Bloomberg Economics Fedspeak: Towards a High Frequency Fed Sentiment Index

Anna Wong, Nick Hallmark, and Ana Galvao

Abstract

When the Federal Reserve speaks, the market moves. For market participants seeking trading opportunities, the challenge is to obtain both a timely and comprehensive read of Fed communications. We fill that gap by constructing a daily, natural language processing-based index – the Bloomberg Economics Fedspeak Index – scored on 60,000 Bloomberg terminal headlines generated from the universe of Fed communications. We score those headlines using a RoBERTa machine learning algorithm, which is trained to decipher five level of FOMC sentiment, ranging from most dovish to most hawkish. Our index improves on the forecasting performance of 3-month, 6-month, and 12-month ahead forecast of fed funds rates and two-year yields.

Keywords

Fedspeak, Sentiment Analysis, Natural Language Processing, Monetary Policy, Probit Model

Introduction

Few spoken words can move markets more than the Federal Reserve's. Financial markets often react instantaneously to any hints of Fed's future moves – forward guidance – well before actual changes in interest rates. The impact on the market, in turn, is a key transmission mechanism of monetary policy. Many Fed watchers therefore scrutinize every word spoken by the FOMC.

A burgeoning academic literature showed that a natural language processing (NLP) and machine learning approach to Fed watching can be a promising supplement to the human read. But there are two key limitation to these studies – the gaps in the coverage of Fed communications, and the timeliness for which the Fed communications could be evaluated. The Fed's public engagement takes many forms. Aside from the post-FOMC presser by the Fed Chairman, policy statements, and FOMC minutes, FOMC officials also give speeches between meetings, conduct video interviews, engage with scrums with the media, tweet, or in the case of regional Fed presidents, post essays on the regional Fed's website.

The challenge for any textual analysis of Fed communications is that they require pre-processing of the communications. Interviews or press conference needs to be transcribed into text first and cleaned; the NLP algorithm needs to parse tweets and essays posted on various websites. Due to the time-lag, and in many cases, lack of availability of transcription of the myriad form of Fed communications, the existing NLP Fed products appear to all be limited by the scope as well as the timeliness of their coverage.

Bloomberg Economics' Fedspeak Index aim to overcome these limitations. We leverage the timely and comprehensive reporting of Bloomberg Fed reporters, who cover the universe of Fed communications, and apply and train an advanced NLP algorithm to score the hawkishness and dovishness of Bloomberg terminal Fed-related headlines.

Out database covers over 60,000 headlines tagged as "Fedspeak" by Bloomberg editors or by the Bloomberg terminal topic assignment algorithm. We then screened the headlines for relevancy, and trained a transformer-based deep learning model (RoBERTa) – using thousands of human-labeled Fed headlines annotations – to classify individual headline as one of the following five scores: highly dovish, dovish, neutral, hawkish, and highly hawkish. The result is a daily index that serves as the basis for a probit model predicting rate changes within various economic contexts. Our index also adds signal and improve the forecasting of two-year Treasury yields. Through this white paper, we present our methodology, findings, and the practical applications of our model.

Literature Review

Here we review recent economic literature on the application of NLP techniques on Fed communications.

The early academic papers that employ NLP techniques on Fed communications typically shared one objective: to identify the transmission of monetary policy surprises to the economy. Analysts apply NLP techniques on various official FOMC documents, many of whom relied on FOMC policy statements (Hansen and McMahon, 2016; Gorodnichenko, Pham and Talavera, 2023; Doh, Song and Yang, 2023), Fed's staff forecasts (Aruoba and Drechsel, 2023), and Fed chairman's opening statement and press conference (De Pooter, 2021).

A second group of studies apply NLP techniques to measure the transparency of Fed official communication

Corresponding author:

Anna Wong, Chief US Economist (awong920@bloomberg.net); Nick Hallmark, Global Economist (nhallmark3@bloomberg.net); Ana Galvao, Senior Economist in Global Modelling (agalvao9@bloomberg.net)

relative to meeting deliberations. Acosta (2023) and Meade and Acosta (2015) examined the semantic persistence in FOMC statements from one meeting to another by calculating the "cosine similarity" of the text, and demonstrated that one can use computation linguistic techniques to uncover whether edits to the post FOMC statement provide a signal or not. Fischer, McCaughrin, Prazad, and Vandergon (2023) found that the sentiment of FOMC transcripts and Tealbooks could have improved on market-implied Fed funds rates forecasts with a six-month forecasting horizon.

Because the primary goal of these studies is to identify monetary policy surprises or measure Fed transparency, they are not concerned with the timeliness or frequency of the Fed communications. There are only eight FOMC policy statements, minutes, and press conference in an entire year, even though Fed officials speak all year long except for the two-week during the blackout window before each FOMC meeting. The FOMC minutes are also released with a three weeks lag from the meeting, and FOMC transcripts are publicly available with an even longer lag of five years.

The aim of our Fedspeak index is to provide trading signals, which requires a more timely source for the text. Swanson and Jayawickrema (2023) show that market reactions to Fed speeches outside official monetary policy announcement days are vital to providing news about monetary policy to financial markets. Addressing some of that shortcoming, another class of studies apply NLP methods exclusively on Fed speeches (Bertsch, Hull, Lumsdaine, Zhang, 2022) and Pfeifer and Marohl (2023).

Still, even these studies don't cover the entire universes of Fedspeak. Fed communications take a myriad of forms outside of official policy statements and speeches: essays, video interviews, scrums with reporters, tweets, Q&A after speeches. Thus, a comprehensive read of Fed communications require inclusion of Fedspeak occurring in all these venues.

Our Fedspeak index contributes to this literature in two key aspects: First, our index covers a wider scope of Fed communications than that in the existing literature, at least to our knowledge; second, the source of our analyzed text is more timely than that in the literature. Additionally, we rely on a deep-learning-based NLP algorithm, which belongs to the relatively more advanced class of models that have been used in the literature. Only a handful of studies, mostly in recent years, have used a RoBERTa model fine-tuned to central bank communications (Bertsch, Hull, Lumsdaine, and Zhang, 2022; Pfeifer and Marohl, 2023), compared to earlier studies that used bag-of-words or dictionary approaches. We show that the machine learning approaches outperforms dictionary approaches in their accuracy.

Data

The key innovation of our approach is our text source: Bloomberg Fed news headlines. These headlines were written by either Bloomberg reporters or editors, and chosen for their newsworthiness and market relevance.

Bloomberg Fed reporters are attuned to the universe of Fed public engagements, some of which are not public. When a Fed official speaks, Bloomberg reporters annotate key parts of the remarks and turn them into headlines that begin with an asterisk, and make them immediately available on the Bloomberg terminal. Bloomberg journalists as well as an automated NLP-powered algorithm within the Bloomberg terminal then tag these headlines as "Fedspeak," along with other labels, including the speaker name and the relevant topics.

Our raw database covers more than 6,200 unique speaking engagements and statements made by FOMC members since 2009, totaling more than 60,000 Bloomberg News headlines. We then designed an algorithm that screened the Fedspeak headlines to remove comments not made by active Fed members and those that reflect procedural information. The screening process also drops headlines that reference Federal Reserve individuals in capacities or events unrelated to their official statements or policy positions. That reduced our sample to about 47,000 headlines through January 2024. Figure 1 shows the distribution of headlines across years in the data set.

Ultimately, the pre-processing stage seeks to refine the data set to an optimally relevant subset of Bloomberg headlines, where each entry is expected to inform the sentiment model about the Federal Reserve's stance on monetary policy. This refinement is crucial to the training process that follows, wherein the cleaned headlines are used to fine-tune the model for ordinal sentiment classification.

Creating the Training Set

Any sentiment analysis model is only as good as how it is trained. Our Fedspeak index is trained on about 6000 manually labeled headlines by either Bloomberg Economics' chief US economist Anna Wong or, under her supervision by global economist Nick Hallmark. The headlines in the training set are randomly sampled from the database of screened headlines.

This process was then divided into two principal phases: assessing the relevance of headlines and assigning sentiment scores to those deemed relevant.

Headline Relevance

The initial phase is to review each headline for relevancy: Is the headline directly relevant to deciphering the Fed's views on economic performance or monetary policy stance? If yes, the headline was marked essentially informative concerning Fedspeak; if not, it would be excluded from the subsequent scoring phase. This process weeded out headlines that, while mentioning the Federal Reserve or its members, did not shed light on policy outlook or economic analysis—such as broad regulatory issues or non-policy-related activities of Fed members.

Sentiment Scoring

Following the relevance filtering, we assigned the headline one of five scores, ranging from -2 (most dovish) to +2 (most hawkish), according to the following criteria:

- **Score -2 (Highly Dovish):** Either explicit recommendation or a statement that clearly implies an intention of conducting accommodative monetary policy

  - *FED'S FISHER SAYS CONSUMPTION IS 'CASCADING DOWNHILL' (February 10, 2009)
  - *POWELL: WILL CONTINUE TO PROVIDE ECONOMY WITH SUPPORT IT NEEDS (March 19, 2021)
- **Score -1 (Moderately Dovish):** Subtler indicators favoring a more supportive economic stance, usually without explicit mention of policy rate changes.
  - *FED'S WILLIAMS: SEEING SOME SLOWING IN LABOR DEMAND (April 11, 2023)
  - *POWELL: ACTIVITY IN HOUSING SECTOR REMAINS WEAK (March 22, 2023)
- **Score 0 (Neutral):** Balanced communications devoid of clear intent towards easing or tightening monetary policy.
  - *FED'S BOSTIC SAYS JOBS DATA DOES NOT CHANGE HIS OUTLOOK AT ALL (January 6, 2023)
  - *POWELL: NEED TO SEE HOW DATA EVOLVE IN COMING MONTHS (December 15, 2021)
- **Score +1 (Moderately Hawkish):** Indirect references to economic conditions typically associated with an inclination toward policy tightening.
  - *WALLER: LABOR MARKET 'VERY ROBUST' DESPITE SIGNS OF COOLING (July 13, 2023)
  - *POWELL: PARTICIPANTS SEE INFLATION RISKS TO THE UPSIDE (December 14, 2022)
- **Score +2 (Highly Hawkish):** Clear indications of restrictive monetary policy actions, usually with explicit references to rate hikes.
  - *POWELL: WOULDN'T HESITATE TO DO LARGER MOVE IF NEEDED (July 27, 2022)
  - *DALY: I HAVE AN OPEN MIND ON WHETHER 75 BPS FOR SEPT NEEDED (August 11, 2022)

Iterative Annotation with BADGE Algorithm

To optimize the annotation process and enhance the model training, the BADGE algorithm (Ash, Zhang, Krishnamurthy, Langford and Agarwal, 2019) was used iteratively during the labeling process. This machine learning technique selects the most valuable headlines for annotation based on model uncertainty, thus extracting maximum informational gain from each labeling cycle. The algorithm intelligently identifies headlines that, once labeled, could most improve the model's predictive accuracy. BADGE algorithm's utilization continued until the growth of model performance began to plateau, which occurred approximately at the 5,000-label threshold.

Model Overview and Fine-Tuning

The backbone of the NLP algorithm for processing the Fedspeak headine is the well-known RoBERTa (Liu, Ott, Goyal, Du, Joshi, Chen, Levy, Lewis, Zettlemoyer and Stoyanov, 2019), short for "Robustly Optimized BERT," originally developed by Meta. These transformer models are known for their ability to capture the complexities of language through deep contextualized representations.

The final prediction incorporates two separately trained models, the first to predict the relevancy of the headline and the second to output a sentiment score for each relevant headline. Both models use the same base pre-trained Roberta model as the backbone for the prediction mechanism. The first model is a typical classification model trained to predict the relevance class using a cross-entropy loss function. The second model, the sentiment model, was trained using a more complex loss procedure described below.

The fine-tuning process of the Roberta model was two-fold: Initially, the model was trained on a broader set of Bloomberg headlines to acclimate it to the financial news domain's language and idiom. Subsequently, it was further refined by Bloomberg Economics on a FEDSPEAK-specific corpus of headlines. This second stage of fine-tuning employed masked training — a technique in which random words in the text are masked, and the model learns to predict them. This effectively enables the model to focus on the particular syntax, diction, and linguistic patterns prevalent in Federal Reserve communications.

Ordinal Loss Classification

To ensure the sentiment model could accurately reflect the ordinal nature of Fedspeak – the inherent ranking in dovishness and hawkishness of the statements – a loss mechanism was implemented, addressing the limitations found in conventional classification loss, such as cross-entropy.

Conventional classification loss functions treat each class as entirely independent of the others, leading to inefficiency in cases where the classes have an intrinsic order. This is clearly suboptimal for analyzing Fedspeak, where the sentiment towards monetary policy is naturally ordinal — a moderately hawkish statement (score 1) is closer in sentiment to a highly hawkish one (score 2) than it is to a highly dovish statement (score -2).

Our loss function is inspired by the CORN (Conditional Ordinal Regression for Neural networks) framework as detailed in Shi, Cao and Raschka (2023). This methodology splits the single class prediction task of K classes into $K-1$ binary model tasks which correspond to the ordinal ranks r. By using conditioned training subsets, the CORN method outputs a series of conditional probabilities the i-th example is greater than rank r_j . CORN offers a significant advantage for our purpose as it not only provides rank consistency among classifications — ensuring the model adheres to the natural order of sentiment scores — but does so without significantly increasing training complexity or limiting neural network expressiveness.

Additionally we weight the sub-elements of the loss function corresponding to each class by the inverse proportion of the class within the training set. This helps to mitigate issues related to the imbalanced data set and ensure the model learns to predict all classes adequately. Fed communications tend to be primarily neutral or subdued and as such the model tends toward predicting neutral statements. This bias is mitigated through the use of the weighted

loss function, aiming to improve model predictions for less prevalent classes, which ultimately hold more meaning for the analysis of Fedspeak.

The loss function for each class subset from Shi et al. (2023) is defined as follows. Let $Z = \{z^{(i)}\}_{i=1}^{N}$ be the logits from the last layer of the model for N examples and $|S_j|$ be the size of the j-th subset of the K classes. These subsets are defined such that $N = |S_1| >= |S_2| >= |S_3| >= ... >= |S_{k-1}|$. For the i-th training example, the total loss for the model is the of losses associated with each conditional training subset S.

$$L(S,y)_j = \Big(\log(\sigma(z^{(i)})) \cdot \mathbb{1}\{y^{(i)} > r_j\}\Big) + \Big((\log(\sigma(z^{(i)})) - z^{(i)}) \cdot \mathbb{1}\{y^{(i)} \leq r_j\}\Big) \tag{1}$$

We add the weighting scheme which captures the relative size of each training subset with respect to the total training set. This weights the loss contribution of each subset $L(S,y)$ by the inverse of its proprotion of the full training set.

$$W_j = \frac{\sum_{j=1}^{K-1}|S_j|}{|S_j|} \tag{2}$$

The complete loss functions is then defined as the weighted sum of of losses for each conditional training subset S then scaled by the total number of training examples.

$$L(Z,y) = -\frac{1}{\sum_{j=1}^{K-1}|S_j|}\sum_{j=1}^{K-1} W_j \cdot L(S,y)_j \tag{3}$$

This conditionally trained set-up ensures that the loss function promotes rank consistency. For instance, the probability estimate for severely dovish (score -2) will always be greater than for merely dovish (score -1), thereby encoding the ordinal relationship directly into the model's training objectives.

To obtain classification predictions from the model, for the i-th example we calculate the cumulative sum of the conditional rank probabilities and classify the example based on a threshold of 0.5. The equation from Shi et al. (2023) is reproduced below.

$$class_i = 1 + \sum_{j=1}^{K-1} \mathbb{1}\left(\hat{P}\left(y_i > r_j\right) > 0.5\right) \tag{4}$$

By implementing this technique, the resulting sentiment model achieves a level of sophistication necessary for accurately interpreting Fedspeak, reflecting not just the direction but the intensity of monetary policy sentiment. The use of the CORN-based loss function is instrumental in preserving the ordinality of Fedspeak sentiment, enabling the model to provide more intuitive and interpretable predictions, crucial for stakeholders reliant on the nuances of central bank communications.

Model Accuracy And Evaluation

Our model outperforms across a variety of accuracy metrics compared to other common NLP models used in the financial sentiment analysis literature.

Table 1 presents a comparative summary of the evaluated models, which includes the Loughran McDonald (LM) Dictionary baseline, enhanced LM model utilizing dependency parsing, RoBERTa Large models without and with Fedspeak pretraining, and an advanced RoBERTa model employing an ordinal loss function – the model underlying the Bloomberg Economics Fedspeak Index.

The accuracy rate for correctly identifying three ordinal classes – dovish, neutral, hawkish – is 72%, much higher than the sub-50% performance by a basic RoBERTa model or the dictionary approaches. The discrepancy between the F1 scores is similarly higher, and the mean absolute error is almost half as small as those models.

One of our benchmark models relies on a dictionary approach to construct the Fedspeak sentiment index. An often-used dictionary model in the financial sentiment literature is the Loughran McDonald (LM) dictionary, which assigns scores to words as positive, negative, or uncertain. The dictionary is a collection of words curated from firms' SEC 10-X filings, and was originally designed to understand financial texts. However, in the context of parsing Fedspeak, our model comparisons indicate that the dictionary approach has low accuracy rate (47%) and F1 score (0.40).

Augmenting the LM model with dependency parsing yielded only slightly more favorable metrics, but ultimately falls short capturing the nuances of Fedspeak sentiment. The domain-general RoBERTa Large model, trained on the common English language, had an even worse performance than the financial text dictionary models.

Our takeaway is that the financial text dictionary approach or LLM models trained on the broad English language proves inadequate when applied to Fedspeak. Complex contexts such as negative speech around high inflation can mislead the model into associating it with negative word connotations with dovish policy predictions.

The model improved significantly once we fine-tuned and trained the RoBERTa model with our human-labeled Fedspeak headlines. Compared to the dependency parsing dictionary approach, the accuracy rate increased by 40.8%, the F1 score improved 56.8%, and the MAE is reduced by 38.6%.

Our final model further augments from the Fedspeak pretrained RoBERTa model with an ordinal loss function. This addition improves model performance 4.3% in accuracy and F1 and 11.4% in MAE over the model trained with cross-entropy loss. This precision in accurately reflecting the sentiment of Fedspeak signifies the practical advantages of modeling ordinal relationships within policy communication sentiment analysis.

The leap in model performance from dictionary models and the basic RoBERTa model to a Fedspeak pre-trained RoBERTa model underscores the uniqueness and complexity of interpreting Fed communications – it cannot be understood simply through the lenses of positive or negative financial sentiment, nor through models simply trained on the English language. To capture the nuances of monetary policy discussions, fine-tuning standard large language model with Fedspeak-specific texts is a must.

The Fedspeak Indexes

Having established a mechanism to classify and score the sentiment of each Fedspeak headline accurately, the next step is to construct a daily index that captures the evolution of the sentiment in Fed communications over time.

Aggregating Scores

To aggregate individual headline scores that the NLP model produced, we take the sum of all scores attached to headlines labeled with the NI code "FEDSPEAK." This results in a raw daily sentiment score that represents the combined sentiment of all Fedspeak communicated on that day. Then, to construct a balanced index and ease interpretation, we scaled the raw sentiment score for each day by the sum of the absolute values of all scores for that day.

In equation form, the index is calculated as follows, where t indicates the day and h is a single scored headline on each day:

$$FS_t = \frac{\sum_{i=1}^{h} Score_h}{\sum_{i=1}^{h} |Score_h|}$$

This scaling serves two key purposes:

1. It ensures that the index value is not biased by the frequency of communication. It also allows for comparability across time, as the measure is by design invariant to Bloomberg reporting coverage of Fed communications.
2. It bounds the daily score between -1 and 1, allowing the score to be interpreted as the sentiment intensity relative to the total potential sentiment expressed each day.

We extend the calculation of the sentiment index to individual FOMC members as well.

Smoothing The Raw Series

The resulting raw daily index is noisy. As our ultimate goal is to predict bonds yields and future fed funds rates changes, we need to smooth the raw series to optimize its usefulness.

Our smoother assumes that the sentiment from the Fed communication lasts about two months, in line with the usual interval between FOMC meetings. The smoother is applied as:

$$FSI_t = 0.975 \cdot FSI_{t-1} + FS_t, t = 2, ..., T, \qquad (5)$$

where the coefficient 0.975 is such that the information in the score FS_t has a half-life of about 1 month.

Forecasting Bond Yields With Fedspeak Index

We now show that the Fedspeak Index, FSI_t, can help forecast bond yields and short-term fed funds rates.

First, we show that including the Fedspeak Index in a yield-curve factor-based model can improve the accuracy of forecasts for the two-year, five-year and ten-year yields compared to the benchmark yield curve model.

We first check whether model-based forecasts can improve on the consensus forecast performance. Table **??** shows the accuracy of the daily consensus (BYFC) forecasts for end-of-quarter values from 3Q21 and 4Q23. The accuracy is measured using the average of the root mean squared forecast error.

The Yield Model computes forecasts only based on yield factors, and it is a version of the Dynamic Nelson-Siegel model. The second model includes Bloomberg Economics daily surprise indexes (growth and inflation), and it is called the Macro-Yield. Daily yield forecasts for both these models are available as tickers for the next two years (and also in BECO MODELS $< GO >$).

The predictive content of the FSI_t is evaluated by adding the index for both of these models, leading to Yield + Fedspeak and Macro-Yield + Fedspeak Models. Table 2 evaluates the performance of daily out-of-sample forecasts. The values are the gains to consensus regarding root mean squared error. We find that the FSI_t improves the performance of both types of models. The best performance is for the two-year yields, as bonds of short maturities are more affected by the information from the Fedspeak Index.

Forecasting Fed Funds Rates with Fedspeak Index

Regression analysis suggested that the scores FS_t are predicted by the Bloomberg Economics growth and inflation surprise indexes.* The growth and inflation surprise indexes measure the information of data release events that were not anticipated by consensus forecasts. As this new information may also contribute to Fed officials' views about the economy, it is natural that hawkish/dovish sentiment captured in the scores FS_t reflects the communication of new economic data.

We use a regression of the daily FS_t on lagged daily growth gs_{t-1} and inflation πs_{t-1} surprises to compute two orthogonal components: MPS_t and IS_t. IS_t is the part of the sentiment score that is predicted by the economic data surprises and MPS_t is the part that is not predicted by surprises. Then, we apply the same smoothing procedure described earlier to obtain FSI_t^{MP} and FSI_t^{I}.

The advantage of this decomposition is that it improves the predictive performance of the fedspeak index, as each component may have different effects. Table 3 shows the results of predictive regressions between the effective fed funds rate and the two-year yields and FSI_t^{MP} and FSI_t^{I}. These are predictive regressions for daily data for the 3-month (65 days), 6-month (130 days), and one-year (260 days) horizons. The dependent variable is the change in the rates over these horizons, that is, $EFR_{t+h} - EFR_{t-1}$ and $y(2)_{t+h} - y(2)_{t-1}$ and the predictor is the indexes available at t. The top panel shows results for the full sample period (2009-2023) and the bottom for the last three years (2021-2023). Table 3 shows the coefficient estimates, the t-statistic to test whether the coefficient is statistically significant, and the R^2.

Using both the monetary policy and the information components leads to a substantial increase in the overall R^2 when predicting the two-year yields. For predicting the effective fed fund rates, the gains are only substantial over

*They are available as tickers BCMPUSGR Index and BCMPUSIF Index.

the most recent sample and for longer forecasting horizons. The information component - measuring how much of the variation of the Fedspeak is explained by data release surprises – is particularly relevant for predicting the two-year yields over the full sample period. For the most recent period, both components have significant effects. It is clear that an increase in FSI_t^{MP} leads to changes in the fed fund rates.

Predicting the Probability of a Fed Rate Cut/Hike

In assessing the likelihood of Federal Reserve rate decisions, a Probit model constitutes an integral part of our analysis. This section elucidates the methodology and insights gleaned from applying such a model, emphasizing its relevance in diverse economic conditions.

The Predicted Probability model's unique methodology offers explicit insights into how shifts in Fedspeak sentiment, as captured by the sophisticated preprocessing and scoring techniques described earlier, translate into action on interest rates. By analyzing historical data and applying our model's outcomes, we are able to discern distinct patterns in the Federal Reserve's responsiveness to economic signals, both from within its communications and from broader economic indicators.

One of the Probit model's key contributions is its ability to quantify the thresholds of sentiment that often precipitate a change in monetary policy, thereby providing stakeholders with a statistically backed anticipation of such moves. The model yields probabilities that inform on the likelihood of rate hikes or cuts under specific conditions, presenting a gauge for market participants on future policy directions.

The final outcome is a set of probabilities mapped against upcoming Federal Reserve meetings, providing a timeline for when market-moving decisions are most probable. These findings enrich the understanding of the Federal Reserve's reaction function and empower market participants with a data-driven perspective on the central bank's future actions.

To predict the probability of a hike/cut of the Fed Fund Rate in the next meeting using only the information on Fedspeak indices, we propose the following regime-switching probit model:

$$
\begin{aligned}
&\text{If } (FSI_{t-1}^{I} + FSI_{t-1}^{MP}) \leq \gamma : \\
&Prob[Cut_t] = \Phi(\beta_0 + \beta_1 FSI_{t-1}^{I} + \beta_2 FSI_{t-1}^{MP}) \\
&\text{If } (FSI_{t-1}^{I} + FSI_{t-1}^{MP}) > \gamma : \\
&Prob[Hike_t] = \Phi(\beta_3 + \beta_4 FSI_{t-1}^{I} + \beta_5 FSI_{t-1}^{MP}),
\end{aligned}
$$

where Φ is the cumulative density function of a standard Gaussian distribution.

The model implies that depending on whether we are in a loosening or tightening policy stance, the predictions will switch between the probability of a cut and the probability of a hike. The monetary policy stance is defined endogenously based on how the previous Fedspeak indexes compare to the threshold γ. The threshold γ and the impact coefficients β are estimated using the Fed Rate decisions in the FOMC meetings from 2009 to 2023 and the previous day observations of both indexes. The model is flexible enough to allow the contribution of the Fedspeak indexes to the policy decision to change with the policy stance, and for contributions from data and policy communication to have different weights.

In the Fedspeak App, we use the regime-switching probit model to display daily probabilities of a hike/cut in the next FOMC meeting.

Concluding Remarks and Future Work

The Bloomberg Economics Fedspeak Index marks a step towards an automated, systematic, and high frequency read of Fed communications. By combining unique Bloomberg terminal data and advanced machine learning techniques, our model is able to generate forecast gains to yields and Fed funds rate predictions.

Looking ahead, we will monitor the performance of the model in real-time, and fine-tune the model as needed. The rapid advancements in NLP techniques will also open avenues for further enhancements.

References

Ash, J. T., Zhang, C., Krishnamurthy, A., Langford, J. and Agarwal, A. (2019). Deep batch active learning by diverse, uncertain gradient lower bounds, *CoRR* **abs/1906.03671**.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V. (2019). Roberta: A robustly optimized BERT pretraining approach, *CoRR* **abs/1907.11692**.

Shi, X., Cao, W. and Raschka, S. (2023). Deep neural networks for rank-consistent ordinal regression based on conditional probabilities, *CoRR* **abs/2111.08851**.

Figure 1. Headlines tagged ”FEDSPEAK” by year

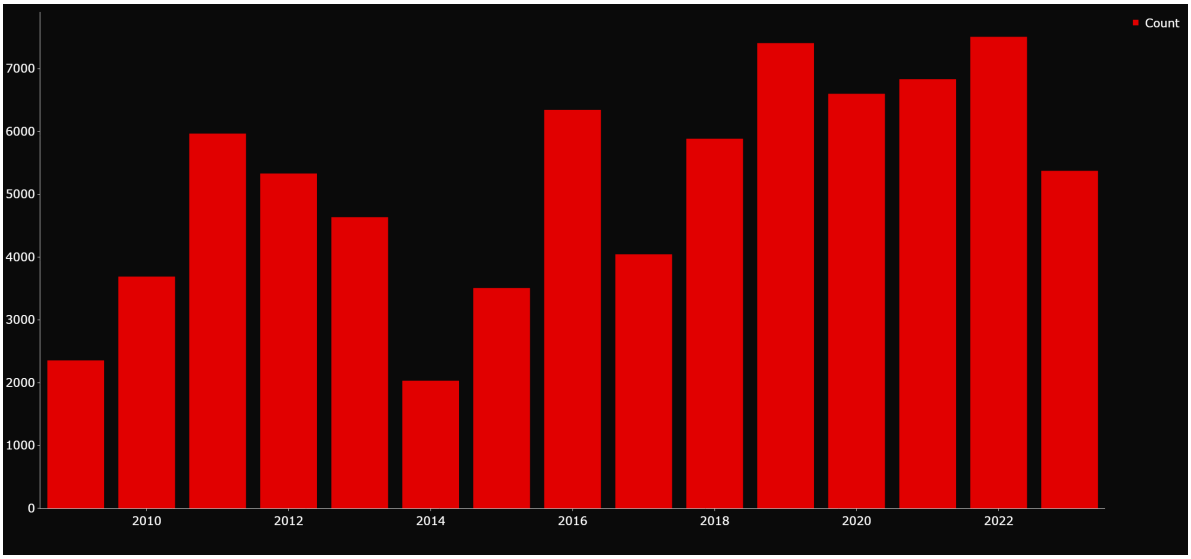

Figure 2. Fedspeak Index

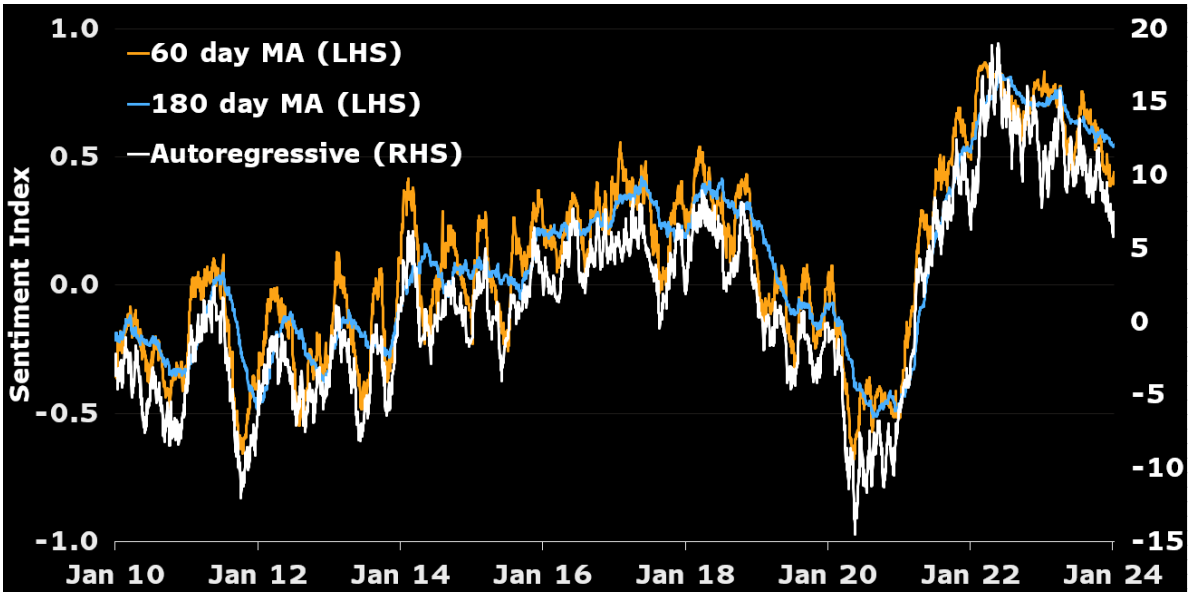

Table 1. Model Performance for Fedspeak Sentiment Analysis

Model	Accuracy	Macro F1	Mean Absolute Error
Loughran McDonald (LM) Dictionary	0.47	0.40	0.59
LM Dictionary w/ Dependency Parsing	0.49	0.44	0.57
RoBERTa Large	0.33	0.26	0.78
RoBERTa Large Fedspeak Pretrained	0.69	0.69	0.35
RoBERTa Large Fedspeak Ordinal Pretrained	**0.72**	**0.72**	**0.31**

Table 2. Fedspeak Predictive Content for Out-of-Sample Yield Forecasting (end-of-quarter from Q321 to Q423)

	2-Year Yields	5-Year Yields	10-Year Yields
consensus - Mean RMSE	1.782	1.220	0.926
	Gains to consensus (%):		
Yield-Only	17.7	16.8	25.0
Macro-Yield	87.0	81.1	77.1
Yield + Fedspeak	65.9	54.9	46.1
Macro-Yield + Fedspeak	132.9	118.3	101.9

Table 3. Fedspeak Predictive Content for Daily Changes of Effective FFR and Two-Year Yields

	Effective FFR			2-Year Yields		
	Full Sample			Full sample		
Horizon:	3-month	6-month	1-year	3-month	6-month	1-year
FSI	0.071	0.080	0.107	0.018	0.040	0.064
tstat	3.210	2.930	8.524	-2.352	3.384	3.872
R2	51.1	46.8	64.5	14.4	24.3	20.6
	Using Subcomponents					
FSI(MP)	0.019	0.039	0.056	0.006	0.016	0.019
tstat	2.897	4.326	3.499	0.952	1.428	0.683
FSI(I)	0.046	0.098	0.217	0.045	0.090	0.162
tstat	2.812	6.519	10.840	3.660	4.088	4.567
R2	41.0	54.6	68.5	27.6	41.7	43.4
	Last 3 years			Last 3 years		
horizon:	3-month	6-month	1-year	3-month	6-month	1-year
FSI	0.028	0.058	0.106	0.025	0.031	0.043
tstat	2.876	4.218	3.984	2.452	2.087	1.668
R2	34.5	45.0	45.2	5.3	9.5	17.0
	Using Subcomponents					
FSI(MP)	0.064	0.067	0.007	-0.008	-0.051	-0.133
tstat	2.804	1.618	0.165	-0.418	-2.819	-4.284
FSI(I)	0.088	0.101	0.227	0.108	0.163	0.253
tstat	2.357	1.589	5.105	4.472	6.464	7.565
R2	53.0	47.7	75.6	42.6	58.8	73.1