The Predictive Power of Fedspeak*

Nick Hallmark[1] and Andrej Sokol[1]

[1]Bloomberg LP

March 18, 2026

[PRELIMINARY AND INCOMPLETE]

Abstract

We show that embeddings derived from Bloomberg News headlines about the Federal Reserve contain meaningful real-time signals for key U.S. macroeconomic variables. Principal components of these embeddings, when incorporated into Bayesian VARs and quantile-regression frameworks, improve point and density forecasts for inflation, unemployment and Treasury yields relative to standard models, at times even outperforming professional forecasters. Our approach also captures shifts in risks that align with policy narratives. The results demonstrate the value of high-frequency central bank communication data for forecasting and enhance our understanding of how monetary policy communication is received by the public.

Keywords: Central Bank Communication; Text-as-Data; Large Language Models; Forecasting; Risk Assessment; Bayesian VAR; Quantile Regression; Inflation; Unemployment; Treasury Yields.

JEL Codes: C32; C45; C53; C55; E37; E52; E58.

**Disclaimer: the views expressed in this paper are those of the authors, and do not necessarily reflect those of Bloomberg LP.* Corresponding author: Andrej Sokol, asokol14@bloomberg.net.

1 Introduction

Central banks today are extraordinarily talkative. While there was a time when monetary policy deliberations hardly transpired into the public, modern central banks conduct press conferences, issue detailed policy statements, publish minutes, reports, forecasts and similar, and also increasingly maintain a deliberate presence on social media. This shift is not accidental: it reflects the way modern monetary institutions were designed in response to the classic time-inconsistency problem and the perceived need for credible, rule-like behaviour (Kydland and Prescott, 1977, Barro and Gordon, 1983, Rogoff, 1985). The move towards inflation targeting, and in particular inflation-forecast targeting, placed published projections and narrative explanations at centre stage, turning them into explicit intermediate targets of policy (Svensson, 1997, Bernanke and Mishkin, 1997). At the same time, a theoretical and policy-oriented literature has analysed the benefits and limits of transparency and communication as tools for managing expectations and ensuring accountability (Geraats, 2002, Woodford, 2005, Blinder et al., 2008, Lustenberger and Rossi, 2020, Dincer et al., 2022, Blinder et al., 2024).

The expansion of central bank communication has been accompanied by a rich empirical literature that aims both to disentangle monetary policy and information shocks (Romer and Romer, 1989, 2004, Nakamura and Steinsson, 2018, Cieslak and Schrimpf, 2019, ter Ellen et al., 2019, Jarocinski and Karadi, 2020, Miranda-Agrippino and Ricco, 2021, Ochs, 2021, Aruoba and Drechsel, 2024) and to construct quantitative indicators of tone[1], content and expectations from it, classify topics or embed their content in high-dimensional spaces (Lucca and Trebbi, 2009, Apel and Blix Grimaldi, 2014, Hansen and McMahon, 2016, Bholat et al., 2015, Ehrmann and Wabitsch, 2022, Moreno Pérez and Minozzo, 2024, Masciandaro et al., 2024, Peia and Romelli, 2024, Araujo et al., 2025, Gorodnichenko et al., 2025, among others).

Our work is related to these efforts and likewise treats text as a high-frequency

[1] For tone, there are many parallels with sentiment-based approaches in macroeconomics more broadly that read news and other large text corpora to extract information about inflation, growth, policy uncertainty etc. (Shapiro et al., 2022, Malliaropulos et al., 2024, Kwon et al., 2025, Baker et al., 2016).

data source on expectations and policy. It aims to extract signals about the future evolution of US unemployment and inflation – the two pillars of the Fed's dual mandate – as well as interest rates. The novelty is that we draw on central bank communications as digested by Bloomberg News and condensed into headlines, capturing the elements of "Fedspeak" that professional editors judge to be salient for investors. We encode these headlines with a RoBERTa (Liu et al., 2019) language model fine-tuned with supervised SimCSE (Gao et al., 2021) and then temporally aggregate them.

We show that this high-frequency, editorially curated dataset allows for real-time analysis of how the Fed's outlook – and therefore policy intentions – evolve and translate into signals that can improve both point forecasts and model-based risk assessments. In the first part of our paper, we build on Araujo et al. (2025) to show that information extracted from the embeddings via principal components substantially improves forecasts of inflation, unemployment and bond yields at horizons of up to one year, making them competitive against a number of model-based benchmarks, and in some cases, even professional forecasts.

We then examine whether Fedspeak also improves density forecasts of macroeconomic outcomes. To this end, we use quantile regression techniques (Koenker and Bassett, 1978, Adrian et al., 2019) to generate predictive densities conditioned on the same information set as in our baseline VAR. We assess their value against standard evaluation metrics, finding meaningful improvements, but also, from a more narrative perspective, show that real-time shifts in the distribution of risks around inflation and unemployment correspond to the intended messaging by the central bank.

Our findings carry important implications for both monetary policy communication and economic forecasting. Because the Fedspeak database is derived from news headlines rather than official transcripts, it captures how central bank messages are received and distilled by the media, and processed by market participants, in real time. From our results, we infer that professional forecasters already appear to internalize the unemployment-relevant content of Fedspeak—embedding-based models do not outperform the consensus in this dimension. However, for inflation and bond yields, our results indicate that professional forecasts can be improved, suggesting that key information is not being fully extracted from available communication.

The methodology we develop is broadly applicable. Similar headline databases exist for other central banks, including the European Central Bank and the Bank of England, and to a lesser extent, other monetary authorities. This opens the door to comparative studies on the effectiveness and transmission of monetary policy communication, as well as spillolvers.

The remainder of the paper is structured as follows. In Section 2 we describe the construction of the Fedspeak headline dataset and the embedding methodology. Section 3 evaluates the predictive content of the embeddings for point forecasts of key macroeconomic variables. Section 4 examines their usefulness for density forecasting. Section 5 concludes.

2 Extracting Signals from Fedspeak Headlines

We rely on a large proprietary database of news headlines about the Federal Reserve that covers, in a timely and comprehensive way, all elements of the central bank's external communications – from post-FOMC press conferences and policy statements to meeting minutes, speeches, interviews, essays, and tweets. The database includes over 70,000 headlines tagged as "Fedspeak" by Bloomberg editors, with daily updates starting in January 2009.

To transform each headline into a quantitative signal we encode the text with a RoBERTa (Liu et al., 2019) language model, fine-tuned with supervised SimCSE (Gao et al., 2021) on 13 years of Bloomberg News and Bloomberg First Word headline–lead pairs (Feb 2007 – Feb 2020). SimCSE's contrastive learning yields context-rich sentence level embeddings, providing greater semantic nuance than other static word embedding approaches. Domain-specific fine-tuning further captures the shorthand and syntax of real-time market reporting. Headline-level embeddings are averaged to monthly frequency. Note that because the fine-tuning text ends in early 2020, our forecast evaluation window — October 2020 through

April 2025 — avoids any potential look-ahead bias (Carriero et al., 2025, Alam et al., 2026).

Each embedding is still a vector of more than 1000 dimensions, but the first 10 principal component scores (Figure 1) explain more than 80% of the variation over time. The resulting scores comove with a broad set of macro variables (Figure 2), suggesting that the vectors distill economically relevant information. Moreover, using UMAP (McInnes et al., 2018) to project the high-dimensional embeddings vectors into a low-dimensional space we show that the headlines – as "summarized" by the principal components – cluster very clearly around the topics that usually feature in the Fed's communication: inflation, unemployment, interest rate policy, quantitative easing, recession risk etc. (Figure 3). Over time those scores also highlight topics that are relevant for the period, so for example, there's an uptick in the share of QE and tapering headlines in 2013, inflation headlines in early 2021, and trade policy and tariffs surge in importance starting in February 2025 (Figure 4).

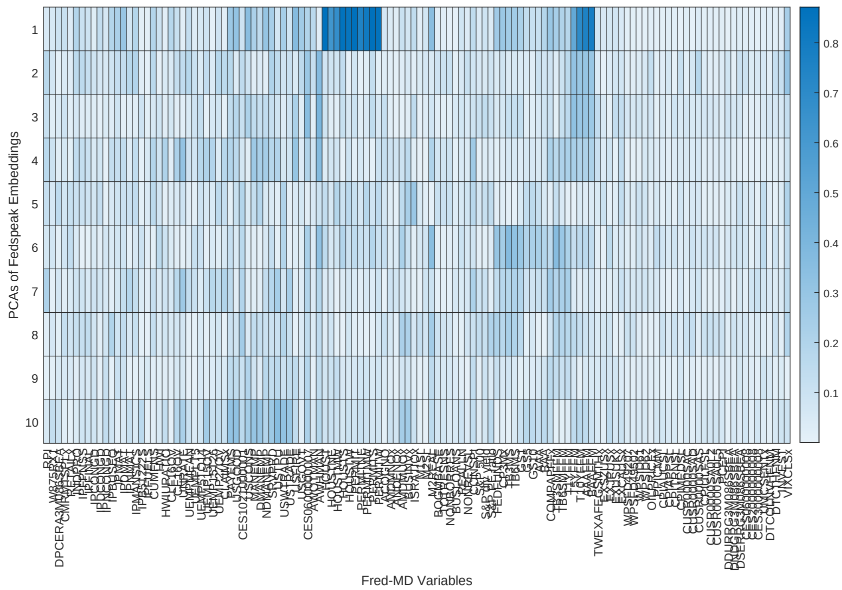
To help interpretability, we also establish a link between the individual PCA scores and the main topics identified by the clustering algorithm using Shapley values (Shapley, 1953, Huettner and Sunder, 2012, Shorrocks, 2013). For each monthly PCA score, we run a regression that relates the score to its lags, the monthly topic-share vector (with the residual "Other" category omitted), controlling for headline volume. We then use an exact Shapley decomposition—computed by averaging each topic's marginal contribution to model fit over all possible topic orderings—to summarize how much of the topics' explanatory contribution is attributable to each topic. Normalizing these contributions to sum to one for each score yields a compact topic-contribution profile that we use to label and group the scores.

The profiles point to a small number of clean semantic dimensions, with the remaining factors capturing mixtures and trade-offs (Figure5). The first group is labor-centered: Score 1 loads mainly on labor-market language with a sizeable monetary-policy component; Score 4 combines labor with trade/tariff narratives; and Score 6 is trade-heavy but still meaningfully tied to both labor and policy,

Figure 1 PCA Components of Fedspeak Embeddings

Note: This figure displays the first ten principal component (PC) scores extracted from the monthly Fedspeak embeddings over the 2009m1-2025m4 period.

Figure 2 Absolute Correlations Between Fedspeak Embeddings and Fred-MD Variables



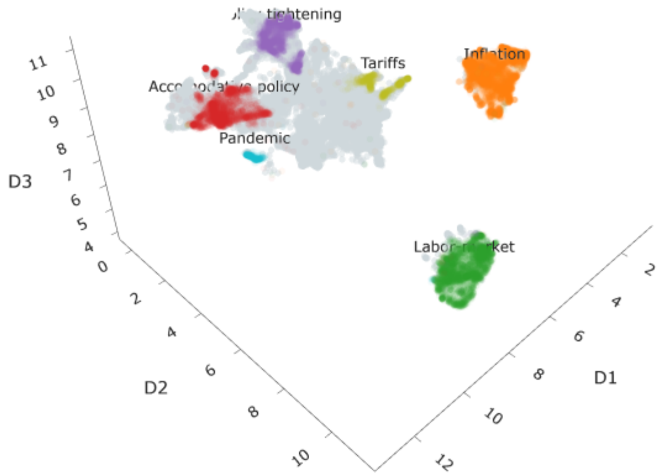
Note: The figure reports the absolute contemporaneous correlations between the first ten principal components of the Fedspeak embeddings and the macroeconomic and financial variables from the FRED-MD dataset (McCracken and Ng, 2016). Darker cells indicate stronger comovement.

consistent with trade shocks spilling into employment and policy discussion. A second group is prices and the policy reaction: Score 3 is a clean inflation factor, while Score 5 blends inflation with monetary policy in a way that naturally reads as a stance/reaction-function dimension. Trade also appears as a distinct axis, with Score 2 strongly identified with trade and tariffs and Score 7 linking trade to growth/outlook risks. Finally, Score 8 primarily reflects monetary policy with an important inflation backdrop (a policy trade-offs dimension), Score 10 is dominated by growth/outlook-and-risks language, and Score 9 is comparatively diffuse across topics, suggesting a residual “macro mix” factor rather than a single-topic label.

3 Forecasting with Fedspeak Embeddings

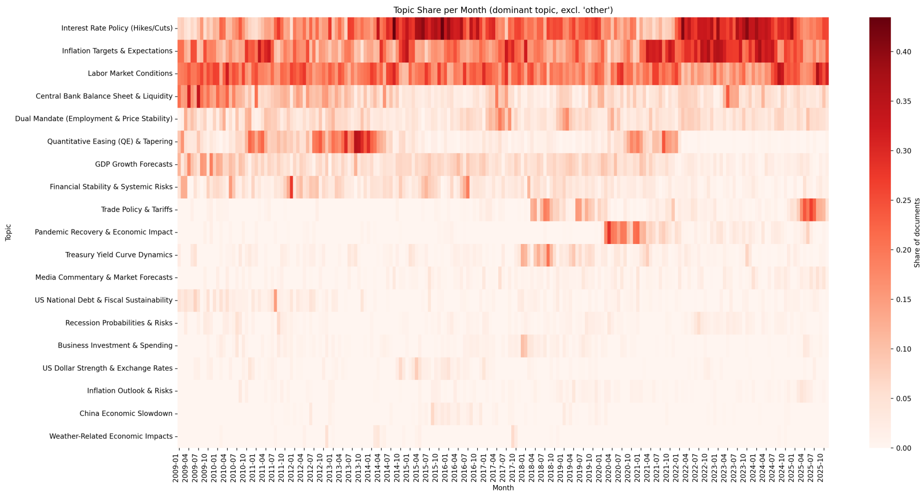
We build on the seminal contribution of Araujo et al. (2025) to show that information extracted from the embeddings via principal components substantially

Figure 3 Clustering of Fedspeak Headlines along First 3 PCA Directions



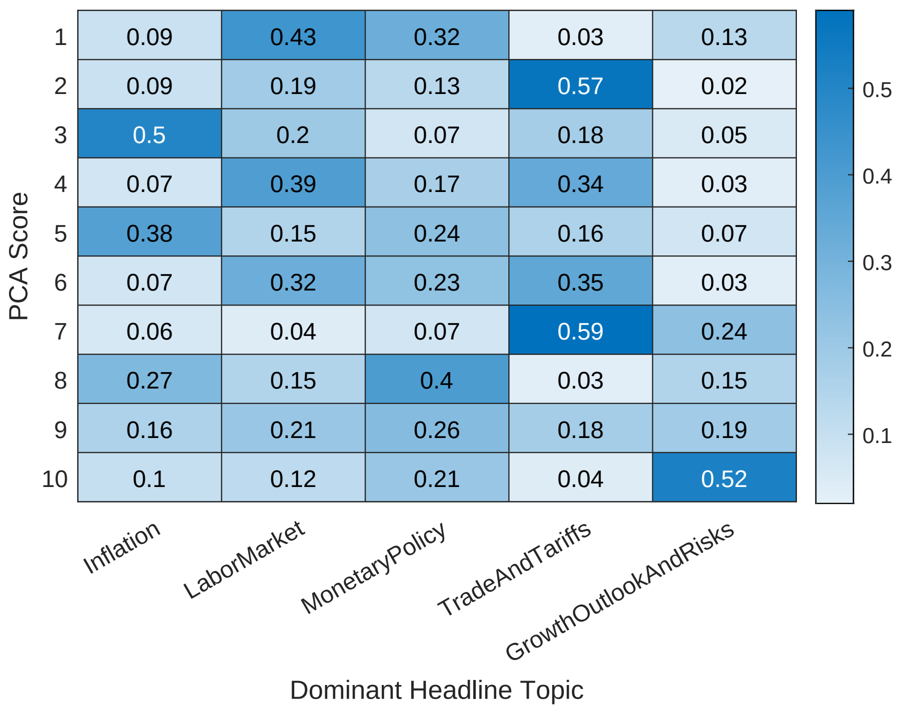
Note: The figure shows the clustering of headline embeddings using UMAP (McInnes et al., 2018) to project the high-dimensional sentence vectors into a low-dimensional space.

Figure 4 Topic Shares of Fedspeak Headlines over Time



Note: The figure displays the evolution of topic shares over time, based on clusters of headline embeddings obtained after projecting the high-dimensional sentence vectors into a low-dimensional space using UMAP (McInnes et al., 2018).

Figure 5 Shapley Decomposition of Incremental Fit: Topic Contributions to PCA Scores



Note: The figure shows, for each PCA score (row), how the explanatory power of the topic-share regressors is allocated across topics using an exact Shapley decomposition (averaging over all possible topic orderings). Values are normalized to sum to 1 within each row, so they should be read as "shares of topic contribution" for that PCA score. Larger values indicate that a given topic accounts for a larger fraction of the topics' incremental fit in the PCA-score regression (which also includes lags of the score and controls for headline volume).

improves forecasts of inflation and unemployment, the two pillars of the Fed's dual mandate, at horizons of up to one year, and even improves Treasury yields forecasts.

We take the first s principal components of the monthly embeddings ($s = 10$ in our baseline application) and include them in a Bayesian VAR model with monthly unemployment, CPI inflation, gasoline prices and 2-year Treasury yields. We estimate a reduced-form VAR(p) of the following form:

$$y_t = c + A(L)\, y_{t-1} + u_t \tag{1}$$

with

$$A(L) = A_1 + A_2 L + \cdots + A_p L^{p-1}$$

where y_t is a vector that contains observations of the four macro variables and the s (filtered) principal components, c a vector of constants, and u_t a vector of reduced-form residuals, assumed to be iid mean-zero Gaussian with full covariance matrix Σ_u.

We estimate the model using standard Bayesian techniques. Specifically, we impose Minnesota and sum-of-coefficients priors (Litterman, 1979, Doan et al., 1984), with hyperpriors selected optimally following Giannone et al. (2015), and including the volatility adjustment of Lenza and Primiceri (2022) to handle the post-pandemic period. All variables enter the model in levels (or log-levels, as appropriate); CPI inflation rates are inferred from the price index forecasts. We use real-time vintages of macroeconomic data from the ALFRED database as of the close of each month in our recursive estimation window, which spans October 2020 until April 2025. We also recompute the principal components from the embeddings for each vintage, also only using embeddings up to the close of the corresponding month.[2]

Our data at the close of each month display a so-called ragged edge: gasoline, yields and embeddings (which are monthly aggregations off daily, even intra-daily,

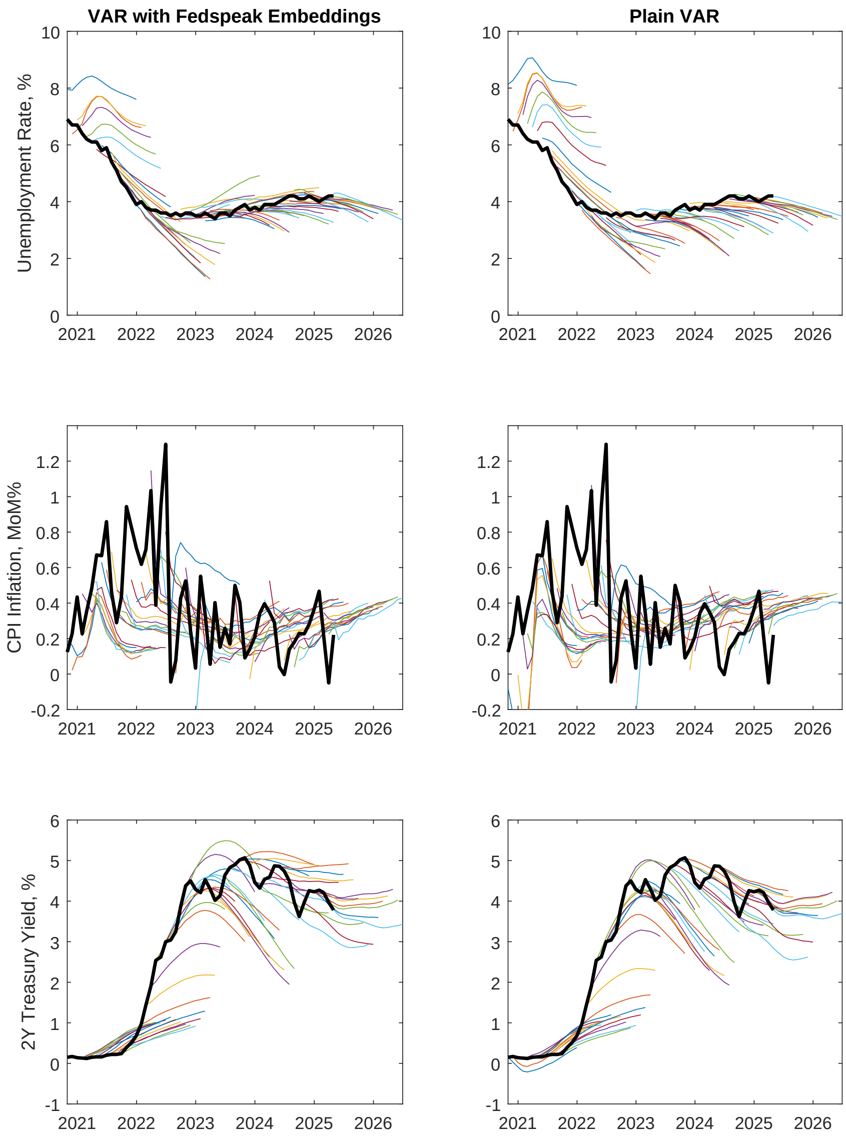
[2]This avoids any potential look-ahead bias arising from the PCA estimate including information that would not have been available at the time of the forecast. See Carriero et al. (2025) for a discussion of this important issue.

variables) are available up to the close of the month, but CPI and the unemployment rate are only available up to the previous month, as their releases typically occur in the first and second week of the *following* month. This feature, typical of "nowcasting" settings (Bańbura et al., 2013), can be handled with standard Kalman filtering and smoothing techniques (Bańbura et al., 2015), which yield nowcasts for the unemployment rate and CPI for the current month, as well as forecasts up to any horizon (in our case, up to 1 year). The left column of Figure 6 shows the full set of out-of-sample forecasts against outturns for the three macro variables of interest.

We compare our out-of-sample forecasts to a number of alternatives: a plain VAR only featuring the four macro variables, which serves as a benchmark (also shown in the right column of Figure 6); a VAR featuring, in addition to the macro variables, the first s factors extracted from the FRED-MD (McCracken and Ng, 2016) database vintage corresponding to each forecast origin; a VAR that instead adds the Bloomberg Economics Fedspeak score (Wong et al., 2023), a sentiment index extracted from the same headlines encoded in our embeddings, but following a different procedure and with a different objective, namely to capture shifts in tone in Fed officials' policy stance; a VAR featuring the San Francisco Fed's Economic Sentiment index (Shapiro et al., 2022); and finally, the median responses to two surveys of professional forecasters administered by Bloomberg – one (ECOS) just ahead of the upcoming releases of the unemployment rate and of CPI inflation, the other (ECFC) at regular intervals and concerning quarterly projections up to several quarters ahead.

The comparison reflects real-time data availability at the close of each month. As already noted, the plain VAR includes gasoline prices and treasury yields up to and including the month of the forecast, while the unemployment rate and the CPI index are only available up to the end of the previous month. Our model with embeddings PCAs, the one featuring the BE Fedspeak score, as well as the one featuring the SF Fed Economic Sentiment index all have those series included up to the month of the forecast, as they are all monthly aggregations of daily data. In contrast, FRED-MD vintages at the time of each forecast only yield factors up to the end of the previous month, as none of the underlying macro data for the current

Figure 6 Out-of-Sample Forecasts, Embeddings VAR and Plain VAR



Note: The figure shows out-of-sample forecasts (colored lines) against the latest vintage of outturns (bold black) for the 3 variables of interest; the left column shows forecasts from the model that includes PCA scores from the Fedspeak embeddings, the right column shows forecasts from the plain VAR benchmark that only includes the 4 macro variables. The vintages span 2020m10 to 2025m4.

Table 1 Out-of-Sample Point Forecast Performance

		Nowcast (Monthly)	1Q Ahead	2Q Ahead	3Q Ahead	4Q Ahead
Unemployment Rate	BVAR	**0.28**	**0.71**	**0.80**	**0.99**	**1.31**
	BVAR + Embeddings	0.78	0.57	0.74	0.81	0.80
	BVAR + FRED-MD Factors	0.89	0.62	0.83	0.92	0.88
	BVAR + BE Fedspeak Score	1.03	1.10	1.17	1.08	1.00
	BVAR + SF Fed Econ. Sentiment	1.01	1.05	1.11	1.07	1.03
	Bloomberg Consensus	0.64	0.36	0.48	0.52	0.50
CPI Inflation	BVAR	**0.21**	**0.79**	**1.58**	**2.41**	**3.15**
	BVAR + Embeddings	0.87	0.92	0.90	0.89	0.90
	BVAR + FRED-MD Factors	0.97	1.00	1.07	1.08	1.07
	BVAR + BE Fedspeak Score	0.98	1.00	1.01	1.02	1.03
	BVAR + SF Fed Econ. Sentiment	0.91	0.92	0.95	0.96	0.97
	Bloomberg Consensus	0.71	1.20	1.16	1.12	1.06
2Y Treasury	BVAR		**0.37**	**0.75**	**1.05**	**1.54**
	BVAR + Embeddings		1.01	0.96	0.96	0.97
	BVAR + FRED-MD Factors		0.97	0.98	1.02	1.01
	BVAR + BE Fedspeak Score		0.98	0.90	0.86	0.90
	BVAR + SF Fed Econ. Sentiment		1.02	1.02	1.02	1.03
	Bloomberg Consensus		1.21	1.12	1.16	1.06

Note: The table shows Root Mean Squared Errors (RMSE) for the Plain BVAR benchmark and relative RMSEs for the competing models, where entries < 1 indicate a smaller RMSE than the benchmark. The evaluation sample is 2020m10:2025m4.

month would have yet been available in real time. ECFC survey responses used for the evaluation of quarterly forecasts are also as of the close of each month, while ECOS responses for the "nowcast" enjoy a slight informational advantage, because most submissions occur only a few days ahead of the corresponding release – so unemployment rate survey responses have about one week's worth of additional information, and CPI survey responses roughly two.

Table 1 shows root mean squared errors (RMSE) for the plain VAR (in bold), and relative RMSEs for the alternatives. For the unemployment rate, the gain in forecast accuracy from adding embeddings PCAs to the model is material: over 20% for most horizons, peaking at 43% for 1-quarter-ahead forecast. That performance is not matched by any of the model-based alternatives. However, professional forecasts submitted to the two Bloomberg surveys do show an additional edge, with gains exceeding 30% throughout, and exceeding 60% 1 quarter ahead. For the nowcast, it's worth recalling that the professional forecasts enjoy a slight informational advantage, as submissions to the ECOS survey can continue past the close of the month, and incorporate informative releases such as the JOLTS and ADP employment change, whereas the information sets for the quarterly forecasts are aligned. Nevertheless, our results suggest that the embeddings capture useful information about labor market prospects.

For CPI inflation, the gains of including embeddings PCAs are somewhat more modest. However, both the monthly nowcast and forecasts several quarters ahead are competitive against all other model-based alternatives. Interestingly, while the Bloomberg consensus for the nowcast (ECOS survey) still fares better than the embeddings-based model, that's not the case for the quarterly inflation forecasts (ECFC survey), which over our evaluation sample perform worse than all model-based alternatives, including the plain VAR. This suggests that longer-term inflation forecasts from professional forecasters might benefit most from the additional information extracted from Fedspeak headlines.

Finally, 2-year Treasury yields forecasts from the model including embeddings PCAs also outperform the consensus survey, as well as most model-based alternatives. The one exception is the BVAR augmented with the BE Fedspeak Score (Wong et al., 2023). The latter result is particularly interesting, because the BE Fedspeak score is derived – albeit with a different methodology – from the same underlying news headlines, underscoring how the same corpus of text can yield different signals depending on how it's processed.

4 Risk Modeling with Fedspeak Embeddings

The information encoded in our Fedspeak embeddings is rich and nuanced, and carries signals whose usefulness goes beyond improving point forecasts. Earlier studies (Barbaglia et al., 2023, Sharpe et al., 2023, Filippou et al., 2023, Adämmer et al., 2024) had already highlighted that both news articles (the typical source of data underlying sentiment indices) and central banks' own textual outputs, such as FOMC transcripts and Greenbook/Tealbook narratives accompanying staff forecasts, can improve density forecasts, notably by highlighting tail risks, which are often easier to articulate in words than quantify. We build on those results to show that the principal components extracted from the embeddings in real time improve density forecasts of CPI inflation and especially the unemployment rate at horizons of up to a year.

To that end, we estimate quantile regressions (Koenker and Bassett, 1978) with the same macro variables as in the preceding section but stationarized as required (so CPI inflation, unemployment rate, 2Y Treasury yield and gasoline price inflation), and considering the same set of additional predictors (Fedspeak embedddings, FRED-MD factors, BE Fedspeak score, SF Fed Economic Sentiment Index). As in the previous section, we reflect the ragged-edge pattern of real-time data availability at the close of each month, exploiting the conditional quantile regression framework proposed by Sokol (2025). We thus in effect estimate "at-risk" models in the spirit of (Adrian et al., 2019), although not using financial conditions as predictors as in their seminal contribution.

Specifically, we first estimate monthly regressions of the following form

$$y_t = \alpha_\tau + \sum_{\ell=1}^{p} \beta'_{\tau,\ell} x_{t-h-\ell} + \gamma'_\tau z_{t-h} + \varepsilon_{\tau,t} \tag{2}$$

using Bayesian methods (Yu and Moyeed, 2001, Kozumi and Kobayashi, 2011, Khare and Hobert, 2012), where y_t is either monthly CPI inflation or the unemployment rate; x_t contains lagged values of the four macro variables, up to lag order p ($p = 3$ in our baseline application); z_t contains variables available up to the period of the forecast, so gasoline price inflation, the 2Y Treasury yield, a pandemic dummy and the embeddings PCAs, FRED-MD factors[3], BE Fedspeak score or SF Fed Economic Sentiment Index; the constant α_τ and the coefficients β_τ and γ_τ are quantile- and horizon-specific, and so is the residual $\varepsilon_{\tau,t}$, which is assumed to follow an Asymmetric Laplace distribution; $h = 0 \ldots H$ is the horizon of the regression, and for CPI inflation and the unemployment rate includes the nowcast.

We then use the estimated parameter draws to obtain predictive quantile draws conditional on the latest available information at each forecast origin. At each horizon, we approximate the implied density by fitting a kernel density to these simulated quantiles and take the resulting monthly predictive distributions as the basis for our evaluation. To construct quarterly distributions, we simulate joint paths for each variable by sampling from the sequence of monthly densities under a Gaussian copula, which captures the serial correlation of our variables of interest.

[3]s FRED-MD factors are introduced in the same way, but they are only available up to $t-1$ in real time.

The simulated monthly outcomes are then aggregated to quarterly frequency, and we fit a second kernel density to the resulting draws to obtain the corresponding quarterly predictive distributions.

We follow Gneiting and Ranjan (2011) and plot average quantile scores for all models over our evaluation sample in Figures 7a to 7c. The quantile score (or tick loss function, see Giacomini and Komunjer, 2005) for each data vintage v is defined as

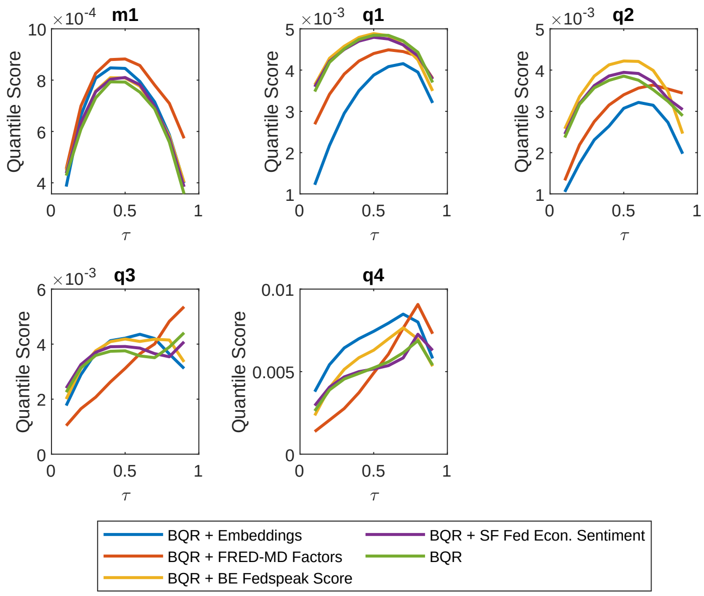
$$QS_{v,\tau,h} = \rho_\tau \left(y_{t_v+h} - \hat{P}^{-1}_{v,h}(\tau) \right) \tag{3}$$

The score penalises outturns that are more extreme (i.e. fall further in the corresponding tail) than the predictive quantile $\hat{P}^{-1}_{v,h}(\tau)$. The figures show that, for the unemployment rate, the model augmented with Fedspeak embeddings delivers similar or lower average quantile scores than the plain benchmark at shorter horizons, with especially pronounced gains for one- to two-quarter-ahead forecasts. For the other specifications, results are more mixed, with improvements that are more sporadic and horizon- or quantile-specific.

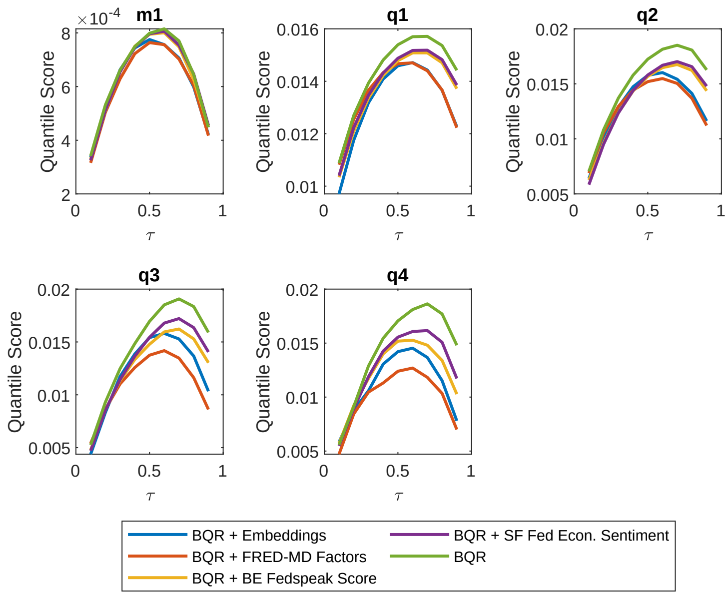
For CPI inflation, improvements are more modest: the embeddings-based model performs similarly for the nowcast, but tends to yield slightly lower quantile scores at intermediate and longer horizons, particularly around the centre of the predictive distribution. Unlike for the unemployment rate, however, all alternative models appear to deliver some improvement over the benchmark for inflation, even if these gains are generally small and uneven across horizons and quantiles. Within this set, the model with FRED-MD factors is broadly competitive (and in places comparable to embeddings), while the BE Fedspeak score mostly tracks the benchmark with mild improvements; by contrast, the SF sentiment index is more likely to underperform at longer horizons in parts of the distribution, showing higher quantile scores in several regions.

Figure 7 Quantile Scores

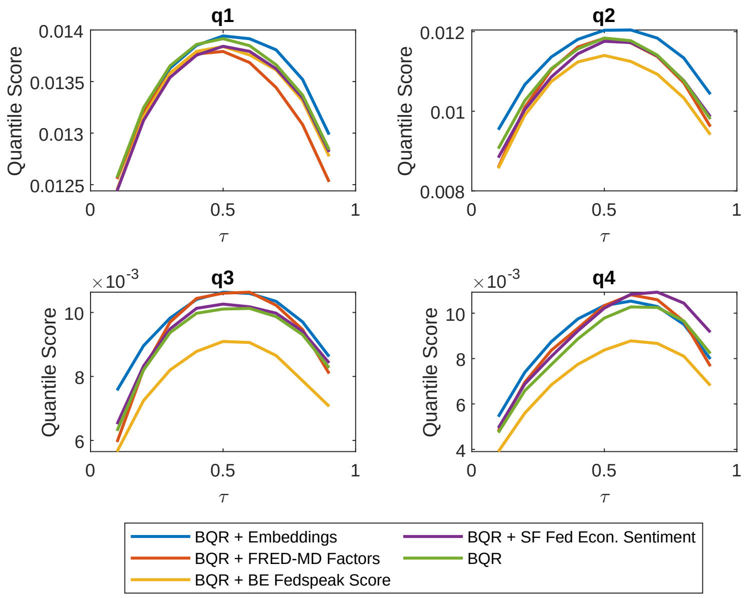
(a) Unemployment Rate



(b) CPI Inflation



(c) 2Y Treasuries



Note: The chart shows average quantiles scores, by quantile and horizon, over the forecast evaluation period 2020m10 to 2025m4. A lower average quantile score denotes better performance for the quantile in question.

For Treasury yields, the overall picture is that the text-based signals are mostly within the pack rather than consistently dominating: the embeddings-based model and the SF Fed sentiment index typically deliver quantile scores that are close to those of the plain BQR benchmark across quantiles and horizons, while the model with FRED-MD factors is likewise broadly competitive. What stands out instead is the BQR model with the BE Fedspeak score, which achieves noticeably lower quantile scores over a wide range of quantiles and horizons—mirroring the earlier point-forecast comparison. This pattern is not too surprising, given that the BE Fedspeak score was constructed with the predictability of medium- and longer-maturity Treasury yields (including 2- and 10-year yields) in mind (**?**).

Gneiting and Ranjan (2011) also propose a set of quantile-weighted versions of continuously ranked probability scores to assess forecasting performance in specific regions of the predictive distribution. The general form of their scores is

$$GR_{v,\tau,h} = \int_0^1 QS_{v,\tau,h} w(\tau)\, d\tau \tag{4}$$

where w are non-negative weight functions on the real line. GR scores are useful summary statistics for comparisons and formal testing, because unlike weighted versions of the traditional log score (Amisano and Giacomini, 2007), they retain propriety (see also Diks et al., 2011) and are amenable to standard statistical testing techniques (Diebold and Mariano, 1995).

For the Bayesian Quantile Regression (BQR) model augmented with the embeddings, Table 2 shows that GR score ratios are close to one for the nowcast but tend to fall below one—indicating better performance than the plain benchmark—at longer horizons, especially when more weight is placed on the centre and on both tails of the distribution (w_1 and w_2). For unemployment, this translates into sizeable gains at one- to three-quarter horizons, with improvements visible across much of the predictive distribution and only some deterioration for the left-tail weights (w_4). Relative to embeddings, the model with FRED-MD factors also improves at intermediate horizons but less uniformly, while the BE Fedspeak score and the SF sentiment index mostly remain close to the benchmark, with only small and uneven gains.

For CPI inflation, the gains from embeddings are more moderate but broad-based, with particularly better performance in the upper tail (w_3), suggesting that the embeddings help refine assessments of upside inflation risks. Inflation is also the case where the alternative models most consistently deliver improvements against the plain BQR: the model with FRED-MD factors shows robust gains that become more pronounced at longer horizons, especially under tail-weighted scores, while the BE Fedspeak score tends to track the benchmark more closely, with at most mild improvements one year ahead. By contrast, the SF sentiment index is broadly comparable at short horizons but tends to deteriorate at longer horizons, particularly when the score puts more emphasis on tail outcomes.

Table 2 Average GR Score Ratios

(a) Unemployment Rate

		Nowcast (Monthly)	1Q Ahead	2Q Ahead	3Q Ahead	4Q Ahead
BQR + Embeddings	Unweighted	1.01	0.72	0.69	0.99	1.30
	Center	1.03	0.76	0.73	1.03	1.35
	Tails	0.96	0.61	0.60	0.86	1.15
	Right tail	1.02	0.83	0.75	1.00	1.20
	Left tail	0.99	0.56	0.60	0.92	1.35
BQR + FRED-MD Factors	Unweighted	1.11	0.90	0.86	0.86	0.96
	Center	1.11	0.91	0.87	0.85	0.97
	Tails	1.13	0.88	0.84	0.91	0.96
	Right tail	1.18	0.97	0.98	1.11	1.16
	Left tail	1.06	0.83	0.72	0.60	0.68
BQR + BE Fedspeak Score	Unweighted	1.00	1.00	1.03	1.02	1.08
	Center	1.00	1.01	1.06	1.05	1.14
	Tails	1.00	0.97	0.98	0.94	0.94
	Right tail	0.99	0.97	0.99	1.03	1.06
	Left tail	1.00	1.01	1.06	0.97	1.04
BQR + SF Sentiment Index	Unweighted	0.96	1.00	0.97	0.98	0.96
	Center	0.97	1.01	0.97	0.97	0.99
	Tails	0.94	0.97	0.97	1.01	0.90
	Right tail	0.95	0.99	0.97	1.03	0.95
	Left tail	0.97	0.99	0.97	0.95	0.95

(b) CPI Inflation

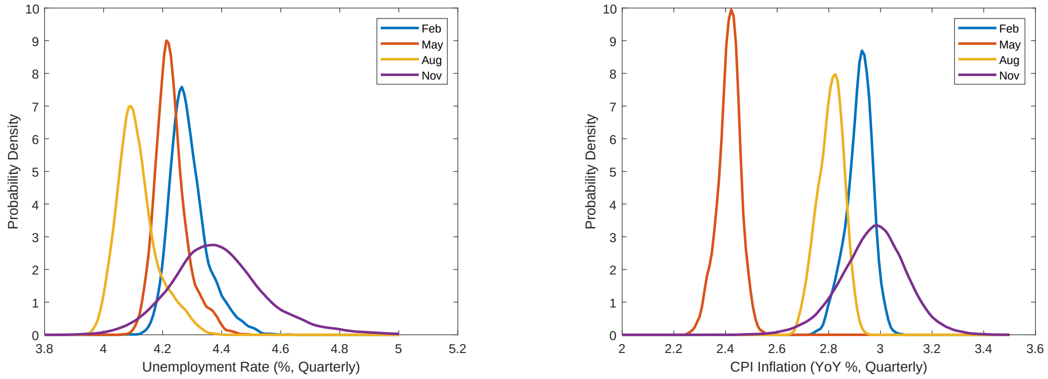
		Nowcast (Monthly)	1Q Ahead	2Q Ahead	3Q Ahead	4Q Ahead
BQR + Embeddings	Unweighted	0.96	0.94	0.94	0.91	0.86
	Center	0.96	0.96	0.96	0.94	0.88
	Tails	0.94	0.90	0.87	0.81	0.79
	Right tail	0.92	0.92	0.86	0.83	0.78
	Left tail	0.99	0.95	1.01	0.98	0.94
BQR + FRED-MD Factors	Unweighted	0.94	0.97	0.93	0.83	0.77
	Center	0.95	0.98	0.95	0.85	0.78
	Tails	0.93	0.95	0.88	0.75	0.72
	Right tail	0.92	0.92	0.84	0.72	0.70
	Left tail	0.96	1.01	1.04	0.95	0.85
BQR + BE Fedspeak Score	Unweighted	0.99	0.99	0.99	0.96	0.95
	Center	0.99	0.99	0.99	0.96	0.96
	Tails	0.98	0.99	1.00	0.96	0.93
	Right tail	0.98	0.99	0.98	0.95	0.91
	Left tail	1.00	0.99	1.01	0.98	0.99
BQR + SF Sentiment Index	Unweighted	1.00	1.04	1.11	1.11	1.14
	Center	1.00	1.04	1.10	1.10	1.12
	Tails	0.99	1.04	1.12	1.12	1.19
	Right tail	0.99	1.04	1.10	1.12	1.20
	Left tail	1.01	1.04	1.13	1.09	1.08

(c) 2Y Treasury

		Nowcast (Monthly)	1Q Ahead	2Q Ahead	3Q Ahead	4Q Ahead
BQR + Embeddings	Unweighted		1.00	1.04	1.07	1.05
	Center		1.00	1.03	1.06	1.05
	Tails		1.00	1.05	1.10	1.04
	Right tail		1.01	1.05	1.04	1.00
	Left tail		1.00	1.04	1.12	1.11
BQR + FRED-MD Factors	Unweighted		0.99	0.99	1.02	1.03
	Center		0.99	0.99	1.03	1.04
	Tails		0.99	0.97	0.98	0.99
	Right tail		0.98	0.99	1.01	1.00
	Left tail		0.99	0.97	1.01	1.04
BQR + BE Fedspeak Score	Unweighted		0.99	0.96	0.88	0.85
	Center		0.99	0.96	0.88	0.85
	Tails		0.99	0.95	0.88	0.84
	Right tail		0.99	0.96	0.86	0.84
	Left tail		0.99	0.96	0.90	0.85
BQR + SF Sentiment Index	Unweighted		0.99	0.99	1.02	1.06
	Center		0.99	0.99	1.01	1.05
	Tails		0.99	0.99	1.02	1.08
	Right tail		1.00	1.00	1.01	1.08
	Left tail		0.99	0.98	1.02	1.04

Note: Average Gneiting and Ranjan (2011) scores relative to the plain BQR for all competing models, under different weighting functions: $w_0 = 1$; $w_1(\tau) = \tau(1-\tau)$; $w_2(\tau) = (2\tau - 1)^2$; $w_3(\tau) = \tau^2$; $w_4(\tau) = (1-\tau)^2$. A ratio > 1 indicates that a model performs worse than the plain BQR.

Figure 8 Evolution of predictive densities for the unemployment rate and CPI inflation in 2025.



(a) Unemployment Rate

(b) CPI Inflation

Note: The charts show the evolution of predictive densities for quarterly unemployment (left panel) and CPI inflation (right panel) over 2025, based on information available at the end of February, May, August and November.

For 2-year Treasury yields, Table 2 suggests that the embeddings-based BQR does not deliver the same systematic improvements as for the macro variables: GR score ratios are often close to one at short horizons, sometimes even above one at longer horizons. The picture is similar for FRED-MD factors and the SF Fed Sentiment Index. The model with the BE Fedspeak score stands out with GR score ratios well below one across weighting functions and horizons, indicating sizeable and robust gains in density forecasting performance for yields, as already evident from the underlying quantile scores.

4.1 Case Study

To illustrate how the embeddings can be used for real-time risk assessment, we study the evolution of predictive densities for quarterly unemployment and CPI inflation over 2025, conditional on information available at the end of the second month of each quarter — February, May, August and November (Figure 8). The distributions shift noticeably over the year in terms of their modes, the degree of dispersion and skewness, tracking changes in the tone and content of Fed communications as captured by the headline embeddings.

For unemployment, the February vintage points to a modal outcome for 1Q around 4.3%, with a moderately dispersed distribution and some probability mass in the lower tail, consistent with concerns about labour-market softness and trade policy. By May, the most likely unemployment rate for 2Q is lower and the density is more concentrated, indicating a more benign and tightly framed outlook. In August, the mode shifts down further but the distribution widens again, with renewed downside risks as Fedspeak increasingly emphasises labour-market uncertainties and the implications of the interest-rate path. By November, the density for 4Q is centred near 4.4% but is noticeably more dispersed than in earlier rounds, reflecting both the absence of October data owing to the government shutdown and heightened uncertainty in communications about the durability of labour-market strength and the appropriateness of further policy easing.

The inflation densities display a similar pattern of evolving risks. In February, the 1Q CPI inflation distribution is relatively wide and centred just below 3%, with a slight tilt towards lower outcomes; by May, the forecast for 2Q is revised down and the associated density becomes much tighter, signalling a more confident view that inflation will remain contained. In August, the modal forecast for 3Q moves back up towards about 2.8%, but the distribution still puts substantial weight on relatively subdued outcomes, in hindsight understating the upside risks. By November, the 4Q distribution is again centred around 3%, yet it is noticeably more spread out than the earlier vintages, indicating increased uncertainty about the near-term inflation path as well as the lingering effects of missing data around the government shutdown.

5 Conclusion

Our analysis demonstrates that Fedspeak, as captured through Bloomberg News headlines, contains meaningful real-time signals about the economic outlook. By transforming headline text into domain-specific embeddings and incorporating their principal components into standard forecasting frameworks, we show that central bank communication not only comoves with underlying macroeconomic conditions but also enhances predictive accuracy for inflation, unemployment, and Treasury yields relative to standard benchmarks. These gains persist across horizons and, for inflation and yields, even surpass those of professional forecasters, suggesting that the latter do not fully internalize all informational content embedded in day-to-day Fedspeak. The results also reveal that communication-driven movements in predictive distributions align closely with policy narratives, indicating that the approach captures features of the intended messaging. Overall, the evidence underscores the value of treating central bank communication as a high-frequency data source that can materially improve economic forecasting and enrich our understanding of how policy signals are transmitted.

References

Adrian, T., N. Boyarchenko, and D. Giannone (2019): "Vulnerable Growth," *American Economic Review*, 109, 1263–1289.

Adämmer, P., J. Prüser, and R. Schüssler (2024): "News media and tail risk forecasts of macroeconomic indicators," *International Journal of Forecasting*, forthcoming.

Alam, M., S. Boyle, H. Li, and T. Sekhposyan (2026): "ChatMacro: Evaluating Inflation Forecasts of Generative AI," Working Paper Series 2026-04, Federal Reserve Bank of San Francisco.

Amisano, G. and R. Giacomini (2007): "Comparing Density Forecasts via Weighted Likelihood Ratio Tests," *Journal of Business & Economic Statistics*, 25, 177–190.

Apel, M. and M. Blix Grimaldi (2014): "How Informative Are Central Bank Minutes?" *Review of Economics*, 65, 53–76.

Araujo, D. K. G., N. Bokan, F. A. Comazzi, and M. Lenza (2025): "Word2Prices: Embedding Central Bank Communications for Inflation Prediction," BIS Working Paper 1253, Bank for International Settlements.

Aruoba, S. B. and T. Drechsel (2024): "Identifying Monetary Policy Shocks: A Natural Language Approach," NBER Working Paper 32417, National Bureau of Economic Research.

Baker, S. R., N. Bloom, and S. J. Davis (2016): "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131, 1593–1636.

Barbaglia, L., S. Consoli, and S. Manzan (2023): "Forecasting with Economic News," *Journal of Business & Economic Statistics*, 41, 708–719.

Barro, R. J. and D. B. Gordon (1983): "Rules, Discretion and Reputation in a Model of Monetary Policy," *Journal of Monetary Economics*, 12, 101–121.

Bańbura, M., D. Giannone, and M. Lenza (2015): "Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections," *International Journal of Forecasting*, 31, 739–756.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013): "Chapter 4 - Now-Casting and the Real-Time Data Flow," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier, vol. 2 of *Handbook of Economic Forecasting*, 195–237.

Bernanke, B. S. and F. S. Mishkin (1997): "Inflation Targeting: A New Framework for Monetary Policy?" *Journal of Economic Perspectives*, 11, 97–116.

Bholat, D., S. Hansen, P. Santos, and C. Schonhardt-Bailey (2015): "Text Mining for Central Banks," Centre for Central Banking Studies Handbook 33, Bank of England.

Blinder, A. S., M. Ehrmann, J. de Haan, and D.-J. Jansen (2024): "Central Bank Communication with the General Public: Promise or False Hope?" *Journal of Economic Literature*, 62, 425–457.

Blinder, A. S., M. Ehrmann, M. Fratzscher, J. de Haan, and D.-J. Jansen (2008): "Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence," *Journal of Economic Literature*, 46, 910–945.

Carriero, A., D. Pettenuzzo, and S. Shekhar (2025): "Macroeconomic Forecasting with Large Language Models," Manuscript.

Cieslak, A. and A. Schrimpf (2019): "Non-monetary news in central bank communication," *Journal of International Economics*, 118, 293–315.

Diebold, F. X. and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

Diks, C., V. Panchenko, and D. van Dijk (2011): "Likelihood-based scoring rules for comparing density forecasts in tails," *Journal of Econometrics*, 163, 215–230.

Dincer, N. N., B. Eichengreen, and P. M. Geraats (2022): "Trends in Monetary Policy Transparency: Further Updates," *International Journal of Central Banking*, 18, 331–348.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric reviews*, 3, 1–100.

Ehrmann, M. and A. Wabitsch (2022): "Central Bank Communication with Non-Experts – A Road to Nowhere?" *Journal of Monetary Economics*, 127, 69–85.

Filippou, I., J. Mitchell, and M. T. Nguyen (2023): "The FOMC versus the Staff: Do Policymakers Add Value in Their Tales?" Working Paper 23-20, Federal Reserve Bank of Cleveland.

Gao, T., X. Yao, and D. Chen (2021): "SimCSE: Simple Contrastive Learning of Sentence Embeddings," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 6894–6910.

Geraats, P. M. (2002): "Central Bank Transparency," *The Economic Journal*, 112, 532–565.

Giacomini, R. and I. Komunjer (2005): "Evaluation and Combination of Conditional Quantile Forecasts," *Journal of Business & Economic Statistics*, 23, 416–431.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior Selection for Vector Autoregressions," *The Review of Economics and Statistics*, 97, 436–451.

Gneiting, T. and R. Ranjan (2011): "Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules," *Journal of Business & Economic Statistics*, 29, 411–422.

Gorodnichenko, Y., T. Pham, and O. Talavera (2025): "Central Bank Communication on Social Media: What, To Whom, and How?" *Journal of Econometrics*, 249, 105869.

Hansen, S. and M. McMahon (2016): "Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication," *Journal of International Economics*, 99, S114–S133.

Huettner, F. and M. Sunder (2012): "Axiomatic Principles for Decomposing the Explanatory Power of a Regression Model," *Journal of Economic Inequality*, 10, 1–19.

Jarocinski, M. and P. Karadi (2020): "Deconstructing Monetary Policy Surprises—The Role of Information Shocks," *American Economic Journal: Macroeconomics*, 12, 1–43.

Khare, K. and J. P. Hobert (2012): "Geometric ergodicity of the Gibbs sampler for Bayesian quantile regression," *Journal of Multivariate Analysis*, 112, 108–116.

Koenker, R. W. and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Kozumi, H. and G. Kobayashi (2011): "Gibbs sampling methods for Bayesian quantile regression," *Journal of Statistical Computation and Simulation*, 81, 1565–1578.

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

Kydland, F. E. and E. C. Prescott (1977): "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 85, 473–492.

Lenza, M. and G. E. Primiceri (2022): "How to Estimate a VAR after March 2020," *Journal of Applied Econometrics*, 37, 688–703.

Litterman, R. B. (1979): "Techniques of forecasting using vector autoregressions," Tech. rep.

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, et al. (2019): "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *arXiv preprint arXiv:1907.11692*.

Lucca, D. O. and F. Trebbi (2009): "Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements," NBER Working Paper 15367, National Bureau of Economic Research.

Lustenberger, T. and E. Rossi (2020): "Does Central Bank Transparency and Communication Affect Financial and Macroeconomic Forecasts?" *International Journal of Central Banking*, 16.

Malliaropulos, D., E. Passari, and F. Petroulakis (2024): "Unpacking commodity price fluctuations: reading the news to understand inflation," Working Papers 334, Bank of Greece.

Masciandaro, D., O. Peia, and D. Romelli (2024): "Central Bank Communication and Social Media: From Silence to Twitter," *Journal of Economic Surveys*, also circulated as BAFFI CAREFIN Working Paper No. 22187.

McCracken, M. W. and S. Ng (2016): “FRED-MD: A Monthly Database for Macroeconomic Research,” *Journal of Business & Economic Statistics*, 34, 574–589.

McInnes, L., J. Healy, and J. Melville (2018): “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction,” *arXiv preprint arXiv:1802.03426*.

Miranda-Agrippino, S. and G. Ricco (2021): “The Transmission of Monetary Policy Shocks,” *American Economic Journal: Macroeconomics*, 13, 74–107.

Moreno Pérez, C. and M. Minozzo (2024): ““Making Text Talk”: The Minutes of the Central Bank of Brazil and the Real Economy,” *Journal of International Money and Finance*, 147, 103133.

Nakamura, E. and J. Steinsson (2018): “High-Frequency Identification of Monetary Non-Neutrality: The Information Effect,” *The Quarterly Journal of Economics*, 133, 1283–1330.

Ochs, A. C. R. (2021): “A New Monetary Policy Shock with Text Analysis,” Cambridge Working Papers in Economics 2148, Faculty of Economics, University of Cambridge.

Peia, O. and D. Romelli (2024): “Central Bank Communication and Social Media: From Silence to Twitter,” *Journal of Economic Surveys*.

Rogoff, K. (1985): “The Optimal Degree of Commitment to an Intermediate Monetary Target,” *Quarterly Journal of Economics*, 100, 1169–1189.

Romer, C. D. and D. H. Romer (1989): “Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz,” *NBER Macroeconomics Annual*, 4, 121–170.

——— (2004): “A New Measure of Monetary Shocks: Derivation and Implications,” *American Economic Review*, 94, 1055–1084.

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

Shapley, L. S. (1953): “A Value for n-Person Games,” *Contributions to the Theory of Games II*, 307–317.

Sharpe, S. A., N. R. Sinha, and C. A. Hollrah (2023): “The power of narrative sentiment in economic forecasts,” *International Journal of Forecasting*, 39, 1097–1121.

Shorrocks, A. F. (2013): "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value," *Journal of Economic Inequality*, 11, 99–126.

Sokol, A. (2025): "Fan Charts 2.0: Flexible Forecast Distributions with Expert Judgement," *International Journal of Forecasting*, 41, 1148–1164.

Svensson, L. E. O. (1997): "Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets," *European Economic Review*, 41, 1111–1146.

ter Ellen, S., V. H. Larsen, and L. A. Thorsrud (2019): "Narrative Monetary Policy Surprises and the Media," Working Paper 19/2019, Norges Bank.

Wong, A., N. Hallmark, and A. Galvao (2023): "Bloomberg Economics Fedspeak: Towards a High Frequency Fed Sentiment Index," Manuscript.

Woodford, M. (2005): "Central Bank Communication and Policy Effectiveness," NBER Working Paper 11898, National Bureau of Economic Research.

Yu, K. and R. A. Moyeed (2001): "Bayesian quantile regression," *Statistics & Probability Letters*, 54, 437–447.