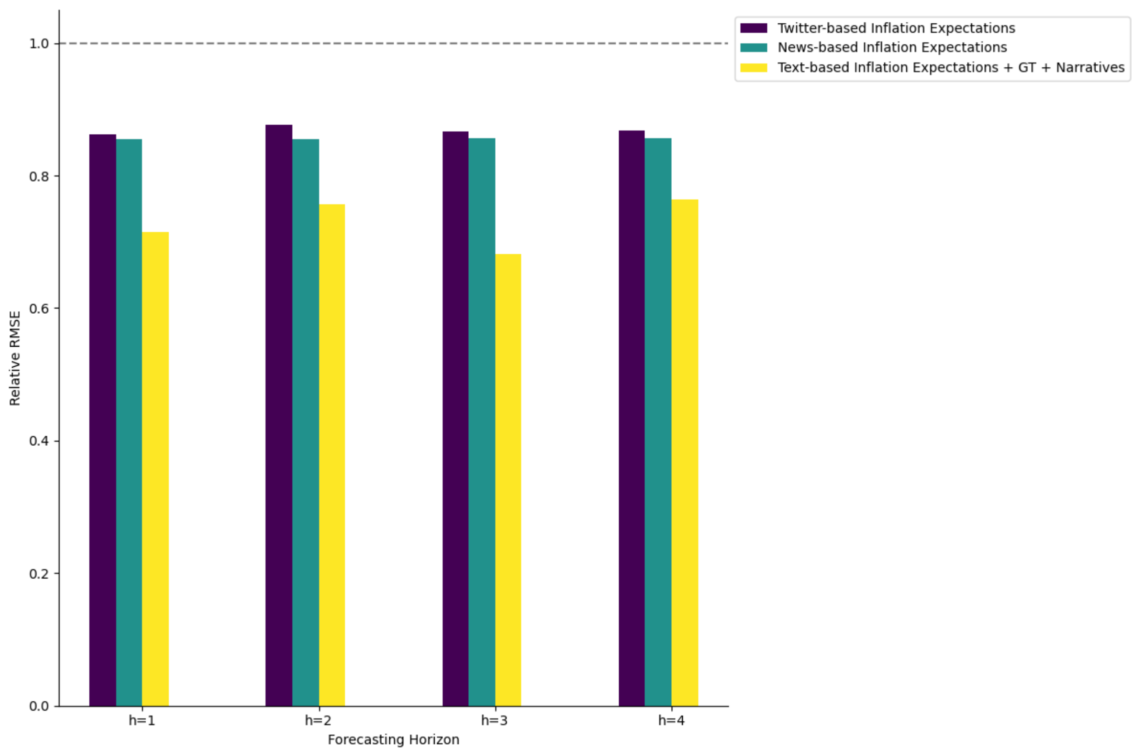


Figure A.18: Comparing Ridge model with different sets of predictors

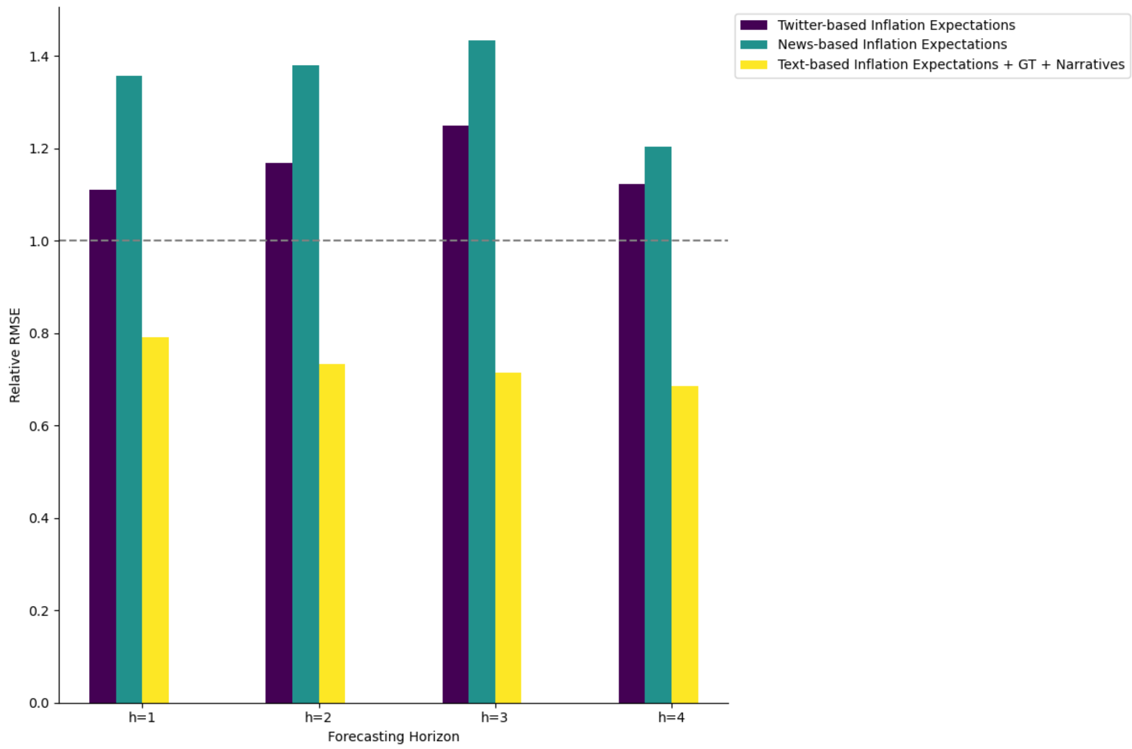


Figure A.19: Comparing Random Forest model with different sets of predictors

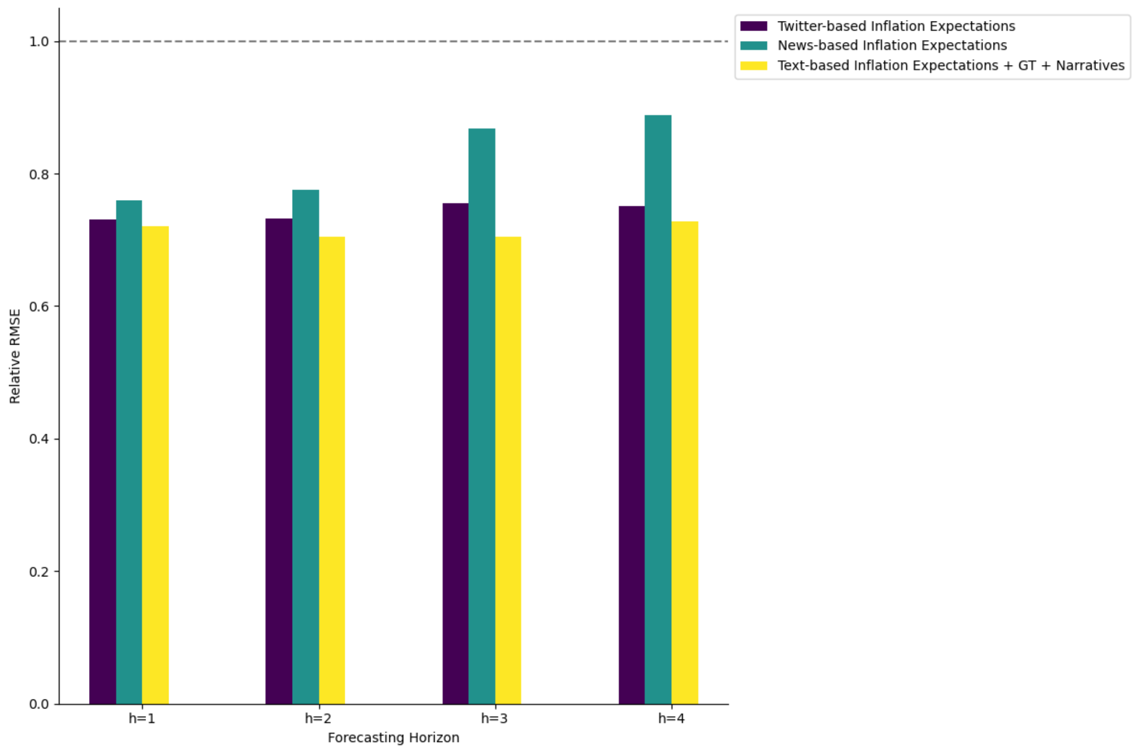


Figure A.20: Comparing Extreme Gradient Boosting model with different sets of predictors

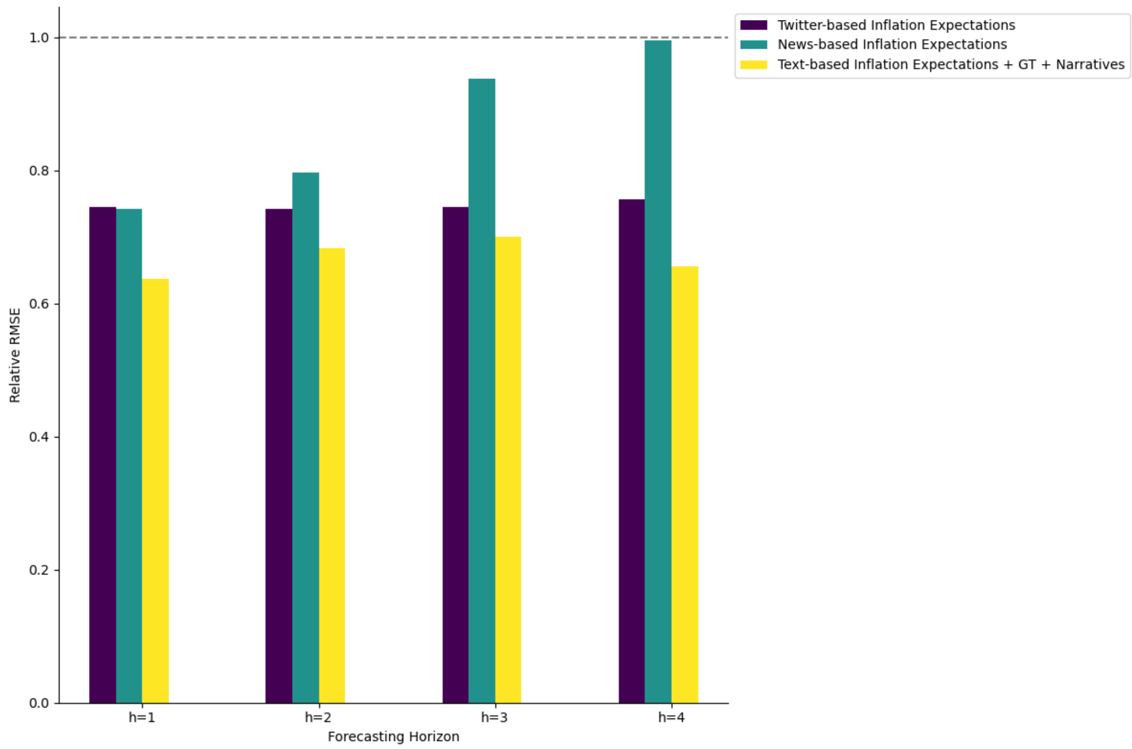


Figure A.21: Comparing Deep Neural Network Model Across Features