PIER

PUEY UNGPHAKORN INSTITUTE
FOR ECONOMIC RESEARCH

News-Based Inflation Expectations: LLM-Assisted Measurement and Forecasting

by

Tanisa Tawichsri, Suppawong Tuarob, Nuwat Nookhwun, and Chinjuta Sa-ngasaeng

May 2026
Discussion Paper
No. 252

News-Based Inflation Expectations: LLM-Assisted Measurement and Forecasting

Tanisa Tawichsri [1,*], Suppawong Tuarob [2],
Nuwat Nookhwun [1], and Chinjuta Sa-ngasaeng [3]

May, 2026

Abstract

We develop a news-based inflation expectations index for Thailand using a scalable workflow that integrates topic modeling, LLM-assisted labeling, and fine-tuned BERT classification. Based on 1.1 million Thai-language news articles from 2015–2024, the index leads both headline inflation and firm inflation expectations. Given that inflation narratives in news are inherently subjective and often ambiguous, we show that prompt design can materially affect downstream economic inference. In out-of-sample forecasting, augmenting autoregressive benchmarks with the news index reduces RMSE by up to 32% for headline inflation and 30% for firm inflation expectations, with gains increasing at longer horizons. SHAP-based decomposition reveals a horizon-dependent information structure: price-specific topics drive short-term forecasts, while macroeconomic narratives dominate at longer horizons. Our findings demonstrate that LLM-assisted text analysis can generate economically meaningful inflation indicators in non-English, emerging-economy settings. The index also performs particularly strong during periods of elevated inflation uncertainty.

Keywords: Inflation expectations, Text-based indicators, Online news data, Large language models (LLMs), Machine learning, Sentiment analysis, Nowcasting and forecasting, Emerging economies,
JEL Codes: E31, E37, D84

[*] *Corresponding author:* `tanisat @ bot.or.th`
[1] Puey Ungphakorn Institute for Economic Research, Bank of Thailand
[2] Department of Linguistic, Chulalongkorn University
[3] Faculty of Information and Communication Technology, Mahidol University
Competing interests and declaration: The authors declare no conflict of interest. The authors used GPTs to assist with language editing and clarity. All outputs from this tool were reviewed, edited as needed, and the authors take full responsibility for the content. **Acknowledgment**: This research receive research grant from Puey Ungphakorn Institute for Economic Research (PIER). We thank Jakphun Nimthong for excellent research assistance. We also thank colleagues at PIER for valuable feedback from research sharing sessions and the PIER grant committee for constructive comments; all remaining errors are our own.

1 Introduction

The news media is one of the most frequented information sources for most people, thereby shaping expectation formation and economic decisions, and eventually economic fluctuations (Shiller, 2017). News coverage may contain information on macroeconomic data releases, financial market developments, or government policies, which in turn influence sentiment and behavior. Empirical studies on information rigidities, notably Carroll (2003), show that news plays a key role in shaping households' inflation expectations. Furthermore, subsequent text-based research finds that news topics and sentiment improve macroeconomic forecasting and provide real-time signals that complement traditional data (Larsen et al., 2021; Shapiro et al., 2022).

The emergence of Large Language Models (LLMs) has been an important advancement in Natural Language Processing (NLP) for analyzing textual data, offering opportunity to enhance the proficiency of textual analyses. Recent applications in macroeconomics illustrate the breadth of this potential: Allard et al. (2024) employ a BERT-based LLM to derive inflation sentiment from U.S. news; Faria e Castro and Leibovici (2024) use Google's PaLM to generate inflation forecasts that outperform traditional approaches at almost all forecast horizons; and Kwon et al. (2025) recover growth and inflation sentiment from financial news while decomposing aggregate sentiment into structural demand and supply drivers. However, existing applications focus largely on advanced economies and English-language sources, leaving open whether these methods generalize to non-English, low-resource settings. Building on this literature, we examine whether Thai online news predicts actual inflation and inflation expectations. As a small open economy exposed to imported inflation via commodity prices and exchange rates and global uncertainty, timely monitoring of inflation narratives is especially relevant for policymakers.

In this paper, we develop a high-frequency, news-based inflation expectations index for Thailand, leveraging a scalable and transparent workflow that integrates topic modeling, LLM-assisted labeling, and supervised text classification. Our news database consists of 1.1 million online news articles from six major news agencies over 2015 to 2024.

The pipeline proceeds in four stages, each offering distinct contributions. First, we apply Latent Dirichlet Allocation (LDA) to identify economically relevant articles. This choice follows established work using topic models for macroeconomic monitoring and forecasting, including Thorsrud (2020), Angelico et al. (2022), and the more recent forecasting application by Adämmer et al. (2025). This filtering step not only improves the signal-to-noise ratio for downstream classification but also provides an interpretable topic structure that we later exploit for SHAP-based decomposition of forecast drivers.

Second, we explore prompt specifications, specifically zero-shot, few-shot, detail-driven, and rule-based factor-driven prompts, to generate LLM-assisted training labels using Gemini 2.5 Pro. While much of the existing literature relies primarily on zero-shot or few-shot prompting, our evidence suggests that prompt design meaningfully influences the economic coherence of the resulting indicators, as judged by stronger alignment with observed inflation dynamics, improved out-of-sample forecasting performance, and more interpretable topic-level decompositions. On this basis, the detail-driven prompt emerges as our preferred specification, delivering stronger downstream economic validation and interpretability than more rigid rule-based prompting, despite slightly lower classification agreement.

Third, we fine-tune three transformer architectures—PhayaThaiBERT, XLM-RoBERTa, and LaBSE—and find that the Thai-specific monolingual model outperforms multilingual and language-agnostic alternatives in our setting. To our knowledge, this is among the first studies to benchmark these architectures for economic news classification in Thai language. Interestingly, in contrast to prior work where multilingual models often perform strongly in low-resource tasks, the updated PhayaThaiBERT architecture performs best here, likely reflecting improvements in handling English loanwords and mixed-language terminology common in financial news.

Finally, we validate our news-based indices to assess their economic meaningfulness, evaluate their predictive performance, and analyze topic-level contributions across forecasting horizons, offering insight into the mechanisms driving inflation dynamics and expectation formation.

This paper makes four main contributions. First, we develop a scalable and transparent pipeline for constructing news-based inflation indices in low-resource language settings, combining LDA-based filtering, LLM-assisted labeling, and fine-tuned BERT classification in a fully replicable workflow. By leveraging LLMs to generate consistent and economically grounded training labels at low cost, the approach helps overcome two barriers that have constrained text-based measurement in emerging market contexts: limited cross-lingual transfer from existing English-domain models in our empirical setting–reflecting both linguistic differences and contextual mismatch between domestic news narratives and the external corpora on which such models are trained–and the prohibitive cost of large-scale expert annotation.

Second, we contribute to the prompting literature by documenting that prompt design can have meaningful downstream implications for economic measurement, even when differences in standard classification metrics are small—complementing and extending the caution raised by Ludwig et al. (2025). Third, we provide new evidence on transformer architecture performance for Thai-language economic text. Relative to prior benchmarking work (Tuarob et al., 2025), we evaluate more recent Thai-specific transformer models, which incorporate improved pretraining and language representations, and assess their performance in an economically grounded classification task. Fourth, we contribute to the literature on information rigidities and expectation formation by documenting the predictive content of economically relevant news for inflation and expectations, and by identifying the topics that drive inflation dynamics and expectation formation across agents and forecast horizons—evidence with direct implications for central bank communication in emerging economies.

Validation exercises show that the news-based index leads both headline inflation and firm one-year-ahead inflation expectations by one quarter to one year, establishing it as a timely leading indicator of inflationary pressures. The index closely tracks major inflation episodes well, including the disinflation episode at the onset of COVID-19, the surge in 2022 following the Russia–Ukraine war, and the subsequent rapid decline. In out-of-sample forecasting using linear and tree-based machine learning models, augmenting autoregressive benchmarks with the news index reduces RMSE by approximately 17–46% for headline inflation and 12–44% for firm inflation expectations across horizons, with gains increasing at longer horizons. Diebold–Mariano tests confirm that these gains are statistically significant.

Improvements for household expectations are more limited and concentrated at medium-term horizons—a finding consistent with the literature on heterogeneous information formation, which shows that households tend to rely more on personal price experience than formal news media, while firms more actively monitor macroeconomic information channels (Carroll, 2003; Weber et al., 2022).

Finally, SHAP-based analysis reveals a horizon-dependent information structure: price-specific topics such as energy and food prices drive short-term forecasts by capturing the magnitude of immediate cost-push shocks, while broader macroeconomic narratives—global economy and Thai financial market news—become the dominant drivers at longer horizons, reflecting the anticipatory nature of macro-level signals. This decomposition offers policymakers a window into the economic mechanisms through which news shapes inflation expectations over time.

Related Literature

This paper relates to several strands of literature. Most closely related are Allard et al. (2024) and Ayivogji (2023), who employ a fine-tuned BERT LLM to extract inflation-related sentiment in news. In the U.S. context, Allard et al. (2024) use GPT-4 Turbo for labeling news articles during the fine-tuning process, and show that their news sentiment index only marginally improves inflation nowcasting accuracy during the pandemic. Meanwhile, Ayivogji (2023) combines newspaper articles with tweets across Canadian provinces, and then extracts inflation sentiment using a RoBERTa-based model trained on manually labeled sentiment. The author finds substantial improvements in predicting both actual and expected inflation. Born et al. (2023), on the other hand, apply a BERT-based model to German tweets, and show that their inflation sentiment index not only aligns well with realized and expected inflation, but also responds to monetary policy shocks and drives household decisions. Relative to these studies, our paper explores an emerging market context while combining manual and LLM-assisted labeling during BERT fine-tuning to simultaneously enhance labeling accuracy and address time and cost constraints.

Other papers resorting to GPT-based LLMs illustrate the broader potential of generative AI for macroeconomic prediction. Faria e Castro and Leibovici (2024) use Google's PALM to generate in-sample conditional inflation forecasts, which outperform those based on traditional methods such as the Survey of Professional Forecasters (SPF) at almost all forecast horizons. Meanwhile, Sun and de Bondt (2025) task ChatGPT with classifying an activity sentiment score based on PMI news releases, and use it to enhance euro area GDP nowcasts, whereas Kwon et al. (2025) use LLM to recover both growth and inflation sentiment from financial news. Taking a more historical perspective, Bybee (2023b) and Bybee (2023a) use LLM GPT-3.5 to simulate economic expectations over various financial and macroeconomic variables based on a sample of news articles. All these contributions highlight the usefulness of LLM models in providing an inexpensive and accurate complementary approach to generating forecasts.

Our paper also relates to earlier contributions leveraging more traditional NLP tools. For example, real-time measures of inflation expectation indicators have been constructed from Italian tweets by combining LDA with a dictionary-based approach (Angelico et al., 2022), whereas Larsen et al. (2021) and Gabrielyan et al. (2020) show that US and UK news topics, derived from LDA, can be good predictors of both inflation and inflation expectations. Meanwhile, Shapiro et al. (2022) show that lexicon-based news sentiment is a strong predictor of Michigan consumer sentiment and can potentially influence output, interest rates, and inflation. Eugster and Uhl (2024) use algorithmically scored sentiment of 730,000 news articles from Refinitiv News Analytics and find that it outperforms random walk or often used benchmarks toward predicting US inflation up to eight quarters.

Finally, this paper connects to the literature on media as a source of information for expectation formation. Carroll (2003) proposes a model where consumers sporadically update their beliefs from media coverage, and shows that more frequent inflation reporting improves the accuracy of consumer expectations. Lamla and Lein (2014) and Dräger and Lamla (2017) find support for this imperfect information model using Germany and US data, respectively. However, Pfajfar and Santoro (2013) show disconnection among news on inflation, consumers' frequency of expectation updating and the accuracy of their expectations. All the papers above rely on the frequency of inflation news reporting as a news measure.

The rest of the paper is organized as follows. Section 2 provides an overview of data and the methods used to construct news-based inflation expectations that span topic modeling, LLM-assisted sentiment labeling, and fine-tuned BERT. Section 3 presents the aggregate inflation index and its topic-level components, followed by a series of validation exercises in Section 4. Then, Section 5 investigates their out-of-sample forecasting performance and feature importance. We conclude with discussions in Section 6.

2 Methodology

This section outlines the construction and validation of our news-based inflation expectations index, which proceeds in five main steps. We first collect and preprocess online news data. Next, we identify economics-related articles through topic modeling. We then classify the filtered corpus by inflation sentiment using LLM-assisted labeling and fine-tuned BERT models. The resulting sentiment scores are aggregated into an index that can be constructed at both daily and monthly frequencies. Finally, we validate the index against benchmark measures of realized inflation and inflation expectations and assess its predictive power, and interpretability of the index.

Our approach integrates traditional topic modeling with modern language models to extract economic signals from textual data in Thai, a low-resource language setting. In particular, we combine interpretable topic structures from LDA with context-sensitive classification using LLM-assisted labeling and fine-tuned BERT models. To ensure robustness and cross-linguistic comparability, we experiment with several transformer-based architectures and evaluate their relative performance in classifying news text by economic relevance and inflation-related sentiment. This hybrid framework contributes both a scalable pipeline for constructing text-based indices in low-resource environments and new evidence on which model types perform best for news corpus and economic applications. Finally, the combination of LLM-assisted labeling and fine-tuned BERT models offers a practical balance between methodological transparency, computational efficiency, and cost-effectiveness.

2.1 Data

This subsection describes the datasets used for index construction and validation, including online news articles and benchmark measures of realized inflation and inflation expectations.

2.1.1 Newspaper Text

Our news dataset spans from 2015 to January 2024, a period that includes phases of low inflation during 2015–19, a brief period of negative inflation rates during COVID-19, and post-pandemic high inflation episodes in Thailand. This period also coincides with several major macroeconomic and policy events, including the COVID-19 pandemic, the Russia–Ukraine war, and subsequent monetary-policy tightening by the Bank of Thailand.

The corpus consists of approximately 1.16 million news articles from six leading online news agencies. Each entry includes the publication date, agency, title, and full text, with an average length of about 152 words. From the full corpus, we retain only economically relevant articles based on our topic classification, resulting in 69,852 news items. Of the six news agencies in our dataset, three specialize in economic and financial reporting, while the remaining three are general-interest outlets. To ensure consistency and relevance, we focus on the specialized agencies as our baseline sample, leaving 47,428 articles of which 23,475 articles are relevant toward conveying inflation sentiment (H, L, or U). We restrict our baseline to these specialized agencies because they offer (i) greater temporal coverage, (ii) thematic relevance to macroeconomic issues, and (iii) higher data consistency across time.[1] Results using the broader six-agency sample are also available (shown in Appendix C); the overall direction and dynamics of the indices remain similar, with minor variations in magnitude.

To evaluate the information content of the news-based index, we benchmark it against official *hard data counterparts*, including realized inflation measures and survey-based inflation expectations of firms and households.

2.1.2 Realized Inflation

We use the year-on-year growth of the monthly headline consumer price index (CPI) published by Thailand's Ministry of Commerce as our primary measure of realized inflation. The CPI is constructed from a nationally representative household consumption basket and captures price changes across a broad range of goods and services consumed by Thai households.

To examine heterogeneity across expenditure categories, we also utilize disaggregated CPI sub-indices. These include raw food, energy, and core inflation. The core inflation is further decomposed into food and non-food components. These subgroup analyses allow us to assess whether the news-based inflation index captures variation in inflation dynamics across distinct consumption categories.

[1] Note that due to technical difficulties in scraping news, we were not able to collect some of the news from the general interest agencies in some time period, resulting in inconsistent temporal coverage for some of these agencies.

2.1.3 Firm Inflation Expectations

Firm-level inflation expectations are drawn from the Bank of Thailand's monthly Business Sentiment Index (BSI) survey. The firm survey asks businesses across a range of sizes and sectors about their expected inflation rate over the next 12 months, i.e., short-term inflation expectations. However, it is important to note that the BSI survey is not designed to be nationally representative of all firms in Thailand. Respondents tend to be larger firms and are disproportionately concentrated in Bangkok and surrounding areas, and the results should therefore be interpreted as reflecting the inflation outlook of relatively large, formal-sector firms rather than the broader population of Thai firms. On average, around 570 firms respond each month over the sample period.

2.1.4 Household Inflation Expectation

Household inflation expectations are drawn from the Consumer Confidence Survey conducted by the Ministry of Commerce, which asks respondents about their expected inflation rate over the next 12 months. The survey also records demographic characteristics—such as age and income—allowing us to decompose inflation expectations across different household subgroups. However, one caveat is that the survey has been launched nationwide since January 2019, and so we only have household expectation data from this period.

2.2 Text Classification and Model Training

This subsection describes the multi-stage text-classification pipeline used to identify relevant articles and determine their inflation-related sentiment.

2.2.1 Topic Classification by LDA

We apply Latent Dirichlet Allocation (LDA) to the full news corpus to cluster articles endogenously based on their underlying textual content. The number of topics is selected through an empirical tuning exercise over a range of 70 to 2,000 topics, with 400 topics yielding the optimal balance between topic coherence score (similarity within topic) and U_{mass} (topic distinctiveness). The resulting 400-topic distributions are screened using their top-ranked words by topic probability to identify economically meaningful content. Relevant topics are selected through expert judgement by two researchers, with supplementary suggestions generated using ChatGPT as a safeguard against potential omissions, and all such suggestions are subsequently reviewed and validated by the researchers.

This process yields a final set of 23 topics related to economic activity, financial markets, and inflation. Topics are labeled based on the topic keywords and word-cloud visualizations and are grouped into five broad thematic categories, namely 1) business, industry and consumer, 2) energy and utilities, 3) food and agriculture, 4) macro, markets and FX, and 5) policy and public services, to enhance interpretability, as shown in Table 1. Word clouds for the five broad thematic categories are presented in Figure 1, while word clouds for the subtopics are provided in Appendix B.

2.2.2 LLM-Assisted Labeling and Fine-tuning BERT

After the initial screening of economically relevant news using topic modeling, we classify articles related to inflation and inflation expectations using LLMs. Articles are first labeled for relevance to inflation expectations (0 for irrelevance), and relevant articles are then

Table 1: Classification of economic and inflation-related news topics

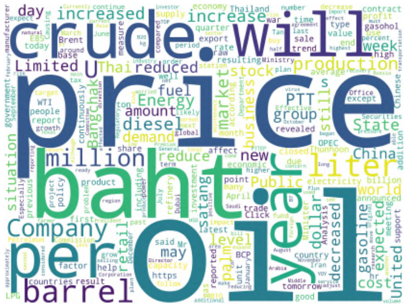
Business, Industry & Consumer	Macro, Markets & FX
Business News	Debt & Bank Loans
Market Trends	Global Economy
Thai SMEs	Thai Financial Market / Interest Rate
Thai E-commerce	Commodity Prices
Energy & Utilities	Economic News / Crisis
Crude Oil	Geopolitics in Asia
Energy Prices	US Monetary Policy
Electricity Rates	**Policy & Public Services**
Retail Oil Prices	Thai Fiscal Policy
Biofuel	Thai Public Policy
Food & Agriculture	Public Transportation
Rubber Prices	Cabinet Meetings
Food Prices	
Corporate & Agribusiness	

assigned a directional sentiment indicating whether inflationary pressures are expected to be high (H), low (L), or uncertain (U). To assign inflation sentiments, we employ a two-stage classification framework. LLM-generated labels, produced using Gemini 2.5 Pro,[2] are used to construct a training dataset of 5,800 news articles for fine-tuning transformer-based models. This approach enhances consistency, scalability, and reproducibility while preserving methodological transparency.

To assess labeling reliability, two expert researchers independently annotate a validation set of 400 news articles. Initial agreement is moderate (70% agreement; macro F1 = 0.58), reflecting the inherent subjectivity of inflation-related narratives in long-form news. Disagreements are resolved through structured adjudication, during which experts also

[2]Gemini 2.5 Pro was selected on the basis of cost-effectiveness following a preliminary evaluation in which it performed comparably to GPT-4 on the validation set. Gemini 2.5 Flash was also evaluated but performed less consistently than the two larger models.

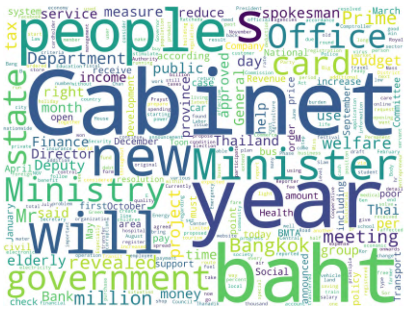
Figure 1: Word clouds of grouped topics



(a) Energy & utilities

(b) Food & agriculture

(c) Macro, markets & FX



(d) Policy & public services

(e) Business, industry & consumer

establish rules for prioritizing economic factors and assigning the uncertain label when conflicting signals arise. The resulting harmonized labels serve as the reference standard for evaluating prompt performance.

We compare four prompt specifications: zero-shot, few-shot, detail-driven, and rule-based factor-driven. While prior studies typically rely on simple zero- or one-shot prompting, we systematically explore richer prompt designs tailored to the economic context. The few-shot prompt provides illustrative examples of high and low inflation expectations, the detail-driven prompt introduces structured definitions and a comprehensive set of inflation drivers (e.g., energy prices, food costs, wages, exchange rates), and the rule-based prompt further incorporates explicit decision rules and factor hierarchies aligned with our adjudication framework (e.g., prioritizing persistent macro pressures over transitory price changes).

We develop the prompt specifications sequentially. We begin with a zero-shot prompt and evaluate its outputs to identify any systematic classification errors. Based on this review, we construct the few-shot prompt by selecting representative examples for each label, prioritizing common cases while also including examples that are frequently misclassified. The detail-driven prompt is then designed by incorporating insights from the few-shot evaluation. In particular, we give a list of factors that influence inflation expectations, and introduce explicit precautions to avoid spurious labeling—-for example, gold prices fall that may reflect risk sentiment rather than disinflation. Finally, we develop the rule-based prompt in alignment with the adjudication guidelines, with the aim of maximizing agreement in settings where multiple factors may simultaneously influence the classification. Selected excerpts of the prompts are presented in Figure 2, with full prompt templates and examples provided in Appendix A.

As expected, the rule-based prompt achieves the highest agreement with human annotations, reflecting its close alignment with expert labeling rules, while the few-shot prompt performs comparably to detail-driven prompt despite its relative simplicity, and both substantially outperform the zero-shot baseline. Notably, this systematic comparison provides a novel contribution by showing how economically structured prompt design shapes classification outcomes beyond standard prompting approaches.

However, classification metrics alone do not determine our preferred specification. In downstream forecasting exercises using aggregate news index (which we will discuss in details in Section 4), the detail-driven prompt performs comparably with rule-based factor-driven prompt. However, the detail-driven prompt produces more economically

Figure 2: Excerpts of Prompt Designs for Inflation-Sentiment Classification

Illustrative Prompt Designs for Inflation-Sentiment Classification

Zero-shot:
"Determine whether each news item contains information that influences inflation expectations in Thailand ... Label each news item with one of the following: 'H', 'L', 'U', or '0'." ...

Few-shot:
"Examples: News Example 1 Label: H; News Example 2 Label: L; News Example 3 Label: U ... This is the news: ..."

Detail-driven You are an economic-news analyst at the Thai central bank. Your task is to classify economic news articles by whether they signal expected inflation trends in Thailand.
... Label "relevance": 1 only if the article contains explicit or strong implied signals about future inflation or price expectations in Thailand. This includes:
- Forward-looking discussions of price pressures, cost changes, or inflation trends
- Anticipated effects from wages, energy, FX, supply shocks, or economic policy
- Global developments only if they are clearly linked to Thai inflation or prices

Label "relevance": 0 if:
- The article discusses only current or past prices
- It discusses trade, demand, or foreign events without clear implications for Thai inflation
- The economic effects are ambiguous or speculative

Rule-based factor-driven:
Consider whether the article contains any of the following factors impacting Thai prices within 12 months ... If the article contains more than 1 factor, identify the group of each factor. The directional label will be determined by the factors in the higher-ranked group (Group A > Group B > Group C > Group D)."

coherent and interpretable SHAP-based topic decompositions. By contrast, the rule-based specification tends to steer ambiguous news toward predefined categories, potentially compressing topic variation and biasing representation.

We therefore adopt the detail-driven prompt as our baseline. This finding aligns with Ludwig et al. (2025), who caution that prompt design can materially affect downstream inference. Our results add a further nuance: even when validation metrics are similar, downstream economic performance may differ. More broadly, rule-based adjudication—whether embedded in prompts or applied by human labelers—can systematically shape topic distributions and influence empirical conclusions.

Aside from prompt specification, we also conduct a model comparison across transformer architectures relevance for Thai-language processing, evaluating three complementary model classes— namely monolingual, multilingual, and language-agnostic—to identify which architectural approach most effectively captures the semantic and contextual structure of Thai economic news.

Specifically, we examine PhayaThaiBERT, XLM-RoBERTa, and LaBSE (Language-Agnostic BERT Sentence Embedding), which represent monolingual, multilingual, and language-agnostic modeling philosophies, respectively. PhayaThaiBERT, the most up-to-date Thai-specific model trained on expanded corpora and code-switched text, has shown strong performance on Thai news classification tasks, while XLM-RoBERTa is a leading multilingual model with strong empirical results on Thai-language benchmarks (Sriwirote et al., 2025; Conneau et al., 2020). LaBSE is included to represent language-agnostic sentence-embedding architectures widely used in cross-lingual NLP (Feng et al., 2022).

Model-level performance comparisons are reported in Table 2. Among the evaluated models, PhayaThaiBERT delivers the most consistent performance and is adopted in the final implementation. We attribute this to the model's Thai-specific pretraining, which likely handles domain-specific economic vocabulary and code-switched text more effectively than multilingual alternatives. The language-agnostic LaBSE model, while offering broader cross-lingual coverage, is substantially more computationally intensive without producing meaningful performance gains in this application. We note that the relative advantage of monolingual models may not generalize across languages and is likely to depend on the availability and quality of language-specific pretraining corpora.

Table 2: Performance of Transformer Models in Classifying Inflation-Expectation Sentiment

Model	**Macro F1**	**Micro F1**	**Accuracy**
PhayaThaiBERT	0.50	0.58	0.62
LaBSE	0.47	0.56	0.60
XLM-RoBERTa	0.47	0.57	0.59

Notes: Performance metrics are computed using the expert-labeled validation set. Classes correspond to non-inflation-related text (O), high inflation expectations (H), low inflation expectations (L), and uncertain or ambiguous narratives (U). The class distribution is unbalanced, with non-inflation-related articles (O) representing the largest share of observations, followed by the H, L, and U categories. All models are fine-tuned with fixed random seeds and deterministic GPU operations; reported results correspond to the checkpoint with the best validation macro F1.

Classification performance against expert labels is moderate: a macro F1 of 0.50 achieved by PhayaThaiBERT represents performance only slightly below the ceiling set by human disagreement itself (macro F1 = 0.58). Such results are consistent with prior multiclass sentiment classification tasks in economics, where moderate macro F1 scores are expected given label subjectivity and class imbalance (Allard et al., 2024; Ayivogji, 2023).

2.3 Index Aggregation

We aggregate article-level directional scores into a composite index of inflation sentiment. The index can be constructed at high frequency, allowing daily tracking of sentiment dynamics. This feature offers potential value for policymakers and analysts seeking alternative, fast-moving indicators of inflation sentiment.

In the analysis that follows, we focus on monthly series derived from simple moving averages over 30- and 60-day windows, while noting that higher-frequency applications remain feasible for future work. We also experiment with alternative functional forms in the aggregation step, including transformations based on the logarithm of news counts and normalization by daily news volume, to assess the robustness of the index to variation in media intensity. In addition, we construct indices using both the full set of news agencies compared to the baseline samples that excludes some agencies to evaluate sensitivity to source composition. The resulting aggregated indices are presented in Section 3.

2.4 Validation and Predictive Performance

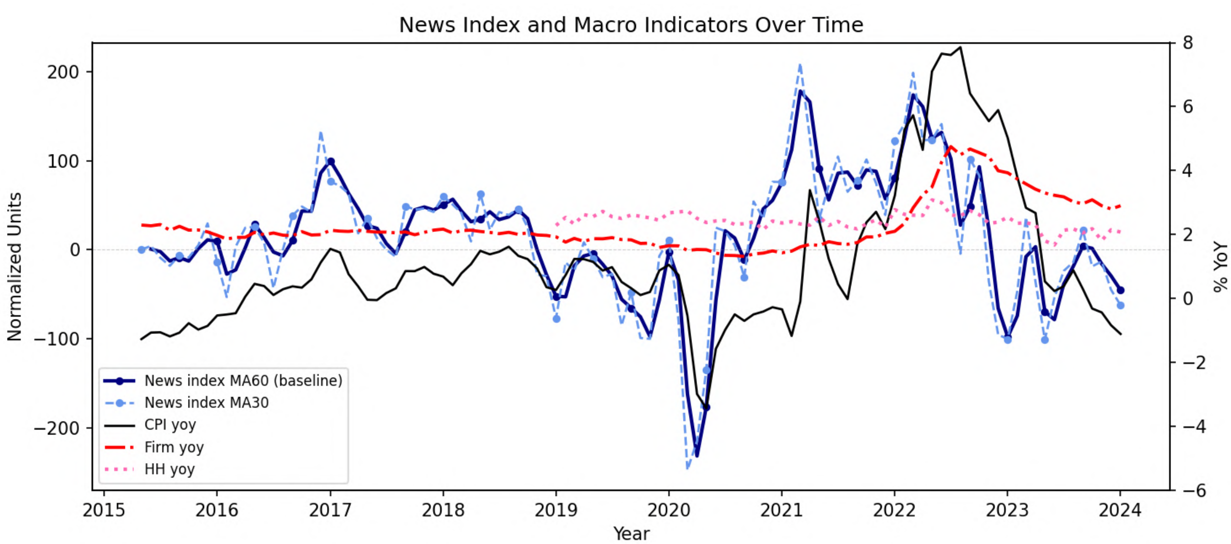
We validate the news-based indices and assess their predictive content in four steps. First, we examine contemporaneous and lead–lag correlations between the index and key benchmark variables—headline inflation, firm inflation expectations, and household inflation expectations—and further explore correlations across CPI subcomponents and firm and household characteristics. Second, we assess the informational value of the index by testing whether it provides explanatory power beyond a univariate autoregressive (AR(1)) process for each target variable, thereby evaluating its incremental content relative to standard time-series dynamics.

Third, we evaluate out-of-sample predictive performance using a range of linear and nonlinear machine-learning models. Forecasts from AR(1) benchmarks are compared with those from models augmented with the news-based index to quantify improvements in predictive accuracy. Finally, we examine model interpretability by analyzing the contribution of topic-level subindices to forecast outcomes using SHAP values. The results of these validation exercises are presented in Section 4.

3 News-based inflation expectation index

3.1 Aggregate news index

Figure 3: News index at monthly frequency compared to realized and expected inflation



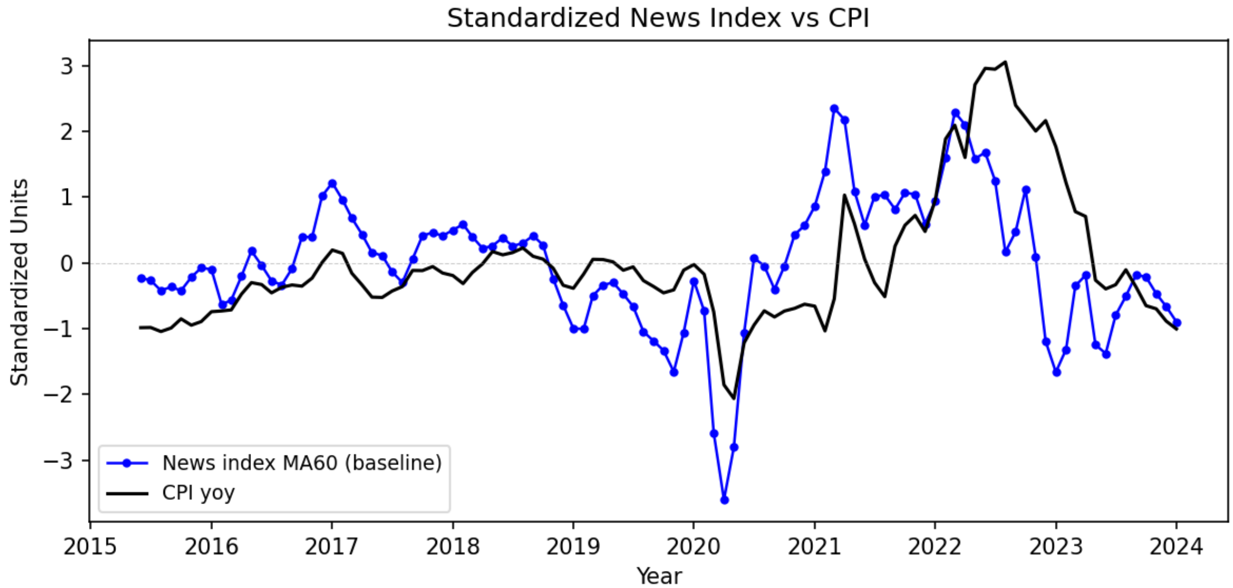
Note: News index normalized to baseline of MA60 in 2017m1 equal to 100. News index on left axis, and headline CPI inflation, firm and household inflation expectation (%yoy) on right axis

Figure 3 presents the aggregate news-based index constructed using 30-day and 60-day moving averages alongside the target variables. Figure 4 reports the standardized baseline index (MA60), which facilitates visual comparison and assessment of co-movement across series.

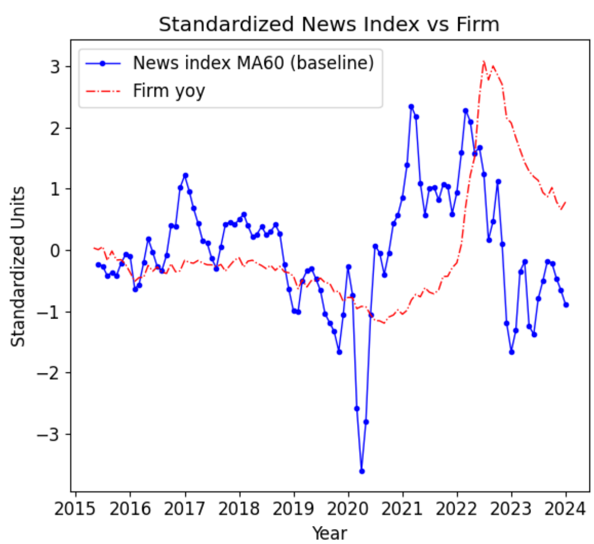
The news-based index and headline inflation exhibit broad co-movement over time, with the index capturing major inflationary and disinflationary episodes (Figure 4a). For example, the index successfully captures the episode of negative inflation at the onset of the COVID-19 pandemic, the surge in 2022 following the Russia–Ukraine war, and the subsequent rapid decline. The news-based measure is more volatile than CPI, consistent with its role as a high-frequency indicator of sentiment rather than realized price changes. The close alignment between the two series suggests that the index tracks underlying inflation dynamics while providing earlier signals, and potentially richer at daily frequency, of shifts in inflation sentiment.

Figures 4b and 4c show the relationship between the standardized news-based index and survey-based measures of firm and household inflation expectations. Firm expectations appear relatively well-anchored ahead of the Russia-Ukraine war, displaying smooth dynamics and small fluctuations while maintaining broad co-movement with the news-based index. During the major inflationary episode in 2022, both the news index and firm expected inflation sharply increase, indicating firms' attention to oil shocks. Firm expectations remain persistently high thereafter despite the news index falling into the negative territory. Meanwhile, household inflation expectations in levels exhibit limited variation; however, after standardization, they become more broadly consistent with the dynamics of the news-based index, especially from 2022.

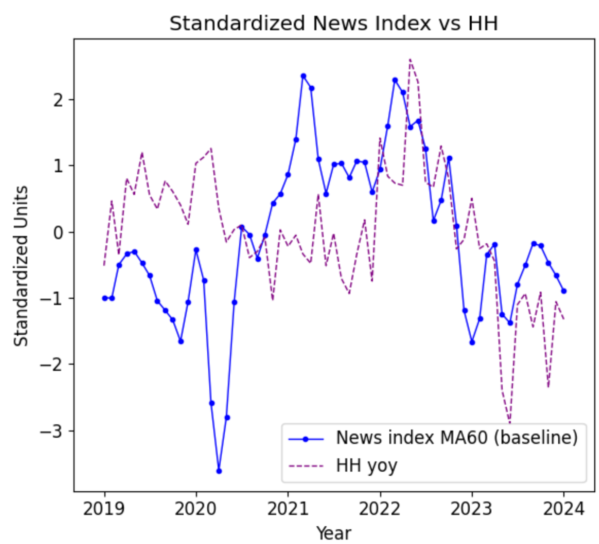
Figure 4: Standardized news index compared to inflation and inflation expectations



(a) News index vs headline inflation



(b) News index vs firm expectation



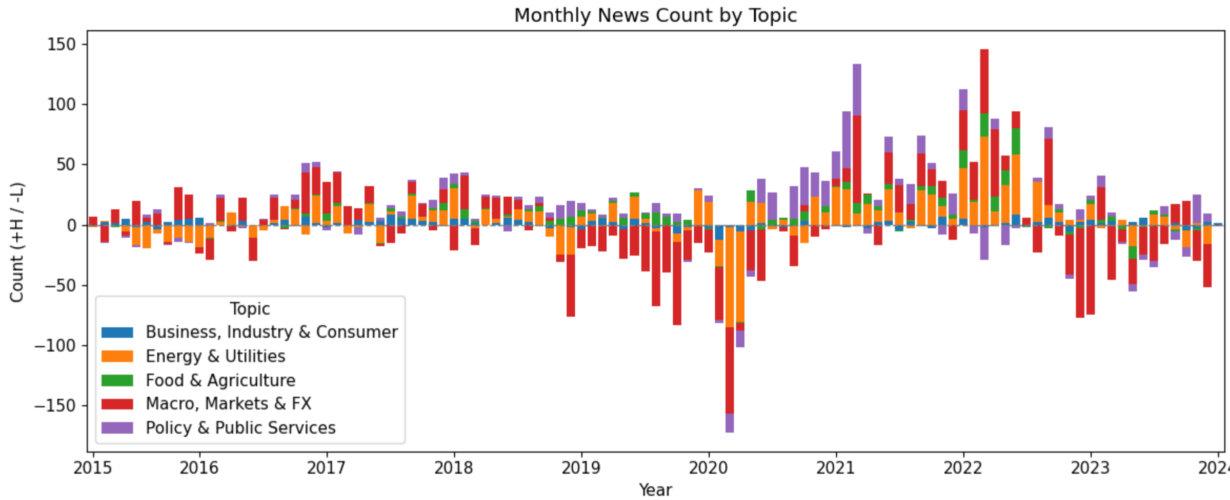
(c) News index vs household expectation

To further isolate unexpected movements from persistence in inflation-related series, we examine the relationship between the news index and AR(1) innovations (the error terms). The news index co-moves with the AR(1) innovations of headline inflation and firm inflation expectations, but shows little association with household expectation innovations. Consistent with this pattern, both Pearson and Spearman correlations are statistically significant for innovations of both headline inflation and firm expectations, while correlations with household expectation innovations are weak and insignificant (Figure C.1), suggesting that the types of news used to construct the index may not fall within households' information sets.

The weaker co-movement with household inflation expectations is not surprising. Unlike firms, households are heterogeneous in their exposure to economic news media, and tend to form inflation beliefs through direct personal experience rather than formal information channels. This divergence between firm and household expectations—and its implications for the informational content of the news index—is examined more formally in Sections 4 and 5.

Finally, we present alternative versions of the index to assess robustness. Indices constructed under different aggregation schemes—including versions that vary the composition of included news sources and normalize by daily news volumes—exhibit qualitatively similar dynamics and quantitatively comparable behavior, with consistent directional movements and some differences in magnitude (see Appendix C).

Figure 5: Inflation sentiment news index by grouped topics



3.2 News index by topic

Next, we examine the index composition by five broad topics (business, energy, food, macroeconomics, and policy). Figure 5 decomposes the inflation sentiment news index by broad topic groups, where contributions are reported as net numbers of news articles between high- and low-inflation sentiment within a broad topic. Energy-related and macro-financial news accounts for the largest share of variation in the index over time, particularly during major inflationary episodes.

Notably, some periods—such as 2021—exhibit elevated signals driven by policy and public-services coverage even in the absence of sustained realized inflation, suggesting that policy-oriented news can generate inflationary narratives independent of actual price outcomes. In contrast, periods of sustained high inflation are primarily associated with stronger contributions from energy and macro-financial news.

4 Validation

In this section we test external-validity of our indices by looking at their correlation and informativeness to realized inflation and surveyed inflation expectations.

4.1 Correlation

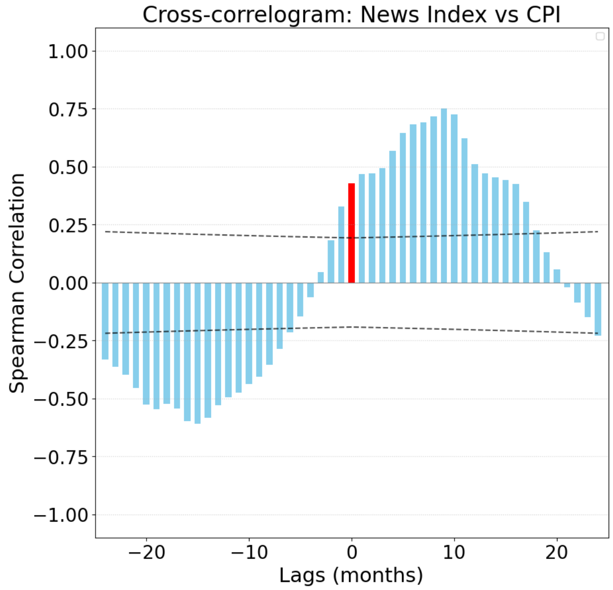
The cross-correlogram shows the Spearman correlation, which is rank-dependent and does not assume linear relationship, of the news index at time t to the lead terms $(t + l)$ or lag terms $(t - l)$ of the benchmarks. Figure 6 confirms that the news index is strongly associated with headline CPI inflation and firm inflation expectation, while its relationship with the household inflation expectation is weak and not statistically significant. It also tends to lead both CPI inflation and firm inflation expectations by one quarter to one year, suggesting that the news-based index captures anticipatory information relevant for subsequent inflation dynamics.

Next, we examine the contemporaneous correlations between the index and CPI components, as well as inflation expectations of firm and household subgroups, noting that these relationships reflect only same-period associations rather than lead–lag dynamics. We find that the news-based index is strongly correlated with the energy component of the CPI, while correlations with other CPI components appear weak (Figure 7a). This may be attributed to the fact that energy prices frequently appear in news media and their movements can have an material impact on economic decisions and actual inflation. Differences in correlations across firm sizes and sectors are less distinct, which may partly reflect the geographic concentration of the sample in the Bangkok metropolitan area (Figure 7b). Although the contemporaneous correlation between the news index and aggregate household inflation expectations is relatively small, the association appears somewhat stronger among older respondents and those with higher income levels. These patterns suggest that certain population subgroups are more exposed to, or more attentive to, information conveyed through news media (Figure 8).

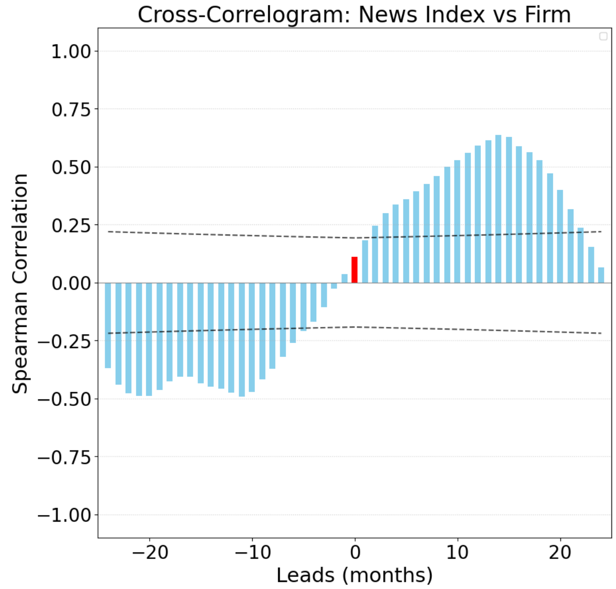
4.2 Informativeness exercise

We next examine whether the news-based inflation sentiment index provides incremental information beyond standard time-series dynamics for Headline CPI inflation and inflation expectations. Specifically, we estimate an AR(1) model for headline inflation and inflation expectations and add the news-based index to test whether it contains incremental information beyond lagged outcomes. All variables are standardized to ensure comparability of coefficient magnitudes across regressors. The results show that the news index provides statistically significant additional information—and increases the R^2 —for CPI inflation and firm inflation expectations. However, the overall improvement in model fit is modest: the R^2 increases by only about 0.01–0.02 (roughly 1–2 percentage points). For household expectations, the index adds some incremental explanatory power, although the coefficient is not statistically significant and the R^2 gains are similarly small. Overall, these findings align with the patterns observed in the earlier analyses.

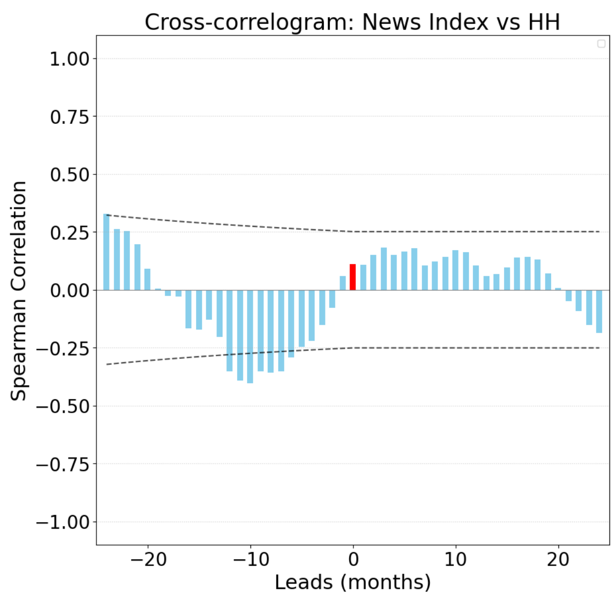
Figure 6: Cross-correlograms of news index with CPI inflation, and firm and household inflation expectations



(a) News index vs. CPI inflation



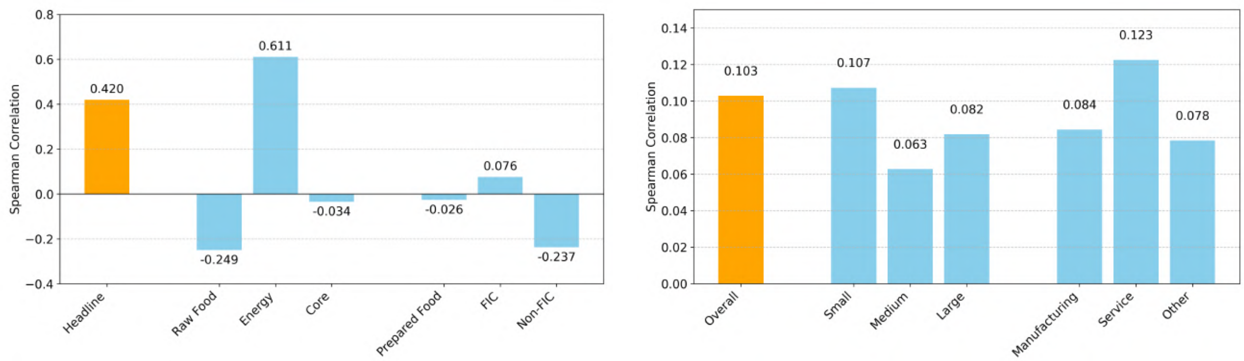
(b) News index vs. firm expectation



(c) News index vs. household expectation

Note: The figure plots cross-correlation coefficients at different lead–lag horizons. Positive lags indicate that the news-based index leads the other variable. Dotted line shows 95% confidence interval.

Figure 7: Correlation of news index to CPI components, and inflation expectations of firms by subgroups

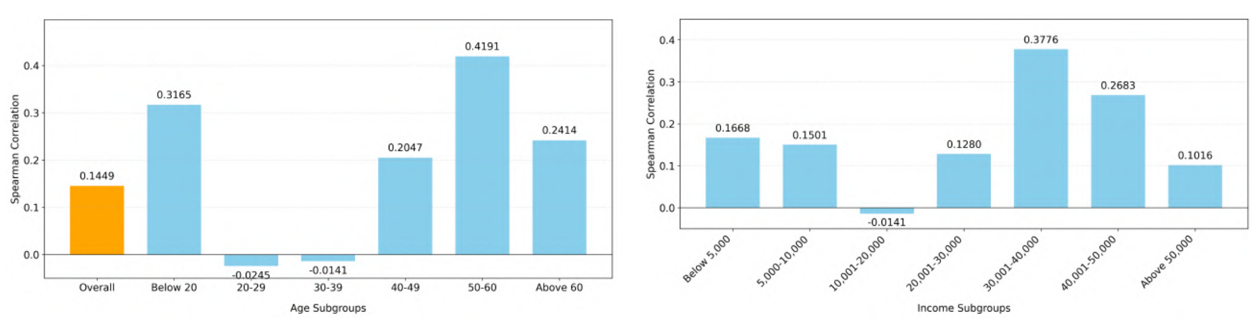


(a) News index vs. CPI inflation

(b) News index vs. firm expectation

Note: FIC is the food component within the core CPI basket.

Figure 8: Correlations of news index to inflation expectations of household subgroups by income and age



(a) News index vs. household expectation by age

(b) News index vs. household expectation by income

Table 3: Informativeness Exercise

	CPI_t	CPI_t	$firm_t$	$firm_t$	HH_t	HH_t
CPI_{t-1}	0.937***	0.880***				
	(0.044)	(0.038)				
$firm_{t-1}$			0.982***	0.983***		
			(0.038)	(0.029)		
HH_{t-1}					0.672***	0.660***
					(0.075)	(0.070)
$NewsIndex_{MA60,t}$		0.183***		0.097***		0.115
		(0.039)		(0.036)		(0.078)
Constant	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.038)	(0.029)	(0.027)	(0.020)	(0.090)	(0.088)
N	103	103	103	103	60	60
Adj. R^2	0.878	0.908	0.965	0.974	0.452	0.465

Coefficients shown with HAC (Newey-West) standard errors in parentheses
*, **, *** denotes p-values of < 0.1, 0.05, and 0.01 respectively

5 Forecast: out-of-sample prediction

In this section, we evaluate the predictive power of our news index for realized inflation and inflation expectations at varying horizons, ranging from nowcasting to forecasting up to 12 months ahead, using monthly data. The forecasting exercise is implemented using an expanding-window approach, ensuring that only information available at each point in time is used to predict future outcomes, thereby preventing any look-ahead bias or data leakage. We begin by assessing the forecasting performance of the aggregate index. Additionally, to understand the contribution of each sub-index and to enhance interpretability of the machine-learning forecasts, we also conduct the forecasting exercise using the topic-specific sub-indexes. We then compute SHAP values to quantify the marginal contribution of each subtopic to the model's predictive performance.

5.1 Out-of-sample prediction

For the forecasting evaluation, the models are trained on pre-2021 data—June 2015 to December 2020 for CPI inflation and firm expectations, and January 2019 to December 2020 for household expectations. Forecasts are then generated for the out-of-sample period from January 2021 through January 2024. Using a direct multi-horizon approach with an expanding window, we evaluate only the baseline MA60 news index and benchmark its performance against a standard AR(1) model without news information.

We select baseline machine learning models with a mix of linear and non-linear models, including Lasso, Ridge, and Elastic Net for the linear class, and Random Forest and XGBoost for the nonlinear class. These models are evaluated against an AR(1) benchmark in out-of-sample forecasting. For linear models, regularization parameters are selected via cross-validation within each expanding window, ensuring that the tuning process uses only information available at the time of each forecast. Regularization parameter selection at each window is an important step in fitting these estimators, as it directly governs shrinkage and variable selection.

In contrast to linear models, tree ensembles perform implicit variable selection through recursive splitting and feature subsampling. Therefore, for the tree-based models, we employ standard default hyperparameter settings (e.g., 500 trees) without per-period re-optimization. This approach follows the macroeconomic forecasting literature: Medeiros et al. (2021) show that forecasting performance is generally robust to alternative tuning choices, while Goulet Coulombe (2024) document that Random Forest exhibits limited sensitivity to hyperparameter selection in macro applications. Using default settings in tree-based models enhances reproducibility, limits the risk of overfitting the tuning procedure to the evaluation sample, and reflects realistic real-time monitoring environments.

We evaluate predictive accuracy using the root mean squared error (RMSE). For a given forecast horizon h, RMSE is defined as

$$RMSE_h = \sqrt{\frac{1}{T_{\text{oos}}} \sum_{t \in \mathcal{T}_{\text{oos}}} \left(y_{t+h} - \hat{y}_{t+h|t}\right)^2},$$

where y_{t+h} is the realized value of the target variable h periods ahead, $\hat{y}_{t+h|t}$ is the forecast made at time t, and T_{oos} is the number of out-of-sample forecasts. We report performance as relative RMSE, defined as the ratio of each model's RMSE to that of the AR(1) benchmark, so that values below one indicate improvements over the benchmark (Tables 4–6). For comparability, we also report the performance of an ARX model: an AR(1) model augmented with our news index. While we do the forecasting exercise at the monthly frequency, we simplify the report as quarterly averages, and use the Diebold-Mariano test to evaluate if the model offers statistically significant prediction power over the benchmark.

Using the aggregate news index for nowcasting tasks, we find that linear models perform marginally better on CPI inflation forecasting and substantially better for firm expectation forecasting. None of the models improve household inflation expectation forecasting.

For multi-horizon forecasting, CPI inflation forecast accuracy improves at most horizons (one to four quarters ahead), with the best relative RMSE ranging from 0.68 to 0.85. The highest-performing model alternates between linear and tree-based model depending on the forecast horizon, with linear model performing better for nowcasting and upto

Table 4: Out-of-sample forecasting exercise for CPI inflation using the aggregated news index: relative RMSE of competing models (reported as quarterly averages)

		Linear ML			Tree ML	
Horizon	**ARX**	**Lasso**	**Ridge**	**Elastic Net**	**Random Forest**	**XGBoost**
$h = 0$	0.9593	0.9550	0.9570	**0.9549**	1.2390	1.5298
Q1	**0.8512*****	0.8529***	0.9949	0.8536***	0.9151***	1.2209
Q2	0.8037***	0.8014***	1.0081	0.7977***	**0.7375*****	0.8335*
Q3	0.7632***	0.7611***	0.8967	0.7743***	0.6908***	**0.6789*****
Q4	0.7791***	0.8095***	0.9142**	0.8138***	**0.6917*****	0.7444***

Note: This table reports the relative RMSE, defined as the ratio of each model's RMSE to that of the AR(1) benchmark. Numbers in boldface indicate the model with the lowest relative RMSE for each forecast horizon. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, based on the Diebold-Mariano test of equal predictive accuracy with Harvey-Leybourne-Newbold (HLN) small-sample correction.

1 quarter ahead, and tree-based models performing better for farther horizons. Firm inflation expectations forecasting also improves across all horizons (h=0 to 4 quarters ahead), with Random Forest consistently achieving the lowest relative RMSE (0.70–0.74) for forecasting. For household expectations, incorporating the news index does not lead to a statistically significant improvement in forecasting accuracy. The AR(1) benchmark performs relatively well, consistent with the high persistence of average household inflation expectations.

Overall, the aggregate index substantially enhances forecasting performance for CPI inflation and firm inflation expectations but offers limited gains for household expectations. Linear models tend to perform best at shorter horizons (nowcasting and one-quarter ahead), while nonlinear models begin to outperform at farther horizons. However, one point to note is that Random Forest and XGBoost are estimated using default hyperparameter settings, whereas the linear models are re-tuned at each rolling window. This

Table 5: Out-of-sample forecasting exercise for firm inflation expectations using the aggregated news index: relative RMSE of competing models (reported as quarterly averages)

		Linear ML			Tree ML	
Horizon	**ARX**	**Lasso**	**Ridge**	**Elastic Net**	**Random Forest**	**XGBoost**
$h = 0$	**0.8460*****	0.8769**	1.0141	0.8776**	1.4725	1.3943
Q1	0.8019***	0.8456***	1.0672	0.8665***	**0.7039*****	0.7312***
Q2	0.8135***	0.8374***	1.0381	0.8456***	**0.7211*****	0.8363***
Q3	0.8183***	0.8325***	0.9670	0.8439***	**0.7386****	0.8629*
Q4	0.7966**	0.8607*	0.8960**	0.8564**	**0.6967****	0.7602*

Note: This table reports the relative RMSE, defined as the ratio of each model's RMSE to that of the AR(1) benchmark. Numbers in boldface indicate the model with the lowest relative RMSE for each forecast horizon. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, based on the Diebold-Mariano test of equal predictive accuracy with Harvey-Leybourne-Newbold (HLN) small-sample correction.

difference in model tuning may partly account for the relatively weaker nowcasting performance of the nonlinear models, particularly at short horizons.

5.2 Explainability

We now examine the contribution of topic-specific sub-indexes to forecasting performance. As a preliminary check, their predictive accuracy—measured by relative RMSE—is broadly similar to that of the aggregate index, with slightly lower RMSE at horizons of one to three quarters ahead for CPI inflation and firm expectations. Notably, topic-level indexes help improve household inflation expectation forecasting slightly (see results in Appendix C.3). We then turn to SHAP values to identify which topics drive these gains and how their contributions vary across horizons.

Since model performance varies across horizons, we select the best-performing specification at each horizon for interpretability analysis-Lasso for nowcasting, where linear

Table 6: Out-of-sample forecasting exercise for household inflation expectations using the aggregated news index: relative RMSE of competing models (reported as quarterly averages)

		Linear ML			Tree ML	
Horizon	**ARX**	**Lasso**	**Ridge**	**Elastic Net**	**Random Forest**	**XGBoost**
$h = 0$	**1.0334**	1.0650	1.0438	1.0658	1.1477	1.1795
Q1	**1.0207**	1.0250	1.1401	1.0291	1.0490	1.1580
Q2	1.0063	1.0164	1.0190	1.0176	**0.9740**	0.9747
Q3	1.0174	1.0005	**0.9947**	1.0002	0.9788	1.0340
Q4	**1.0664**	1.1140	1.0949	1.1111	1.1030	1.2736

Note: This table reports the relative RMSE, defined as the ratio of each model's RMSE to that of the AR(1) benchmark. Numbers in boldface indicate the model with the lowest relative RMSE for each forecast horizon. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, based on the Diebold-Mariano test of equal predictive accuracy with Harvey-Leybourne-Newbold (HLN) small-sample correction.

models dominate, and Random Forest for multi-horizon forecasting, where nonlinear models consistently outperform.

For CPI inflation forecasting, the nowcasting results indicate that lagged headline inflation is the most important predictor, followed by food prices, commodity prices, and electricity prices. As we move along farther forecasting horizons, lagged inflation no longer dominates. At horizons of three and six months ahead, electricity prices, commodity prices, and corporate and agribusiness news contribute most strongly, while at longer horizons of nine and twelve months, broader macroeconomic topics—such as global economy news and Thai financial market news—become the primary drivers (see Figure C.3 for details). This pattern is consistent with energy and food prices acting as fast-moving cost-push signals, while global macroeconomic conditions shape inflation trends over longer horizons.

For firm inflation expectations, the SHAP values highlight strong persistence: the lagged expectation is the most influential feature in nowcasting and remains the top contributor up to six months ahead. Additional important sub-indexes include food prices, commodity prices, and market and business news. Similar to CPI inflation, the contributions shift toward global economy news at the 12-month horizon (Figure C.4). The persistence of lagged expectations as a dominant feature suggests that firm beliefs are highly anchored to recent outcomes, with news providing incremental updating rather than sharp revisions.

For household inflation expectations, the key drivers include electricity prices, corporate and agribusiness news, and topics related to debt and bank loans. Several domestic news categories also play an important role, including Thai financial market news, rubber prices, Thai fiscal policy, and SME-related news (Figure C.5). The more dispersed feature importance for household expectations reflects the weaker and less consistent relationship between the news index and household beliefs documented earlier, suggesting that households draw on a broader and possibly less news-centric information set.

We next draw on these results to characterize the broader information structure embedded in the news index and its influence on inflation and expectation dynamics across forecast horizons.

5.3 Horizon-Dependent Information Structure

News can convey two distinct types of information relevant to inflation forecasting: signals about the *magnitude* of current price pressures, and signals about the *timing* of future inflationary trends. The SHAP-based value analysis in Section 5.2 suggests that the topic-level sub-indexes map naturally onto this distinction across forecast horizons. Price-specific topics—such as “Electricity Rates”, “Food Prices,” and “Commodity Prices”—capture immediate cost-of-living shocks whose effects are felt quickly and accordingly dominate at short horizons. Broader macroeconomic narratives—such as “Global Economy” and “Thai Financial Market” topics—reflect more diffuse and forward-looking signals and become the primary drivers at longer horizons. This systematic shift in feature importance across horizons is not merely a statistical pattern; it reflects the underlying economics of how different types of news enter the inflation process at different lags.

While the transition is gradual rather than discrete, and some topics such as commodity prices retain relevance across multiple horizons, the overall pattern is consistent across all three target variables. This horizon-dependent structure illustrates the key value of the topic-level decomposition: even where predictive accuracy is broadly comparable to the aggregate index, the sub-index analysis reveals which news topics are driving forecasts at each horizon. For policymakers, this is informative in its own right — it identifies whether near-term inflation pressures are being driven by energy price shocks, food prices, or broader macroeconomic developments, and how the balance of these drivers shifts across horizons. These results underscore the value of topic-level decomposition—not just for interpretability, but for uncovering the mechanisms linking news to inflation dynamics and inflation expectations over time.

6 Discussion

This paper develops a timely news-based inflation expectations index using a scalable and transparent workflow that integrates topic modeling, LLM-assisted labeling, and supervised text classification. A key strength of this integrated approach is that each stage serves a distinct purpose within a coarse-to-fine architecture: LDA topic modeling acts as an interpretable retrieval layer that filters economically relevant news and organizes articles into coherent thematic clusters, providing cleaner input for downstream classification as well as a transparent structure for topic-level decomposition; LLM-assisted labeling generates high-quality training data efficiently and at scale; and supervised fine-tuning of a BERT model provides semantic precision, transparency, and reproducibility across the full corpus. Together, these components form a cost-effective pipeline that is transparent—without relying excessively on proprietary black boxes—and adaptable to other low-resource language settings.

This architecture builds on an established strand of text-based inflation monitoring work that uses LDA-based topic models as an interpretable dimensionality-reduction and retrieval device (see Angelico et al. (2022)), while extending that framework with modern language models and supervised classification. At the same time, we view the framework as a benchmark that can be further refined as newer methods emerge. In particular, future work could explore embedding-based semantic retrieval or neural topic models such as BERTopic and dynamic topic models, which may capture evolving narratives and richer semantic structure while preserving interpretability. More broadly, an end-to-end semantic retrieval-classification pipeline offers a promising extension. We view such approaches as complements and refinements to, rather than substitutes for, the interpretable LDA-based benchmark developed here.

The emerging market context also constitutes a distinct contribution. Existing applications of LLM-assisted text analysis to inflation measurement focus almost exclusively on advanced economies and high-resource languages, including the United States (Allard et al., 2024), Canada (Ayivogji, 2023), and Germany (Born et al., 2023). Emerging market economies, however, present a particularly strong case for real-time news-based monitoring. Inflation expectations in emerging economies tend to be more weakly anchored and more sensitive to shocks than in advanced economies, with higher pass-through of exchange rate and commodity price movements into domestic prices (Ha et al., 2019). This makes timely and forward-looking sentiment indicators especially valuable for policymakers seeking to monitor nascent inflationary pressures. By demonstrating that this framework delivers economically meaningful and predictive results in a non-English, emerging market setting, this paper extends the applicability of LLM-assisted macroeconomic analysis beyond the advanced economy contexts in which it has predominantly been developed, and offers a replicable template for similar applications in other emerging markets and low-resource language environments.

The prompting framework also deserves particular attention. We evaluate four prompt specifications — zero-shot, few-shot, detail-driven, and rule-based factor-driven prompts. While the rule-based factor-driven prompt achieves the highest consistency with human-annotated labels, and the few-shot prompt performs comparably to more elaborate alternatives, we find that the detail-driven prompt—which provides explicit definitions of economically relevant factors—produces the most interpretable and economically coherent results in downstream forecasting tasks. This highlights an important insight: prompt selection should not be evaluated solely on classification metrics but also on the quality and interpretability of the resulting economic indicators.

The weaker relationship with household expectations is an intuitive finding that connects to a broader literature on heterogeneous information sets and expectation formation. As Carroll (2003) emphasized, households update their beliefs sporadically and selectively, drawing on accessible and low-cost signals. Our finding that the news-based index correlates more strongly with the expectations of older and higher-income households is consistent with this view: these groups are more likely to regularly consume economic news and to have the financial literacy to process macroeconomic information (Weber et al., 2022). More broadly, our news indices constructed from the filtered news corpus—consisting predominantly of economic and inflation-relevant articles from specialized agencies—is more likely to fall within the information sets of firms and more economically engaged households than of the general population. This heterogeneity in exposure to economic news media represents a natural channel through which rational inattention manifests across agent types.

In forecasting applications, the news index delivers substantial improvements in predictive accuracy for realized inflation and firm inflation expectations. Relative to the AR(1) benchmark, augmenting models with the news index reduces RMSE by approximately 17–46% for CPI inflation and 12–44% for firm inflation expectations, with larger gains generally at longer horizons. These improvements are statistically significant across most horizons and model specifications, as confirmed by the Diebold-Mariano test. Gains for household inflation expectations are more modest and less consistent, with improvements concentrated at the two- and three-quarter-ahead horizons, consistent with the weaker informational link documented above.

We note some caveats regarding the forecasting exercise. The out-of-sample period spans January 2021 to January 2024, which coincides with the highly unusual post-COVID inflation episode; forecasting gains during this period may partly reflect the index's ability to capture an exceptional inflationary surge rather than its performance under more typical or stagnant conditions. The training sample for household expectations is also considerably shorter than for CPI inflation and firm expectations, given that the household survey was launched nationwide only in January 2019, which may further constrain model performance. Addressing these limitations—through a longer evaluation window and a broader inflation cycle—represents a natural direction for future work if given data availability.

Finally, the SHAP-based explainability analysis reveals the economic mechanisms through which news contributes to inflation dynamics at different horizons. Near-term forecasts are dominated by lagged outcomes and price-related topics such as energy, food, and commodities prices, while broader macroeconomic and global narratives become increasingly important at longer horizons. This horizon-dependent structure is economically interpretable, reflecting the distinct roles of cost-push shocks and forward-looking macro signals in shaping inflation dynamics. For policymakers, the decomposition offers a practical tool to monitor which inflationary pressures dominate the news narrative and how these signals evolve over time, complementing traditional survey indicators.

7 Conclusions

This paper develops a transparent and scalable framework for constructing a news-based inflation expectations index using topic modeling, LLM-assisted labeling, and supervised text classification. Applied to a non-English emerging market setting, the resulting index delivers economically interpretable signals and improves the forecasting accuracy of CPI inflation and firm inflation expectations.

Beyond predictive gains, the findings highlight how news narratives transmit differently across agent types and forecasting horizons. By demonstrating that LLM-assisted macroeconomic measurement can be implemented credibly in low-resource language environments. The resulting index offers central banks and other institutions a timely and transparent tool for monitoring inflation narratives, understanding the mechanisms of expectation formation across heterogeneous agents, and improving inflation forecasting.

References

Adämmer, P., J. Prüser, and R. A. Schüssler (2025). Forecasting macroeconomic tail risk in real time: Do textual data add value? *International Journal of Forecasting 41*(1), 307–320.

Allard, M.-A., P. Teiletche, and A. Zinebi (2024). Enhancing inflation nowcasting with llm: Sentiment analysis on news.

Angelico, C., J. Marcucci, M. Miccoli, and F. Quarta (2022). Can we measure inflation expectations using twitter? *Journal of Econometrics 228*(2), 259–277.

Ayivogji, F. (2023). High-frequency inflation expectations from the big data: A natural language approach. Technical report, Mimeo.

Born, B., H. Dalal, N. Lamersdorf, and S. Steffen (2023). Monetary policy in the age of social media: A twitter-based inflation analysis. Technical Report 4567305, SSRN.

Bybee, J. L. (2023a). The ghost in the machine: Generating beliefs with large language models. *arXiv preprint arXiv:2305.02823*, 349–389.

Bybee, L. (2023b, May). Surveying generative ai's economic expectations. Papers 2305.02823, arXiv.org.

Carroll, C. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics 118*(1), 269–298.

Conneau, A., K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov (2020). Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pp. 8440–8451.

Dräger, L. and M. J. Lamla (2017, December). Imperfect Information and Consumer Inflation Expectations: Evidence from Microdata. *Oxford Bulletin of Economics and Statistics 79*(6), 933–968.

Eugster, P. and M. W. Uhl (2024, None). Forecasting inflation using sentiment. *Economics Letters 236*(C), None.

Faria e Castro, M. and F. Leibovici (2024). Artificial intelligence and inflation forecasts. *Federal Reserve Bank of St. Louis Review 106*(4), 249–262.

Feng, F., Y. Yang, D. Cer, N. Arivazhagan, and W. Wang (2022). Language-agnostic bert sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 878–891.

Gabrielyan, D., J. Masso, and L. Uusküla (2020). Mining news data for the measurement and prediction of inflation expectations. In *Theory and Applications of Time Series Analysis: Selected Contributions from ITISE 2019 6*, pp. 253–271. Springer.

Goulet Coulombe, P. (2024). The macroeconomy as a random forest. *Journal of Applied Econometrics 39*(3), 401–421.

Ha, J., M. A. Kose, and F. Ohnsorge (Eds.) (2019). *Inflation in Emerging and Developing Economies: Evolution, Drivers, and Policies*. Washington, DC: World Bank Publications.

Kwon, B., T. Park, P. Rungcharoenkitkul, and F. Smets (2025, Oct). Parsing the pulse: decomposing macroeconomic sentiment with llms. BIS Working Papers 1294, Bank for International Settlements.

Lamla, M. and S. Lein (2014). The role of media for consumers' inflation expectation formation. *Journal of Economic Behavior & Organization 106*(C), 62–77.

Larsen, V. H., L. A. Thorsrud, and J. Zhulanova (2021). News-driven inflation expectations and information rigidities. *Journal of Monetary Economics 117*, 507–520.

Ludwig, J., S. Mullainathan, and A. Rambachan (2025). Large language models: An applied econometric framework. Working Paper 33344, National Bureau of Economic Research.

Medeiros, M. C., G. F. R. Vasconcelos, Á. Veiga, and E. Zilberman (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics 39*(1), 98–119.

Pfajfar, D. and E. Santoro (2013, September). News on Inflation and the Epidemiology of Inflation Expectations. *Journal of Money, Credit and Banking 45*(6), 1045–1067.

Shapiro, A. H., M. Sudhof, and D. J. Wilson (2022). Measuring news sentiment. *Journal of econometrics 228*(2), 221–243.

Shiller, R. J. (2017, April). Narrative economics. *American Economic Review 107*(4), 967–1004.

Sriwirote, P., A. T. Rutherford, J. Thapiang, and V. Timtong (2025). Phayathaibert: Enhancing a pretrained thai language model with unassimilated loanwords. *ACM Transactions on Asian and Low-Resource Language Information Processing 24*(11), 1–17.

Sun, Y. and G. de Bondt (2025, Jun). Enhancing gdp nowcasts with chatgpt: a novel application of pmi news releases. Working Paper Series 3063, European Central Bank.

Thorsrud, L. A. (2020, April). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics 38*(2), 393–409.

Tuarob, S., P. Tatiyamaneekul, S. Pongpaichet, et al. (2025). Beyond administrative reports: A deep learning framework for classifying and monitoring crime and accidents leveraging large-scale online news. *Neural Computing and Applications 37*, 7183–7205.

Weber, M., F. D'Acunto, Y. Gorodnichenko, and O. Coibion (2022). The subjective inflation expectations of households and firms: Measurement, determinants, and implications. *Journal of Economic Perspectives 36*(3), 157–184.

Appendix

A Prompting strategy

We started developing our prompts by first establishing zero- and few-shot baselines to observe the natural classification behavior of the model before introducing any structured guidance. Misclassification patterns revealed where the model misunderstood economic signals, prompting us to clarify task definitions and highlight features that distinguish inflation-relevant content. By analyzing why the model made specific errors, we distill the correct economic reasoning and incorporate it directly into the prompt, enabling systematic refinement and more consistent classification performance.

We developed four prompt versions: a zero-shot baseline, a few-shot baseline, a detail-driven prompt, and a factor-driven prompt. In the zero-shot version, the model receives only the task instruction without any examples, allowing us to observe its unguided classification behavior. The few-shot version provides a small set of labeled examples to give the model an initial pattern for how inflation-relevant news should be classified. These two baselines serve as reference points for evaluating the benefits of more structured approaches. The detailed version—used throughout this paper—extends the label definitions by specifying which economic factors signal an effect (or no effect) on inflation expectations and includes examples and explanations based on news types the model frequently misclassifies. The factor-driven version requires the model to first identify the economic factors mentioned in the article, rank their importance, and then apply factor-specific labeling criteria to determine the final classification, while also providing a confidence score for each label to indicate the model's certainty in its assessment. Full prompt texts are provided in the subsection A.1-A.4 for reference. All prompt-based experiments use greedy decoding (temperature = 0) to ensure deterministic and reproducible model outputs.

Table 7 summarizes the performance of the four prompt specifications. Across overall classification metrics (Macro F1, Micro F1, and accuracy), the rule-based factor-driven prompt achieves the highest scores. This result is unsurprising, as the prompt was explicitly designed to incorporate the same adjudication rules used by the two expert labelers in cases of ambiguous news. By embedding these decision rules directly into the prompt structure, the model effectively replicates the human resolution strategy, which improves measured classification performance in a setting where news interpretation is inherently subjective and context-dependent.

Table 7: Performance of Prompting Techniques in Classifying Inflation-Expectation Sentiment

Prompt	Macro F1	Micro F1	Accuracy	Prec (H)	Rec (H)	F1 (H)	Prec (L)	Rec (L)	F1 (L)
Zero-Shot	0.57	0.63	0.63	0.40	0.83	0.54	0.46	0.87	0.60
Few-Shot	0.61	0.67	0.67	0.45	0.81	0.58	0.54	0.78	0.64
Detail-Driven	0.55	0.66	0.66	0.49	0.77	0.60	0.58	0.56	0.57
Factor-Driven	0.76	0.81	0.81	0.58	0.83	0.68	0.80	0.78	0.79

Notes: Performance metrics are computed using the expert-labeled validation set. Classes correspond to non-inflation-related text (O), high inflation expectations (H), low inflation expectations (L), and uncertain or ambiguous narratives (U). The class distribution is unbalanced, with non-inflation-related articles (O) representing the largest share of observations, followed by the H, L, and U categories.

We also experimented with alternative approaches aimed at addressing the subjective nature of inflation-related news. In particular, we tested ensemble prompt strategies intended to mimic multiple labelers and aggregate across prompt variations. However, ensemble prompts performed similarly to the detailed-driven and few-shot prompts, without delivering meaningful gains in F1 scores. Consequently, we refined the rule-based factor-driven prompt as a benchmark for maximizing classification metrics.

Beyond aggregate metrics, we place particular emphasis on performance within the High (H) and Low (L) sentiment classes, as these directional labels directly determine the magnitude and movement of our inflation expectations index. Within these classes, precision is critical to ensure label reliability, while recall is important for broad coverage across diverse economic topics.

When comparing subclass performance, we evaluate downstream forecasting results using both the detailed-driven prompt and the factor-driven prompt. Although both specifications generate similar improvements in out-of-sample forecasting performance, the SHAP decompositions from the detailed-driven prompt yield more coherent and economically interpretable topic contributions. By contrast, the rule-based factor-driven prompt tends to assign ambiguous news systematically toward predefined "top-tier" factors, potentially distorting the downstream topic distribution.

Given this trade-off, we ultimately adopt the detailed-driven prompt. While its classification metrics are lower than the rule-based factor-driven specification, it avoids imposing structural adjudication rules that may influence downstream topic composition and interpretability. We therefore view its performance as sufficiently strong, particularly given the inherently subjective and complex nature of news-based inflation sentiment classification. Beyond prompt design, this result highlights a broader methodological concern: rule-based adjudication—whether implemented algorithmically or by human labelers—may mechanically shape topic distributions and thereby influence downstream empirical conclusions. Careful attention to labeling rules is therefore essential when constructing text-based economic indices.

A.1 Zero-shot prompt

Q: You are a research assistant. Your task consists of two parts:

1. Determine whether each news item contains information that influences inflation expectations in Thailand.

o If no influence, label it as: ’0’

2. If the news does influence inflation expectations, classify the direction of influence as one of the following:

o ’H’ – Likely to influence inflation expectations to be higher

o ’L’ – Likely to influence inflation expectations to be lower

o ’U’ – The direction is unclear, inflation is likely to remain stable, or the news implies high uncertainty about future inflation

Label each news item with one of the following: ’H’, ’L’, ’U’, or ’0’. Answer only with the label.

This is the news: ...

A:

A.2 Few-shot prompt

Q: You are a research assistant. Your task consists of two parts:

1. Determine whether each news item contains information that influences inflation expectations in Thailand.

o If no influence, label it as: ’0’

2. If the news does influence inflation expectations, classify the direction of influence as one of the following:

o ’H’ – Likely to influence inflation expectations to be higher

o ’L’ – Likely to influence inflation expectations to be lower

o ’U’ – The direction is unclear, inflation is likely to remain stable, or the news implies high uncertainty or mixed signal about future inflation

Label each news item with one of the following: ’H’, ’L’, ’U’, or ’0’. Answer only with the label.

Examples:

1. News Example 1 Label: H
2. News Example 2 Label: L
3. News Example 3 Label: U

This is the news: ...

A:

A.3 Detail-driven prompt

You are an economic-news analyst at the Thai central bank. Your task is to classify economic news articles by whether they signal expected inflation trends in Thailand. Follow the instructions strictly. Think step-by-step, but provide only the final dictionary in the output section.

—

INSTRUCTIONS

1. RELEVANCE Label "relevance": 1 only if the article contains explicit or strong implied signals about future inflation or price expectations in Thailand. This includes:

- Forward-looking discussions of price pressures, cost changes, or inflation trends
- Anticipated effects from wages, energy, FX, supply shocks, or economic policy
- Global developments only if they are clearly linked to Thai inflation or prices

Label "relevance": 0 if:

- The article discusses only current or past prices
- It discusses trade, demand, or foreign events without clear implications for Thai inflation
- The economic effects are ambiguous or speculative

2. DIRECTION (if relevant) Choose the expected direction of Thai inflation:

- "H" = Signals higher future inflation. Includes:
  - Rising input costs, wage or labor tightness, strong demand
  - Commodity or energy price surges
  - Bullish global demand that may tighten Thai supply or raise export prices
- "L" = Signals lower future inflation. Includes:
  - Declining input costs, import prices, or weakening demand
  - Monetary tightening, policy shifts that ease pressure
  - Clearly disinflationary developments (not just volatility or easing tensions)
- "U" = Mixed, unclear, or conflicting signals

Do not label as ”L” just because:
- Gold prices fall (may reflect risk sentiment, not disinflation)
- Tariff increases are postponed (not canceled)
- Foreign inflation data is weak without a link to Thailand

3. IF UNCERTAIN, return:
{”relevance”: ”unsure”, ”direction”: ”unsure”}
—
EXAMPLES
Example 1: ... → Reason → Output: {”relevance”: 1, ”direction”: ”H”}
Example 2: ... → Reason → Output: {”relevance”: 1, ”direction”: ”H”}
Example 3: ... → Reason → Output: {”relevance”: 0}
Example 4: ... → Reason → Output: {”relevance”: 0}
—
NEWS ITEM: ...
—
FINAL OUTPUT
Provide only this dictionary (no explanation):
{”relevance”: ..., ”direction”: ...}

A.4 Factor-driven prompt

You are an economic-news analyst at the Thai central bank. Classify an economic news article into the following categories:

- Relevance (0 or 1) — whether it is related to future inflation/consumer prices in Thailand within 12 months.
- Confidence scores (0–100) for relevance and direction
- Factor — ”A1”, ”A2”, “A3”, “B1”, “B2”, “B3”, “C1”, “C2” or ”D1”
- Direction — ”H” (higher prices), ”L” (lower prices), or ”U” (uncertain/mixed)
- Evidence spans — 1–3 verbatim excerpts (≤30 words)

—
You must follow these steps:

1. Consider whether the article contains any of the following factors impacting Thai prices within 12 months. Review every detail and background, not only the main idea:

Group A

A1. Major CPI commodities: oil, fuels, LPG, electricity, public transport, staple foods (only rice, pork).

A2. Major economy conditions (ONLY USA) that could affect Thai CPI-related imports/exports, tourism, demand, or capital flows. Non-CPI financial assets (stocks, bonds, crypto, stock market index) are not included.

A3. Thai economy conditions: the article **explicitly** mentions the growth or slowdown of the Thai economy.

Group B

B1. Energy/commodity shocks (OPEC, drought, ASF, floods, supply shifts): drought/flood news affecting crops must have a clear link to 'affecting crops,' and only main crops impacting a large number of farmers (ONLY rice, rubber, cassava, sugarcane, eggs, and vegetables) are considered.

B2. Policy changes (tax, subsidies, price caps, FT, wages, interest rates)

B3. Production costs: articles mentioning rising/falling production costs (e.g., logistics, raw materials, electricity, wages) if clearly linked to inflation.

GROUP C

C1. Major economy conditions (ONLY China/EU) that could affect Thai CPI-related imports/exports, tourism, demand, or capital flows.

C2. Trade war/trade tension developments (e.g. USA-China)

Group D

D1. FX movements.

NOTE:

- Seasonal or regional-only incidents (e.g., short-lived local floods, harvest season effects, one province's market prices) are not included unless clearly connected to nationwide inflation.

- Project or intervention considered to affect inflation expectations must be definite, near-term actions. Broad, long-term projects are not included.

- Disregard Gold prices/ prices of specific stocks ('relevance' =0), unless factors A-D are mentioned

- Disregard promotion news ('relevance' = 0)

2. If the article does not contain any of the factors above, assign 'relevance' = 0. If it does, assign 'relevance' = 1 with the confidence score based on the following rubric:

Relevance Confidence

- 90–100 = Explicit match
  - o Direct mention of CPI-relevant factors with clear present/future effect.
- 70–89 = Clear, one inference step
  - o Strong CPI factor movement but need one step of inference.
- 50–69 = Weak, multi-step inference
  - o Indirect reasoning needed
- <50 = Speculative/weak
  - o Very indirect or long-term effect (>12 months) only.

3. If the article contains more than 1 factor: Identify the group of each factor. The directional label will be determined by the factors in the higher-ranked group (Group A > Group B > Group C > Group D). Factors in lower groups are ignored.

4. Record the final selected factor in the output field 'Factor'.

5. Label the direction based only on the excerpts explaining the final factor chosen in Step 4. Ignore all other content in the article. Assign the direction with a confidence score according to the following rubric:

Rubric — Direction (H / L / U)

- H (Higher prices) = Clear upward pressure (oil (OPEC, US or traders)↑, food ↑, energy ↑, demand ↑, wages ↑, US/China/EU recovery).
- L (Lower prices) = Clear downward pressure (oil (OPEC, US or traders)↓, energy ↓, food ↓, demand ↓, subsidy/tax cut reducing prices, US/China/EU slowdown).
- U (Uncertain/mixed) =
  - o FX movement without clear CPI impact.
  - o Oil prices not driven by OPEC, US or traders.
  - o Trade war/geopolitics with uncertain effects on Thai inflation.
  - o Government actions that only stabilize prices (stabilize prices, fix prices) → not H. Mostly label as U.
  - o Policy changes, or interest rate hike/drop that report only policy details with no clear link to economy or prices

Trade war

- If mentioning clear impact: H for explicit deals easing boosting global demand or Thai exports; L for explicit tariffs hurting them.
- U for talks only, uncertainty, or mentioning no clear impact.

NOTE:

- Rule – Conflicting factors: If multiple factors from the same tier (e.g. A2, A3) imply different directions, assign U.

Direction Confidence

- 90–100 = Explicit price direction
  - o Example: "Diesel will rise next month."
- 70–89 = Clear implied direction
  - o Factor movement strongly implies direction.
- 50–69 = Mixed/uncertain
  - o Competing signals or multi-step reasoning.
- <50 = No reliable match
  - o Speculative.

NEWS: ...

—

OUTPUT JSON ONLY:

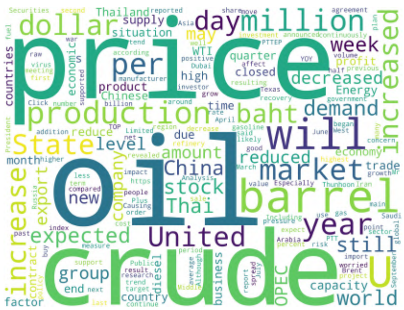
{ "relevance": 0 or 1,
"confidence relevance": 0-100,
"factor": "A1" | "A2" | "A3" | "B1" | "B2" | "B3" | "C1" | "C2" | "D1" | null,
"direction": "H" | "L" | "U" | null,
"confidence direction": 0-100 | null,
"evidence spans": ["...", "..."]
}

Rules for nulls:

- If "relevance"=0, then "direction": null and "confidence direction": null.

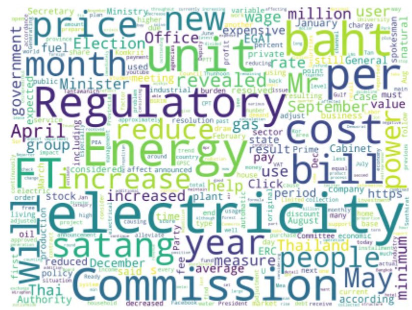
B Word clouds by topics

Figure B.1: Word clouds of news topics: energy and utilities

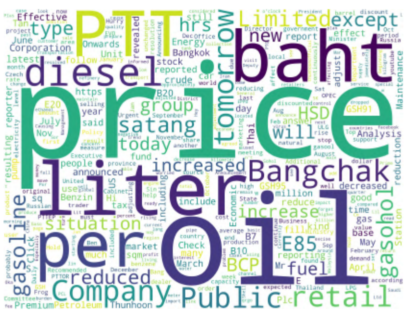


(a) Crude oil

(b) Energy prices



(c) Electricity prices



(d) Retail oil prices

(e) Biofuel

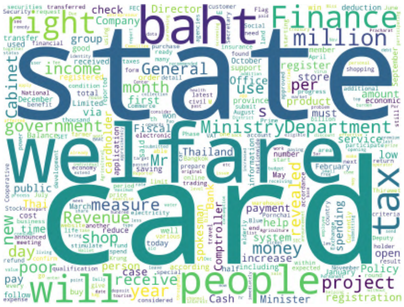
Figure B.2: Word clouds of news topics: food and agriculture

(a) Rubber prices

(b) Food prices

(c) Corporate and agribusiness

Figure B.3: Word clouds of news topics: policy and public services



(a) Thai fiscal policy

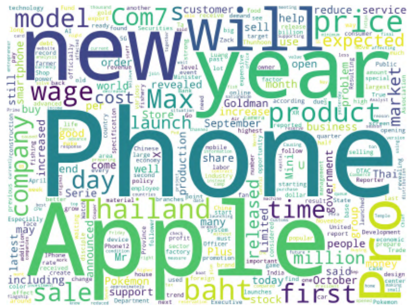
(b) Thai public policy

(c) Public transportation

(d) Cabinet meeting

Figure B.4: Word clouds of news topics: business, industry and consumer

(a) Business news



(b) Market trends

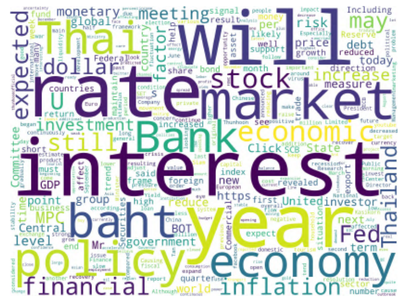
(c) Thai SMEs

(d) Thai e-commerce

Figure B.5: Word clouds of news topics: macro, markets, and FX

(a) Debt and bank loans

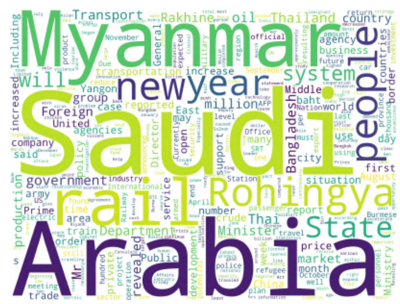
(b) Global economy



(c) Thai market/ interest rate

(d) Commodity prices

(e) Economic news/ crisis



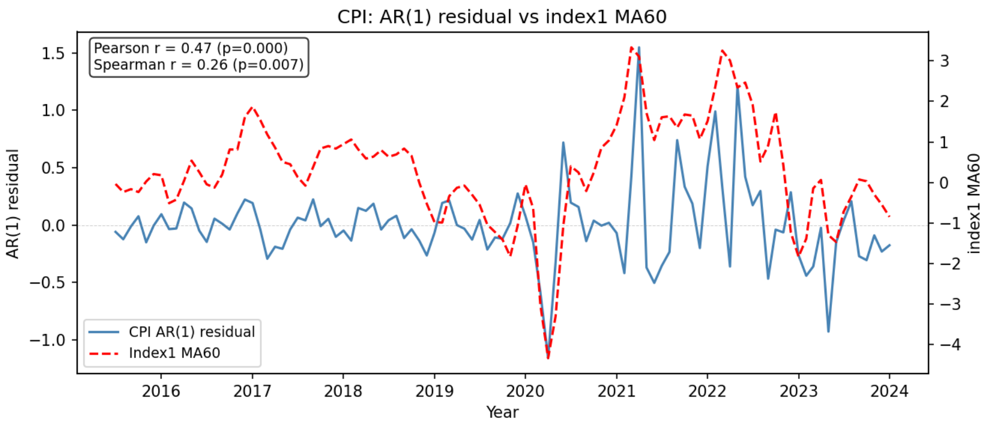
(f) Geopolitics in Asia

(g) U.S. monetary policy

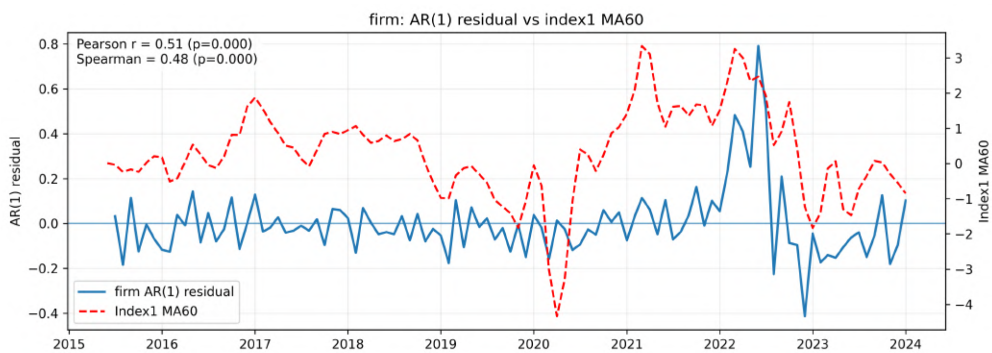
C News Index Robustness and Explainability

C.1 AR(1) innovations and news index

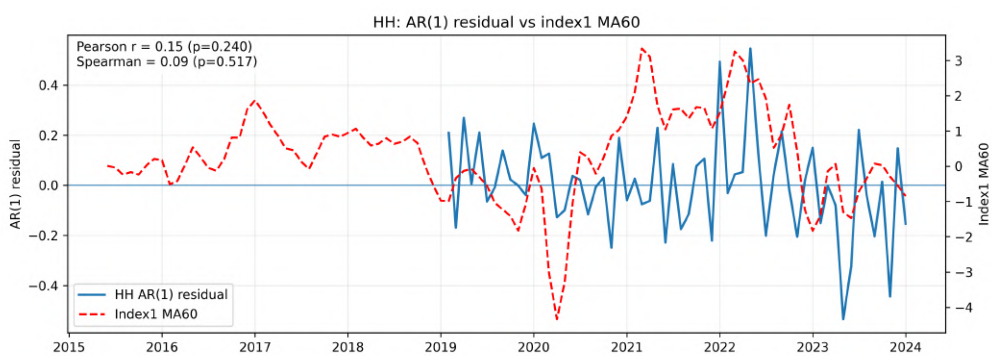
Figure C.1: AR(1) innovations and news index



(a) News index vs CPI inflation error terms



(b) News index vs firm expectation error terms



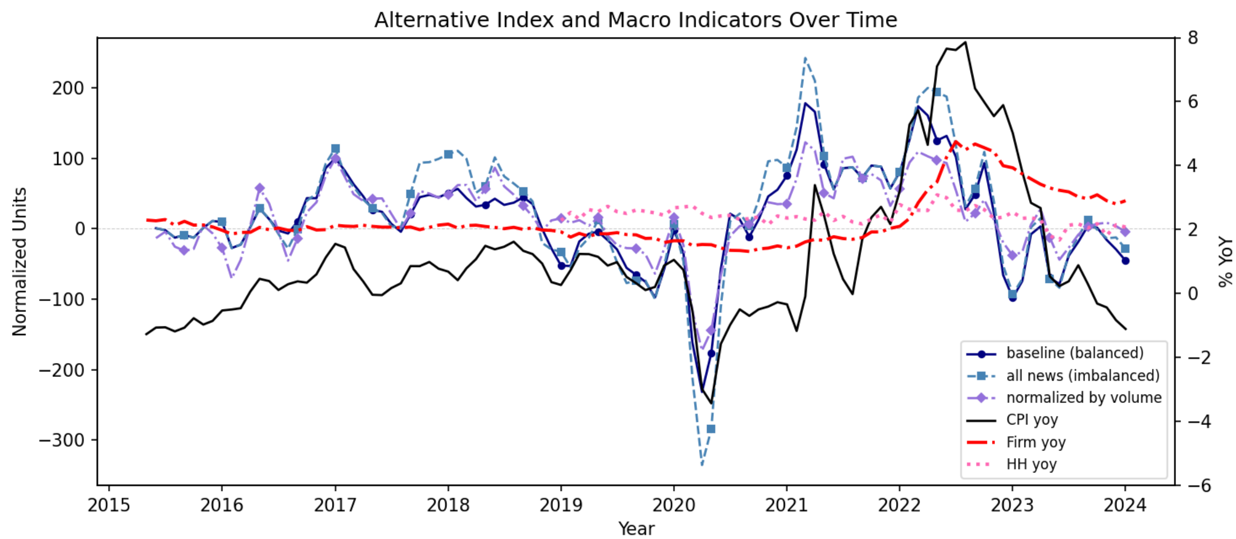
(c) News index vs household expectation error terms

Note: For each benchmark series, we estimate an AR(1) process, $y_t = \rho y_{t-1} + \epsilon_t$, and extract innovations ε_t. The figure plots the resulting AR(1) residuals together with the news index at time t.

C.2 Index robustness

We construct several alternative indexes to assess the robustness of the variable. Figure C.2 compares three specifications: (i) the baseline index, based on a balanced panel of news from three agencies; (ii) an index using all news collected from six agencies, which is imbalanced; and (iii) an index from the same six agencies, adjusted by normalizing for daily news volume. Across these alternatives, we find the indexes to be both qualitatively and quantitatively similar. Their overall trends align, with only minor differences in the magnitude of changes.

Figure C.2: News Index at monthly frequency compared to CPI inflation and inflation expectations



For comparability, both the 3-agency and 6-agency news indexes are scaled to match the MA60 value of the baseline balanced panel in January 2017 (set to 100). The index normalized by daily news volume is scaled to 100 using its own January 2017 level as the baseline. CPI, firm, and household inflation expectations are shown as year-on-year percentage changes on the right axis.

C.3 Explainability

Out-of sample prediction results using features from all 23 topics (sub-indices by subtopics with lag up to two months) are reported in Table C.1 to C.3.

Next, we analyze the SHAP values of selected machine-learning models to examine the importance of each sub-index and its contribution to forecasting performance across horizons. For the nowcasting exercise, AR(1) models with sub-index features and the Lasso model perform best, while Random Forest models tend to outperform other specifications

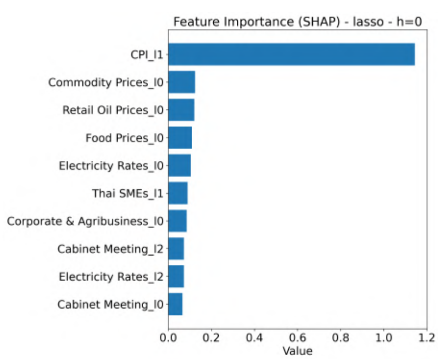
Table C.1: Out-of-sample forecasting exercise for CPI inflation using topic-level features: relative RMSE of competing models (reported as quarterly averages)

		Linear ML			Tree ML	
Horizon	**ARX**	**Lasso**	**Ridge**	**Elastic Net**	**Random Forest**	**XGBoost**
$h = 0$	1.4393	**1.1036**	1.3766	1.2062	1.5449	1.3724
Q1	1.3149***	**0.8057*** **	1.0160	1.0971	0.9367	1.0393
Q2	1.0606*	0.7569**	0.8012**	1.0323	0.7222***	**0.7049** **
Q3	1.1130	0.8040*	0.8158**	0.9826	**0.5807** **	0.6993**
Q4	1.1903	0.9091*	0.8121**	0.9449	**0.6834** *	0.8906*

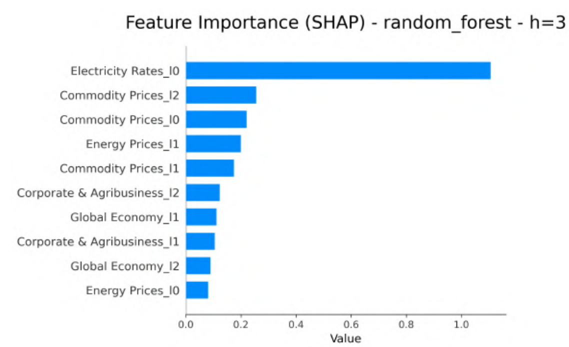
Note: This table reports the relative RMSE, defined as the ratio of each model's RMSE to that of the AR(1) benchmark. Numbers in boldface indicate the model with the lowest relative RMSE for each forecast horizon. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, based on the Diebold-Mariano test of equal predictive accuracy with Harvey-Leybourne-Newbold (HLN) small-sample correction.

in the multi-horizon forecasting tasks. Accordingly, for interpretability, we report SHAP values from the Lasso model for nowcasting, and SHAP values from the Random Forest model for all forecasting horizons ($h = 3, 6, 9, 12$ months ahead) in Figures C.3- C.5.

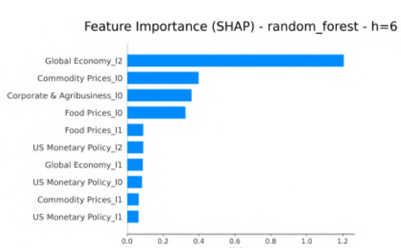
Figure C.3: SHAP values from for CPI inflation forecasting using sub-indexes



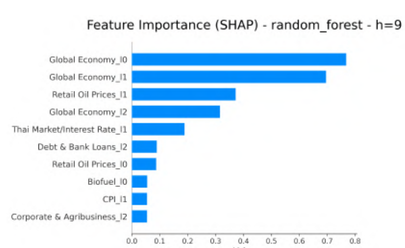
(a) lasso, $h = 0$



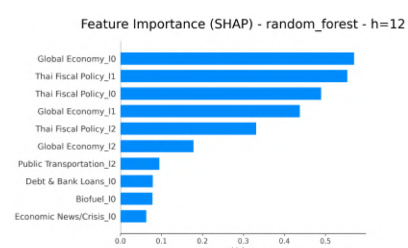
(b) random forest, $h = 3$



(c) random forest, $h = 6$



(d) random forest, $h = 9$



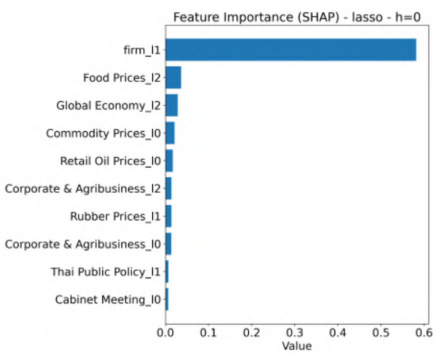
(e) random forest, $h = 12$

Table C.2: Out-of-sample forecasting exercise for firm inflation expectations using topic-level features: relative RMSE of competing models (reported as quarterly averages)

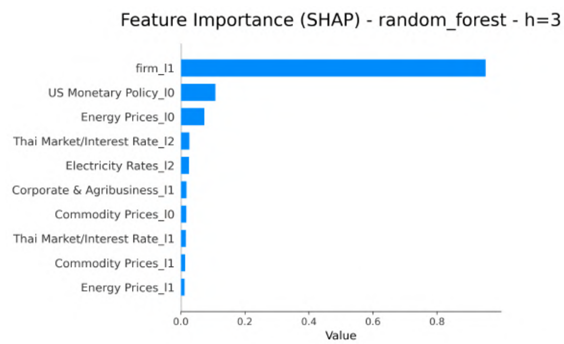
		Linear ML			Tree ML	
Horizon	ARX	Lasso	Ridge	Elastic Net	Random Forest	XGBoost
$h = 0$	1.0593	**0.8589**	2.6239	4.2339	2.6395	1.3452
Q1	0.6091***	**0.5420*** **	0.9748	1.9137	0.9714	0.9857
Q2	0.6916*	0.6358**	**0.6176** **	1.2772	0.6483***	0.7698**
Q3	1.0318	0.7488*	0.6957**	1.1160	**0.6380** **	0.6761**
Q4	1.2638	0.7862*	0.7817**	1.0429	**0.6967** *	0.7674*

Note: This table reports the relative RMSE, defined as the ratio of each model's RMSE to that of the AR(1) benchmark. Numbers in boldface indicate the model with the lowest relative RMSE for each forecast horizon. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, based on the Diebold-Mariano test of equal predictive accuracy with Harvey-Leybourne-Newbold (HLN) small-sample correction.

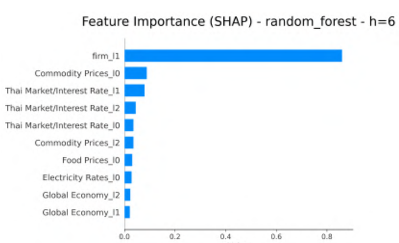
Figure C.4: SHAP values for firm inflation expectation forecasting using sub-indexes



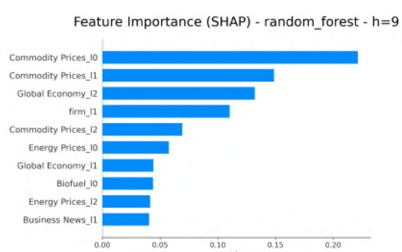
(a) lasso, $h = 0$



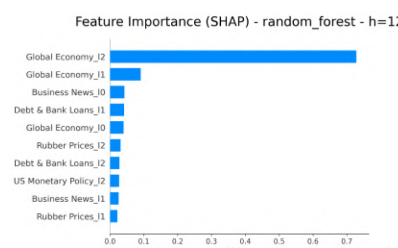
(b) random forest, $h = 3$



(c) random forest, $h = 6$



(d) random forest, $h = 9$



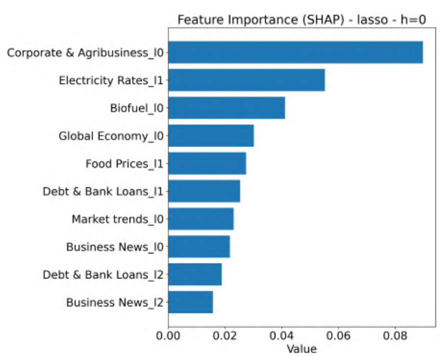
(e) random forest, $h = 12$

Table C.3: Out-of-sample forecasting exercise for household inflation expectations using topic-level features: relative RMSE of competing models (reported as quarterly averages)

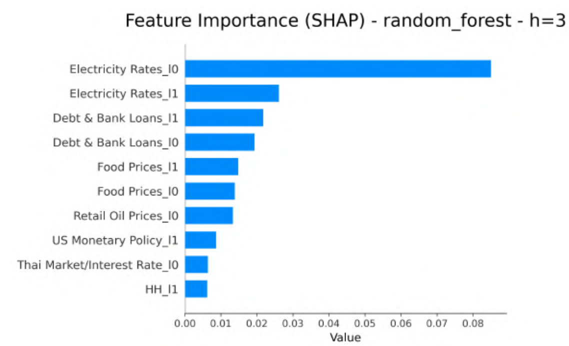
		Linear ML			Tree ML	
Horizon	**ARX**	**Lasso**	**Ridge**	**Elastic Net**	**Random Forest**	**XGBoost**
$h = 0$	6.9184	1.0079	**0.9892**	1.3561	1.1058	1.0485
Q1	17.4974***	1.1246***	0.9671	1.1662	0.9647	**0.9253**
Q2	15.4094*	0.9849**	0.9201**	1.0182	**0.8931*****	0.9364**
Q3	13.5140	1.0067*	**0.9506****	0.9986	0.9565**	0.9810**
Q4	11.7840	1.0915*	1.0494**	1.1095	**0.9942***	1.1517*

Note: This table reports the relative RMSE, defined as the ratio of each model's RMSE to that of the AR(1) benchmark. Numbers in boldface indicate the model with the lowest relative RMSE for each forecast horizon. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, based on the Diebold-Mariano test of equal predictive accuracy with Harvey-Leybourne-Newbold (HLN) small-sample correction.

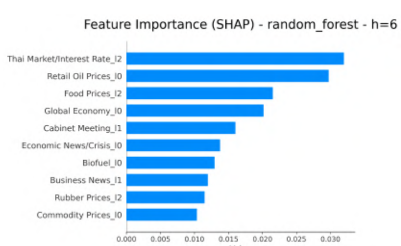
Figure C.5: SHAP values for household inflation expectation forecasting using sub-indexes



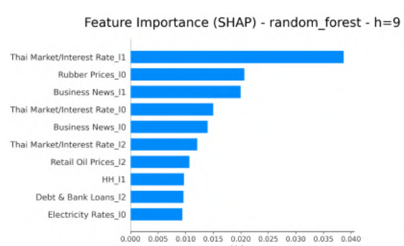
(a) lasso, $h = 0$



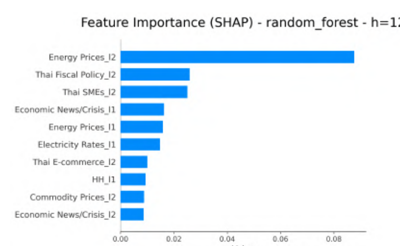
(b) random forest, $h = 3$



(c) random forest, $h = 6$



(d) random forest, $h = 9$



(e) random forest, $h = 12$