ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Mind Your Language: Market Responses to Central Bank Speeches

Authors	Maximilian Ahrens, Deniz Erdemlioglu, Michael McMahon, Christopher J. Neely, and Xyie Yang
Working Paper Number	2023-013B
Revision Date	February 2024
Citable Link	https://doi.org/10.20955/wp.2023.013
Suggested Citation	Ahrens, M., Erdemlioglu, D., McMahon, M., Neely, C.J., Yang, X., 2024; Mind Your Language: Market Responses to Central Bank Speeches, Federal Reserve Bank of St. Louis Working Paper 2023-013. URL https://doi.org/10.20955/wp.2023.013

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

Mind Your Language: Market Responses to Central Bank Speeches*

Maximilian Ahrens
Oxford-Man Institute of Quantitative Finance, Department of Economics, University of Oxford

Deniz Erdemlioglu
Department of Finance, IESEG School of Management and CNRS

Michael McMahon
Department of Economics, University of Oxford

Christopher J. Neely
Research Division, Federal Reserve Bank of St. Louis

Xiye Yang
Department of Economics, Rutgers University

February 21, 2024

Abstract

Researchers have carefully studied post-meeting central bank communication and have found that it often moves markets, but they have paid less attention to the more frequent central bankers' speeches. We create a novel dataset of US Federal Reserve speeches and develop supervised multimodal natural language processing methods to identify how monetary policy news affect financial volatility and tail risk through implied changes in forecasts of GDP, inflation, and unemployment. We find that news in central bankers' speeches can help explain volatility and tail risk in both equity and bond markets. Our results challenge the conventional view that central bank communication primarily resolves uncertainty and indicate that markets attend to speech signals more closely during *abnormal* GDP and inflation regimes. Our analysis also reveals that the *views* of Fed members (i.e., hawkish versus dovish) tend to play a marginal role in terms of the strength of the speech signals. Looking at the speeches by the Fed Chair, we find that the Chair signals produce a larger tail risk compared to non-Chair signals, and the estimated magnitude of the market responses depends on the position of the officials (i.e., the Fed Chair or other Fed member).

Keywords: Central Bank Communication, Multimodal Machine Learning, Natural Language Processing, Speech Analysis, High-Frequency Data, Volatility, Tail Risk.

JEL: E52, C45, C53, G12, G14

*We thank Andrea Ajello, Ilias Chronopoulos, Michael Ehrmann, Klodiana Istrefi, Galina Potjagailo, and participants at the ECONDAT 2023 conference for valuable comments and discussions. We are particularly grateful to Galina Potjagailo for excellent suggestions and discussion on the earlier version of this paper. Michael McMahon gratefully acknowledges financial support from the European Research Council (Consolidator Grant Agreement 819131). Maximilian Ahrens is thankful for support from the Oxford-Man Institute for Quantitative Finance. The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

1 Introduction

A large branch of monetary policy research seeks to explain how central bank communication (CBC) steers market dynamics and expectations (Blinder, 2018). Theory suggests that if central bank announcements and speeches convey information on economic and monetary conditions, market participants will update their beliefs as reflected in their portfolio choices. Central bank communication can thus contribute to revaluing assets and stabilizing market conditions by reducing uncertainty (Bernanke et al., 2005). Empirical research largely corroborates this theoretical prediction and establishes a consensus that central bank communication influences asset prices through its effects on market participants' expectations about economic outlook and policy decisions (Bernanke and Kuttner, 2005; Ramey, 2016). Monetary policy communication also appears to influence investors' risk aversion and hence the risk premium (Hanson and Stein, 2015; Cieslak and Schrimpf, 2019; Swanson, 2021).

Despite these findings, there are still at least two unresolved issues: (i) how to identify monetary policy news in central bank communication, and (ii) how to identify effects of such news on market uncertainty, i.e., volatility and tail risk. Official central bank announcement dates, such as those of FOMC announcements, occur rather infrequently (every 6-8 weeks). However, policy makers and researchers have suggested that markets continually revise their understanding of central bank information as policy makers give speeches (Neuhierl and Weber, 2019). Although recent developments in natural language processing (NLP) have allowed economists to analyse text with machine learning methods (see e.g., Bholat et al., 2015; Hansen et al., 2018; Ahrens and McMahon, 2021), researchers have paid only limited attention to speeches so far[1], partly because their content is difficult to quantify and the field still lacks easily accessible datasets of central bank speeches.

In this paper, we develop a novel multimodal NLP method to identify macroeconomic news in central bank speeches and we assess their impact on market volatility and tail risk. To the best of our knowledge, we are the first to do so. Some earlier research has focused on how central bank communication affects volatility in financial markets (see e.g., Bekaert et al., 2013; Cieslak and Schrimpf, 2019; Ehrmann and Talmi, 2020; Gómez-Cram and Grotteria, 2022), while only Hattori et al. (2016) has studied tail risk.[2] Moreover, there is an extensive literature that studies the effects of central bank communication about the economic outlook on asset price surprises. Signals about the economic situation can have a multitude of different effects. The classic channel as emphasised in, for example, Romer and Romer (2000) and Nakamura and Steinsson (2018), is an information effect. The central bank, either explicitly or implicitly through its policy decision, releases superior information about the economy and this information is then incorporated in updated private sector forecasts. An alternative channel is one in which the central bank's information is not considered superior; releasing an alternative assessment of the state of the economy,

[1] Recently, Neuhierl and Weber (2019) have investigated the tone of speeches by central bank chairs and vice-chairs while Petropoulos and Siakoulis (2021) use a mixture of machine learning and dictionary methods to calculate sentiment indices from central bank speeches. The latter authors argue that this sentiment predicts financial turmoil. Swanson (2023) highlights the importance of Fed Chair speeches using an event-study surprise decomposition, and Cieslak and McMahon (2023) focus on the communication of Fed stance and its effects on the risk premium.

[2] We focus on measuring market uncertainty rather than uncertainty about monetary policy (see e.g., Bauer et al., 2022; Husted et al., 2020; Ozdagli and Velikov, 2020; Tillmann, 2020), or uncertainty of monetary policy makers Cieslak et al. (2023).

that the market do not believe, could heighten concerns about the possibility of a monetary policy mistake which would make the economy more volatile (Caballero and Simsek, 2022; Cieslak and McMahon, 2023). The central bank may communicate, as part of its outlook, their view of uncertainty which can influence private views about uncertainty (Hansen et al., 2019). Finally, a cacophony of economic assessments, even if just reflecting different views on the outlook for the economy, might itself signal greater uncertainty surrounding the outlook which can increase the uncertainty of market participants about the economic and the policy outlook (Ahrens and McMahon, 2021).

Our methodological framework has two parts. First, we use machine learning methods from the field of multimodal natural language processing to infer implied macroeconomic forecast revisions from Fed officials' public speeches. Our training dataset consists of Greenbook texts and their respective forecasts, which allows us to learn a mapping from central bank language to central bank forecasts (see Ahrens and McMahon, 2021). In our test dataset, we then apply the learned mapping to central bank speeches to infer how news signals in speeches can predict revisions of public macroeconomic forecasts. Second, we investigate the high-frequency (intradaily) responses of market volatility and tail risk to speech-implied revisions in CPI, GDP, and unemployment outlooks.[3]

Our paper contributes to the literature in several ways. Most importantly, we show that central bankers' speeches have a statistically significant impact on volatility and tail risk in financial markets. In order to show this, we develop a new, multimodal methodological framework for identifying monetary policy news about GDP growth, CPI, and unemployment outlooks. We compare and contrast the performance of an extensive array of modern machine learning methods for multimodal NLP on our empirical datasets of Greenbook texts and forecasts as well as on FOMC members' speeches. We show that our speech-implied forecast revisions predict future changes in Survey of Professional Forecasters (SPF) forecasts substantially better than models that use purely tabular data and ignore the textual content of the speeches. It is these speech-implied macroeconomic news signals that explain a sizeable part of realized volatility and tail risk in financial markets. Furthermore, our findings suggest that markets 'listen' or react more strongly to news in central bank speeches during *abnormal* GDP and inflation regimes. In order to contribute to future examinations of Federal Reserve speeches, we make our comprehensive dataset on Federal Reserve speeches accessible to other researchers.

The remainder of the paper is organized as follows. In the next section, we review the related literature. Section 3 describes the data and section 4 introduces our methodological framework. In section 5 and 6, we present the empirical results pertaining to our analyses of speech-implied news and high-frequency market responses. Section 8 concludes the paper.

[3]High-frequency market analysis is common in monetary research; see, for example, Gurkaynak et al. (2005); Gertler and Karadi (2015); Nakamura and Steinsson (2018); Jarociński and Karadi (2020) and Miranda-Agrippino and Ricco (2021).

2 Related Literature

Central Bank Communication Effects on Market Volatility and Tail Risk

Our paper is most closely related to studies of the high-frequency effects of CBC on market uncertainty and volatility. Cieslak and Schrimpf (2019) study the high-frequency effects of the non-monetary news component of communication on volatility. Leombroni et al. (2021) explore how CBC influences credit risk premia through high-frequency changes in yield curve. Ehrmann and Talmi (2020) measure textual differences between central bank announcements and find that higher levels of textual similarity to the previous announcement statement are usually associated with lower market volatility after the announcement date. Relying on a one-day event window, Hansen et al. (2019) analyse the Bank of England's Inflation Reports via topic modelling and find that communication of uncertainty plays an important role in shaping long-run interest rates. Bekaert et al. (2013) find evidence that looser policy reduces risk aversion and uncertainty. Gómez-Cram and Grotteria (2022) explore the price discovery process for several asset classes on FOMC announcement days. Bauer et al. (2022) develop a policy uncertainty measure based on financial derivatives, and show that FOMC (uncertainty cycle) announcements reduce uncertainty. Finally, Hattori et al. (2016) study the impact of Unconventional Monetary Policy (UMP) on stock market and bond market tail risk. UMP increases (decreases) the realized volatility of stocks (bonds), but lowers the tail risk in both markets. Forward guidance (and hence communication) appears to have stronger "dampening effects", compared to other UMP events.

We extend this line of research in two ways. First, these aforementioned studies often overlook extreme market responses when assessing the effects of news. For example, the main result of Hattori et al. (2016) that UMP decreases the tail risk in stock and bond markets does not appear to hold when we move outside the cycles of FOMC press releases. Unlike Hattori et al. (2016), we focus on the intraday market responses to speeches, which can occur at any time, rather than only the times of FOMC announcements, and measure the *realized* tail risk instead of the *implied* tail risk from derivatives. In contrast with Hattori et al. (2016), we find that speeches *increase* realized tail risk. This type of CBC does not appear to reduce uncertainty and calm financial markets.[4]

Second, prior research on monetary policy news has commonly employed jump-diffusion models with Poisson jumps to capture responses to news. The approach of Bauer et al. (2022) relies on such a representation for "FOMC jumps". Despite its simplicity, these jump models are not compatible with the stylized facts of jump occurrences, as news-induced tail responses are persistent in the presence of heterogeneous investors interpreting the content of speeches. Consequently, these studies underestimate the realized tail risk. Departing from this conventional approach, we consider a more flexible model that allows for *time-varying* tails. This allows us to separate extreme volatility responses from the tail responses and,

[4] In the context of forward guidance, Ehrmann et al. (2019) put forward a model where forward guidance can amplify the reaction of expectations to macroeconomic news. Empirically, they show that the type and horizon of forward guidance—time-contingent, state-contingent, open-ended, short or long horizon—influences the sensitivity of bond yields to news and degree of disagreement among forecasters. For example, while long-horizon forward guidance reduces interest rate sensitivity to macroeconomic news, short-horizon guidance amplifies it. Similarly, state-contingent forward guidance limits bond price responses to macro news but open-ended forward guidance essentially has no statistically significant effect on the response.

more importantly, to identify the speeches that create *tail cascades*. Unlike the previous studies treating jumps as one-shot events, we accommodate the stochastic intensity of jumps that potentially occurs from heterogeneous interpretation of news by market participants. Our high-frequency event study approach is hence more flexible methodologically and better captures the dynamics of intradaily volatility and tail risk.

Regime Dependence of Monetary Policy Effects

Both theory and data suggest that monetary policy is regime dependent. Mandler (2012) uses a threshold vector autoregression (VAR) framework to analyse the effectiveness of classical monetary policy shocks, depending on the respective inflationary regime in the US economy between 1965-2007. He finds that monetary policy shocks have markedly different effects in low and high inflation regimes. Such inflation regime differences can be theoretically motivated. Sizeable deviations from inflation target levels might affect a central bank's credibility and its ability to credibly signal. Similarly, substantial off-target inflation levels might affect private sector inflation expectations, altering the Philips curve and inflation dynamics (Mandler, 2012).

Tenreyro and Thwaites (2016) examine GDP regime dependence of monetary policy shock effects, derived from the unexpected component of interest rate changes. The empirical results of Tenreyro and Thwaites suggest that medium- to long-run monetary policy shock effects on the real economy strongly depend on the state of the business cycle. GDP growth is the most consistent factor determining monetary policy effectiveness, and shocks seem to have a more pronounced effect during economic upswings than during downswings.[5] They also find that contractionary shocks have greater impact than expansionary ones, with both being equally represented during recessions and booms. Desired effects of policy rate changes might be subdued during recessions and central bankers might rely more strongly on unconventional monetary policy near the effective lower bound (ELB). To the best of our knowledge, we are the first to investigate regime dependence — with regards to both inflation and GDP growth — of the effectiveness of unconventional monetary policy and central bank communication.

Text Analysis for Monetary Policy

Lastly, we are part of a burgeoning literature that uses natural language processing to analyse monetary policy. Various text analysis methods have been tested in this field. For example, researchers have used topic models (Hansen et al., 2019), combined dictionary methods with classic machine learning models such as XGBoost (Petropoulos and Siakoulis, 2021), and have deployed deep neural network models such as transformers (Cai et al., 2021). In our work, instead of choosing a specific NLP algorithm a priori, we decide to take a more model-agnostic, data-driven approach to reduce modeler bias. That is, we train a variety of NLP models and choose the algorithm that works best in our validation set.

Similarly, researchers have employed various frameworks and datasets to identify monetary policy news. In particular, researchers have often studied the market effects of central bank policy announcements. For

[5] Tenreyro and Thwaites (2016) further emphasize the historical evidence that fiscal policy measures have been more important in times of recession, while fiscal and monetary policy have historically reinforced one another during booms.

instance, Lucca and Trebbi (2009) and Hansen and McMahon (2016) both leverage approaches from computational linguistics within a VAR framework to asses the effect of the content in FOMC statements on macroeconomic variables. Lucca and Trebbi (2009) find CBC to be a more important factor than contemporaneous policy rate decisions. Hansen and McMahon (2016) conclude that shocks to forward guidance have a stronger effects on markets than communication of current economic conditions. Handlan (2020) uses a deep neural network architecture to identify text-based shocks in FOMC announcements, assessing their impact on Fed funds futures. She finds that shocks derived from forward guidance wording of FOMC statements account for four times more variation in Fed funds future prices than direct announcements of changes in the target federal funds rate. Gómez-Cram and Grotteria (2022) apply a video analysis on words mentioned during central bank press conference videos. Nesbit (2020) proposes a word count based instrumental variable framework to identify monetary policy shocks in FOMC transcripts. Aruoba and Drechsel (2022) use NLP techniques to analyse FOMC meetings in order to measure the information set of the FOMC at the time of policy decisions. They then use these measures to generate estimates of FOMC monetary policy shocks.

Although each of these studies use different methods, they all utilise text to help us to identify effects of monetary policy. However, official central bank announcements, such as FOMC announcements, occur only infrequently (every 6-8 weeks). We therefore shift our focus on central bankers' speeches which happen in much higher frequency. Researchers have paid only limited attention to speeches, partly because their content is difficult to quantify. At the same time, central bank deliberation and communication is continuous (Neuhierl and Weber, 2019). Thus, it is important to frequently measure CBC effects.

A few notable papers move in this direction. Neuhierl and Weber (2019) find that the tone of US Fed chair and vice-chair speeches, measured via word count methods, can explain stock market price dynamics. Using a mixture of machine learning and dictionary methods, Petropoulos and Siakoulis (2021) derive sentiment indices from central bank speeches and find that the sentiment predicts financial turmoil. We use a two-step macroeconomic news identification framework, in which we first learn a mapping from central bank language to central bank forecasts with Greenbook data, and then infer how FOMC member speeches imply revisions to GDP, inflation, and unemployment forecasts — an approach which is motivated by Ahrens and McMahon (2021).[6]

To identify the news content of a speech, we must control for market expectations. Ellen et al. (2022), for example, construct a monetary news series from the difference in narrative between central bank statements and news media coverage. The results of Ellen et al. (2022) highlight the pivotal role of news media as catalysts in the process of forming market expectations and confirm earlier findings in the literature that monetary policy shocks cause measurable macroeconomic responses. Similarly, Cai et al. (2021) analyse FOMC announcements using BERT (Devlin et al., 2019) and identify monetary policy and information shocks, controlling for market expectations by analysing relevant New York Times articles with NLP methods. Instead of inferring market expectations from noisy news media coverage, we take

[6] See also Gáti and Handlan (2022), who use regularized regressions to map the wording of FOMC statements to Greenbook forecasts of output growth, unemployment and the federal funds rate. They argue that the statement wording implies FOMC expectations fairly well, with the exception of short-run inflation expectations, although these patterns have changed over time with Fed Chairs. In addition, disagreement about the Fed's communication rule causes beliefs to diverge.

the latest forecast measures from the widely viewed Survey of Professional Forecasters (SPF) conducted by the Federal Reserve Bank of Philadelphia. SPF forecasts directly measure expected GDP growth, inflation, and unemployment. We then define a macroeconomic news shock as the difference between a speech-implied forecast revision and the most recent SPF forecast for that variable available at the time of the speech.

3 Federal Reserve and Markets Data

The data used in our paper consists of several types: FOMC member speeches, Greenbook text, Greenbook forecasts, SPF forecasts, and intraday volatility and tail risk measures of US stock and bond markets. We use Greenbook forecasts and the respective Greenbook text sections that describe them to map central bank language to central bank forecasts. We then apply our learned mapping to FOMC member speeches and assess how speech-implied forecast revisions affect volatility and tail risk in financial markets.

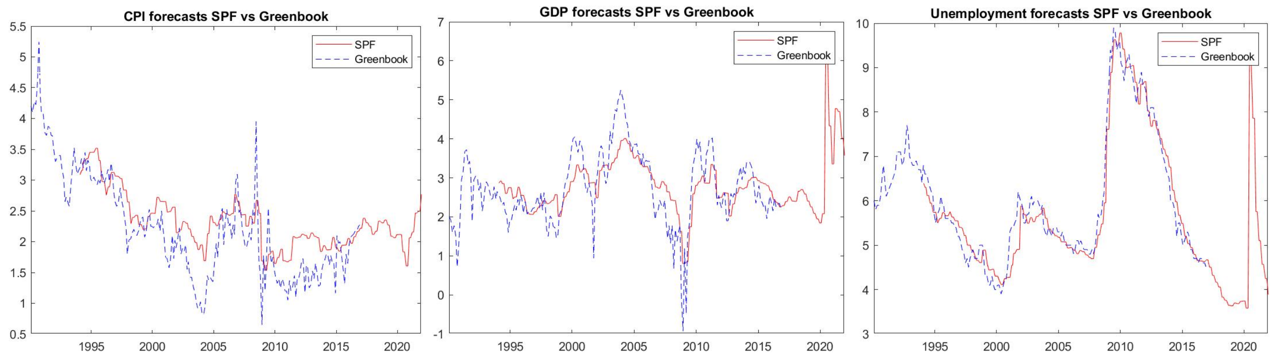
3.1 Federal Reserve Speech and Forecast Data

The central bank data is split into a training and a test set. We describe these datasets below.

Training set: In the training phase, we learn the mapping of the Fed's Greenbook texts associated with the descriptions of GDP growth, CPI, and unemployment outlooks to the change in the Greenbook forecasts of those variables from the previous forecast period. That is, we target the difference in a current period's one-quarter-ahead Greenbook forecast to the previous quarter's forecast, such that for any of our macroeconomic key figures of interest, y, we define $\Delta y_m = y_m - y_{m-1}$, where m indicates the date of the Greenbook forecast. We also tested a one-year-ahead horizon, although this was less informative as one-year forecasts tend to revert to long-run values. The training sample spans 145 Greenbook documents, from January 1, 1995 to December 31, 2013. We only consider the 8,155 Greenbook sections that directly relate to GDP growth, CPI, and unemployment (see Appendix D for a detailed list of section allocations). The average Greenbook section in our dataset has about $3,000$ words; the longest section consists of $31,000$ words and the shortest section contains around 140 words. At any date, we concatenate all Greenbook sections that relate to the same forecasting variable.

Test set: Training the NLP models consists of estimating complex mappings from Greenbook text on each date, for each variable, to the associated revisions to the one-quarter-ahead Greenbook forecasts on each date, for each variable. Once the models are trained, we apply the learned mappings to a test set consisting of FOMC members' speeches made from January 1, 2014 to December 31, 2021. The applied mappings imply one-quarter-ahead forecast revisions for GDP growth, CPI, and unemployment. We assume that central bankers' speeches convey news from the Fed's information set that can alter the economic outlook of private agents. The Fed's information set could contain private or superior information about economic conditions, superior or alternative analysis (as in Byrne et al., 2023), or new information about the Fed's own preferences for monetary policy.

Figure 1: Comparison of Greenbook and SPF forecasts



Notes: The figure displays the Greenbook and SPF forecasts over time for CPI (left panel), GDP (middle panel) and unemployment (right panel). SPF forecasts are the mean across SPF participants. The two forecasts match quite closely for the majority of the inspected time-series.

The target variables in the test set are the one-quarter-ahead respective changes in GDP growth, CPI, and unemployment in the SPF forecasts. The SPF is a publicly available and widely referenced source for economic forecasts. We use the mean SPF forecasts across SPF participants as our proxy for market expectations, rather than the next Greenbook forecasts, because Greenbook forecasts are released to the public with a 5-year delay. We expect that central bank speeches should have similar predictive power for Greenbook and SPF forecast revisions. Figure 1 corroborates the assumption that the SPF forecasts match the Greenbook forecasts quite well during 1993 to 2016. We assume that this pattern also holds post 2016, for which there was no public Greenbook data available when the data for this paper was collected. We release our dataset of central bank speeches, time-stamped on the minute of release, on our Github repository.[7]

3.2 High-Frequency Market Data

We use high-frequency transaction prices for 22 Dow Jones Industrial Average (DJIA) stocks, together with 2-year, 5-year, and 10-year U.S. Treasury note and bond futures traded on the Chicago Board of Trade (CBOT). Appendix E lists the individual stocks and bonds. Wharton Research Data Services (WRDS) and Tick Data LLC provide data for individual stocks and bond futures, respectively. As is standard in the literature, we exclude U.S. holidays, Christmas periods, and weekends from our sample. We only consider trading hours from 9:30 EST−16:00 EST and 7:30 CT−14:00 CT, for stock and bond markets, respectively. To reduce the potential impact of market microstructure noise, we filter out *bouncebacks* and irregular quotes that typically occur in ultra high-frequency data. Using our adjusted data, we create equally-spaced 15-second observations, which is an appropriate frequency to implement our response measures. Our sample runs from January 1, 2014 through December 31, 2021.

[7] github.com/MaximilianAhrens/data/tree/main/central_bank_speeches

4 Methodological Framework

Our methodological framework can be broken down into two parts. Section 4.1 explains our multimodal NLP framework used to estimate the mapping from central bank language to forecasts. We test and compare our estimation framework with a variety of machine learning algorithms. Section 4.2 then describes the measurements of the asset price dynamics and their relationship with the speech signals.

4.1 Multimodal NLP Framework

We seek to estimate how new information revealed in central bank speeches influences financial markets. To do so, we map central bank language to macroeconomic forecasts, controlling for the macroeconomic conditions at the time.

The macroeconomic conditionality is important because the effect of a given forecast revision on financial markets depends on initial economic conditions. This economic context requires the multimodal modelling approach. For example, a speech that raised forecast inflation would be a positive signal of improving conditions if inflation was below its desired level. However, the same speech would convey a negative signal if inflation was substantially above target. We employ multimodal machine learning approaches that allow us to use both text and tabular data when mapping central bank language to central bank forecasts and then predicting output, inflation, and unemployment outlook revisions.

4.1.1 Learning Mapping from Central Bank Language to Forecasts

We learn the mapping from the Fed's Greenbook text to the respective Greenbook forecasts. The Greenbooks contain dedicated sections on the Fed's forecasts of GDP growth, CPI, and unemployment, including the rationales for the forecasts. These sections allow us to map the Greenbook text - ergo central bank language - to central bank forecasts.

In the training phase, we estimate a separate mapping for each of the three variables, i.e., the one-quarter-ahead forecast change in CPI, GDP growth, or unemployment. We measure the change from the previous $(m-1)$ Greenbook to the current (m) in the one-quarter-ahead forecasts (q_1). CPI is denoted by π, GDP growth by g, and unemployment by u. Hence, our three target variables are: $\Delta\pi_{q_1,m}$, $\Delta g_{q_1,m}$, and $\Delta u_{q_1,m}$. For ease of notation in the following equations of our modelling framework, let y serve as a placeholder variable for any of the CPI, GDP growth, and unemployment variables. Hence, we denote our placeholder target variable as $\Delta y_{q_1,m}$.

To capture the economic context, we control for both change and level of the CPI, GDP, and unemployment of the previous Greenbook report, denoted as X_{m-1}. We fit a function, f, to learn how the respective Greenbook text maps into forecasts, controlling for macroeconomic conditions. The equations for CPI, GDP growth, and unemployment have the same explanatory variables, except for the text input, which is specific to the respective Greenbook forecast section. That is, θ_π represents the text features for the CPI corpus, while θ_g represents GDP-related text, and θ_u unemployment-related text. We use θ_y as a placeholder for any of the three text inputs. With this notation, $\theta_{y,k}$ represents the k^{th} text feature for the respective target variable y. Let us define f as the function that takes text and tabular data as inputs and maps them to the target output y, given parameters Ω, which are to be learned. We can now write out our regression equation as

$$\Delta y_{q_1,m} = f\left(X_{m-1}, \theta_{y_m}; \Omega\right). \tag{1}$$

If we assume linearity in function f, the regression equation can be written as follows:

$$\begin{aligned}\Delta y_{q_1,m} = {} & \omega_\pi \pi_{q_1,m-1} + \omega_g g_{q_1,m-1} + \omega_u u_{q_1,m-1} \\ & + \omega_{\Delta u} \Delta u_{q_1,m-1} + \omega_{\Delta \pi} \Delta \pi_{q_1,m-1} + \omega_{\Delta g} \Delta g_{q_1,m-1} \\ & + \sum_{k=1}^{K} \omega_k \theta_{y,k,m} + \epsilon_m. \end{aligned} \tag{2}$$

Here, the ωs represent the regression parameters and ϵ is the measurement error. We use the first 80% of the Greenbook dataset for training and the remaining 20% for validation. The data is furthermore de-meaned and standardized based on training set values. We did not randomly split the training and validation set to acknowledge the time-series characteristics (and therefore the potential for information leakage) in the data. We then train the machine learning models to map central bank texts and control variables to the respective target variables. We treat this as a regression problem and use a least squares error loss function, commonly used in economics and monetary policy econometrics.

4.1.2 Identifying Information Signals in Central Bank Speeches

In the test phase, we apply the trained models for each of the macroeconomic variables (CPI, GDP growth, unemployment) to the central bank speeches to infer macroeconomic forecast revisions. The text data is now the central bank speech content. The tabular data points on current economic conditions are the most recent SPF forecast levels and changes on GDP growth, CPI, and unemployment.[8] This procedure maps each central bank speech into an implied revision of the forecasts for CPI, GDP growth, and unemployment.

4.1.3 Calculating News Signals

Markets should only react to relevant news that have not yet been incorporated into asset prices. If a central bank speech does not change the expected macroeconomic path, then the speech has no news component. We proxy market expectations with the latest public SPF forecast for each target variable. We then calculate the difference between the most recent SPF forecast change ($\Delta y_{SPF,s}$) available at the time of each speech and the implied forecast change in each speech ($\Delta \hat{y}_{\text{speech},s}$). This difference is our forecast revision news, ν, for target variable, y, and speech event, s, such that

$$\nu_{y,s} = \Delta y_{\text{SPF},s} - \Delta \hat{y}_{\text{speech},s}. \tag{3}$$

[8] As previously shown in Figure 1, the SPF forecasts track the Greenbook forecasts quite closely.

For GDP, a positive difference, $\nu_{y,s}$, is bad news, because a positive value means that the central bank speech implies lower GDP growth than does the most recent SPF forecast. The opposite is true for unemployment. Here, a positive difference is good news, as the speech implies that the central bank expects unemployment rates to fall faster (or rise less quickly) than previously anticipated.

For CPI, the categorisation into good and bad news depends on the relation of the current inflation level to the target. The Fed aims for an inflation rate of around 2%, as do most central banks of advanced economies.[9] Therefore, a positive $\nu_{\pi,s}$ — i.e., an implied downward forecast revision — is good news when the forecast of inflation is above target. This means inflation will revert faster back to target than anticipated (or won't rise as fast as anticipated). Conversely, when forecast of inflation rate is below target, a negative $\nu_{\pi,s}$ is good news. A later analysis will assess how financial market volatility and tail risk react to these implied forecast-revisions.

4.1.4 Machine Learning Methods

We do not know, a priori, which statistical learning model would best approximates the function, f, in equation (1). We have relatively few data points compared to many machine learning projects (e.g. hundreds or thousands rather than millions or billions of data points). Each data point itself is rich in information, however, consisting of a high dimensional feature set. That is, each set of text can be several thousand words long, which presents a problem for many modern language models such as transformer family models (e.g. BERT-based models), which can usually only handle up to around 100-1,000 tokens per data point (Das et al., 2021). Some extensions based on sparse transformers have been proposed such as Child et al. (2019); Zaheer et al. (2020), which can handle sequences of a couple of thousand tokens. However, document lengths of $20,000+$ words would still pose a challenge. Lacking reason to favour a specific class of models, we deploy a range of models, to search broadly for the best model and reduce the a priori modeler bias of favouring one model over alternatives.

We therefore deploy an extensive array of multimodal machine learning algorithms to approximate function f and to learn parameters Ω. We use the multimodal machine learning benchmark suite, AutoGluon (AutoGL) (Erickson et al., 2020), and we add to it the class of multimodal supervised topic models (Card et al., 2018; Ahrens et al., 2021).

AutoGluon

AutoGL is an automated machine learning (AutoML) framework that has been developed to fuse multimodal features such as text, images, and tabular data. We chose this AutoML framework because it outperformed competing frameworks in multimodal benchmark tasks (see Erickson et al., 2020).

Base models: AutoGL fits machine learning *base models* and then combines them through ensembling and stacking to boost performance. AutoGL allows us to apply hyperparameter optimization over all models. The *base models* in AutoGL span the following broad machine learning algorithm classes:

[9] The FOMC targets a 2% rate of change for the personal consumption expenditure price index (PCE), not the CPI. The two inflation rates are very highly correlated, however, which makes it reasonable to use information about implied CPI forecasts to proxy for PCE forecasts.

1. **K-nearest neighbours** (Dudani, 1976): AutoGL uses two variations of k-nearest neighbours (KNN) that differ in their weighting approaches. One allocates uniform weights to all points while the other weights points according to the inverse of their respective distances.

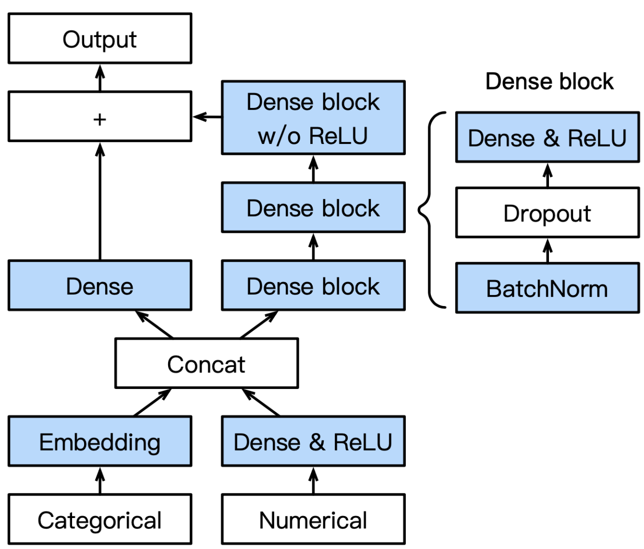
2. **Random forests** (Breiman, 2001): AutoGL again deploys two variations of this algorithm class. One option uses the information gain of nodes for the assessment of the split quality. The other option uses Gini impurity instead.

3. **Extremely randomized trees** (Geurts et al., 2006): For the random tree class, AutoGL deploys both an implementation resorting to information gain and another option that uses Gini impurity for the assessment of split quality.

4. **Boosted decision trees**: AutoGL runs (where applicable to the task) Extreme Gradient Boosting (Chen and Guestrin, 2016), Light Gradient Boosting (Ke et al., 2017), Categorical Boosting (Prokhorenkova et al., 2018).

5. **Neural networks**: Figure 2 schematically outlines AutoGL's neural network architecture, which Erickson et al. (2020) details. The architecture has been specifically designed for the multimodal use of categorical (text, images) and numerical data. It uses variable-specific embeddings for each of the categorical features. These are then concatenated with the numerical features into one overall input vector. This vector is in turn fed through a 3-layer feed-forward network as well as through a linear skip-connection (for details see Erickson et al., 2020). Model ensembling and stacking can be applied and are optimally chosen in the validation process.

Figure 2: AutoGL schematic neural network architecture



Notes: The figure displays the AutoGluon schematic neural network architecture, based on the design by Erickson et al. (2020), p. 3. Layers with learnable parameters coloured in blue.

Text representation options: We must also choose how to represent the text in machine-readable format. We define the following approaches:

1. **AutoTab**: Only tabular features are used. Text is excluded. AutoTab is our tabular data baseline next to an OLS regression that only uses tabular data.[10]

2. **AutoTab + tfidf**: Use tf-idf weighted word counts of the text as features. Standard text cleaning procedures of removing stopwords and punctuation have been applied.

3. **AutoTab + topics**: Use topic shares from supervised topic models as features (using rSCHOLAR without tabular data for the topic estimation).

4. **AutoMM transformer**: Use the AutoGL's multimodal modelling infrastructure that is based on a large language model (we use Roberta-base (Liu et al., 2019)) for multimodal fine-tuning. Tabular data can be fused into this process as well.[11]

5. **AutoTab + embed**: Use AutoMM transformer as well as AutoTab models that featurize text data as n-grams and ensemble over this zoo of models.[12]

4.2 Asset Price Dynamics

4.2.1 Underlying Continuous-Time Model

We model the intraday behaviour of asset prices with the following continuous-time model: The log-price X of each asset (stock or bond) follows an Itô semimartingale defined on a filtered space (Ω, $\mathcal{F}_t$, $(\mathcal{F}_t)_{t\in[0,T]}$, $\mathbb{P}$) over an interval $[0,T]$. The Grigelionis decomposition (see e.g., Erdemlioglu and Yang, 2022; Boswijk et al., 2018; Dungey et al., 2018) implies that X_t has the following specification:

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \delta * (\mu_t - \psi_t) + (\delta - h(\delta)) * \mu_t, \quad (4)$$

where b_s is the drift term, σ_s is the stochastic volatility component, W is a standard Brownian motion, δ is a predictable function, h is a truncation function (e.g., $h(x) = x1_{\{\|X\|\leq 1\}}$), μ is the jump measure of X, and ψ is its jump compensator, which adopts the decomposition

$$\psi_t(dt, dx) = [f_t(x)\lambda_t dx]dt$$

where the function, $f_t(x)$, controls the jump size distribution and λ_t denotes the jump intensity as in Erdemlioglu and Yang (2022) and Boswijk et al. (2018). We focus on the *tail* component of this jump

[10] AutoGL's *TabularPredictor* approach.
[11] AutoGL's *MultimodalPredictor* approach.
[12] AutoGL's *TabularPredictor* approach with the *hyperparameter* option being set to *multimodal*.

compensator or λ_t, which captures the jump intensity dynamics.[13] We can specify λ_t as

$$\lambda_t = \lambda_0 + \int_0^t b'_s ds + \int_0^t \sigma'_s dW_s + \int_0^t \sigma''_s dB_s + \delta' * \mu_t + \delta'' * \mu_t^{\perp}, \tag{5}$$

where B is a standard Brownian motion independent of W, $\mu_t^{\perp}$ is orthogonal to μ_t, and δ', δ'' are predictable. This model, given by equations (4) and (5), satisfies no-arbitrage conditions and leaves the volatility and jump components unrestricted. We now present our volatility and tail risk measures from this model.

4.2.2 High-Frequency Measurement of Volatility and Tail Risk

Given the price dynamics in equations (4) and (5), let us define the ith intradaily return on a trading day as $r_{i,t} = X_{i,t}$ - $X_{i-1,t}$. We can write the daily realized volatility (RV) as the square root of realized variance, which is the sum of the squared intraday returns $(1, \ldots, M)$. That is,

$$RV = \sqrt{\sum_{i=1}^{M} r_i^2}. \tag{6}$$

It is well-known that realized variance converges to quadratic variation (see e.g., Andersen et al., 2003, 2001 and Barndorff-Nielsen and Shephard, 2002 for in-depth discussion).

Turning to the estimation of $\lambda_{i,t}$ in equation (5), we define the post-signal realized intensity (RI) measure as

$$RI = \frac{\Delta_n^{\varpi \widehat{\beta}_i}}{k_n \Delta} \sum_{j=1}^{k_n} g\left(\frac{|r_i|}{\alpha \Delta^{\varpi}}\right) \frac{\alpha^{\widehat{\beta}}}{C_{\widehat{\beta}_i}(k_n)}, \tag{7}$$

where Δ is incremental change between observations, $\alpha \Delta^{\varpi}$ is threshold to retain only large jumps, $g(\cdot)$ admits a specific functional form, k_n is a constant which admits $(1/K \leq k_n \Delta^{\rho} \leq K)$ for $(0 < \rho < 1)$ and $(0 < K < \infty)$, and β_i is the estimator of jump activity index that controls the vibrancy of sharp fluctuations. In equation (7), $g(\cdot)$ as an auxiliary function that separates jump-type movements from the diffusive volatility, based on an α deviation (e.g., α = 2, 3, 6) from the continuous component of the model.[14] We use RI as a proxy for time-varying (high-frequency) *tail risk* (TR), which is considerably accurate at high frequency, similar to the measures adapted in Bollerslev et al. (2015).[15]

In our context, using our tail risk measure RI (equation (7)) has several advantages. First, RI captures the tail of intradaily return distributions. We compute this quantity to estimate the tail behaviour of returns within a window after the speeches. Second, measurement of return tails in continuous-time is a non-trivial task because the tail-type return movements can also be attributed high-frequency volatility

[13] See Andersen et al. (2020), who exploit jump intensity process to measure tail risk and assess its equity premium implications.

[14] See e.g., Erdemlioglu and Yang (2022), Boswijk et al. (2018) and Dungey et al. (2018) for implementation details, particularly on the selection of the functional form for $C_{\widehat{\beta}_i}(k_n)$ in (7).

[15] Our tail risk indicator RI is also quite similar to the estimator of Hill (1975). See also Aït-Sahalia and Jacod (2009) for a related discussion on the role of β_i in (7).

(such as realized volatility). This challenge leads to an econometric identification problem, as realized volatility movements and realized tail movements potentially mingle with each other at high frequency. Consequently, it becomes a tedious task to separate different response forms (i.e., volatility versus tails). We use RI to measure tail responses accurately and disentangle them from volatility responses. Third, RI accounts for time-varying volatility, clustering in extreme price changes (jump clustering) and accommodates tail (jump) activity of the price variation around speeches. It does not require strong assumptions about the underlying asset pricing process and it is relatively easy to implement (see Appendix B.1 for the estimation steps). While large values of RI computed for a given window and speech indicate that the returns generate heavy tails, small RI values show weak evidence for tail behavior.[16]

In summary, we quantify two types of responses to CBC. First, communication likely creates sudden surges in market volatility. We assess these surges with realized volatility. Second, CBC can cause asset price jumps and persistently elevated jump intensity. Our approach allows us to first detect the speech-implied jumps, and then assess the 'intensity' of the jump responses. As Bollerslev et al. (2018) document, heterogeneous investors often release private information as they trade in the wake of such jumps, creating large price moves, which amplify high-frequency TR.[17]

4.2.3 Identifying Association Between News and Market Reactions

The final step in our methodological framework is to measure how realized volatility and tail risk in both equity and bond markets react to central bankers' speeches. To this end, we regress the market reactions on the forecast revision implied by the corresponding speech. As the forecast revision itself is a linear combination of the central bank signal and the latest public forecast, we already control for the partial correlation between the SPF forecasts and the market reactions.[18] The same holds true for all control variables used in the creation of the speech signals. We do not include additional low-frequency macroeconomic control variables because market prices should already incorporate such publicly available information.

5 Results: Language Mapping and SPF Prediction

The first step of our method is to learn the mapping from central bank language to central bank forecasts. We train our model on the first 80% of the Greenbook sample, holding out the last 20% of observations for validation. In our validation set, we assess how well a model can map Greenbook language to Greenbook forecasts. For each machine-learning class, we select the best performing model from the validation set and then assess its performance on the test set. The test sample is the post-2013 sample of speeches in which we assess how well the speech signals predict subsequent changes in SPF forecasts. Given the results in the Tables 1, 2, and 3, we have reason to believe that the identified signals in the central bank speeches

[16] It is perhaps worth emphasizing that the term *intensity* in RI refers to the stochastic intensity of the jump process. While RV in equation (6) is an estimator for the stochastic volatility, RI is an estimator for the stochastic intensity.

[17] We aggregate the information in measures by equally weighting the stocks in the panel. We apply the measures to all stocks, obtain the estimates of response measures, equally weight and use the cross-sectional average for a given speech.

[18] See e.g., Frisch and Waugh (1933) and Lovell (1963) for Frisch-Waugh-Lovel theorem.

carry relevant information to change market expectations and hence public macroeconomic forecasts. The tables report the R^2 associated with predictions of SPF forecast revisions.

For example, the second row of Table 1 indicates that the multimodal neural topic model (MM NTM non-linear) has an R^2 of 0.67 in predicting CPI forecast revisions in the Greenbook training set, 0.83 in the Greenbook validation set, and 0.735 in the test set (speeches). Appendix F presents all tested machine learning approaches.[19]

For each of the three macroeconomic target variables, the best multimodal NLP models markedly outperform models that only use tabular data. Specifically, the multimodal neural topic model (MM NTM) class performs best both in the validation and in the test set. For CPI, Table 1 shows that the MM NTM (non-linear) model has an R^2 of 0.735 in the test set, which is 15% better than MM NTM (linear) and 44% better than the R^2 of the next best method. Likewise, Table 2 shows that MM NTM (non-linear) has an R^2 of 0.797 in the test set, which is right behind MM NTM (linear)'s R^2 of 0.825. Finally, Table 3 shows that MM NTM (non-linear) performs best again for unemployment, with an R^2 of 0.208, which is markedly better than the second best R^2 of 0.131, achieved by AutoTab.

Interestingly, AutoGL's models underperform an OLS regression for CPI inflation and GDP growth. There might be several explanations for this underperformance. The datasets at hand contain relatively few data points — a common challenge in macroeconomics and macro-finance, especially for 'data hungry' machine learning methods. AutoGL's machine learning models might therefore struggle to converge or might easily overfit on the limited training data. Second, macroeconomic forecasts (or the revisions to them) might be well approximated by a linear model, since such models are a very common design choice in monetary economics, macroeconomics, and macroeconometrics. Hence, perhaps the relatively strong performance of an OLS regression compared to the AutoGL models.

Table 1: Central bank language to forecast mapping - CPI Q1

Metric: R^2	**train (GB)**	**val (GB)**	**test (speeches)**
OLS	0.288		0.510
MM NTM (linear)	0.600	0.650	0.640
MM NTM (non-linear)	0.670	**0.830**	**0.735**
AutoTab	0.565	0.302	0.475
AutoTab + tfidf	0.953	0.305	0.299
AutoTab + topics	0.370	0.284	0.358
AutoTab + embed	0.573	0.139	0.132
AutoMM transformer	-0.155	-†	-0.292

Notes: The table reports R^2 for training, validation, and test sets for each of the models. Best performing model in validation and test set in bold. †: Model only reports MSE for validation set.

[19] Of course, it is worth highlighting that Greenbooks (speeches) are written (given) by staff and FOMC members. In fact, it is rather plausible to think that they are not directed at the same audiences. Nevertheless, this type of feature does not necessarily imply that the mappings from those two types of text have significantly different mappings to forecasts. The R^2 values that we obtain from the test data confirm that the mappings must be indeed similar. Moreover, one may also expect Greenbook text and speeches to have significant commonality. This is mainly because the economic topics are similar or identical, and the two types of text use the same types of vocabulary and even phrases.

Table 2: Central bank language to forecast mapping - GDP Q1

Metric: R^2	**train (GB)**	**val (GB)**	**test (speeches)**
OLS	0.301		0.785
MM NTM (linear)	0.372	**0.426**	**0.825**
MM NTM (non-linear)	0.483	0.371	0.797
AutoTab	0.497	0.304	0.380
AutoTab + tfidf	0.752	0.240	0.268
AutoTab + topics	0.730	0.253	0.285
AutoTab + embed	0.587	0.220	0.142
AutoMM transformer	0.013	-†	-0.044

Notes: The table reports R^2 for training, validation, and test sets for each of the models. Best performing model in validation and test set in bold. †: Model only reports MSE for validation set.

Table 3: Central bank language to forecast mapping - unemployment Q1

Metric: R^2	**train (GB)**	**val (GB)**	**test (speeches)**
OLS	0.231		-0.377
MM NTM (linear)	0.197	0.109	0.066
MM NTM (non-linear)	0.285	**0.457**	**0.208**
AutoTab	0.191	0.058	0.131
AutoTab + tfidf	0.577	0.113	-0.045
AutoTab + topics	0.278	0.053	-0.010
AutoTab + embed	0.415	0.145	-0.044
AutoMM transformer	-0.737	-†	-1.177

Notes: The table reports R^2 for training, validation, and test sets for each of the models. Best performing model in validation and test set in bold. †: Model only reports MSE for validation set.

6 Results: Intraday Market Effects

We use the model that performed best in the validation set (Greenbook data) to estimate the speech-implied information on GDP, CPI, and unemployment forecast revisions in the test set (speech data). The news on forecast revisions, as outlined in section 4.1.3, are defined as the difference between the speech-implied forecast for CPI, GDP, and unemployment outlook and the respective most recent SPF forecast. We then fit an OLS regression where we use the speech-implied news as independent variables. Market volatility and tail risk are the respective dependent variables. We first show our estimation results across regimes in section 6.1. In section 6.2, we then segment our speech dataset into low, normal, and high GDP and CPI regimes, respectively. Section 6.3 shows the news effect analysis by CPI regime. Section 6.4 covers the same analysis by GDP regime.

6.1 News Effects Across Regimes

We use the estimated realized volatility (RV) and tail risk (TR) in the 30-minute window after a speech as as our dependent variables. We regress both RV and TR on all absolute speech-implied news across all regimes. That is, we expect larger forecast revision news (in absolute value) to raise volatility and tail risk. The data is de-meaned and standardized. For each speech s, denote its CPI news component as $\nu_{\pi,s}$, GDP news as $\nu_{g,s}$, and unemployment news as $\nu_{u,s}$. The regression equations for realized volatility and tail risk are then

$$RV_s = \beta_0|\nu_{\pi,s}| + \beta_1|\nu_{g,s}| + \beta_2|\nu_{u,s}| + \epsilon_{RV} \tag{8}$$

$$TR_s = \rho_0|\nu_{\pi,s}| + \rho_1|\nu_{g,s}| + \rho_2|\nu_{u,s}| + \epsilon_{TR}. \tag{9}$$

We estimate both equations for both equity and bond markets.

Equity Markets

The positive and statistically significant coefficients in the top panel of Table 4 reveal that larger absolute forecast revision news, i.e., larger absolute differences between the implied forecast and the most recent SPF forecast, are associated with higher realized equity volatility. All three types of forecast revisions are highly statistically significant at the 10% level. The bottom panel of Table 4 indicates that the magnitude of speech-implied forecast revisions to CPI and unemployment has a statistically significant association with higher tail risk in equity markets. GDP news have no statistically significant effect.

Table 4: Association between absolute speech-implied forecast revision news and volatility (top panel) and tail risk (bottom panel) in equity markets across all regimes

Target variable: RV_e	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	0.1675	0.022	7.585	0.000	0.124	0.211
|GDP news|	0.0780	0.043	1.800	0.072	-0.007	0.163
|U news|	0.1967	0.024	8.078	0.000	0.149	0.244
R^2: 0.722	Adj. R^2: 0.718	n. obs.: 191	Heteroscedasticity robust standard errors			
Target variable: TR_e	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	2.2613	0.483	4.677	0.000	1.314	3.209
|GDP news|	1.1819	0.990	1.193	0.233	-0.759	3.123
|U news|	2.4452	0.484	5.056	0.000	1.497	3.393
R^2: 0.526	Adj. R^2: 0.519	n. obs.: 191	Heteroscedasticity robust standard errors			

Notes: The table shows the association between speech-implied forecast revision news in absolute value about CPI, GDP, and unemployment and realized volatility (top panel) and tail risk (bottom panel). The estimation results are reported for the U.S. equity market.

Bond Markets

Tables 5, 6, and 7 show the results for the 2-, 5-, and 10-year bond futures markets. The bond market results are similar to those of the equity market. Larger absolute speech-implied forecast revision news are strongly associated with higher realized bond price volatility and tail risk across maturities.

Table 5: Association between absolute speech-implied forecast revision news and volatility (top panel) and tail risk (bottom panel) in bond markets (2-year maturity) across all regimes

Target variable: $RV_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	0.0149	0.003	5.643	0.000	0.010	0.020
|GDP news|	0.0110	0.005	2.121	0.034	0.001	0.021
|U news|	0.0166	0.003	5.412	0.000	0.011	0.023
R^2: 0.672	Adj. R^2: 0.667	n. obs.: 175	Heteroscedasticity robust standard errors			
Target variable: $TR_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	3.7368	0.809	4.619	0.000	2.151	5.322
|GDP news|	5.4056	1.022	5.288	0.000	3.402	7.409
|U news|	3.3025	0.887	3.725	0.000	1.565	5.040
R^2: 0.508	Adj. R^2: 0.500	n. obs.: 175	Heteroscedasticity robust standard errors			

Notes: The table shows the association between speech-implied forecast revision news in absolute value about CPI, GDP, and unemployment and realized volatility (top panel) and tail risk (bottom panel). The estimation results are reported for 2-year maturity U.S. Treasury bond futures.

Table 6: Association between absolute speech-implied forecast revision news and volatility (top panel) and tail risk (bottom panel) in bond markets (5-year maturity) across all regimes

Target variable: $RV_{b,5y}$	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	0.0298	0.006	4.866	0.000	0.018	0.042
|GDP news|	0.0238	0.013	1.852	0.064	-0.001	0.049
|U news|	0.0354	0.006	5.900	0.000	0.024	0.047
R^2: 0.592	Adj. R^2: 0.588	n. obs.: 175	Heteroscedasticity robust standard errors			
Target variable: $TR_{b,5y}$	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	2.3726	0.744	3.189	0.001	0.914	3.831
|GDP news|	3.6080	1.500	2.405	0.016	0.667	6.549
|U news|	1.4576	0.684	2.132	0.033	0.118	2.797
R^2: 0.424	Adj. R^2: 0.413	n. obs.: 175	Heteroscedasticity robust standard errors			

Notes: The table shows the association between speech-implied forecast revision news in absolute value about CPI, GDP, and unemployment and realized volatility (top panel) and tail risk (bottom panel). The estimation results are reported for 5-year maturity U.S. Treasury bond futures.

Table 7: Association between absolute speech-implied forecast revision news and volatility (top panel) and tail risk (bottom panel) in bond markets (10-year maturity) across all regimes

Target variable: $RV_{b,10y}$	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	0.0574	0.010	5.687	0.000	0.038	0.077
|GDP news|	0.0443	0.021	2.132	0.033	0.004	0.085
|U news|	0.0614	0.010	6.000	0.000	0.041	0.082
R^2: 0.650	Adj. R^2: 0.644	n. obs.: 175	Heteroscedasticity robust standard errors			
Target variable: $TR_{b,10y}$	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	1.8245	0.644	2.833	0.005	0.562	3.087
|GDP news|	3.0200	1.413	2.137	0.033	0.250	5.790
|U news|	1.3404	0.555	2.414	0.016	0.252	2.429
R^2: 0.434	Adj. R^2: 0.424	n. obs.: 175	Heteroscedasticity robust standard errors			

Notes: The table shows the association between speech-implied forecast revision news in absolute value about CPI, GDP, and unemployment and realized volatility (top panel) and tail risk (bottom panel). The estimation results are reported for 10-year maturity U.S. Treasury bond futures.

6.2 Economic Regime Definitions

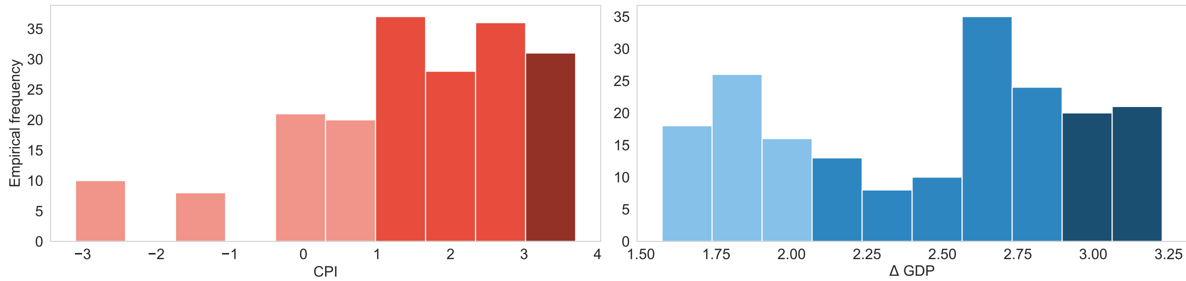
We also assess whether the effects of speech-implied forecast revisions depend on the GDP and inflation regimes. We do not separately analyse unemployment regimes. We divide our GDP and CPI datasets into a *high*, *normal*, and *low* regime (see Table 8). The categorisation is based on the Federal Reserve's inflation target and the historic distributions of the respective variables as depicted in Figure 3. Figure 4 shows the two time-series of the regime indicators.

Table 8: Categories of economic regimes

	CPI	**ΔGDP**
High	$\pi > 3\%$	$g > 3\%$
Normal	$1\% < \pi < 3\%$	$2\% < g < 3\%$
Low	$\pi < 1\%$	$g < 2\%$

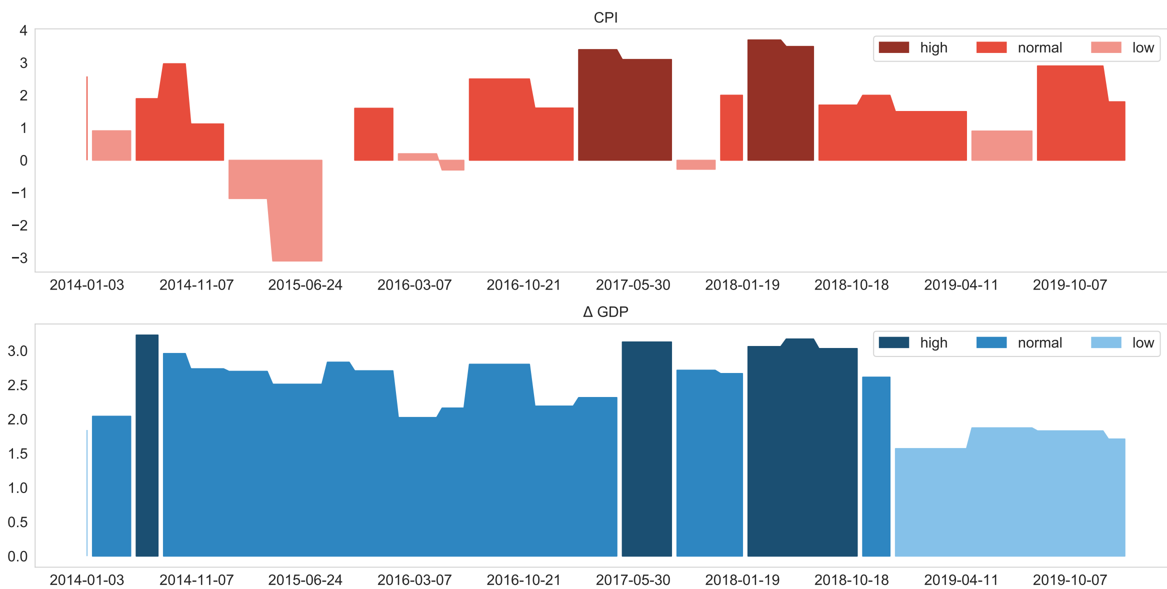
Notes: The table presents the classification of different economic regimes (high, normal, low) for GDP and CPI.

Figure 3: Empirical distribution of CPI and GDP growth target variables



Notes: The figure shows the empirical distribution of CPI and GDP regimes. CPI: low regime (light red), normal regime (mid red), high regime (dark red). GDP: low regime (light blue), normal regime (mid blue), high regime (dark blue).

Figure 4: Time-series of CPI and GDP growth regimes



Notes: The figure displays the evolution of different economic regimes over time. CPI (upper panel): low regime (light red), normal regime (mid red), high regime (dark red). GDP (lower panel): low regime (light blue), normal regime (mid blue), high regime (dark blue).

Conditional on the regime classification, we categorise the speech-implied news into *good* and *bad* news for the market. The division in the GDP-regime is straightforward. In any GDP regime, speeches that imply higher (lower) GDP-growth than the most recent SPF forecast are good (bad) GDP news. Similarly, lower (higher) unemployment forecast revisions are good (bad) news. The story for the CPI regime is more complex: If a speech implies that inflation will move closer to the 2% target than the most recent SPF forecast, it is considered good news. If a speech implies that inflation will move further from the target, it is bad news. So, a speech that implies an increase in inflation would be good news if inflation is below target but bad news if inflation is above target. Table 9 outlines the news classifications.

Table 9: Central bank GDP news classification

	Good news	**Bad news**
High GDP	$g_{\text{cb}} > g_{\text{spf}}$	$g_{\text{cb}} < g_{\text{spf}}$
Normal GDP	$g_{\text{cb}} > g_{\text{spf}}$	$g_{\text{cb}} < g_{\text{spf}}$
Low GDP	$g_{\text{cb}} > g_{\text{spf}}$	$g_{\text{cb}} < g_{\text{spf}}$

Notes: The table presents the classification of *good* versus *bad* GDP news for different levels of GDP.

Table 10: Central bank CPI news classification

	Good news	**Bad news**
High CPI	$\pi_{\text{cb}} < \pi_{\text{spf}}$	$\pi_{\text{cb}} > \pi_{\text{spf}}$
Normal CPI (slightly above target)	$\pi_{\text{cb}} < \pi_{\text{spf}} \vert \pi_{\text{spf}} > 2\%$	$\pi_{\text{cb}} > \pi_{\text{spf}} \vert \pi_{\text{spf}} > 2\%$
Normal CPI (slightly below target)	$\pi_{\text{cb}} > \pi_{\text{spf}} \vert \pi_{\text{spf}} < 2\%$	$\pi_{\text{cb}} < \pi_{\text{spf}} \vert \pi_{\text{spf}} < 2\%$
Low CPI	$\pi_{\text{cb}} > \pi_{\text{spf}}$	$\pi_{\text{cb}} < \pi_{\text{spf}}$

Notes: The table presents the classification of *good* versus *bad* CPI news for different levels of CPI.

6.3 News Effects by CPI Regime

We now analyse the effects of speech-implied forecast revision news by CPI regime. We separate god news from bad news to assess whether asymmetric speech-implied news effects exist. The regression equations for realized volatility (RV) and tail risk (TR) in the 30 minutes after each speech are as follows:

$$RV_s = \beta_0|\nu_{\pi,s,good}| + \beta_1|\nu_{\pi,s,bad}| + \beta_2|\nu_{g,s,good}| + \beta_3|\nu_{g,s,bad}| + \beta_4|\nu_{u,s,good}| + \beta_5|\nu_{u,s,bad}| + \epsilon_{RV} \quad (10)$$

$$TR_s = \rho_0|\nu_{\pi,s,good}| + \rho_1|\nu_{\pi,s,bad}| + \rho_2|\nu_{g,s,good}| + \rho_3|\nu_{g,s,bad}| + \rho_4|\nu_{u,s,good}| + \rho_5|\nu_{u,s,bad}| + \epsilon_{TR}. \quad (11)$$

The variables have the same meaning as before. That is, for each speech s, denote its CPI news component as $\nu_{\pi,s}$, GDP news as $\nu_{g,s}$, and unemployment news as $\nu_{u,s}$. However, for each macroeconomic news component, we now have a *good news* variable and a *bad news* variable (both in absolute values), denoted by *good* and *bad* subscripts. We estimate the volatility regression for both the equity and the bond markets for each CPI regime: low, normal, and high. The tail risk equation is estimated by CPI regime for equity markets only, due to scope limitations of this paper.

Equity Markets

Table 11 reports the effects of speech-implied forecast revisions on realized volatility and tail risk in equity markets, broken down by CPI regime. Appendix G details these results for each CPI regime and target variable.

Table 11: Association between speech-implied forecast revisions and volatility in equity markets across CPI regimes

	High CPI regime		**Low CPI regime**		**Normal CPI regime**	
	RV	TR	RV	TR	RV	TR
|News CPI good|	+***	-	+***	-	-	+*
|News CPI bad|	+***	-	+**	-	-	-
|News GDP good|	-	-	+**	+***	-	-
|News GDP bad|	-	-	-	+*	-	-
|News U good|	+***	+***	-	-	-	-
|News U bad|	+**	-	+***	-	+***	+***
n. obs.	36		59		70	

Notes: + = positive association. *= $p \leq 0.1$, **= $p \leq 0.05$, ***= $p \leq 0.01$. − = no statistically significant results.

High CPI regime: When CPI is high, speech-implied forecast revisions to CPI and unemployment forecasts have a statistically significant, positive association with realized volatility in equity markets in the 30 minutes after the speech (see the columns labeled *RV*). This holds true both for positive and negative news. Tail risk dynamics (see the columns labeled TR) are less strongly associated with central bank speech news signals in the high CPI regime.

Low CPI regime: A similar picture emerges in the low CPI regime. Speech-implied forecast revisions to CPI, good and bad, are strongly associated with increased equity market volatility. Low CPI regimes occur exclusively with normal or low GDP regimes (see Figure 4). Therefore, it is not surprising to see that speech-implied forecast revisions to GDP have a slightly stronger association with market volatility than during high CPI regimes, which almost exclusively co-occur with high GDP regimes. We interpret this as indicating that when the economy is in full swing, market sentiments tend to be optimistic and less 'attention' might be given to central bank announcements. Tail risk in the low CPI regime seems to be sensitive to both positive and negative speech-implied forecast revisions to GDP.

Normal CPI regime: Normal CPI times are defined as periods when the inflation rate is close to 2%. During these periods, there are no longer statistically significant associations between speech-implied forecast revisions of any kind and market volatility, except for negative unemployment news. Again, we would interpret these results as indicating that markets 'listen' less attentively to central bank communication when the economy is in normal or good times compared to periods of undesirably high or low inflation. Table 11 shows similar patterns for the prediction of equity volatility and tail risk in the normal CPI regime.

Bond Markets

Table 12 summarizes how speech-implied forecast revisions affect bond futures volatility across CPI regimes. Appendix I details the regression tables for each CPI regime and target variable combination. Bond markets produce patterns similar to those in equity markets: large speech-implied forecast revisions are more significantly associated with higher bond volatility when CPI is far from the target.

Table 12: Association between speech-implied forecast revisions and volatility in bond markets across CPI regimes

	High CPI regime			**Low CPI regime**			**Normal CPI regime**		
	2y	5y	10y	2y	5y	10y	2y	5y	10y
|News CPI good|	-	-	-	+***	+***	+***	-	-	-
|News CPI bad|	-	+*	-	-	+*	+*	-	-	-
|News GDP good|	+*	-	-	+***	-	-	-	-	-
|News GDP bad|	-	-	-	-	-	-	-	-	-
|News U good|	n/a	n/a	n/a	-	-	-	-	-	-
|News U bad|	+**	+**	+***	-	+*	+*	+***	+***	+***
n. obs.		33			42			52	

Notes: + = positive association. *= $p \leq 0.1$, **= $p \leq 0.05$, ***= $p \leq 0.01$. $-$ = no statistically significant results. 'n/a' = no observations available.

6.4 News Effects by GDP Regime

We now estimate equations (10) and (11) by different GDP regimes: low, normal, and high.

Equity Markets

Table 13 reports speech-implied forecast revision effects on realized volatility and tail risk in equity markets, broken down by GDP regime. Appendix H details these results for each CPI regime and target variable.

Table 13: Association between speech-implied forecast revisions and volatility in equity markets across GDP regimes

	High GDP regime		**Low GDP regime**		**Normal GDP regime**	
	RV	TR	RV	TR	RV	TR
|News CPI good|	-	-	+***	-	-	+**
|News CPI bad|	-	-	+***	-	-	-
|News GDP good|	-	-	+***	+***	+*	-
|News GDP bad|	-	-	+***	+*	-	-
|News U good|	-	n/a	+**	+**	-	-
|News U bad|	+**	+**	+***	+**	-	-
n. obs.	36		44		81	

Notes: + = positive association. *= $p \leq 0.1$, **= $p \leq 0.05$, ***= $p \leq 0.01$. $-$ = no statistically significant results. 'n/a' = no observations available.

High and normal GDP regimes: In high GDP times, negative speech-implied-forecast revisions to unemployment raise equity RV and TR. Similarly, positive speech-implied revisions to CPI forecasts raise TR during normal GDP periods.

Low GDP regime: In low GDP times, all speech-implied forecast revisions influence equity RV and all GDP and unemployment revisions influence equity TR. That is, RV and TR are a substantially more sensitive to forecast revisions during periods of low economic activity.

Overall, markets 'listen' most carefully in times of economic distress. In normal or good times, news in central bank speeches have less impact on RV and TR in equity markets.

Bond Markets

Table 14 shows speech-implied forecast revision effects on realized volatility in bond futures markets, broken down by GDP regime. Appendix J details these results for each GDP regime. Bond markets are also most sensitive to central bank speeches in extreme GDP regimes. Low GDP regimes witness the most significant association between GDP and unemployment forecast revisions and bond volatility. But markets also appear to be more sensitive to central bank speeches in high GDP regimes than in periods of normal economic growth.

Table 14: Association between speech-implied forecast revisions and volatility in bond markets across GDP regimes

	High GDP regime			**Low GDP regime**			**Normal GDP regime**		
	2y	5y	10y	2y	5y	10y	2y	5y	10y
|News CPI good|	+***	-	-	n/a	n/a	n/a	+***	-	-
|News CPI bad|	+*	+*	-	-	-	-	-	-	-
|News GDP good|	-	-	-	+***	+***	+***	-	-	-
|News GDP bad|	-	-	-	-	-	-	-	-	-
|News U good|	n/a	n/a	n/a	-	+**	+**	-	-	-
|News U bad|	+*	-	+*	+***	+***	+***	-	-	-
n. obs.		35			42			52	

Notes: + = positive association. *= $p \leq 0.1$, **= $p \leq 0.05$, ***= $p \leq 0.01$. − = no statistically significant results. 'n/a' = no observations available.

7 Extensions and Discussion

In this section, we consider various extensions and robustness checks of our framework. We examine the characteristics of speeches, implied signals and reassess the market responses, based on the speeches by the Fed Chair versus speeches by other (non-Chair) Fed members.

7.1 Characteristics of the Speech Data

We start by assessing the characteristics of the speeches in terms of the name of the speaker and the word characteristics of the statements. Table 15 reports the summary statistics.

Table 15: Summary statistics of the speeches by the Fed officials

Speaker	# Speeches	First speech	Last speech	Sum # words	Mean	Max	Min	Median
Charles I. Plosser	23	1/4/2014	2/17/2015	64358	2798.17	3660	1620	2928
Daniel K. Tarullo	26	2/6/2014	4/4/2017	100563	3867.81	4698	330	4420.5
Dennis Lockhart	34	1/13/2014	2/14/2017	78053	2295.68	3197	1141	2274
Janet L. Yellen	58	2/11/2014	11/29/2017	152887	2635.98	5124	517	1984.5
Jeremy C. Stein	5	1/3/2014	5/6/2014	16631	3326.20	4842	784	3423
Jerome H. Powell	82	3/13/2014	10/6/2020	195736	2387.02	5140	462	2108.5
Lael Brainard	81	12/2/2014	12/17/2020	247136	3051.06	5014	312	3325
Michelle W. Bowman	19	2/11/2019	12/4/2020	39792	2094.32	3869	609	1899
Patrick T. Harker	80	10/2/2015	12/2/2020	155850	1948.13	3738	435	1956.5
Randal K. Quarles	48	11/30/2017	12/11/2020	136744	2848.83	4922	783	2865.5
Richard H. Clarida	30	10/25/2018	11/16/2020	81526	2717.53	5091	556	2390.5
Richard W. Fisher	18	1/14/2014	3/9/2015	47573	2642.94	4932	626	2810
Robert S. Kaplan	27	11/18/2015	9/29/2020	64639	2394.04	4698	82	2908
Sandra Pianalto	2	2/26/2014	3/27/2014	5738	2869.00	3402	2336	2869
Stanley Fischer	45	7/10/2014	9/28/2017	142387	3164.16	4878	779	3276

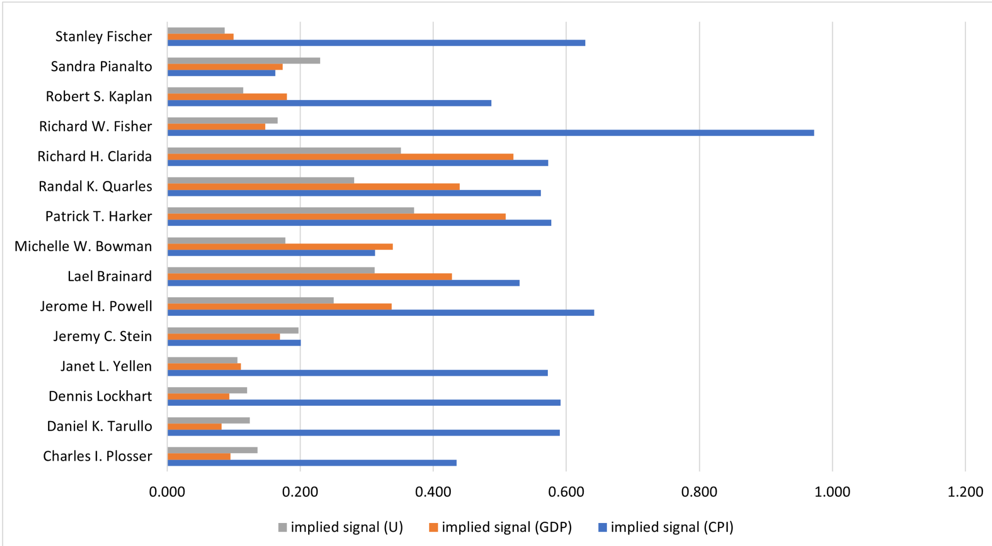
Notes: The table reports the summary statistics of the statements and speeches by the Fed officials (FOMC members, Fed Chair) in our text dataset. The table presents the name of the speaker, number of speeches, first and last speeches, sum of the number of words in the speech as well as mean, maximum, minimum and median number of words in the statements.

Several features are worth noting. For instance, among the speeches by fifteen Fed officials over the sample period, the speeches by Jerome H. Powell, Janet L. Yellen and Richard H. Clarida include the maximum number of words. Jerome H. Powell gives the most speeches and his speeches are among the longest ones in terms of the sum of the number of words in the statements. We also observe that the periods of speeches (i.e., the time between first and last speech) vary across speakers.

7.2 Forecast Revision News and Implied Speech Signals

We now take a closer look at the link between forecast revision news and the implied speech signals. To proceed, we compute the mean absolute values of the implied signals by speakers, measured based on our model. We implement the analysis for all three macro factors (CPI, GDP, unemployment). Figure 5 shows whose speech reveals the strongest and weakest signal about each macro indicator.

Figure 5: Implied speech signals and forecast revision news

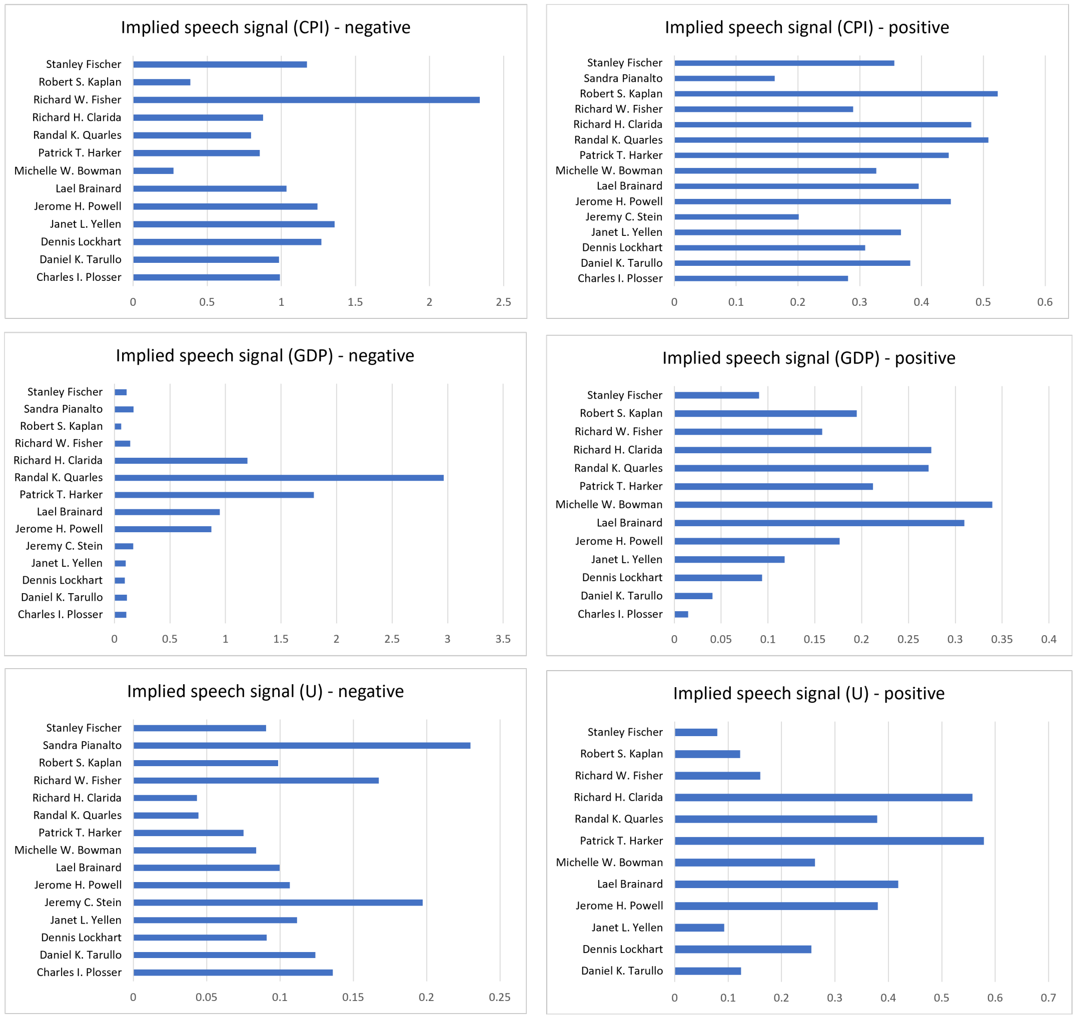


Notes: The figure shows the mean absolute values of the implied speech signals of speakers, measured based on our model implemented for three macro factors (CPI, GDP, unemployment). For each macro factor, horizontal bars indicate the strength of the signal (strong versus weak) stemming from the speeches by the Fed Chair and other Fed members.

Assessing the signals across speakers, based on three main macro factors, we observe that implied signals about inflation are significantly stronger than the signals about GDP and unemployment. This pattern holds regardless of the name of speakers. Richard W. Fisher, Jerome H. Powell and Stanley Fisher are the top three Fed members, sending the strongest forecast revision news among all members. While the mean absolute signal estimates of the speeches by Janet L. Yellen and Jerome H. Powell are close to each other (0.57 and 0.64, respectively), the speech signals by Powell on GDP and unemployment are larger than those that come from Yellen. The figure displays that the weakest inflation signals stem from the speeches by Sandra Pianalto, Jeremy C. Stein and Michelle W. Bowman.

To evaluate the role of signal *direction*, we further decompose the total implied speech signals into positive and negative signals, and implement the same analysis. As we can clearly see in Figure 6, there is a considerable amount of heterogeneity in terms of the sign of the forecast revision news. For example, the first pattern that emerges from the figure is that "negative" inflation signals conveyed by the speeches (top left) are, on average, stronger than the "positive" inflation signals (1.04 versus 0.36). A similar pattern holds for the GDP signals, but not for the unemployment signals (middle left and right panels). Focusing on the inflation signals, we also notice that the speeches by Richard W. Fisher, who is often considered the Federal Open Market Committee's (FOMC) most hawkish member, send the largest negative signal (with estimated implied signal of 2.34), followed by Janet L. Yellen and Jerome H. Powell. As we compare the signal strength of the speeches by Yellen versus Powell, the largest divergence occurs for the "negative" implied signals with respect to GDP: the estimated negative forecast revision signals by Powell for the GDP are, on average, larger than those by Yellen (0.87 and 0.10 in the middle-left panel).

Figure 6: Signal direction: forecast revision news and negative versus positive implied speech signals



Notes: The figure displays the mean absolute values of the negative and positive implied speech signals of speakers (left and right panels, respectively). To generate the signal plots, we decompose the total implied speech signals into positive and negative signals for each macro factor. Horizontal bars indicate the strength of the signal (strong versus weak) stemming from the speeches by the Fed Chair and other Fed members.

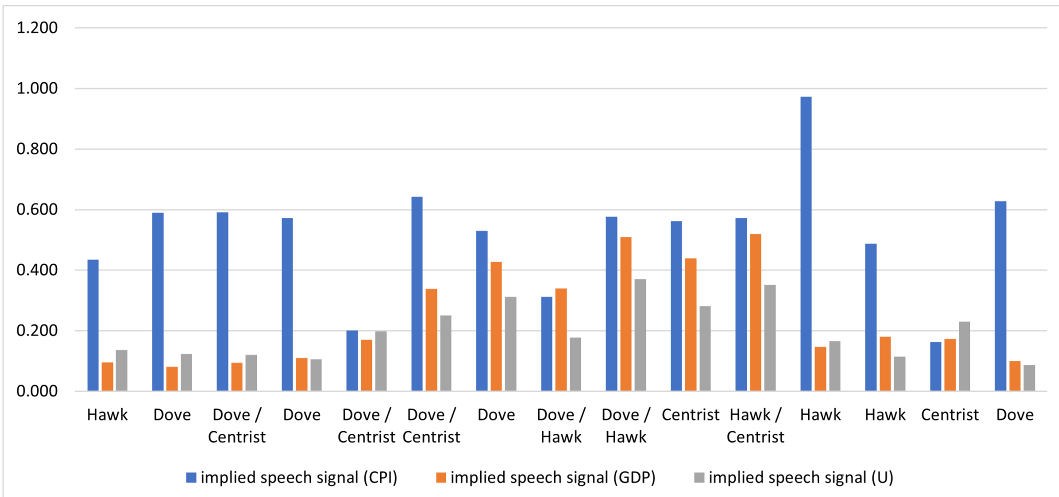
By looking at these implied signals, we can also indirectly examine the heterogeneity (or potential *disagreement*) among officials in terms of the implied signals that they convey through their speeches. For each macro factor, we also compute the standard deviation of the implied positive and negative signals of the speakers. The largest heterogeneity or disagreement occurs for the negative signals about GDP factor (0.86) followed by negative signals on CPI (0.50). For these two factors, speeches reveal relatively more consensus through the lens of positive signals, however. The highest level of consensus (i.e., the lowest signal heterogeneity) occurs for the negative unemployment signals.

7.3 Hawks versus Doves

Another noteworthy analysis would be to explore whether the *view* of Fed members affects the strength of speech signals. In other words, does it matter to be dovish or hawkish when sending signals via speeches?

While we leave an in-depth investigation to future research, we examine in this section the relationship between the view of officials (hawkish, centrist, dovish) and their estimated forecast revision signals. To carry out this analysis, we start by identifying the views and search among several sources including Reuters, Financial Times, Business Insider, Deutsche Bank, Marketplace and Mitsubishi UFJ Financial Group, Inc. (MUFG). For each speaker, we then rely on multiple measures of dovishness and hawkishness. Based on this approach, we create the following categories: centrist, dove, hawk, dove/centrist, hawk/centrist, and dove/hawk. After assigning these view labels to Fed officials, we compute the mean absolute implied signals for each category.

Figure 7: Hawks versus doves: disaggregated implied speech signals based on the Fed views

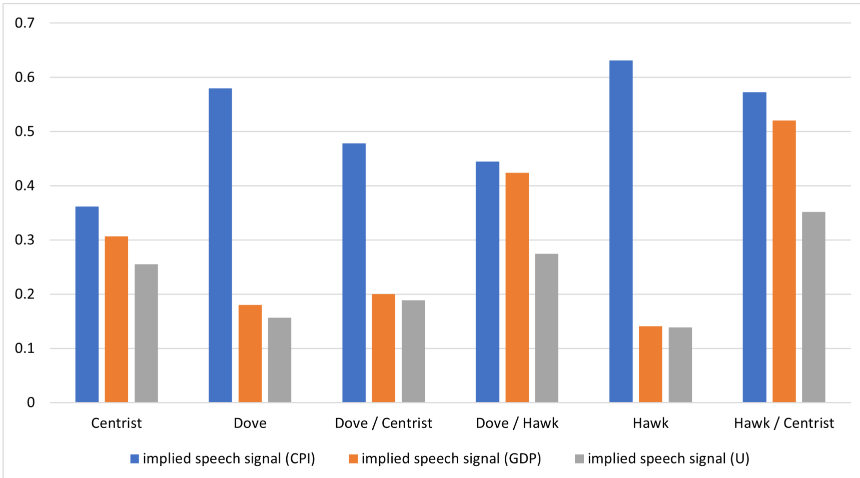


Notes: The figure illustrates the mean absolute values of the implied speech signals, based on the views of Fed members. For the three macro news factors (CPI, GDP and unemployment (U)), vertical bars indicate the strength of the speech signal (strong versus weak) by each Fed member. X-axis displays the views based on the hawkish, dovish, centrist views and their combinations. From left to right in the X-axis: "Hawk": Charles I. Plosser, "Dove": Daniel K. Tarullo, "Dove/Centrist": Dennis Lockhart, "Dove": Janet L. Yellen, "Dove/Centrist": Jeremy C. Stein, "Dove/Centrist": Jerome H. Powell, "Dove": Lael Brainard, "Dove/Hawk": Michelle W. Bowman, "Dove/Hawk": Patrick T. Harker, "Centrist": Randal K. Quarles, "Hawk/Centrist": Richard H. Clarida, "Hawk": Richard W. Fisher, "Hawk": Robert S. Kaplan, "Centrist": Sandra Pianalto, "Dove": Stanley Fischer

Two main patterns emerge from this assessment (Figures 7 and 8). First, looking at the signals at the disaggregated level (i.e., based on the names of Fed officials), we do not find clear evidence that officials with hawkish *or* dovish view generate systematically the strongest or weakest signals. For example, Richard W. Fisher, who is considered the Federal Open Market Committee's (FOMC) most hawkish member, gives speeches that create the largest implied inflation signals. However, other members with dovish/centrist view (e.g., Stanley Fischer, Daniel K. Tarullo, Janet L. Yellen, Jerome H. Powell) also send considerably high level of signals about inflation. Similar regularities hold when we consider other macro factors (GDP and unemployment): there is no evidence that dovish view dominates the hawkish view, or vice versa, in terms of conveying forecast revision signals. Indeed, we observe that both dovish and hawkish views could be associated with low levels of signals, although leaning towards centrist view seems to have relatively larger levels of signals (e.g., the speeches by Randal K. Quarles, Richard H. Clarida, Jeremy C. Stein, and

Jerome H. Powell).

Figure 8: Hawks versus doves: aggregated implied speech signals based on the Fed views



Notes: The figure shows the mean absolute values of the implied speech signals under different view categories (i.e., hawkish, dovish, centrist) and combinations. For the three macro news factors (CPI, GDP and unemployment (U)), vertical bars indicate the strength of the speech signal (strong versus weak) when we aggregate the signals based on the views of the Fed members.

Second, as we aggregate the signals based on the views of the Fed members, we observe similar patterns (Figure 8). For example, both dove and hawk views tend to be associated with high level of signals for inflation (0.58 and 0.63, respectively). For other macro factors (GDP and unemployment), the implied signals by both dovish and hawkish members are very weak and close to each other. It is, however, worth mentioning that the signal dispersion across three macro factors is the lowest for *centrist* views. In other words, officials with centrist view are likely to convey similar level of signals about the state of the economy based on the CPI, GDP, and unemployment factors. Overall, based on the evidence we have, it is hard to draw a conclusion that a specific view (e.g., hawkish) dominates the other (e.g., dovish) in terms of the strength of the speech signals. The underlying drivers of different levels of speech signals (i.e., strong, moderate, weak) seem to be related to the macro factor (CPI versus GDP and unemployment) and the sign of the signals (positive versus negative). Of course, we refrain from explicitly delving into the roles of hawks and doves in the language processing stage here. We defer this intriguing avenue for exploration to future research.

7.4 Reassessing Market Responses: To Chair, or Not to Chair?

Do the speeches by the Fed Chair speak louder than the speeches by other Fed members? As highlighted earlier, our analysis reveals that the Chair Jerome H. Powell conveys the strongest forecast revision news among all speakers after Richard Fisher. While the mean absolute signal estimates of the speeches by Janet L. Yellen and Jerome H. Powell are similar, the speech signals by Powell on GDP and unemployment are larger than those that come from Janet L. Yellen. Our results also indicate that the *dispersion* among officials in terms of the strength of their signals is the largest for the negative (downward) forecast revision signals about GDP and CPI. Looking at the positive signals, we observe more *consensus* among officials. Based on the findings of our assessment, the answer to the question of whether being a Chair matters for conveying (stronger) speech signals seems to depend on the intended direction of the signal (i.e., negative versus positive) and the underlying macro factors.

To explore this aspect further, we employ a regression analysis and rerun our baseline regressions by using now *Chair signals* and *non-Chair signals*. Considering realized volatility and realized tail risk as response measures, we carry out this analysis for both equity and bond markets. Tables 16-19 report these regression results.

Table 16: Speech signals by the Fed Chair, equity market volatility and tail risk

Target variable: RV_e	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	0.2391	0.036	6.725	0.000	0.169	0.309
|GDP news|	0.0515	0.048	1.083	0.279	-0.042	0.145
|U news|	0.1792	0.041	4.368	0.000	0.099	0.260
R^2: 0.693	Adj. R^2: 0.681	n. obs.: 77	Heteroscedasticity robust standard errors			
Target variable: TR_e	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	3.2833	0.840	3.909	0.000	1.637	4.930
|GDP news|	0.6968	1.456	0.479	0.632	-2.156	3.550
|U news|	2.4698	1.132	2.181	0.029	0.251	4.689
R^2: 0.522	Adj. R^2: 0.503	n. obs.: 77	Heteroscedasticity robust standard errors			

Notes: The table reports the regression results for the association between speech-implied forecast revision news (in absolute value about CPI, GDP, unemployment) and realized volatility (upper panel) and tail risk (lower panel). We consider the speech signals by the Fed Chair. The estimation results are reported for the U.S. equity market.

Looking at the volatility effects of "Chair signals" first (upper panel of Table 16), we find that forecast revisions to CPI and unemployment news are still very significant (at 1% level) and they increase stock market volatility. The results also indicate that the implied signals of Chair speeches on CPI news have a larger effect on volatility (0.239 versus 0.167), although Chair signals for GDP news have no significant effect in this case. Three Chair-speech signal factors, when combined, explain 0.681 of the variation in intradaily realized volatility, which is close to the adjusted R^2 of our baseline volatility regression results (0.718). Turning to the impact on equity tail risk (lower panel of Table 16), our regression results indicate that the estimated coefficients are similar in size to those from our baseline regression results. Chair speeches for the CPI news factor have a larger effect on tail risk, however, compared to the tail risk effects of all speeches (3.28 versus 2.26).

Table 17: Speech signals by the non-Chair Fed members, equity market volatility and tail risk

Target variable: RV_e	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	0.1246	0.025	5.001	0.000	0.076	0.173
|GDP news|	0.1211	0.076	1.592	0.111	-0.028	0.270
|U news|	0.1993	0.033	6.043	0.000	0.135	0.264
R^2: 0.761	Adj. R^2: 0.754	n. obs.: 114	Heteroscedasticity robust standard errors			
Target variable: TR_e	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	1.6157	0.538	3.001	0.003	0.561	2.671
|GDP news|	1.9146	1.135	1.688	0.092	-0.309	4.138
|U news|	2.3643	0.568	4.160	0.000	1.250	3.478
R^2: 0.548	Adj. R^2: 0.535	n. obs.: 114	Heteroscedasticity robust standard errors			

Notes: The table reports the regression results for the association between speech-implied forecast revision news (in absolute value about CPI, GDP, unemployment) and realized volatility (upper panel) and tail risk (lower panel). We consider the speech signals by the Fed members and exclude those by the Fed Chair. The estimation results are reported for the U.S. equity market.

Going one step further, we extend this assessment and compare the equity market responses stemming from only Chair signals versus non-Chair signals (Table 16 and Table 17, respectively). Focusing on the CPI news factor, we find that the non-Chair signals have a smaller impact on volatility (upper panel), compared to the effects of Chair signals (0.12 versus 0.23, respectively). As the lower panel of the table indicates, a similar pattern holds with respect to the tail risk effects (1.61 versus 3.28).

Table 18: Speech signals by the Fed Chair, bond market volatility and tail risk

Target variable: $RV_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	0.0246	0.005	5.102	0.000	0.015	0.034
|GDP news|	0.0090	0.008	1.205	0.228	-0.006	0.024
|U news|	0.0148	0.006	2.354	0.019	0.002	0.027
R^2: 0.713	Adj. R^2: 0.700	n. obs.: 70	Heteroscedasticity robust standard errors			
Target variable: $TR_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	7.3717	1.616	4.561	0.000	4.204	10.539
|GDP news|	6.6960	1.847	3.625	0.000	3.076	10.317
|U news|	1.4508	1.765	0.822	0.411	-2.009	4.911
R^2: 0.591	Adj. R^2: 0.573	n. obs.: 70	Heteroscedasticity robust standard errors			

Notes: The table reports the regressions results for the association between speech-implied forecast revision news (in absolute value about CPI, GDP, unemployment) and realized volatility (upper panel) and tail risk (lower panel). We consider the speech signals by the Fed Chair. The estimation results are reported for the 2-year maturity U.S. Treasury bond futures.

Table 19: Speech signals by the non-Chair Fed members, bond market volatility and tail risk

Target variable: $RV_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	0.0104	0.002	4.225	0.000	0.006	0.015
|GDP news|	0.0145	0.008	1.787	0.074	-0.001	0.031
|U news|	0.0167	0.004	4.046	0.000	0.009	0.025
R^2: 0.673	Adj. R^2: 0.664	n. obs.: 105	Heteroscedasticity robust standard errors			
Target variable: $TR_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
|CPI news|	3.1841	0.923	3.448	0.001	1.374	4.994
|GDP news|	1.6228	2.233	0.727	0.467	-2.754	5.999
|U news|	4.6046	1.003	4.589	0.000	2.638	6.571
R^2: 0.498	Adj. R^2: 0.483	n. obs.: 105	Heteroscedasticity robust standard errors			

Notes: The table reports the results for the association between speech-implied forecast revision news (in absolute value about CPI, GDP, unemployment) and realized volatility (upper panel) and tail risk (lower panel). We consider the speech signals by the Fed members and exclude those by the Fed Chair. The estimation results are reported for the 2-year maturity U.S. Treasury bond futures.

As we conduct the analysis for the bond market, we find that, compared to the signals by other Fed members, Chair signals tend to generate a larger tail risk, although the results for volatility remain quite stable (Table 18 and Table 19). For instance, the estimated coefficient of the implied Chair signals for assessing tail risk impact is 7.37 whereas it is 3.18 for the non-Chair signals (lower panels of Table 18 and Table 19). These volatility and tail risk results also hold for other maturities (unreported for brevity but available upon request), particularly for the forecast revisions pertaining to the CPI news.

In sum, reassessing the market responses by decomposing the signals into Chair and non-Chair versions, we find evidence that considerably underpins our main results in terms of the significance of effects. Nevertheless, the estimated magnitude of the impact in both stock and bond markets appears to depend on the position of the speaker (i.e., Chair or not) and the "Chair effect" is particularly pronounced for the CPI news. This analysis can be viewed in parallel with the study of Swanson and Jayawickrema (2023), who compare high-frequency changes in interest rates after Fed Chair versus Fed Vice Chair speeches and find that Chair speeches have a much higher impact.

8 Conclusion

We use supervised multimodal natural language processing methods to map central bank language to forecasts of macroeconomic variables. We benchmark an extensive array of machine learning methods on this task. Finally, we apply this approach to a dataset of time-stamped speeches from Federal Reserve FOMC members in order to create a novel monetary policy news series by taking the difference between central bank speech-implied forecast revisions and market expectations which we approximate with the latest available figures from the Survey of Professional Forecasters.

Our results indicate that news signals derived from central bank speeches can help explain volatility and tail risk in both equity and bond markets. Speech-implied news seem to carry information to which markets react - particularly in *abnormal* GDP and inflation regimes. We find no evidence that speeches resolve uncertainty. These findings underpin the importance of analysing the *continuous flow* of central bank communication with markets such as through FOMC member speeches.

Our empirical analysis also reveals that *hawkish* versus *dovish* views do not necessarily dominate each other in terms of the strength of the speech signals. Instead, we find that the magnitude of signals mostly depends on the macro news factor (CPI versus GDP and unemployment) and the direction of the signals (i.e., positive versus negative). Based on our framework, we further assess whether the speeches by the Fed Chair produce different signals and market responses. We show that the Chair signals tend to generate greater tail risk and volatility compared to the signals conveyed by other Fed members. In fact, this result can be viewed in parallel with the study of Swanson and Jayawickrema (2023), who document that the speeches of the Fed Chair have a higher market impact. Understanding the implications of Fed Chair speeches in affecting high-order market uncertainty over the short and long horizon would be of interest, and we leave this direction to future research.

References

Ahrens, M., Ashwin, J., Calliess, J.P., Nguyen, V., 2021. Bayesian topic regression for causal inference, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 8162–8188.

Ahrens, M., McMahon, M., 2021. Extracting economic signals from central bank speeches, in: Proceedings of the Third Workshop on Economics and Natural Language Processing, pp. 93–114.

Aït-Sahalia, Y., Jacod, J., 2009. Estimating the degree of activity of jumps in high-frequency data. The Annals of Statistics 37, 2202–2244.

Andersen, T., Fusari, N., Todorov, V., 2020. The pricing of tail risk and the equity premium: Evidence from international option markets. Journal of Business and Economic Statistics 38, 662–678.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61, 43–76.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Aruoba, B., Drechsel, T., 2022. Identifying monetary policy shocks: A natural language approach. CEPR Discussion Papers 17133. C.E.P.R. Discussion Papers.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64, 253–280.

Bauer, M.D., Lakdawala, A., Mueller, P., 2022. Market-based monetary policy uncertainty. The Economic Journal 132, 1290–1308.

Bekaert, G., Hoerova, M., Duca, M.L., 2013. Risk, uncertainty and monetary policy. Journal of Monetary Economics 60, 771–788.

Bernanke, B., Boivin, J., Eliasz, P.S., 2005. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. The Quarterly Journal of Economics 120, 387–422.

Bernanke, B.S., Kuttner, K.N., 2005. What explains the stock market's reaction to Federal Reserve Policy? Journal of Finance 60, 1221–1257.

Bholat, D., Hansen, S., Santos, P., Schonhardt-Bailey, C., 2015. Text mining for central banks. Available at SSRN 2624811 .

Blinder, A.S., 2018. Through a crystal ball darkly: The future of monetary policy communication. AEA Papers and Proceedings 108, 567–71.

Bollerslev, T., Li, J., Xue, Y., 2018. Volume, volatility, and public announcements. The Review of Economic Studies 85, 2005–2041.

Bollerslev, T., Todorov, V., Xu, L., 2015. Tail risk premia and return predictability. Journal of Financial Economics 118, 113–134.

Boswijk, H.P., Laeven, R.J.A., Yang, X., 2018. Testing for self-excitation in jumps. Journal of Econometrics 203, 256–266.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Byrne, D., Goodhead, R., McMahon, M., Parle, C., 2023. The central bank crystal ball: Temporal information in monetary policy communication. Discussion Papers 17930. Centre for Economic Policy Research.

Caballero, R.J., Simsek, A., 2022. Monetary policy with opinionated markets. American Economic Review 112, 2353–92.

Cai, Y., Camara, S., Capel, N., 2021. It's not always about the money, sometimes it's about sending a message: Evidence of Informational Content in Monetary Policy Announcements , 1–24.

Card, D., Tan, C., Smith, N.A., 2018. Neural models for documents with metadata, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics. pp. 2031–2040.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.

Child, R., Gray, S., Radford, A., Sutskever, I., 2019. Generating long sequences with sparse transformers. Technical Report.

Cieslak, A., Hansen, S., McMahon, M., Xiao, S., 2023. Policymakers' Uncertainty. Mimeograph.

Cieslak, A., McMahon, M., 2023. Tough talk: The Fed and the risk premia. Mimeograph.

Cieslak, A., Schrimpf, A., 2019. Non-monetary news in central bank communication. Journal of International Economics 118, 293–315.

Das, S., Giggins, C., He, J., Karypis, G., Krishnamurthy, S., Mahajan, M., Prabhala, N., Slack, D., van Dusen, R., Yue, S., Zha, S., Zheng, S., 2021. Context, language modeling, and multimodal data in finance. The Journal of Financial Data Science Summer .

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL HLT, 4171–4186.

Dudani, S.A., 1976. The distance-weighted k-nearest-neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics , 325–327.

Dungey, M., Erdemlioglu, D., Matei, M., Yang, X., 2018. Testing for mutually exciting jumps and financial flights in high frequency data. Journal of Econometrics 202, 18–44.

Ehrmann, M., Gaballo, G., Hoffmann, P., Strasser, G., 2019. Can more public information raise uncertainty? the international evidence on forward guidance. Journal of Monetary Economics 108, 93–112.

Ehrmann, M., Talmi, J., 2020. Starting from a blank page? Semantic similarity in central bank communication and market volatility. Journal of Monetary Economics 111, 48–62.

Ellen, S.T., Larsen, V.H., Thorsrud, L.A., 2022. Narrative monetary policy surprises and the media. Journal of Money, Credit and Banking 54, 1525–1549.

Erdemlioglu, D., Yang, X., 2022. News arrival, time-varying jump intensity, and realized volatility: Conditional testing approach. Journal of Financial Econometrics (nbac015), 1–38.

Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., Smola, A., 2020. Autogluon-tabular: Robust and accurate automl for structured data.

Frisch, R., Waugh, F.V., 1933. Partial time regressions as compared with individual trends. Econometrica 4, 387–401.

Gertler, M., Karadi, P., 2015. Monetary policy surprises, credit costs, and economic activity. American Economic Journal: Macroeconomics 7, 44–76.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine learning 63, 3–42.

Gómez-Cram, R., Grotteria, M., 2022. Real-time price discovery via verbal communication: Method and application to fedspeak. Journal of Financial Economics 143, 993–1025.

Gurkaynak, R.S., Sack, B.P., Swanson, E.T., 2005. Do actions speak louder than words? The response of asset prices to monetary policy actions and statements. International Journal of Central Banking 1.

Gáti, L., Handlan, A., 2022. Monetary communication rules. Working Paper Series 2759, European Central Bank.

Handlan, A., 2020. Text shocks and monetary surprises: Text analysis of FOMC statements with machine learning. Published Manuscript .

Hansen, S., McMahon, M., 2016. Shocking language: Understanding the macroeconomic effects of central bank communication. Journal of International Economics 99, 114–133.

Hansen, S., McMahon, M., Prat, A., 2018. Transparency and deliberation within the FOMC: A computational linguistics approach. The Quarterly Journal of Economics 133, 801–870.

Hansen, S., McMahon, M., Tong, M., 2019. The long-run information effect of central bank communication. Journal of Monetary Economics 108, 185–202.

Hanson, S.G., Stein, J.C., 2015. Monetary policy and long-term real rates. Journal of Financial Economics 115, 429–448.

Hattori, M., Schrimpf, A., Sushko, V., 2016. The response of tail risk perceptions to unconventional monetary policy. American Economic Journal: Macroeconomics 8, 111–136.

Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics 3, 1163–1174.

Husted, L., Rogers, J., Sun, B., 2020. Monetary policy uncertainty. Journal of Monetary Economics 115, 20–36.

Jarociński, M., Karadi, P., 2020. Deconstructing monetary policy surprises-The role of information shocks. American Economic Journal: Macroeconomics 12, 1–43.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems 30.

Leombroni, M., Vedolin, A., Venter, G., Whelan, P., 2021. Central bank communication and the yield curve. Journal of Financial Economics 141, 860–880.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., Allen, P.G., 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692 .

Lovell, M., 1963. Seasonal adjustment of economic time series and multiple regression analysis. Journal of the American Statistical Association 58, 993–1010.

Lucca, D.O., Trebbi, F., 2009. Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements. Technical Report 15367. National Bureau of Economic Research, Inc.

Mandler, M., 2012. Inflation-regime dependent effects of monetary policy shocks. Evidence from threshold vector autoregressions. Economics Letters 116, 422–425.

Miranda-Agrippino, S., Ricco, G., 2021. The transmission of monetary policy shocks. American Economic Journal: Macroeconomics 13, 74–107.

Nakamura, E., Steinsson, J., 2018. High-frequency identification of monetary non-neutrality: The information effect. Quarterly Journal of Economics 133, 1283–1330.

Nesbit, J., 2020. Text as Instruments (Job Market Paper) click here for latest version. Technical Report.

Neuhierl, A., Weber, M., 2019. Monetary policy communication, policy slope, and the stock market. Journal of Monetary Economics 108, 140–155.

Ozdagli, A., Velikov, M., 2020. Show me the money: The monetary policy risk premium. Journal of Financial Economics 135, 320–339.

Petropoulos, A., Siakoulis, V., 2021. Can central bank speeches predict financial market turbulence? Evidence from an adaptive NLP sentiment index analysis using XGBoost machine learning technique. Central Bank Review 21, 141–153.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. Catboost: unbiased boosting with categorical features. Advances in Neural Information Processing Systems 31.

Ramey, V.A., 2016. Macroeconomic shocks and their propagation, in: Handbook of Macroeconomics. volume 2, pp. 71–162.

Romer, C.D., Romer, D.H., 2000. Federal Reserve Information and the Behavior of Interest Rates. American Economic Review 90, 429–457.

Swanson, E.T., 2021. Measuring the effects of federal reserve forward guidance and asset purchases on financial markets. Journal of Monetary Economics 118, 32–53.

Swanson, E.T., 2023. The importance of fed chair speeches as a monetary policy tool. AEA Papers and Proceedings 113, 394–400.

Swanson, E.T., Jayawickrema, V., 2023. Speeches by the Fed Chair are more important than FOMC announcements: An improved high-frequency measure of US monetary policy shocks. Working Paper.

Tenreyro, S., Thwaites, G., 2016. Pushing on a string: US monetary policy is less powerful in recessions. American Economic Journal: Macroeconomics 8, 43–74.

Tillmann, P., 2020. Monetary policy uncertainty and the response of the yield curve to policy shocks. Journal of Money, Credit and Banking 54, 803–833.

Zaheer, M., Guruganesh, G., Dubey, A., Ainslie, J., Alberti, C., Ontanon, S., Pham, P., Ravula, A., Wang, Q., Yang, L., Ahmed, A., 2020. Big bird: Transformers for longer sequences, in: Advances in Neural Information Processing Systems.

Appendix

A Description of the Machine Learning Models

In this section, we provide a detailed description of the machine learning models spanned by our base AutoGL framework. These classes are K-nearest neighbours (KNN), Random Forest, Extremely Randomized Trees, Boosted Decision Trees and Neural Networks.

K-nearest neighbours (KNN)

The K-nearest neighbours (KNN) class that we consider is a widely-used machine learning algorithm, belonging to the family of instance-based, non-parametric learning. It operates on the simple principle of feature similarity, assuming that similar data points can be found near each other in feature space. In both classification and regression, KNN works by finding the k closest training samples to a new data point and then predicts the output based on these neighbours. For classification, the algorithm typically assigns the class most common among its k nearest neighbours, while in regression, it usually takes the average of their values. In fact, KNN is easy to implement and understand, but its performance can significantly decline with high-dimensional data (the curse of dimensionality) and large datasets (due to computational cost).

Random Forest

The other machine learning algorithm that we implemented for performance comparison is the technique called Random forest. This machine learning method is versatile and powerful that operates by constructing multiple decision trees during training and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. This ensemble learning technique, particularly effective for large datasets, enhances predictive accuracy and controls over-fitting by averaging or *voting* across various trees. Each tree in the forest is built from a sample drawn with replacement (i.e., a bootstrap sample) from the training set. Furthermore, when splitting each node during the construction of a tree, the best split is found either from all input features or a random subset of them. This randomness, along with the ensemble approach, ensures the model's robustness against overfitting, making Random Forest an appealing choice for many applications in diverse domains ranging from finance to healthcare. We utilize the Random Forest algorithm under the AutoML framework.

Extremely Randomized Trees

Extremely randomized trees (ERT), also known as extra trees, is an ensemble learning technique that constructs a multitude of decision trees at training time. Similar to Random Forests, it operates by averaging predictions for regression tasks or using a majority vote in classification. However, it introduces additional randomness in the way splits are computed: instead of searching for the most discriminating thresholds, thresholds are drawn at random for each candidate feature and the best of these randomlygenerated thresholds is picked as the splitting rule. This randomness leads to more diversified trees and typically faster training than Random Forest, often with comparable performance.

Boosted Decision Trees

Boosted decision trees involve an ensemble learning technique that combines multiple weak decision tree learners to form a strong predictive model. Unlike methods like Random Forests which build trees in parallel, boosting builds them sequentially. Each tree is trained on the dataset with an emphasis on correctly predicting instances that were misclassified by previous trees. This is achieved through iterative updates to the weights of data points. The final prediction is made based on a weighted vote (in classification) or sum (in regression) of the predictions from individual trees. This method often results in high accuracy, especially for complex datasets, but requires careful tuning to avoid overfitting.

Neural Networks

Neural networks, as our base machine learning model that we put forward in our study, are a foundational model in machine learning, inspired by the structure and function of the human brain. At their core, neural networks consist of layers of interconnected nodes, or *neurons*, each performing simple computations. The network typically includes an input layer to receive the data, one or more hidden layers that process the data, and an output layer that produces the prediction. Each neuron in a hidden layer transforms the values from the previous layer with a weighted linear summation followed by a *non-linear* activation function. These weights are learned during training through a process called backpropagation, which iteratively adjusts the weights to minimize the difference between the network's prediction and the actual data outcomes. Deep neural networks, with many hidden layers, can model complex patterns and relationships in data. They are highly versatile, being applied in fields such as image and speech recognition and natural language processing, as we adopt and extend in our study via multimodal setting.

B Procedures for the Response Measures

In this section, we present the specifics of our procedures with respect to our high-frequency market response measures. To proceed, we first outline the estimation steps of the realized intensity as a high-frequency tail risk measure. We then present a method to assess the accuracy of parameters estimates and stability for both realized volatility and realized intensity. Finally, we present the estimated responses.

B.1 Estimation Steps of the Realized Intensity

We proceed with the details on the estimation of our RI measure (equation (7)) as follows.

Step 1: *Start by defining the jump activity index β:*

$$\beta =: \inf\{r \geq 0; \sum_{0 \leq s \leq t} |\Delta_s X|^r < \infty\}, \tag{12}$$

where $\Delta_s X = X_s - X_{s-}$ is the jump size at time s, and r is the power variation parameter.

Step 2: *Compute the jump activity index β in equation (7):*

$$\widehat{\beta}(t, \varpi, \theta, \theta') := \log \frac{V(\varpi, \theta, g)_t^n}{V(\varpi, \theta', g)_t^n} / \log(\frac{\theta'}{\theta}), \tag{13}$$

for which select $0 < \theta < \theta'$, $0 < \varpi < 1/2$ and

$$V(\varpi, \theta, g)_t^n := \sum_{i=1}^{[t/\Delta_n]} g\left(\frac{|\Delta_i^n X|}{\alpha \Delta_n^{\varpi}}\right), \tag{14}$$

where $g(t)$ is the weight function, choose a form that needs to satisfy the condition $g(x) = |x|^p$ if $|x| \leq a$ for some constant $a > 0$ and even integer $p > 2$.

Step 3: *Choose values for the tuning parameters ϖ, k_n and α in equation (7).*

Step 4: *Compute the g function in equation (7) to disentangle volatility component from the jump component.*

Step 5: *Identify the release times (minutes and seconds) of speeches.*

Step 6: *For each speech, select a window length (e.g., one hour) and estimate RI in equation (7) by using high-frequency returns in this window.*

B.2 Accuracy Assessment

To evaluate the accuracy of the estimated parameters of the response measures, we proceed with the realized intensity first. Let us use TR for $\widehat{\lambda}(k_n)_{t_p}$, instead of instead of λ and continue from this stage. We have

$$\sqrt{\frac{k_n \Delta_n}{\Delta_n^{\varpi\beta}}} \left(\widehat{TR} - TR\right) \xrightarrow{L_{st}} N\left(0, TR \frac{\alpha^\beta C_\beta(2)}{\left(C_\beta(1)\right)^2}\right),$$

where

$$C_\beta(k) = \int_0^\infty \left(g(x)\right)^k / x^{1+\beta} dx.$$

Therefore, the 95% confidence interval for $\widehat{\lambda}(k_n)_{t_r}$ is given by

$$\widehat{TR} \pm \text{c.v.} \times \sqrt{\frac{\widehat{TR}(\alpha \Delta_n^{\varpi})^\beta C_\beta(2)}{\left(C_\beta(1)\right)^2 k_n \Delta_n}},$$

for which we can use critical value such as c.v. = 1.96. The average of the lower and upper bound gives us the estimated intensity.

For spot realized volatility, we have

$$\sqrt{k_n^\sigma}\left(\hat{c}_{t_r} - c_{t_r}\right) \xrightarrow{L_{st}} N(0, 2c_{t_r}^2),$$

and the 95% confidence interval is

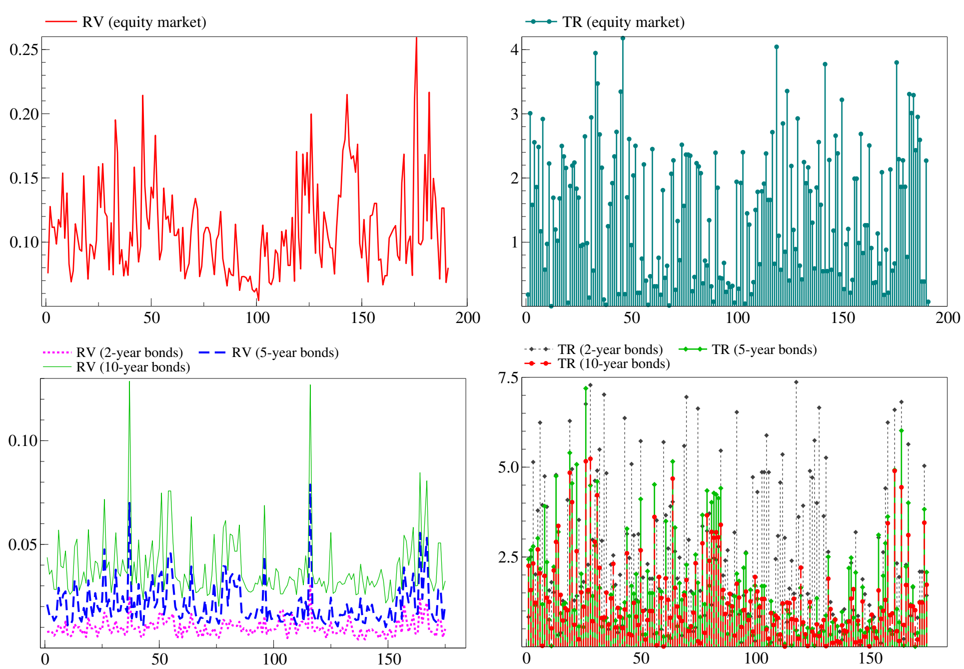
$$c_{t_r} \pm \text{c.v.} \times \sqrt{\frac{2}{k_n^\sigma}} c_{t_r}.$$

In light of these constructed confidence intervals, we assess the fit of the estimates, considering the lower and upper bounds.

B.3 Estimated Response Measures: Realized Volatility and Tail Risk

As we describe in the main text, we use high-frequency data and identify market responses in the forms of realized volatility and tail risk (computed based on realized intensity). Figure 9 displays the estimates of these quantities for each speech in our full sample for both equity and bond markets (upper and lower panels, respectively).

Figure 9: Estimated market response measures for central bank speeches



Notes: The figure shows the estimated response measures for each central bank speech (X-axis) in our dataset. Given the speech release, we compute realized volatility and tail risk—based on the realized intensity (labels RV and TR in the figure). The figure displays the quantities for the equity market (upper panels) and bond market (lower panels). For the equity market, RV and TR estimates are the cross-sectional averages of the individual stocks. For the bond market, the figure shows the estimated RV and TR separately for the 2-year, 5-year and 10-year bond futures.

The figure exhibits a number of features. First, both realized volatility and tail risk vary across central bank speeches. Second, looking at the response patterns of the bond market, we see noticeable differences between the reactions of short- and long-term bonds. That is, while the realized volatility of 2-year bond futures is clearly lower than the realized volatility of 5-year and 10-year bond futures (lower left panel), the realized tail risk identified through 2-year bonds is the highest across all maturities (lower right panel). Finally, central bank speeches tend to create distinct effects on bond and equity markets, which potentially reflects the importance of information signals embedded in the speeches.

C Further Considerations

Remark 1. It is worth emphasizing that the speeches have a much wider content beyond those key macro indicators (CPI, GPD, unemployment) that we rely on in our study. Nevertheless, we do not observe differences in terms of financial market effects mainly because we train our multimodal NLP model, test its out-of-sample performance, and construct the implied speech signals, entirely based on these three macro factors. Our proposed model processes the topics under this setting by utilizing both tabular (macro) data and text (speech) data. Therefore, our framework helps select the most important topics and those that do not carry significant explanatory information are directly excluded. This approach brings an advantage, rather than a setback, as it prevents us from incorrect measurement of market response to other generally important yet irrelevant speeches. Of course, it is possible to extend our model and feed the model by focusing also on other variables beyond macro factors.

Remark 2. When we identify the implied speech signals through our multimodal NLP model, we rely on a time frame for which we evaluate the information content in the entire period. During this process, we "synchronize" the time stamps of the speech and the SPF releases so that when we create the signal, the signal utilizes the information up to the *same* calendar time. Regardless of the time difference between the SPF release time and the speech release time, the time stamp of the signal is the time stamp of the speech and it remains the same as long as both SPF release and speech release fall in the same time frame. In fact, proceeding this way ensures that the process is a *martingale*. That is, the "speech release time" is the time that conditional expectations will be formed, based on all available information (including SPF news) up to speech time. This holds regardless of the past values and the time distance between SPF release and speech release.

Our high-frequency approach allows us to examine the impact of speech immediately after the public release by quantifying the changes in market volatility and market tail risk withing seconds and minutes. When a central bank speech is released *a few weeks* after an SPF release, investors still tend to use the most updated information available to them, perhaps related to market efficiency, so they wait for the release of the central bank speech. As soon as the speech is released and it becomes publicly available, we quantify the market response through our measures. So, the response already incorporates the information content in the SPF news, as investors wait for the new SPF release. As another situation, even if a speech is released, for example, *two days* after an SPF release, the reaction time that we rely on remains the same and hence it is still the speech release time. In this situation, while it is true that investors have relatively short period of time to *digest* the content of the SPF release, the period is sufficient for those monitoring markets at intradaily levels.

Remark 3. One argument would be that only relevant speeches matter and hence irrelevant speeches should not convey important signals. To test this conjecture, we conduct a simple, yet insightful, robustness check. We first rank the speeches in our database in order of their implied signal levels. We then identify the speeches that have the highest and lowest signal estimates (i.e., top ten and bottom ten). We observe that the highest implied signals often derive from the statements about topics on monetary policy, financial stability, economic conditions, and economic outlook. In contrast, the signals with the lowest values are often associated with statements that are indirectly related to macro environment, financial markets, or monetary policy. For example, these low signal speeches are about the situation of middle-income families (unemployment factor), consumer behavior in credit and payment markets, and small business (GDP factor). Of course, these statements are not necessarily redundant, as they are made by the Fed members and the Chair. However, they are not directly relevant and hence their signal levels that we measured using our model turn out to be low.

In light of this assessment, we also find that the name of the speaker (e.g., Chair or not) does not play an important role, as we see that Chair speeches can be associated with both lowest and highest signals. This regularity holds for all three news factors (CPI, GDP, unemployment) and for all other Fed members. Therefore, it is our understanding that, by looking at the name and whether the speaker is Chair, it is hard to draw a direct conclusion about which signals should matter. This is largely in line with our additional analyses on speech characteristics. Statements that look similar in terms of the speaker name, time, title of the talk have different levels of implied forecast revision signals.

Remark 4. One may also argue that forecasts at different horizons are potentially correlated. In fact, it is rather unlikely that there will be one trend for three months, but it will reverse four months from now. To assess the role forecast horizon further, we conduct an analysis for the horizon assessment, which suggests that the choice does not play a critical role (e.g., one-quarter forecast versus one year forecast), as we achieve steady state in both horizon choices.

Furthermore, regarding the policy implications of forecast horizon, our results indicate that the signals embedded in the *language* that central bankers use actually seem to generate *similar* market response for relatively short and long horizons (such as one-year-ahead forecasts). There might be several potential explanations for this result. First, *words* do not systematically carry a specific signal about the interest rate policy (such as potential changes in Fed Fund rates). In other words, while *actions* may allow to separate the short-term (policy) versus long-term (real economy) effects, for example, in the context of term structure or yield curve, *words* do not help unravel such asymmetric effects. Another reason could be related to the signals that we seek to identify. In fact, even though we can modify the horizon for the forecast revisions (e.g., monthly, quarterly, yearly), our focus is on three macro factors: CPI, GDP growth and unemployment. For these three macro indicators, we find no clear evidence that the horizon selection of the speech-implied forecast revision is critical. However, the evidence may change substantially for other important indicators, such as policy rates. We believe this would be an important line of research, as an extension of our framework.

D List of Relevant Greenbook Sections

Table 20: Considered Greenbook sections per economic indicator

GDP	**CPI**	**Unemployment**
Ec.GDP	Ec.Prices	Ec.Labor
For.Ec.Overview	For.CostPrice	For.Labor
For.Ec.Summary	Ec.Wages	
For.Outlook		
For.HH		
For.G		
For.Inven		
For.BusInvest		
For.Trade		

Notes: In the table, EC = Economic Conditions Section, For = Forecasts Section.

E Lists of Stocks and Bonds

Table 21: Stock tickers and names

AAPL	Apple	AXP	American	BA	Boeing	CAT	Caterpillar
CSCO	Cisco	CVX	Chevron	DIS	Disney	HD	Home
IBM	IBM	INTC	Intel	JNJ	Johnson	KO	Coca-Cola
MCD	McDonald's	MMM	3M	MRK	Merck	MSFT	MSFT
NKE	Nike	PFE	Pfizer	UNH	UnitedHealth	VZ	Verizon
WMT	Wal-Mart	XOM	Exxon				

Notes: The table lists the tickers and descriptions of the individual stocks used in our empirical analysis.

Table 22: Bond names and maturities

US Treasury Note Futures:	2-Year	5-Year	10-Year

Notes: The table lists the tickers and descriptions of the U.S. Treasury bond futures used in our empirical analysis.

F Additional Results: Language to Forecast Mapping

Table 23: CPI mapping and fit performance

Model \Predictive R^2	**score test**	**score val**	**score train**	**data source**
MM Neural Topic Model (non-lin)	**0.735**	**0.830**	0.670	joint MM tabular + topics
MM Neural Topic Model (linear)	**0.640**	**0.650**	0.600	joint MM tabular + topics
ExtraTreesMSE_BAG_L1	0.588	0.084	0.880	tabular
RandomForestMSE_BAG_L1	0.584	0.052	0.622	tabular + topics
ExtraTreesMSE_BAG_L1	0.584	0.089	0.595	tabular + topics
RandomForestMSE_BAG_L1	0.568	0.047	0.876	tabular
KNeighborsUnif_BAG_L1	0.559	0.141	0.460	tabular + topics
KNeighborsDist_BAG_L1	0.549	0.128	0.798	tabular + topics
KNeighborsUnif_BAG_L1	0.520	0.152	0.439	tabular + tfidf
KNeighborsDist_BAG_L1	0.519	0.146	1.000	tabular + tfidf
KNeighborsUnif_BAG_L1	0.516	0.142	0.442	tabular
NeuralNetFastAI_BAG_L1	0.515	0.233	0.251	tabular + topics
KNeighborsDist_BAG_L1	0.513	0.121	1.000	tabular
OLS	0.512		0.288	tabular
NeuralNetFastAI_BAG_L1	0.494	0.272	0.594	tabular
RandomForestMSE_BAG_L1	0.482	0.103	0.883	tabular + tfidf
WeightedEnsemble_L2	0.475	0.302	0.565	tabular
CatBoost_BAG_L1	0.386	0.200	0.698	tabular
CatBoost_BAG_L1	0.384	0.170	0.905	tabular + tfidf
XGBoost_BAG_L1	0.377	0.169	0.595	tabular + topics
XGBoost_BAG_L1	0.374	0.155	0.937	tabular + tfidf
LightGBMXT_BAG_L1	0.373	0.126	0.295	tabular
XGBoost_BAG_L1	0.368	0.152	0.770	tabular
WeightedEnsemble_L2	0.358	0.284	0.370	tabular + topics
LightGBMLarge_BAG_L1	0.357	0.080	0.646	tabular + tfidf
LightGBM_BAG_L1	0.327	0.136	0.294	tabular
WeightedEnsemble_L2	0.299	0.305	0.953	tabular + tfidf
LightGBM_BAG_L1	0.289	0.138	0.245	tabular + topics
NeuralNetTorch_BAG_L1	0.269	0.210	0.128	tabular + topics
NeuralNetTorch_BAG_L1	0.262	0.247	0.401	tabular
XGBoost_BAG_L1	0.260	0.056	0.783	tabular + embeddings
LightGBMXT_BAG_L1	0.252	0.092	0.348	tabular + tfidf
LightGBM_BAG_L1	0.252	0.131	0.368	tabular + tfidf
LightGBMLarge_BAG_L1	0.251	0.139	0.302	tabular
LightGBMLarge_BAG_L1	0.202	0.156	0.323	tabular + topics
ExtraTreesMSE_BAG_L1	0.193	0.143	0.889	tabular + tfidf
LightGBMLarge_BAG_L1	0.191	0.074	0.440	tabular + embeddings
CatBoost_BAG_L1	0.177	0.250	0.525	tabular + topics
LightGBMXT_BAG_L1	0.162	0.140	0.192	tabular + topics
NeuralNetFastAI_BAG_L1	0.148	0.280	0.912	tabular + tfidf
WeightedEnsemble_L2	0.132	0.139	0.573	tabular + embeddings
CatBoost_BAG_L1	0.126	0.116	0.633	tabular + embeddings
LightGBMXT_BAG_L1	0.116	0.001	0.520	tabular + embeddings
LightGBM_BAG_L1	0.112	-0.018	0.338	tabular + embeddings
NeuralNetTorch_BAG_L1	0.095	0.153	0.500	tabular + tfidf
NeuralNetTorch_BAG_L1	-0.030	0.076	0.161	tabular + embeddings
AutoGluon Multimodal Transformer	-0.292		-0.155	multimodal embeddings

Notes: The table reports the performance (predictive R^2) of different models for the language mapping analysis of the CPI.

Table 24: GDP mapping and fit performance

Model \Predictive $\mathbf{R}^2$	**score test**	**score val**	**score train**	**data source**
MM Neural Topic Model (lin)	**0.825**	**0.426**	0.372	joint MM tabular + topics
MM Neural Topic Model (non-lin)	**0.797**	**0.371**	0.483	joint MM tabular + topics
WeightedEnsemble_L2	0.380	0.304	0.497	tabular
OLS	0.785	0.301		tabular
NeuralNetFastAI_BAG_L1	0.480	0.270	0.443	tabular
WeightedEnsemble_L2	0.285	0.253	0.730	tabular + topics
WeightedEnsemble_L2	0.268	0.240	0.752	tabular + tfidf
WeightedEnsemble_L2	0.142	0.220	0.587	tabular + embeddings
CatBoost_BAG_L1	0.249	0.211	0.552	tabular
RandomForestMSE_BAG_L1	0.302	0.204	0.892	tabular + tfidf
RandomForestMSE_BAG_L1	0.348	0.202	0.892	tabular + topics
ExtraTreesMSE_BAG_L1	0.408	0.193	0.891	tabular
ExtraTreesMSE_BAG_L1	0.381	0.192	0.890	tabular + topics
ExtraTreesMSE_BAG_L1	0.111	0.188	0.891	tabular + tfidf
CatBoost_BAG_L1	0.207	0.187	0.671	tabular + tfidf
LightGBMXT_BAG_L1	0.203	0.178	0.322	tabular
LightGBM_BAG_L1	0.154	0.172	0.367	tabular
XGBoost_BAG_L1	0.141	0.171	0.580	tabular + topics
CatBoost_BAG_L1	0.006	0.169	0.531	tabular + topics
CatBoost_BAG_L1	0.101	0.169	0.552	tabular + embeddings
LightGBM_BAG_L1	0.099	0.162	0.704	tabular + embeddings
NeuralNetTorch_BAG_L1	0.461	0.160	0.341	tabular
LightGBM_BAG_L1	0.101	0.159	0.734	tabular + tfidf
KNeighborsUnif_BAG_L1	0.253	0.158	0.402	tabular + tfidf
LightGBMLarge_BAG_L1	0.245	0.155	0.598	tabular
KNeighborsDist_BAG_L1	0.256	0.151	1.000	tabular + tfidf
NeuralNetTorch_BAG_L1	0.049	0.150	0.553	tabular + tfidf
LightGBMXT_BAG_L1	0.120	0.150	0.348	tabular + tfidf
RandomForestMSE_BAG_L1	0.394	0.150	0.885	tabular
LightGBMLarge_BAG_L1	0.111	0.149	0.536	tabular + topics
LightGBMLarge_BAG_L1	0.181	0.149	0.665	tabular + embeddings
XGBoost_BAG_L1	0.119	0.142	0.567	tabular
NeuralNetFastAI_BAG_L1	0.060	0.136	0.797	tabular + tfidf
KNeighborsDist_BAG_L1	0.255	0.132	1.000	tabular
KNeighborsUnif_BAG_L1	0.248	0.130	0.407	tabular
LightGBM_BAG_L1	0.111	0.126	0.496	tabular + topics
LightGBMXT_BAG_L1	0.105	0.125	0.505	tabular + embeddings
NeuralNetTorch_BAG_L1	-0.071	0.123	0.275	tabular + embeddings
NeuralNetTorch_BAG_L1	0.151	0.108	0.497	tabular + topics
XGBoost_BAG_L1	-0.015	0.107	0.663	tabular + embeddings
LightGBMLarge_BAG_L1	0.108	0.095	0.581	tabular + tfidf
XGBoost_BAG_L1	0.041	0.083	0.564	tabular + tfidf
KNeighborsUnif_BAG_L1	0.286	0.081	0.400	tabular + topics
KNeighborsDist_BAG_L1	0.274	0.074	1.000	tabular + topics
LightGBMXT_BAG_L1	0.097	0.049	0.318	tabular + topics
TextPredictor_BAG_L1	-0.077	-0.123	-0.103	tabular + embeddings
NeuralNetFastAI_BAG_L1	0.407	-0.126	0.438	tabular + topics
AutoGluon Multimodal Transformer	-0.044		0.013	multimodal transformer

Notes: The table reports the performance (predictive R^2) of different models for the language mapping analysis of the GDP.

Table 25: Unemployment mapping and fit performance

Model \Predictive R^2	**score_test**	**score_val**	**score_train**	**data source**
MM Neural Topic Model (non-lin)	**0.208**	**0.457**	0.285	joint MM tabular + topics
WeightedEnsemble_L2	-0.044	0.145	0.415	tabular + embeddings
NeuralNetTorch_BAG_L1	-0.152	0.122	0.313	tabular + embeddings
WeightedEnsemble_L2	-0.045	0.113	0.577	tabular + tfidf
MM Neural Topic Model (linear)	0.066	0.109	0.197	joint MM tabular + topics
CatBoost_BAG_L1	-0.055	0.104	0.690	tabular + tfidf
LightGBMXT_BAG_L1	-0.068	0.074	0.336	tabular + tfidf
NeuralNetTorch_BAG_L1	-0.029	0.070	0.394	tabular + tfidf
WeightedEnsemble_L2	0.131	0.058	0.191	tabular
WeightedEnsemble_L2	-0.010	0.053	0.278	tabular + topics
NeuralNetFastAI_BAG_L1	0.124	0.047	0.237	tabular
CatBoost_BAG_L1	0.021	0.041	0.411	tabular + embeddings
NeuralNetTorch_BAG_L1	0.106	0.033	0.098	tabular
LightGBM_BAG_L1	0.006	0.027	0.349	tabular + embeddings
LightGBM_BAG_L1	-0.035	0.025	0.316	tabular + tfidf
CatBoost_BAG_L1	-0.003	0.021	0.260	tabular + topics
CatBoost_BAG_L1	0.019	0.010	0.095	tabular
RandomForestMSE_BAG_L1	-0.072	0.008	0.868	tabular + tfidf
NeuralNetTorch_BAG_L1	-0.004	0.006	0.022	tabular + topics
XGBoost_BAG_L1	-0.112	0.006	0.883	tabular + tfidf
LightGBMLarge_BAG_L1	-0.001	0.001	0.594	tabular + embeddings
LightGBMLarge_BAG_L1	0.002	-0.003	0.109	tabular + topics
ExtraTreesMSE_BAG_L1	-0.045	-0.003	0.868	tabular + tfidf
LightGBMXT_BAG_L1	-0.001	-0.005	0.084	tabular
LightGBMXT_BAG_L1	0.000	-0.006	0.009	tabular + topics
LightGBM_BAG_L1	0.000	-0.007	0.015	tabular + topics
LightGBMXT_BAG_L1	-0.005	-0.024	0.292	tabular + embeddings
XGBoost_BAG_L1	-0.043	-0.027	0.495	tabular + topics
LightGBM_BAG_L1	-0.002	-0.028	0.170	tabular
LightGBMLarge_BAG_L1	0.013	-0.034	0.094	tabular
NeuralNetFastAI_BAG_L1	0.002	-0.036	0.565	tabular + tfidf
XGBoost_BAG_L1	-0.061	-0.041	0.624	tabular + embeddings
LightGBMLarge_BAG_L1	-0.045	-0.044	0.519	tabular + tfidf
NeuralNetFastAI_BAG_L1	-0.016	-0.058	0.025	tabular + topics
RandomForestMSE_BAG_L1	-0.005	-0.101	0.855	tabular + topics
XGBoost_BAG_L1	-0.048	-0.126	0.277	tabular
ExtraTreesMSE_BAG_L1	0.008	-0.144	0.849	tabular
ExtraTreesMSE_BAG_L1	0.049	-0.163	0.848	tabular + topics
KNeighborsUnif_BAG_L1	-0.013	-0.185	0.188	tabular + tfidf
KNeighborsUnif_BAG_L1	-0.004	-0.187	0.186	tabular
KNeighborsUnif_BAG_L1	-0.048	-0.187	0.195	tabular + topics
TextPredictor_BAG_L1	-0.067	-0.190	-0.070	tabular + embeddings
KNeighborsDist_BAG_L1	-0.003	-0.191	1.000	tabular + tfidf
RandomForestMSE_BAG_L1	-0.034	-0.192	0.842	tabular
KNeighborsDist_BAG_L1	-0.030	-0.210	1.000	tabular + topics
KNeighborsDist_BAG_L1	0.003	-0.215	1.000	tabular
OLS	-0.377		0.231	tabular
AutoGluon Multimodal Transformer	-1.177		-0.737	multimodal transformer

Notes: The table reports the performance (predictive R^2) of different models for the language mapping analysis of the unemployment.

G Additional Results: Equity Market, CPI Regimes

G.1 High CPI Regime

Table 26: Association between news, market volatility, and tail risk: equity market, high CPI regime

Target variable: RV_e	coef	std err	z	P> |z|	[0.025	0.975]
| CPI news pos. |	0.2740	0.070	3.936	0.000	0.138	0.410
| CPI news neg. |	0.1437	0.052	2.780	0.005	0.042	0.245
| GDP news pos. |	0.0820	0.164	0.499	0.618	-0.240	0.404
| GDP news neg. |	0.0118	0.087	0.136	0.892	-0.159	0.183
| U news pos. |	9.1621	2.098	4.368	0.000	5.051	13.273
| U news neg. |	0.1683	0.076	2.215	0.027	0.019	0.317
R^2: 0.917	Adj. R^2: 0.901	n. obs.: 36	Heteroscedasticity robust standard errors			

Target variable: TR_e	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	3.1074	2.184	1.423	0.155	-1.173	7.388
| News CPI neg. |	2.7033	1.791	1.509	0.131	-0.808	6.215
| News GDP pos. |	-1.5404	4.349	-0.354	0.723	-10.064	6.983
| News GDP neg. |	0.8172	1.466	0.557	0.577	-2.056	3.690
| News U pos. |	187.3136	13.601	13.772	0.000	160.657	213.970
| News U neg. |	2.3664	2.479	0.955	0.340	-2.492	7.225
R^2: 0.683	Adj. R^2: 0.619	n. obs.: 36	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility and tail risk regressions (upper panel and lower panels, respectively) for the equity market under the high CPI regime.

G.2 Low CPI Regime

Table 27: Association between news, market volatility, and tail risk: equity market, low CPI regime

Target variable: RV_e	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.1657	0.048	3.457	0.001	0.072	0.260
| News CPI neg. |	0.1305	0.064	2.046	0.041	0.005	0.256
| News GDP pos. |	0.4317	0.170	2.546	0.011	0.099	0.764
| News GDP neg. |	0.1279	0.158	0.812	0.417	-0.181	0.437
| News U pos. |	0.1730	0.160	1.084	0.278	-0.140	0.486
| News U neg. |	0.1008	0.029	3.459	0.001	0.044	0.158
R^2: 0.774	Adj. R^2: 0.748	n. obs.: 59	Heteroscedasticity robust standard errors			

Target variable: TR_e	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	1.5924	1.068	1.491	0.136	-0.500	3.685
| News CPI neg. |	1.2883	1.368	0.942	0.346	-1.393	3.970
| News GDP pos. |	10.3541	3.365	3.077	0.002	3.759	16.949
| News GDP neg. |	4.6929	2.575	1.823	0.068	-0.354	9.740
| News U pos. |	3.5833	3.297	1.087	0.277	-2.880	10.046
| News U neg. |	-0.2576	0.663	-0.388	0.698	-1.557	1.042
R^2: 0.622	Adj. R^2: 0.580	n. obs.: 59	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility and tail risk regressions (upper and lower panels, respectively) for the equity market under the low CPI regime.

G.3 Normal CPI Regime

Table 28: Association between news, market volatility, and tail risk: equity market, normal CPI regime

Target variable: RV_e	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.1013	0.070	1.447	0.148	-0.036	0.238
| News CPI neg. |	0.2412	0.161	1.494	0.135	-0.075	0.558
| News GDP pos. |	0.2766	0.199	1.392	0.164	-0.113	0.666
| News GDP neg. |	0.1507	0.243	0.620	0.536	-0.326	0.627
| News U pos. |	0.7982	0.909	0.878	0.380	-0.984	2.580
| News U neg. |	0.1983	0.059	3.369	0.001	0.083	0.314
R^2: 0.771	Adj. R^2: 0.749	n. obs.: 70	Heteroscedasticity robust standard errors			
Target variable: TR_e	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	1.5422	0.892	1.729	0.084	-0.206	3.291
| News CPI neg. |	5.2256	3.888	1.344	0.179	-2.395	12.847
| News GDP pos. |	2.6501	2.156	1.229	0.219	-1.576	6.876
| News GDP neg. |	-0.1986	2.686	-0.074	0.941	-5.463	5.066
| News U pos. |	4.6007	14.375	0.320	0.749	-23.574	32.775
| News U neg. |	2.6626	0.624	4.264	0.000	1.439	3.886
R^2: 0.593	Adj. R^2: 0.555	n. obs.: 70	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility and tail risk regressions (upper and lower panels, respectively) for the equity market under the normal CPI regime.

H Additional Results: Equity Market, GDP Regimes

H.1 High GDP Regimes

Table 29: Association between news, market volatility, and tail risk: equity market, high GDP regime

Target variable: RV_e	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.1078	0.068	1.586	0.113	-0.025	0.241
| News CPI neg. |	0.0011	0.089	0.012	0.990	-0.173	0.175
| News GDP pos. |	0.3347	0.292	1.148	0.251	-0.237	0.906
| News GDP neg. |	0.1446	0.106	1.358	0.174	-0.064	0.353
| News U pos. |	0.2226	0.216	1.032	0.302	-0.200	0.645
| News U neg. |	0.2192	0.100	2.200	0.028	0.024	0.414
R^2: 0.578	Adj. R^2: 0.545	n. obs.: 36	Heteroscedasticity robust standard errors			
Target variable: TR_e	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.9807	1.686	0.582	0.561	-2.324	4.286
| News CPI neg. |	0.1379	1.242	0.111	0.912	-2.297	2.573
| News GDP pos. |	2.0496	3.835	0.534	0.593	-5.467	9.566
| News GDP neg. |	0.9372	2.394	0.391	0.695	-3.756	5.630
| News U neg. |	4.2181	1.666	2.531	0.011	0.952	7.484
R^2: 0.652	Adj. R^2: 0.596	n. obs.: 36	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility and tail risk regressions (upper and lower panels, respectively) for the equity market under the high GDP regime.

H.2 Low GDP Regime

Table 30: Association between news, market volatility, and tail risk: equity market, low GDP regime

Target variable: RV_e	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.2141	0.049	4.330	0.000	0.117	0.311
| News CPI neg. |	0.1031	0.035	2.980	0.003	0.035	0.171
| News GDP pos. |	0.5916	0.060	9.840	0.000	0.474	0.709
| News GDP neg. |	0.1953	0.071	2.767	0.006	0.057	0.334
| News U pos. |	0.5219	0.272	1.918	0.055	-0.011	1.055
| News U neg. |	0.1513	0.026	5.795	0.000	0.100	0.202
R^2: 0.796	Adj. R^2: 0.778	n. obs.: 44	Heteroscedasticity robust standard errors			
Target variable: TR_e	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	-0.4354	1.289	-0.338	0.735	-2.962	2.091
| News CPI neg. |	0.2284	1.619	0.141	0.888	-2.944	3.401
| News GDP pos. |	7.3587	1.744	4.219	0.000	3.940	10.777
| News GDP neg. |	5.0831	2.958	1.718	0.086	-0.715	10.881
| News U pos. |	8.7712	4.091	2.144	0.032	0.752	16.790
| News U neg. |	1.7789	0.894	1.991	0.047	0.028	3.530
R^2: 0.565	Adj. R^2: 0.517	n. obs.: 44	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility and tail risk regressions (upper and lower panels, respectively) for the equity market under the low GDP regime.

H.3 Normal GDP Regime

Table 31: Association between news, market volatility, and tail risk: equity market, normal GDP regime

Target variable: RV_e	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.1482	0.095	1.560	0.119	-0.038	0.334
| News CPI neg. |	0.1780	0.183	0.974	0.330	-0.180	0.536
| News GDP pos. |	0.5693	0.335	1.700	0.089	-0.087	1.226
| News GDP neg. |	0.3184	1.055	0.302	0.763	-1.749	2.386
| News U pos. |	0.8327	0.593	1.405	0.160	-0.329	1.994
| News U neg. |	0.1523	0.179	0.853	0.394	-0.198	0.502
R^2: 0.858	Adj. R^2: 0.811	n. obs.: 81	Heteroscedasticity robust standard errors			
Target variable: TR_e	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	2.4965	1.133	2.204	0.028	0.276	4.717
| News CPI neg. |	1.3217	2.863	0.462	0.644	-4.290	6.934
| News GDP pos. |	6.0009	5.160	1.163	0.245	-4.112	16.114
| News GDP neg. |	2.0084	5.287	0.380	0.704	-8.354	12.370
| News U pos. |	2.6617	3.787	0.703	0.482	-4.760	10.084
| News U neg. |	1.9410	2.169	0.895	0.371	-2.311	6.193
R^2: 0.546	Adj. R^2: 0.496	n. obs.: 81	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility and tail risk regressions (upper and lower panels, respectively) for the equity market under the normal GDP regime.

I Additional Results: Bond Market, CPI Regimes

I.1 High CPI Regime

Table 32: Association between news and market volatility: bond market, high CPI regime

Target variable: $RV_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.0008	0.008	0.101	0.920	-0.015	0.017
| News CPI neg. |	-0.0069	0.008	-0.862	0.389	-0.023	0.009
| News GDP pos. |	0.0730	0.039	1.849	0.064	-0.004	0.150
| News GDP neg. |	0.0014	0.021	0.066	0.947	-0.039	0.042
| News U pos. |	0	0	nan	nan	0	0
| News U neg. |	0.0242	0.012	2.042	0.041	0.001	0.047
R^2: 0.830	Adj. R^2: 0.802	n. obs.: 33	Heteroscedasticity robust standard errors			

Target variable: $RV_{b,5y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	-0.0378	0.027	-1.418	0.156	-0.090	0.014
| News CPI neg. |	-0.0383	0.023	-1.668	0.095	-0.083	0.007
| News GDP pos. |	0.1472	0.118	1.245	0.213	-0.085	0.379
| News GDP neg. |	-0.0155	0.052	-0.299	0.765	-0.117	0.086
| News U pos. |	0	0	nan	nan	0	0
| News U neg. |	0.0876	0.039	2.248	0.025	0.011	0.164
R^2: 0.715	Adj. R^2: 0.655	n. obs.: 33	Heteroscedasticity robust standard errors			

Target variable: $RV_{b,10y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	-0.0410	0.036	-1.126	0.260	-0.112	0.030
| News CPI neg. |	-0.0499	0.037	-1.335	0.182	-0.123	0.023
| News GDP pos. |	0.2627	0.179	1.465	0.143	-0.089	0.614
| News GDP neg. |	-0.0118	0.087	-0.136	0.892	-0.182	0.158
| News U pos. |	0	0	nan	nan	0	0
| News U neg. |	0.1311	0.057	2.309	0.021	0.020	0.242
R^2: 0.799	Adj. R^2: 0.733	n. obs.: 33	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility regressions for the bond market (2-year, 5-year, 10-year bonds) under the high CPI regime.

I.2 Low CPI Regime

Table 33: Association between news and market volatility: bond market, low CPI regime

Target variable: $RV_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI neg. |	-0.0161	0.015	-1.075	0.283	-0.045	0.013
| News GDP pos. |	0.0334	0.007	4.882	0.000	0.020	0.047
| News GDP neg. |	0.0115	0.014	0.850	0.395	-0.015	0.038
| News U pos. |	0.0468	0.031	1.508	0.131	-0.014	0.108
| News U neg. |	0.0250	0.006	4.519	0.000	0.014	0.036
R^2: 0.70	Adj. R^2: 0.660	n. obs.: 42	Heteroscedasticity robust standard errors			
Target variable: $RV_{b,5y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI neg. |	-0.0360	0.030	-1.186	0.236	-0.095	0.023
| News GDP pos. |	0.0789	0.015	5.352	0.000	0.050	0.108
| News GDP neg. |	0.0342	0.031	1.106	0.269	-0.026	0.095
| News U pos. |	0.1268	0.063	2.005	0.045	0.003	0.251
| News U neg. |	0.0521	0.012	4.296	0.000	0.028	0.076
R^2: 0.691	Adj. R^2: 0.649	n. obs.: 42	Heteroscedasticity robust standard errors			
Target variable: $RV_{b,10y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI neg. |	-0.0657	0.043	-1.511	0.131	-0.151	0.020
| News GDP pos. |	0.1629	0.026	6.207	0.000	0.111	0.214
| News GDP neg. |	0.0695	0.048	1.448	0.148	-0.025	0.164
| News U pos. |	0.2607	0.102	2.550	0.011	0.060	0.461
| News U neg. |	0.0835	0.018	4.598	0.000	0.048	0.119
R^2: 0.767	Adj. R^2: 0.735	n. obs.: 42	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility regressions for the bond market (2-year, 5-year, 10-year bonds) under the low CPI regime.

I.3 Normal CPI Regime

Table 34: Association between news and market volatility: bond market, normal CPI regime

Target variable: $RV_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.0069	0.006	1.165	0.244	-0.005	0.019
| News CPI neg. |	0.0112	0.018	0.624	0.533	-0.024	0.046
| News GDP pos. |	0.0102	0.013	0.785	0.433	-0.015	0.036
| News GDP neg. |	0.0088	0.028	0.319	0.750	-0.045	0.063
| News U pos. |	0.0719	0.063	1.140	0.254	-0.052	0.196
| News U neg. |	0.0221	0.006	3.702	0.000	0.010	0.034
R^2: 0.811	Adj. R^2: 0.716	n. obs.: 52	Heteroscedasticity robust standard errors			

Target variable: $RV_{b,5y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.0155	0.017	0.910	0.363	-0.018	0.049
| News CPI neg. |	0.0356	0.041	0.878	0.380	-0.044	0.115
| News GDP pos. |	0.0160	0.035	0.457	0.648	-0.053	0.085
| News GDP neg. |	0.0248	0.081	0.305	0.760	-0.134	0.184
| News U pos. |	0.1808	0.204	0.887	0.375	-0.219	0.580
| News U neg. |	0.0409	0.011	3.712	0.000	0.019	0.062
R^2: 0.737	Adj. R^2: 0.703	n. obs.: 52	Heteroscedasticity robust standard errors			

Target variable: $RV_{b,10y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.0324	0.025	1.276	0.202	-0.017	0.082
| News CPI neg. |	0.0752	0.059	1.276	0.202	-0.040	0.191
| News GDP pos. |	0.0473	0.052	0.908	0.364	-0.055	0.150
| News GDP neg. |	0.0316	0.124	0.255	0.799	-0.211	0.274
| News U pos. |	0.2868	0.349	0.822	0.411	-0.397	0.970
| News U neg. |	0.0705	0.019	3.744	0.000	0.034	0.107
R^2: 0.767	Adj. R^2: 0.735	n. obs.: 52	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility regressions for the bond market (2-year, 5-year, 10-year bonds) under the normal CPI regime, based on 20 basis point buffer to each extreme regime.

J Additional Results: Bond Market, GDP Regimes

J.1 High GDP Regime

Table 35: Association between news and market volatility: bond market, high GDP regime

Target variable: $RV_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.0130	0.005	2.653	0.008	0.003	0.023
| News CPI neg. |	0.0142	0.008	1.865	0.062	-0.001	0.029
| News GDP pos. |	0.0250	0.028	0.890	0.373	-0.030	0.080
| News GDP neg. |	0.0034	0.024	0.145	0.885	-0.043	0.050
| News U neg. |	0.0156	0.009	1.734	0.083	-0.002	0.033
R^2: 0.783	Adj. R^2: 0.747	n. obs.: 35	Heteroscedasticity robust standard errors			
Target variable: $RV_{b,5y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.0162	0.014	1.135	0.257	-0.012	0.044
| News CPI neg. |	0.0358	0.020	1.825	0.068	-0.003	0.074
| News GDP pos. |	0.0327	0.060	0.546	0.585	-0.085	0.150
| News GDP neg. |	0.0015	0.052	0.028	0.977	-0.099	0.102
| News U neg. |	0.0335	0.024	1.394	0.163	-0.014	0.081
R^2: 0.641	Adj. R^2: 0.581	n. obs.: 35	Heteroscedasticity robust standard errors			
Target variable: $RV_{b,10y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.0242	0.021	1.161	0.246	-0.017	0.065
| News CPI neg. |	0.0486	0.032	1.505	0.132	-0.015	0.112
| News GDP pos. |	0.0834	0.111	0.753	0.452	-0.134	0.301
| News GDP neg. |	0.0079	0.096	0.082	0.934	-0.181	0.197
| News U neg. |	0.0710	0.039	1.797	0.072	-0.006	0.148
R^2: 0.731	Adj. R^2: 0.687	n. obs.: 35	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility regressions for the bond market (2-year, 5-year, 10-year bonds) under the high GDP regime.

J.2 Low GDP Regime

Table 36: Association between news and market volatility: bond market, low GDP regime

Target variable: $RV_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI neg. |	-0.0161	0.015	-1.075	0.283	-0.045	0.013
| News GDP pos. |	0.0334	0.007	4.882	0.000	0.020	0.047
| News GDP neg. |	0.0115	0.014	0.850	0.395	-0.015	0.038
| News U pos. |	0.0468	0.031	1.508	0.131	-0.014	0.108
| News U neg. |	0.0250	0.006	4.519	0.000	0.014	0.036
R^2: 0.700	Adj. R^2: 0.660	n. obs.: 42	Heteroscedasticity robust standard errors			
Target variable: $RV_{b,5y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI neg. |	-0.0360	0.030	-1.186	0.236	-0.095	0.023
| News GDP pos. |	0.0789	0.015	5.352	0.000	0.050	0.108
| News GDP neg. |	0.0342	0.031	1.106	0.269	-0.026	0.095
| News U pos. |	0.1268	0.063	2.005	0.045	0.003	0.251
| News U neg. |	0.0521	0.012	4.296	0.000	0.028	0.076
R^2: 0.691	Adj. R^2: 0.649	n. obs.: 42	Heteroscedasticity robust standard errors			
Target variable: $RV_{b,10y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI neg. |	-0.0657	0.043	-1.511	0.131	-0.151	0.020
| News GDP pos. |	0.1629	0.026	6.207	0.000	0.111	0.214
| News GDP neg. |	0.0695	0.048	1.448	0.148	-0.025	0.164
| News U pos. |	0.2607	0.102	2.550	0.011	0.060	0.461
| News U neg. |	0.0835	0.018	4.598	0.000	0.048	0.119
R^2: 0.767	Adj. R^2: 0.735	n. obs.: 42	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility regressions for the bond market (2-year, 5-year, 10-year bonds) under the low GDP regime.

J.3 Normal GDP Regime

Table 37: Association between news and market volatility: bond market, normal GDP regime

Target variable: $RV_{b,2y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.0212	0.006	3.432	0.001	0.009	0.033
| News CPI neg. |	0.0032	0.014	0.220	0.826	-0.025	0.031
| News GDP pos. |	0.0406	0.035	1.154	0.248	-0.028	0.110
| News GDP neg. |	0.0215	0.037	0.579	0.563	-0.051	0.094
| News U pos. |	0.0051	0.017	0.301	0.763	-0.028	0.038
| News U neg. |	0.0162	0.015	1.058	0.290	-0.014	0.046
R^2: 0.658	Adj. R^2: 0.613	n. obs.: 52	Heteroscedasticity robust standard errors			

Target variable: $RV_{b,5y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.0163	0.013	1.245	0.213	-0.009	0.042
| News CPI neg. |	0.0010	0.034	0.030	0.976	-0.065	0.067
| News GDP pos. |	0.0247	0.094	0.264	0.792	-0.159	0.208
| News GDP neg. |	0.0459	0.081	0.567	0.571	-0.113	0.205
| News U pos. |	0.0345	0.039	0.874	0.382	-0.043	0.112
| News U neg. |	0.0486	0.041	1.180	0.238	-0.032	0.129
R^2: 0.516	Adj. R^2: 0.453	n. obs.: 52	Heteroscedasticity robust standard errors			

Target variable: $RV_{b,10y}$	coef	std err	z	P> |z|	[0.025	0.975]
| News CPI pos. |	0.0415	0.026	1.607	0.108	-0.009	0.092
| News CPI neg. |	0.0005	0.056	0.009	0.993	-0.110	0.111
| News GDP pos. |	0.0595	0.153	0.390	0.696	-0.239	0.358
| News GDP neg. |	0.0811	0.143	0.566	0.572	-0.200	0.362
| News U pos. |	0.0600	0.073	0.821	0.412	-0.083	0.203
| News U neg. |	0.0808	0.067	1.205	0.228	-0.051	0.212
R^2: 0.543	Adj. R^2: 0.483	n. obs.: 52	Heteroscedasticity robust standard errors			

Notes: The table reports the estimation results of volatility regressions for the bond market (2-year, 5-year, 10-year bonds) under the normal GDP regime, based on 20 basis point buffer to each extreme regime.