The Voice of Monetary Policy

Yuriy Gorodnichenko
University of California, Berkeley, and NBER

Tho Pham
University of Reading

Oleksandr Talavera
University of Birmingham

First draft: October 31, 2020

This draft: June 14, 2022

Abstract: We develop a deep learning model to detect emotions embedded in press conferences after the Federal Open Market Committee meetings and examine the influence of the detected emotions on financial markets. We find that, after controlling for the Fed's actions and the sentiment in policy texts, a positive tone in the voices of Fed chairs leads to significant increases in share prices. Other financial variables also respond to vocal cues from the chairs. Hence, *how* policy messages are communicated can move the financial market. Our results provide implications for improving the effectiveness of central bank communications.

Keywords: monetary policy, communication, voice, emotion, text sentiment, stock market, bond market.

JEL: E31, E58, G12, D84

Acknowledgements: We thank Annette Vissing-Jorgensen, Michael Ehrmann, William English, Isaiah Andrews, anonymous referees, and seminar and conference participants at the University of Reading, the University of Birmingham, the University of East Anglia, UC Davis, FED Cleveland, FED Richmond, Economic Policy Working Group – Hoover Institution, NBER SI 2021, ESCoe Economic Measurement 2021 conference, Advanced Analytics: New Methods and Applications for Macroeconomic Policy conference, Non-traditional Data, Machine Learning and Natural Language Processing in Macroeconomics conference, and the Workshop on Empirical Monetary Economics for their comments. We are grateful to Eric Swanson for sharing his series of monetary policy shocks.

"How can a president not be an actor?" *Ronald Reagan (1980)*

"As Chairman, I hope to foster a public conversation about what the Fed is doing to support a strong and resilient economy. And one practical step in doing so is to have a press conference like this after every one of our scheduled FOMC meetings. ... [This] is only about improving communications." *Jerome Powell (2018)*[1]

"Monetary policy is 98% talk and 2% action, and communication is a big part."

Ben Bernanke (2022)[2]

1 Introduction

In a famous analysis, Mehrabian (1971) posited a 7-55-38 rule of communication: the words convey 7 percent of a message, the body language (gestures, facial expressions, etc.) accounts for 55 percent, and the tone delivers 38 percent. While the debate on exact percentages for each channel is open, it is clear that effective communication has to involve more than just words. Central banks have been relying increasingly on communication-based tools (e.g., forward guidance) to manage the public's expectations, but do central bankers utilize communication to its full potential?

Textual analyses of policy statements, minutes, and transcripts (e.g., Rozkrut et al., 2007; Hansen and McMahon, 2016; Hansen et al., 2018; Cieslak et al., 2019; Ehrmann and Talmi, 2020) suggest that central bankers' words carry considerable weight,[3] but little is known about the effects of their non-verbal communication. To shed further light on this issue, we use deep learning methods to quantify tone (vocal emotions) embedded in the answers given by Federal Reserve chairs during press conferences. To the best of our knowledge, our study is the first to examine the effects of central bank communications through the vocal dimension. In other words, we move beyond text analysis and study *how* policy messages are voiced and whether emotions channeled through voice tone can move financial markets. This offers a new tool for communicating with the public and for managing expectations.

[1] https://www.federalreserve.gov/mediacenter/files/FOMCpresconf20180613.pdf (Accessed on 21 July 2021)
[2] https://www.brookings.edu/events/ben-bernanke-the-fed-from-the-great-inflation-to-covid-19/ (Accessed on 24 May 2022)
[3] More generally, central banks have significant power to influence the macroeconomy and expectations. For example, a large number of studies have documented the effectiveness of policy announcements in moving financial markets (e.g., Kuttner, 2001; Gurkaynak et al., 2005) or shaping firm and household inflation expectations (e.g., Coibion et al., 2019; Enders et al., 2019; Lamla and Vinogradov, 2019).

We focus on policy communication during press conferences for several reasons. First, press conferences have been commonly used as an important communication tool. As suggested by Ehrmann and Fratzche (2007) and emphasized by Powell (2018)[4], press conferences, particularly the Q&A sessions, play a key role in helping financial markets and the general public to understand policy outlook and the interpretation of current economic conditions. Especially during times of high uncertainty, market participants tend to seek further guidance and clarification through press conferences' Q&A sessions. Second, press conferences allow policymakers to go off script and to communicate soft information via non-verbal channels, thus potentially influencing investors' decision making.[5] Finally, because video-audio recordings of press conferences are available in a consistent format, we can measure the tone of communication in a consistent manner and provide a systematic analysis of how voice tone can influence economic outcomes.

Specifically, we split a given FOMC (Federal Open Market Committee) press conference during the April 2011 – June 2019 period into audio segments corresponding to the response of the speaker to each question raised during the event. The split audios are then run through a machine learning model, which is trained to recognize emotions from voice variations. Each answer is rated as being one of three emotion classes: positive (happy or pleasantly surprised emotions), negative (sad or angry emotions), and neutral. After aggregating the tone of the answers for a given press conference, we examine how the tone affects a variety of financial variables at high frequencies. We find that the tone can materially move financial markets. For example, making the voice tone more positive by one standard deviation could raise S&P 500 returns by approximately 75 basis points. This order of magnitude is comparable to what one can observe after a one-standard deviation shock to forward guidance. In other words, the voice component can generate economically significant effects on the stock market. We also find that inflation expectations and exchange rates respond to variations in voice tone, e.g., a more positive tone leads to a decrease in expected inflation. At the same time, the evidence for the bond market is more mixed in our sample. These results suggest that policy communication is more nuanced than

[4] https://www.federalreserve.gov/mediacenter/files/FOMCpresconf20180613.pdf (Accessed on 21 July 2021)

[5] Caffi and Janney (1994), Visschedijk et al. (2013), Dricu et al. (2017), and others document that voice conveys information beyond the verbal content and that information contained in voice can affect decision making.

reading and posting prepared remarks and speeches. It appears that a certain level of acting skill may be helpful for ensuring that the public receives the policy message fully and accurately.

In addition to the vast literature on policy communication (see Blinder et al., 2008 and Coenen et al., 2017 for comprehensive surveys) and high-frequency analyses of monetary policy shocks (e.g., Kuttner, 2001, Gurkaynak et al., 2005, and many others), our study relates to research investigating the economic impacts of vocal cues. Using a sample of CEO speeches made during earnings conference calls, Hobson et al. (2012) find that the vocal markers of cognitive dissonance can predict the likelihood of irregular restatements of earnings reports. In a related study, Mayew and Venkatachalam (2012) show that market participants and analysts react to the affective states of managers expressed through vocal cues, such as happy or unhappy voices. For example, a positive affect is positively related to changes in stock recommendations and future unexpected earnings. These results suggest that the affective states contain useful information about a firm's fundamentals. In a more recent study, Hu and Ma (2020) find that positivity about a startup, shown through visual, verbal, and vocal dimensions, increases the likelihood of being funded, even if the startup's quality is low. Apart from having a different focus (central banking communication vs. CEO/manager communication), our study differs in terms of the tools employed to quantify the variation in tone. Earlier studies use commercial software or pre-trained machine learning algorithms for voice analysis, while we develop a customized deep learning model for detecting speech emotion. Our approach offers several advantages in terms of flexibility and the potential for further development and implementations. For example, we can fine-tune the model's parameters to achieve a higher accuracy rate, which is not a feature available to commercial software. Similarly, the customized model also allows us to adjust the number and class of emotions, which cannot be done with commercial software and pre-trained algorithms.

Curti and Kazinnik (2021), a concurrent paper that is closest in spirit to our work, examine the responses of the financial market to variations in the chair's facial expressions during post-FOMC press conferences. Using intra-day data, they find that negative facial expressions lead to lower stock-market returns. We view their results as reinforcing our message that non-verbal communication can move the financial markets, and hence, the non-verbal component is a potentially important channel of communication for economic players.

The rest of this paper is organized as follows. In the next section, we provide anecdotal evidence to justify the importance of the non-verbal channels in monetary policy communication. Section 3 describes the deep learning algorithms used to analyze the tone of voice and the sentiment of policy texts. In Section 4, we discuss our main results and their robustness to a series of robustness checks. We also provide some tentative interpretations and explanations of our results in this section. Finally, Section 5 concludes and discusses the implications of the results.

2 Non-verbal communication and monetary policy

Why would non-verbal communication matter? One explanation is that, due to asymmetric information between the public and the central bank, market participants tend to seek additional information through the aspects which are not explicitly "scripted", such as the choice of words, tone of voice, or the body language of the Fed chair. As with words, the non-verbal elements of communication can signal the Fed's perspective on the current and future economic outlook and the future course of monetary policy. The Fed chairs understand that press conferences are high-stake interactions with the public and the media and that communication is a complex process. For example, in her closing remarks for the FOMC meeting on 16 December 2015, Janet Yellen said:

> "*Okay. Boxed lunches will be available. If anybody wants to watch TV in the Special Library and see me get skewered at the press conference, please feel free. I will do my best to communicate the points that have been made here.* END OF MEETING."

Unsurprisingly, the Fed invests significant resources into preparation for press conferences, as well as post-conference analysis.

We also know that investors and the media watch and listen to FOMC press conferences, analyze the chair's voice, and attempt to interpret what it (i.e., the voice tone, emotion, etc.) implies. For example, in the International Quest Means Business program aired on 22 June 2011, Felicia Taylor, a business reporter for CNN, said while covering an FOMC press conference:

> *"The press conference, though, that is coming up in just a few minutes is where traders are really going to be looking for every little nuance. They want to see how [Ben Bernanke] is going to read into everything. The tone of his voice, his body language, his inflection, for any clue about the direction the markets are still looking for."*

This quote suggests that the press and financial market investors appear to pay attention to non-verbal communication.[6] Indeed, it is not uncommon for media reports to assess the non-verbal elements of press conferences. For example, the Wall Street Journal[7] reported Ben Bernanke's voice as either shaky or quavering during the first FOMC press conference on 27 April 2011. The tone of Jerome Powell's voice at the press conference on 16 September 2020 was perceived to be consistent with previous press conferences, which was interpreted as a signal of downplaying his dovish position.[8]

These anecdotes suggest that non-verbal communication could be an important channel. To study this channel systematically—in an objective, reproducible fashion—we build on recent advances in voice recognition technology and classify the voice tone of the Fed chairs into a spectrum of emotions. We, then, study how variations in voice tone (emotions) can affect financial variables.

3 FOMC speeches: voice and linguistic analysis

Our sample runs from April 2011 (when the first FOMC press conference was held) to June 2019. During this period, 68 meetings and 36 press conferences were held. For each meeting, the FOMC statement and the transcript of the press conference are obtained from the Federal Reserve (Fed) website. The press conference videos are downloaded from the Fed's official YouTube channel. We use only the audio component of these videos.[9] As the Q&A session is the only part of the press conference that is not scripted, our analysis focuses on the answers of the chair during the Q&A.

3.1 Voice Tone

In this section, we describe how we train a neural network model (a deep learning algorithm) to recognize emotions and refer the readers to Appendix B for more detail. Conceptually, it is

---

[6] See, e.g., https://cutt.ly/FzbnCBo; https://cutt.ly/fzbn1le (Accessed on 21 July 2021)
[7] https://www.wsj.com/articles/BL-REB-13815 (Accessed on 21 July 2021)
[8] https://cutt.ly/RzbnAm5 (Accessed on 21 July 2021)
[9] Video recordings can be used to study body language (e.g., facial expressions, gestures). However, videos are harder to analyze because cameras are moving (different speakers, different angles). In this respect, audio tracks offer a more consistent method of measurement.

necessary to create a mapping from various measures of sound waves into emotions. We also present descriptive statistics for variations in voice tone during post-FOMC press conferences.

3.1.1 Emotion detection using neural networks

Voice can be characterized by various parameters such as pitch (indicating the level of highness/lowness of a tone) and frequency (indicating the variation in the pitch) which are useful for determining the emotion of a speaker. Building on earlier research on voice recognition (e.g., Pan et al., 2012; Gao et al., 2017; Likitha et al., 2017; Bhavan et al., 2019), we use Librosa, a Python package, to extract the following vocal features. First, we extract 128 Mel Spectrogram Frequencies (Mel), which allows us to determine the level of loudness of a particular frequency at a particular time. Second, a chromagram with 12 chroma coefficients is extracted. The chromagram reflects the distribution of energy along 12 chroma bands (i.e., C, C#, D, D#, E, F, F#, G, G#, A, A#, and B) over time and, hence, can capture melodic and harmonic characteristics of audio. Finally, we extract 40 Mel-frequency cepstral coefficients (MFCCs), which are discrete cosine transformations of the Mel frequency spectrogram. Although MFCCs can be derived from Mel Spectrogram Frequencies, we find that using both types of features helps to improve the accuracy of the model. Note that the number of Mel spectrogram coefficients, MFCCs, and chroma coefficients can be adjusted to achieve more accurate predictions.

Once the inputs from an audio track are constructed, we need measures of emotions corresponding to the audio track so that we can build a model to classify sounds into emotions. We use two data sets where emotions can be matched to audio tracks. The first is the Ryson Audio-Visual Database of Emotional Speech and Song (RAVDESS). To create these data, 12 actors and 12 actresses spoke two statements in a neutral North American accent using eight different emotions (calm, happy, sad, angry, fearful, surprised, disgusted, and neutral). The second data set is the Toronto Emotional Speech Set (TESS). To create these data, two actresses spoke a set of 200 words using 7 emotions (happy, sad, angry, fearful, pleasantly surprised, disgusted, and neutral). These data sets are widely used in the computer science literature to build the speech emotion/expression systems (see, e.g., Verma and Mukhopadhyay, 2016; Gao et al., 2017; Choudhury et al., 2018; Bhavan et al., 2019; Andersson, 2020). As the fearful or disgust emotions are unlikely to arise during the Q&A sessions, we only use audios for five emotions: happy, (pleasantly) surprised, neutral, sad, and angry. We use 80% of RAVDESS and TESS as the training sample and the remaining 20% are used for testing.

After extracting the vocal features from each recording in RAVDESS and TESS, we use Keras, a deep learning API run on top of Google's machine learning platform, TensorFlow, to build a neural network model, i.e., a computing system consisting of different layers where each layer is a collection of different neurons (nodes). We illustrate the mechanism behind a neural network comprised of three fully connected layers in Figure 1. The first layer in this network is the input layer with three nodes and each node is an audio feature. The second layer is a hidden layer consisting of four nodes (HK_k, k=[1;4]) which are the activation functions of the input features IN_i (i=[1;3]). Particularly, a node HK_k is connected with the input through weight ($w_{k,i}$) and bias (b_k): $\sum_{i=1}^{3} IN_i \times w_{k,i} + b_k$. The weighted summations are passed through an activation function such as a binary step function, linear activation function, or non-linear activation function, to obtain the outputs O_k (k=[1;2]). Applying the same procedure to these outputs gives us the final output (i.e., the classification of emotions).

In this study, we build a fully connected network with an architecture specified as follows. The first layer takes 180 vocal features (128 Mel coefficients, 40 MFCCs, and 12 chroma coefficients) as inputs to produce 200 nodes as outputs. The second layer has 200 nodes which are connected with 200 nodes in the first layer through the linear activation function. The third layer has 200 nodes which are connected with 200 nodes in the previous layer through the linear activation function. The output layer has five nodes representing five emotions (happy, pleasantly surprised, neutral, sad, and angry). Given that our task is a multi-group classification task, we use the softmax activation function (i.e., normalized exponential function) to connect the nodes in this layer with 200 nodes in the second hidden layer.[10] To minimize overfitting, we add three dropout layers with a dropout rate of 0.3 after the input layer and each of the hidden layers. This means that 30% of the inputs are randomly set to 0 at each step during the training time (hence, only 70% of the inputs are retained for training). After training the model, we use the accuracy score to evaluate the model's performance:

[10] In other words, a linear model is applied to all layers except the output layer where a multinominal logistic model is applied. Thus, the output in the output layer is a probability distribution over five emotion classes and the sum of all probabilities is equal 1. The predicted emotion is the emotion class which has the highest probability.

$$Accuracy(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n} 1(\hat{y}_i = y_i) \tag{1}$$

where y and $\hat{y}$ are the true emotion and the predicted emotion, respectively, and n is the number of audio files in the testing dataset. The trained model gives us an overall accuracy score of 84%. Applying the same formula for each emotion, we obtain accuracy levels for predicting angry, sad, neutral, surprised, and happy emotions of 87%, 84%, 74%, 87%, and 80%, respectively.

A key advantage of our approach to classifying voice tone into emotions is its objectivity and reproducibility. Indeed, any interested researcher can apply this tool—or its variations—to trace every step of the approach and to measure sensitivity to various assumptions and data points. In contrast, using emotions obtained from human classification is not only much more costly but is also likely to involve judgement, biases, conflicting interpretations, and potentially too great a reliance on “reading between the lines”. However, humans may be better at detecting subtle tone variations.

3.1.2 FOMC audio input

We manually processed and split each FOMC press confference’s audio into smaller audio segements where each segement is a chair’s answer to a question. As an illustration of our approach, consider the first press conference given by Ben Bernanke on April 27, 2011. After giving opening remarks, he invited the audience to ask questions (11:23 mark of the press conference). The first question (about a weak GDP forecast) ended at 11:49. Bernanke’s response lasted until 12:59. Hence, the first audio segment runs for 1 minute and 10 seconds from 11:49 to 12:59. We process other segments in the same manner. In the end, our audio sample contains 692 answers from three speakers (Ben Bernanke, Janet Yellen, and Jerome Powell). The lengths of the audio segments vary from 10 seconds to more than 5 minutes, but most answers are between 1 and 3 minutes long. The number of audio segments per press conference ranges from 12 to 26, with an average of 19.

3.1.3 Emotion detection output

After training the neural network to recognize emotions from variations in vocal features (MFCCs, Chromagram, and Mel Spectrogram), we feed the audio tracks of the policymakers’ answers into the neural network. Each audio segment receives one predicted emotion (happy, pleasantly surprised, neutral, sad, or angry). We then classify each audio segment into positive (predicted emotion is "happy" or "pleasantly surprised"), negative (predicted emotion is "sad" or "angry"), or neutral. We aggregate the tone to the press conference level as follows:

$$VoiceTone = \frac{Positive\ answers - Negative\ answers}{Positive\ answers + Negative\ answers} \tag{2}$$

where $VoiceTone$ ranges from -1 (negative emotions) to +1 (positive emotions). We report descriptive statistics in Panel A of Table 1 and scores for each Q&A session in Appendix Table A1.

We find that Ben Bernanke, on average, had more positive emotions in his voice than Janet Yellen, who, in turn, had more positive emotions in her voice than Jerome Powell. Bernanke had five Q&A sessions with only positive emotions in his voice. In contrast, Jerome Powell had five Q&A sessions with only negative emotions. Janet Yellen's sessions always had a mix of positive and negative emotions. The average tone across these central bankers is close to zero. There is considerable within-speaker variation in tone, with Jerome Powell exhibiting the largest variation.

Although we do not have a standard benchmark to validate $VoiceTone$, we can use reports in the media and other external information to get a sense of how $VoiceTone$ aligns with other sources. For example, Ben Bernanke tends to get positive score values of $VoiceTone$ for most of the meetings, but his score is unusually low (close to -1) for the press conference on September 18, 2013. We can recall that the preceding press conference (June 19, 2013) resulted in the "taper tantrum" when Bernanke communicated that the Fed planned to scale back quantitative easing and surprised the markets. Between this press conference and the next (September 18), Bernanke and other Fed officials walked back on that message. In the September 18 press conference, Bernanke was on the defensive when he was repeatedly asked about the "mistake" made in the previous press conference and his personal responsibility for it.[11] He also had to say that the Fed was working with untested policies. In addition, Bernanke had to answer questions about his own future

[11] For example, Binyamin Appelbaum (New York Times) asked, "To what extent do you regard yourself as responsible for the tightening in financial conditions that you noted? Was it a mistake to talk about tapering in the way that you did in June and do you stand by your guidance that it will be appropriate?".

(i.e., Janet Yellen was expected to be officially nominated to chair the Fed) and his regrets. Thus, it is natural that he sounded relatively negative.

Consider now the press conference on September 17, 2015, when Janet Yellen had an unusually positive score (0.83). During this press conference, Yellen voiced her view that the economy has a lot more space to grow despite a relatively low unemployment rate of 6.3%, close to the natural rate perceived at the time, and a spike in inflation. The media reported that "…Yellen seemed slightly unconvinced by the Fed's insistence that the economy is improving"[12], "[Yellen] expressed scepticism that the official headline unemployment rate (6.3%) was an adequate reflection of labor market tightness"[13], and "Yellen struck a dovish tone"[14]. Perhaps in her attempts to signal her more positive outlook for the economy, Yellen sounded more positive than usual.

Powell's first press conference on March 21, 2018 had a low score (-1) when, according to the Financial Times[15], "Mr Powell seemed anxious to underline the uncertainty hanging over the outlook rather than sending up too many hawkish warning flares". A year later (March 20, 2019), Powell had one of his most positive scores in our sample (0.92). The media coverage suggested then that Powell "at every turn managed to out dove expectations"[16].

Obviously, these cross-checks should be interpreted as tentative. However, they suggest that there is some consistency between our scores on the one hand and the nuances detected by the media on the other hand.

3.2 Textual Analysis

Successful policy communication should utilize multiple channels to guide the public in a desired direction. To avoid a cacophony and, thereby, confusion, verbal and non-verbal messages should be congruent and reinforce each other. However, if policy communication is a concerted effort, how can one hear the voice (tone) of monetary policy? In the absence of exogenous variation in how

---

12 https://www.theguardian.com/commentisfree/2014/jun/18/fed-janet-yellen-economy-inflation (Accessed on 16 January 2022)
13 https://www.businessinsider.com/janet-yellen-dove-2014-6 (Accessed on 16 January 2022)
14 https://www.ftadviser.com/2015/10/06/investments/fixed-income/markets-no-longer-listening-to-fed-vGD6k9R3rbm3YQYknBjsTN/article.html (Accessed on 16 January 2022)
15 https://www.ft.com/content/1047a2a8-6f7c-11e8-92d3-6c13e5c92914 (Accessed on 16 January 2022)
16 https://www.afr.com/markets/equity-markets/asx-advances-on-dovish-fed-20190322-p516gq (Accessed on 16 January 2022)

policy messages are telegraphed to the public, identifying the effects of voice tone has to rely on controlling for the text sentiment of policy messages as well as policy actions. To this end, we employ state-of-the-art natural language processing (NLP) tools to quantify the sentiment of FOMC texts.

3.2.1 Sentiment analysis using BERT embeddings

To extract the word embeddings[17] of policy texts, we adopt Bidirectional Encoder Representations Transformer (BERT), the NLP algorithm developed by Google AI, which has several advantages over other tools. First, unlike context-free models (e.g., Word2Vec, GloVE), BERT generates an embedding representation of a word based on its surrounding context (contextual representation). Second, in contrast to unidirectional contextual models (e.g., ELMo, ULMFit), which create a word's representation based on previous words in the text, BERT is a bidirectional model which takes into account both preceding and subsequent context to generate the embeddings of a word.[18] As a result of these features, BERT has very high accuracy in interpreting texts (Devlin et al., 2018) and has been increasingly applied in economic research (see, e.g., Chava et al., 2019; 2021).

Although pre-trained BERT models can assign certain interpretations to texts (e.g., positive or negative), previous works on the textual analysis of policy communication have focused on the dovish-hawkish spectrum. To bridge our work to earlier studies, we need to undertake an additional step. Specifically, we create a customized training dataset that includes all FOMC statements released between 1997 and 2010 and the individual sentences in these statements. Each text in the training data is independently scored by several research assistants. The scores run from -10 (very hawkish) to +10 (very dovish). We calculate the average score for each text and classify a text as dovish (the average score ≥ 0.5), hawkish (the average score ≤ -0.5), or neutral (the average score is between -0.5 and +0.5). We then use the word embeddings obtained from BERT

---

[17] Simply put, word embeddings (mappings of words to vectors of real numbers) capture the semantics of the words (i.e., the meanings of the words) and the syntactic relationships between them (i.e., the grammatical structure).

[18] There are different versions of BERT embeddings, depending on the training data and the architecture. In this study, we use word embeddings generated from the BERT-base model (12 layers, 768 hidden states, 12 heads, and 110M parameters). As a robustness check, we also use the embeddings obtained from the Robustly Optimized BERT Pre-training Approach (RoBERTa) model (12 layers, 768 hidden states, 12 heads, and 125M parameters).

as the inputs for a neural network[19] tailored to identify the sentiment of monetary policy texts. In other words, we map texts processed with BERT into three categories: dovish, hawkish, and neutral. Applying formula (1), we obtain an overall accuracy score of 81% and the accuracy scores in predicting hawkish, neutral, and dovish sentiments are 85%, 79%, and 77%, respectively.

After applying this procedure (i.e., the BERT embedding model and the trained neural network for sentiment classification) to the text in the 2011-2019 sample, we aggregate the sentiment of the text from an FOMC statement, remarks, and Q&A as follows:

$$TextSentiment = \frac{Dovish\ text - Hawkish\ text}{Dovish\ text + Hawkish\ text} \tag{3}$$

where $Dovish\ text$ and $Hawkish\ text$ are the counts of respective paragraphs in the FOMC statements and opening remarks, as well as the counts of respective answers when a press conference is held. By construction, $TextSentiment$ is in the range of [-1;1] and a positive value indicates that an expansionary monetary policy is expected or is being implemented.

3.2.2 Sentiment analysis output

We find that the sentiment of texts during the terms of Yellen and Bernanke was generally more dovish than Powell's (Panel B of Table 1).[20] This pattern likely reflects that policy rate increases dominated during Powell's period in our sample. The within-speaker variation in the text sentiment is broadly similar across the Fed chairs. The correlation between sentiment for statements and remarks is discernibly positive ($\rho = 0.37$), while the correlation between the text sentiment of responses during Q&A sessions and the text sentiment of statements is slightly lower ($\rho = 0.21$). At the same time, the text sentiments for Q&A and remarks are correlated at $\rho = 0.13$. To measure

[19] The architecture of this neural network is as follows. The input layer is the output of the last layer of the BERT embedding model, which has the input dimension of 512 × 768. In the first hidden layer, a long short term memory (LSTM) layer is wrapped with a bidirectional layer (bidirectional LSTM). The LSTM layer has 512 units and is connected with the input layer through the hyperbolic tangent activation function. The 2-directional output of this bidirectional LSTM layer is passed on a global average pooling 1D layer to transform into 1-dimensional data. The third hidden layer is a dense layer with 512 nodes and the fourth hidden layer is a dense layer with 128 nodes. A dropout layer (the dropout rate of 0.1) is added after the first, second, and third hidden layers. The output layer has three nodes representing three sentiment categories (hawkish, neutral, and dovish). Similar to the neural network used for voice emotion classification, we use the softmax activation function to connect the output layer with the third hidden layer. See Appendix C for more detail.

[20] The text sentiment scores for each meeting are listed in Appendix Table A2.

the totality of the sentiment, we compute $TextSentiment$ using all three sources (statement, remarks, and Q&A responses). Given that we have a limited number of events in our sample, this approach allows us to save degrees of freedom, but our results are robust to using $TextSentiment$ measured for each source separately or in other combinations.

3.3 Co-movement in policy actions, words, and tone

To what extent do text sentiment and voice tone co-move? Although one might think that text and voice should be highly congruent, Figure 2 demonstrates that the relationship between these two channels of communication is more nuanced. Specifically, the positive messages conveyed in the tonality of voice are associated with more dovish statements in the accompanying texts. Hence, we observe congruence in words and tone, but this relationship is not perfect. For example, the correlation between the text sentiment in statements and the voice tone in the corresponding Q&A sessions is $\rho = 0.37$ (Spearman correlation is equal to 0.30). Similarly, the correlation between the Q&A tone and the text sentiment in remarks or Q&A is 0.48 and 0.29, respectively. Figure 2 shows that it is not uncommon to observe dovish texts and negative tonality. These results suggest that the tone of Q&A responses may generate variation in policy communication that is unrelated to the content of the texts of policymakers' statements, remarks, or even the Q&A responses themselves.

In a similar spirit, the variation in tone appears to be only weakly correlated with actual policy shocks (Panels A-C in Figure 3) as identified in Swanson (2021): a shock to the policy rate (FFR shock), a forward guidance (FG) shock, or an asset purchase (AP) shock. There is a slightly stronger correlation between voice tone and the stage of the policy cycle. Specifically, the correlation between the shadow rate (as measured in Wu and Xia, 2016)[21] and voice tone is -0.23 (for comparison, the correlation with FFR shocks is -0.19), i.e., the tone of voice becomes more negative as the policy rate increases.

These results suggest that communication is done via multiple channels and it is important to control for these channels if we are to isolate the effect of voice tone. However, these correlations are far from perfect, and thus we have independent variation in voice tone. In part, these imperfect correlations can reflect the nature of the sample period. For example, one would

[21] The updated series of the shadow rate are available at: https://sites.google.com/view/jingcynthiawu/shadow-rates. (Accessed on 21 July 2021)

expect that Fed funds rate movements (a standard measure of policy) and voice tone variations should be congruent. But the Fed funds rate was at the zero lower bound for most of the sample period and, hence, the correlation is naturally lower than one would expect in normal conditions. On the other hand, this period can offer us a greater chance of detecting the effects of policy communication because it is less clouded by potentially confounding factors such as changes in the Fed funds rate.[22]

4 Empirical analysis

In this section, we investigate whether voice tone can move various financial indicators. In particular, we estimate the following specification in the spirit of Jordà (2005):

$$\begin{aligned} Outcome_{t,t+h} = {} & b_0^{(h)} + b_1^{(h)} VoiceTone_t + b_2^{(h)} TextSentiment_t \\ & + b_3^{(h)} FFRShock_t + b_4^{(h)} FGShock_t + b_5^{(h)} APShock_t \\ & + b_6^{(h)} ShadowRate_t + b_7^{(h)} \mathbb{I}\{NoPressConference_t\} + error_t^{(h)} \end{aligned} \tag{4}$$

where t dates FOMC meetings. $VoiceTone_t$ measures the voice tone of the Q&A session at date t. $TextSentiment_t$ indicates the sentiment in the policy statement, remarks, and Q&A responses. $FFRShock_t$, $FGShock_t$, and $APShock_t$ are policy shocks identified using intraday data with a three-factor model by Swanson (2021). These policy shocks are normalized to have unit variance over a "typical" period (e.g., the FFR shock is normalized to have unit variance for the period that excludes zero lower bound). $ShadowRate_t$ is the shadow policy rate from Wu and Xia (2016). Policy shocks and the shadow rate control for "actions" of the Fed so that we can more cleanly identify the effects of voice tone on outcome variables. Note that we code $VoiceTone$ as equal to zero for FOMC meetings without Q&A sessions but our results are robust to focusing only on meetings with press conferences. We include the indicator variable $\mathbb{I}\{NoPressConference_t\}$ which is equal 1 when an FOMC meeting did not have a press conference.

[22] We do not make any normative statements about whether the observed variation in voice was helpful or not because we do not have a complete picture about the objectives of the Fed chairs in their press conferences. In other words, we do not know if a certain variation in voice was intentional or made in error. If the latter is the case, one may be concerned that voice control was inadequate and that unnecessary volatility has been introduced.

We estimate specification (4) for each horizon h ($h = [0; 15]$) separately and plot the estimated coefficients, e.g., $\left\{\hat{b}_1^{(h)}\right\}_{h=0}^{H}$, to illustrate the dynamics of the response to a form of policy action or communication. Note that while high-frequency analyses tend to find clear responses to policy announcements at the intraday frequencies (e.g., Kuttner, 2001, Swanson, 2021), we use the daily frequency which, given the dramatic volatility of some financial indicators, often yields statistically insignificant estimates (see, e.g., Gorodnichenko and Weber, 2016). However, one could expect that the response may build over time, consistent with the notion of "slow-moving" capital proposed by Duffie (2010) and Fleckenstein et al. (2014). Using daily series allows us to examine responses at longer horizons, which may be important for identifying policy actions and communication tools with durable effects. We will use estimate responses to policy shocks ($\left\{\hat{b}_3^{(h)}\right\}_{h=0}^{H}$, $\left\{\hat{b}_4^{(h)}\right\}_{h=0}^{H}$, $\left\{\hat{b}_5^{(h)}\right\}_{h=0}^{H}$) and text sentiment ($\left\{\hat{b}_2^{(h)}\right\}_{h=0}^{H}$) to benchmark responses to voice tone variations.

The outcome variables are daily financial indicators available from Thomson Reuters and other popular sources, including Yahoo Finance and Tiingo. We generally use prices of exchange-traded funds (ETFs) that track popular indices. For example, we use the SPY, an ETF fund that tracks the S&P 500 index, to measure the reactions of the stock market to policy shocks. We measure returns on ETF funds or similar securities as the log close price at date $t + h$ minus log open price at date t, e.g., $Outcome_{t,t+h}^{SPY} = \log\left(SPY_{t+h}^{close}\right) - \log\left(SPY_{t}^{open}\right)$. Hence, the return on the day of an FOMC meeting is the log difference between close and open prices.

As the sample is relatively small (68 FOMC meetings), we estimate specification (4) using nonparametric (accelerated) bootstrap methods to correct for possible biases in the estimates, as well as to construct confidence intervals with good coverage. Specifically, 90 percent bias-corrected confidence intervals are reported. As a robustness check, we estimate specification (4) with $VoiceTone$ as the only regressor. In further robustness checks, we will also explore the sensitivity of the estimates to including additional controls and other variations in the specification.

4.1 *Stock market reactions*

When we use the SPY ETF to measure the reactions of the stock market to policy actions and communications, we find that a more positive voice tone leads to an increase in share prices (Panel B of Figure 4). Specifically, the impact response (i.e., $h = 0$) of the stock market is weak and not statistically significant. Over time, the response builds up and after five days, the return on SPY reaches approximately 100 basis points for a unit increase in voice tone. The response levels off after the first few days and stays statistically significant at 10 percent. We observe this pattern irrespective of whether we include controls (Panel B) or not (Panel A) in specification (4).

The sentiment of the policy texts does not appear to have a statistically significant effect on the SPY in our sample, although the point estimates are positive, suggesting that a more dovish sentiment leads to a boom in the stock market. This finding is qualitatively in line with the results documented in the literature. For example, employing the high-frequency event study approach, Rosa (2011b) shows that surprise hawkish FOMC statements lead to a reduction in equity returns. However, using monthly data over the 1998 – 2014 period, Hansen and McMahon (2016) find a statistically insignificant reaction of stock markets to FOMC statements which focus on strong economic conditions.

The FFR shock does not have a statistically significant effect on the stock market, which likely reflects the fact that the sample period is dominated by the zero lower bound and that changes in the short-term policy rate may have provided a relatively limited outlook for the stance of monetary policy. Changes in the pace of asset purchases by the central bank (AP shock) also do not have a clear effect on the stock market, a finding consistent with Swanson (2021). Note that our sample does not include the first round of quantitative easing in 2009, which led to a strong stock market reaction (see, e.g., Chodorow-Reich, 2014; Krishnamurthy and Vissing-Jorgensen, 2011). For this sample period, however, a forward guidance shock leads to a persistent decrease in stock prices, in line with the intra-day responses estimated by Swanson (2021). This response is consistent with the signaling effect suggested by Campbell et al. (2012): an FG shock reveals that the Fed could be pessimistic about the state of the economy. The magnitude of the stock market response to a unit decrease in voice tone is approximately equal to the response we observe after a one-standard-deviation forward-guidance shock. When we use the Shapley decomposition of the R^2, we find that the absolute contribution of voice tone to R^2 is around 20 percent, which is slightly larger than that of the text sentiment indicator and is similar to the contribution of the FG shock.[23] Thus, the variation in voice tone has economically significant effects.

To understand the reaction of the stock market to policy actions and messages, we examine the response of the CBOE Volatility Index VIX (Figure 5), a popular measure of the stock market's expectations about future volatility. We also study the responses of futures on the VIX to provide us with a more refined sense of how policy can influence the outlook for volatility. Specifically, we use VIXY (Figure 6; VIX Short-Term Futures) and VIXM (Appendix Figure A7; VIX Mid-Term Futures) ETFs. We find that Fed actions (FFR/FG/AP shocks) tend to raise the volatility in the stock market. Consistent with this result, Ehrmann et al. (2019) document the greater responsiveness of treasury yields to macro news during weak forms of forward guidance which can be interpreted as evidence of greater uncertainty in relation to rate paths. Both a more positive tone of voice and more dovish text sentiment lead to a decrease in current and anticipated volatility. This result is in line with the notion that central bank communication can shape uncertainty about future economic conditions (Hansen et al., 2019). The variation in voice tone has economically significant effects: a unit decrease in the tone increases the volatility by an amount that is roughly equal to the increase after a one-standard-deviation shock to forward guidance.

Relatedly, monetary policy can convey information about the path of interest rates and thus reduce the interest rate risk (Hattori et al., 2016). To quantify the importance of this channel, we measure the interest rate risk with the following spread: $Outcome_{t,t+h} = log\left(\frac{P^{LQD}_{t+h,close}}{P^{LQD}_{t,open}}\right) - log\left(\frac{P^{LQDH}_{t+h,close}}{P^{LQDH}_{t,open}}\right)$, where P^{LQD} is the price of LQD ETF (investment grade corporate bonds) and P^{LQDH} is the price of interest rate hedged corporate bond LQDH ETF. A decrease in this measure indicates a decline in the perceived interest rate risk. Our results (Figure 7) suggest that a more positive tone leads to a reduction in investor expectations about interest rate risk. Consistent with this interpretation, and in line with the existing studies which document the impact of policy actions on bond risk premia (e.g., Hattori et al., 2016), we find that a forward guidance shock reduces uncertainty about the future path of interest rates. A unit decrease in voice tone and a one-standard-deviation increase in the FG shock generate similar responses of the spread, again

[23] Appendix Figure A9 plots the contribution by horizon.

pointing to the economic significance of the tone of voice. In contrast, an AP shock could signal a lower amount of interest rate risk in investor portfolios in the future, and thus, increase the perceived current interest rate risk, which is consistent with the analysis in Gorodnichenko and Ray (2017). Intuitively, asset purchases are a form of discretionary policy and the deployment of such a tool increases uncertainty about the future path of policy.

4.2 Bond market reactions

Kuttner (2001), Swanson (2021), and a number of others document a strong reaction of the bond market to monetary policy shocks. Consistent with these earlier works, we find (Figure 8) that the price of GOVT ETF (a fund covering U.S. government nominal debt) decreases in response to a forward guidance shock (i.e., yields rise) and increases in response to an asset-purchase shock (i.e., yields fall). FFR shocks do not lead to a statistically significant response in GOVT prices, which likely reflects the prominence of the zero-lower bound (ZLB) in our sample. In contrast to the strong responses of the stock market, the responses of the bond market to voice tone are not statistically significant (although we later document that voice tone can move spreads between nominal and real bonds). Similarly, text sentiment does not move GOVT prices materially. These findings are consistent with Cieslak and Pang (2020) and Ehrmann and Talmi (2020), who document that the bond market reaction to Fed communications is weak. Using ETFs for government debt with different maturities, we also examine if there could be a differential response across the maturity space. We find qualitatively similar responses for all maturities (Appendix Figures A1-A6), although the magnitudes of the responses to FFR/FG/AP shocks tend to be smaller for shorter maturities. While the responses of the bond market appear to be somewhat decoupled from the responses of the stock market, differentiated responses have been documented in the previous literature. For example, Lucca and Moench (2014) find that there is a pre-FOMC announcement drift in the stock market but a similar effect is not found for U.S. Treasuries.

An important dimension of monetary policy transmission is how policy can influence the interest rates faced by the corporate sector. While the bond market is highly integrated, the pass-through from U.S. government debt to corporate debt may be limited and nuanced. In our first attempt to address this question, we use the LQD ETF (a fund covering investment grade corporate bonds) and find that policy actions (FFR/FG/AP shocks) tend to move yields in the same direction as they move yields for U.S. government debt (Figure 9). Text sentiment does not have a statistically significant effect on LQD prices. A positive voice tone appears to elevate LQD prices (i.e., yields fall) for a few days after an FOMC meeting, but this effect is short-lived and statistically insignificant. The results are broadly similar when we use the IVR ETF (real estate investment trust; Figure 10) to gauge the responses of the real estate sector.

4.3 Inflation expectations

Management of inflation expectations is a key element of monetary policy (see Coibion et al., 2020 for a survey). To evaluate the success of policymakers in this matter, we use two popular metrics. The first one is the spread between nominal and inflation-protected U.S. Government bonds. Specifically, we use $Outcome_{t,t+h} = log\left(\frac{P^{GOVT}_{t+h,close}}{P^{GOTV}_{t,open}}\right) - log\left(\frac{P^{TIP}_{t+h,close}}{P^{TIP}_{t,open}}\right)$ as a measure of the spread, where P^{GOVT} is the price of GOVT ETF (nominal bonds) and P^{TIP} is the price of TIP ETF (inflation-protected bonds). An increase in this spread can be interpreted as a decrease in expected inflation. The second is the GLD ETF, a fund that tracks the gold spot price. This ETF is used as a hedge against inflation: an increase in the price of GLD signals higher expected inflation. Although neither of these measures is perfect (e.g., the spread varies not only due to changes in inflation expectations, but also with changes in liquidity conditions; gold prices can move for reasons unrelated to inflation), these two measures are consistently available and are based on reasonably deep markets.

We find that the responses of the GOVT-TIP spread (Figure 11) and GLD (Figure 12) paint a similar picture. As before, the FFR shock does not have a clear impact. The FG shock lowers inflation expectations, while the AP shock raises inflation expectations. More dovish text sentiment appears to raise inflation expectations (the GLD price increases), but this response is not statistically significant. Moreover, it does not seem to have support from the GOVT-TIP spread, which appears to increase (i.e., expected inflation is lower) and the effect is statistically significant. The impact response of the GOVT-TIP spread to a positive tone of voice is close to zero, but the spread gradually increases (thus, signaling lower expected inflation) and peaks after about 10 days. The GLD price has similar dynamics (i.e., lower expected inflation), but the estimates are less precise. Hence, voice tone seems to have an independent effect on inflation expectations. One may conjecture that a positive tone plays a signaling role: a happy tone of the Fed chair indicates satisfaction with future inflation dynamics.

4.4 Exchange rate

The exchange rate is an important channel for monetary policy transmission in the increasingly globalized economy. To shed further light on how policy actions and communication can work via this channel, we examine the responses of two key exchange rates: dollar/yen (JPY; Figure 13) and dollar/euro (EURO; Figure 14).[24] We find that policy actions generally lead to mixed reactions across currencies in our sample. For example, after an FFR shock (monetary tightening), the dollar depreciates against the euro (although the effect is not statistically significant) and appreciates against the yen (for the first five days after the FOMC meetings).[25] After a more dovish text sentiment (the opposite to monetary tightening), the dollar appreciates against the euro while the response of the dollar/yen exchange is close to zero. Similarly, a more positive tone of voice leads to an appreciation of the dollar against the euro, but there is neither a statistically nor economically significant response for the dollar/yen exchange rate. While somewhat unexpected, the relatively lower level of reaction of the dollar/yen exchange rate to the monetary policy shocks has also been observed in other studies (e.g., Fatum and Scholnick, 2008; Rosa, 2011a).

4.5 Robustness checks

To assess the sensitivity of our findings to additional measurements and assumptions, we perform a series of robustness checks that may be grouped into two categories. To isolate the effect of policy shocks on financial and macroeconomic variables, it is important to control for the information set available to economic agents at the time when policy changes are announced. To this end, the first category explores the robustness of our results to additional controls or alternative specifications. As we have a small sample size, we typically include one additional control at a time. Given the importance of verbal communication, the second category focuses on alternative

[24] We also report results for the pound/dollar (GBP) exchange rate in Appendix Figure 8.

[25] Analyses using intraday data (e.g., Swanson, 2020) find that the dollar appreciates against the euro and yen after a FFR shock.

measures of sentiment in policy texts. To save space, we present results for the stock market responses, but we reach similar conclusions for other variables.

4.5.1 Additional controls

While most of the FOMC announcements did not overlap with the releases of other macro data over the examined period, there are certain exceptions. In particular, five FOMC announcements (none of which was accompanied by a press conference) were released on the day when gross domestic product (GDP) data were unveiled. In addition, seven FOMC announcements (six of which were followed by a press conference) were made on the same day as consumer price index data releases. As macroeconomic news can move financial markets (Gurkaynak et al., 2005), we introduce an additional control in specification (4), the Citigroup Economic Surprise Index, which aggregates macroeconomic surprises in released data into one indicator. We find that voice tone continues to move the stock market (Panel A of Figure 15). We also do not observe important changes in the estimates when we control for the volume of corporate earnings data releases (Panel B, Figure 15).[26]

One may be concerned that important variations in voice tone coincide with the media cycles that can affect the stock market. To address this potential confounder, we collect FOMC-related news coverage from Nexis Uni, a popular news database. We search for all news items which contain "FOMC" and one of the following keywords: "interest rate", "monetary", "federal funds rate", or "fed funds rate". The search results are restricted to English news generated by U.S. media outlets. We further exclude government sources that simply announce the events and the press conference transcripts such as press releases. After screening and cleaning, the final dataset consists of 23,275 news articles covering the 01/01/2011 – 15/07/2019 period. The news dataset provides us with information on the publication date of the news articles, as well as their content. We then use our BERT-based sentiment algorithm to process the news and construct a measure of media sentiment about monetary policy for each day t:

$$MediaSentiment_t = \frac{Dovish\ articles_t - Hawkish\ articles_t}{Articles_t}$$

[26] Data on the number of corporate earnings announcements of the U.S. listed firms during the 2011 – 2019 period were scrapped from Yahoo Finance's earnings calendar (https://finance.yahoo.com/calendar/earnings).

To smooth out noise, we compute 5-day averages of $MediaSentiment$ before each FOMC meeting. Including this control does not materially change our results (Panel C, Figure 15).

As we indicated above, Fed chairs vary in their voice tones. As their tenures coincided with different phases of the policy cycle (recall that we use the shadow policy rate to control for the phase of the policy cycle), one may be concerned that the phases may confound the voice tone variation. To address this potential issue, we use fixed effects for the Fed chairs. Although these fixed effects are very demanding on the data given our small sample size, we find that using within-chair variation in voice tone continues to result in voice tone moving the market (Panel D). Relatedly, given the negative media bias in covering monetary policy (Berger et al., 2011), the financial market responses may be driven by negative news, which may also correlate with the policy cycle (e.g., raising interest rates may be interpreted as bad news for the economy). To assess the quantitative importance of this potential asymmetry, we modify specification (4) to give us separate regressors for positive and negative voice tone measures. While the estimates are noisier, we find that the responses to positive and negative voice tones are similar in the first 10 or so days, but the effects then appear to be stronger for negative news (Panel E).

4.5.2 Alternative textual sentiment

Although our baseline analysis uses BERT, a highly accurate natural language processing (NLP) tool, we want to explore if using alternative tools to quantify the sentiment of policy texts can affect our estimates for responses to voice tone. First, we employ the Robustly Optimized BERT Pre-training Approach (RoBERTa), a modified version of BERT that was developed by Facebook AI (Liu et al., 2019). Compared to the original BERT, RoBERTa is trained on more data, larger batches, longer sequences, and dynamically changes the masking pattern applied to the training data while removing the next sentence prediction objective. At the cost of being significantly more computationally expensive, RoBERTa can yield a modest improvement in accuracy. Second, as mentioned earlier, the pre-trained BERT sentiment classification is based on a corpus of training texts that is not tailored for monetary policy. As a result, we have to fine-tune the model to convert BERT's word embeddings into the hawkish/dovish spectrum. As an alternative to our approach, we use FinBERT (Araci, 2019), a pre-trained BERT model which was fine-tuned for the sentiment analysis of financial texts (rather than general texts).

Third, we use the search-and-count approach which has been widely used in the literature (e.g., Apel and Grimaldi, 2014; Neuhierl and Weber, 2019). In short, we build lists (dictionaries) of nouns, adjectives, and verbs that can indicate the stance of monetary policy (hawkish or dovish) and the strength of economic outlook. We then compute the frequency of words in these dictionaries (Appendix D1 provides more detail). While this approach is less sophisticated than BERT, it is particularly transparent and easy to implement. Fourth, we ask a group of research assistants to score texts in policy statements, remarks, and Q&A sessions for the 2011-2019 sample. The scores vary from -10 (very hawkish) to +10 (very dovish). For each FOMC meeting, we compute the average score across texts and research assistants. While this approach may be more accurate in detecting the nuances of human communication, it is more subjective (and hence potentially less reproducible) than BERT.

Panels F-I of Figure 15 show that using alternative measures of text sentiment does not affect our conclusion that voice tone can move the financial markets. In Appendix D2, we report results for additional robustness checks. In these checks we apply an approach similar to Kozlowski et al. (2019) and Jha et al. (2021) to measure the *intensity* of dovishness/hawkishness of the policy texts and we experiment with allowing non-linear terms in text sentiment. In short, we find similar results.

One could also argue that our measure of voice tone simply captures some features of the policy texts which were not fully accounted for in our text sentiment measure. While we cannot rule out this alternative explanation completely, we note that the results on the tone of voice are robust to different measures of text sentiment, ranging from human classification to state-of-the-art methods in machine learning. This robustness suggests that the voice tone measure could capture additional information which goes significantly further than the message captured by the text sentiment. As a result, voice tone and other forms of non-verbal communication can expand the toolkit for managing the expectations of the public.

4.5.3 High-frequency analysis

Our results suggest that the effect of voice tone builds up gradually with only relatively small responses observed on impact at the daily frequency. We use a number of regressors to control for possible confounders (e.g., a piece of macroeconomic news is released on the day of a press conference). In this section, we use intra-day data to zoom in on the high-frequency movements of asset prices. While this approach can sharpen our identification, it also likely amplifies noise as voice tone measured with error and financial markets can take time to process non-verbal cues from the Fed (recall that voice tone is a flow rather than a stock).

With this caveat in mind, we generate the precise timing (down to a second) for each answer during press conferences. We then match SPY ETF prices to the timings. Finally, we estimate the following specification:

$$Outcome_{t\to t+h,m,s} = b_0^{(h)} + b_1^{(h)} VoiceTone_{t,m,s} + b_2^{(h)} TextSentiment_{t,m,s}$$

$$+\lambda_m + \gamma_s + error_t^{(h)} \qquad (5')$$

where t is the start time of the answer to the s^{th} question in meeting m, $t + h$ is h minutes after the end of the answer, λ_m is the fixed effect for meeting m, and γ_s is the fixed effect for the order of questions (i.e., question number s in a press conference). $VoiceTone$ takes values -1 (negative), 0 (neutral), and 1 (positive). $TextSentiment$ takes values -1 (hawkish), 0 (neutral), and 1 (dovish). $Outcome_{t\to t+h,m,s}$ is the price change for SPY between t and $t + h$.

We also estimate a modified version of specification (5'):

$$Outcome_{t_0\to t+h,m,s} = b_0^{(h)} + b_1^{(h)} VoiceTone_{t_0\to t,m,s} + b_2^{(h)} TextSentiment_{t_0\to t,m,s}$$

$$+\lambda_m + \gamma_s + error_t^{(h)} \qquad (5'')$$

where t_0 is the start of the answer for the first question in a press conference. $Outcome_{t_0\to t+h,m,s}$ measures the return between t_0 and h minutes after time t when the answer to question s in press conference m ends. $VoiceTone_{t_0\to t,m,s}$ measures the cumulative tone of answers (calculated as in equation (2)) between t_0 and t. $TextSentiment_{t_0\to t,m,s}$ measures the cumulative text sentiment of answers (calculated as in equation (3)) between t_0 and t. As there is clear dependence in the variables by construction, we use meetings as strata for bootstrap.

Specification (5'), which we call the "flow" specification, examines the reaction of a financial variable answer by answer. In contrast, the "cumulative" specification (5'') focuses on the response of a financial variable to information conveyed since the start of a press conference. The main advantage of the "flow" specification is that the unit of analysis is an answer. However, each answer may be a noisy measure of policy stance and, hence, one may also want to use the "cumulative" specification which likely attenuates the noise. Note that both specifications include meeting fixed effects meaning that we control for a broad range of factors including the macroeconomic environment and the Fed chair's personality.

As shown in Figure 16, we observe a small and statistically significant response of stock prices to voice tone on impact: a positive voice tone raises the stock market by approximately one basis point. For the minute that follows the answer, we cannot reject the null of no response and we cannot reject the null of the stable response (i.e., the impact response is equal to subsequent responses). The "cumulative" specification suggests that, as the voice tone becomes clearer during the course of a press conference, the point estimates gradually increase with the horizon, but the estimates continue to be imprecise. The magnitude of the response is smaller than the magnitude observed in the analysis with daily data. This is consistent with our conjecture that it takes time for the market to interpret signals from the tone of voice.

4.6 Discussion and additional analysis

In general, our findings shed new light on the effectiveness of press conferences as a central bank communication tool. We show that, just as the actions of the Fed move financial markets, so too does the vocal aspect of FOMC press conferences. The vocal dimension of the central bank communication appears to convey information beyond that found in the content of the text, and market participants form their expectations and make their decisions based on that information.

4.6.1 What is communicated?

The estimated responses suggest that a more positive voice tone of a Fed chair leads to rate risk reduction, lower expected volatility, depressed inflation expectations, and increased stock prices. The exact information that is communicated is open to further inquiry, but tentative interpretations are possible.

One interpretation is that the tone of voice effectively works as a form of forward guidance. For example, a positive voice tone could signal that the Fed is unlikely to change the policy stance in the near future. If rate risk is attenuated due to a no-change-in-policy signal, then volatility due to policy shocks is diminished, which is reflected by VIX and futures on VIX. Given that our sample is dominated by the zero lower bound, a lower rate risk then means lower inflation expectations (i.e., the Fed does not see a need to raise interest rates to fight inflation). If interest rates are unlikely to increase and the perceived future volatility is lower, the discount factor for future earnings could be lower, thus, pushing up the value of stocks. Furthermore, a decrease in policy uncertainty can have a direct positive effect on the economy (e.g., Husted et al., 2020). In line with this interpretation, the estimated responses to forward guidance shocks and to voice tone are qualitatively similar for many variables.

Alternatively, these responses reflect some forms of the Fed information effect (Romer and Romer, 2000), i.e., the notion that the Fed has superior information about the current or future state of the economy. For instance, the Fed chair may be satisfied with the inflation dynamics and the pace of the economic recovery after the Great Recession. Through a positive voice tone, they can communicate that monetary tightening aimed at fighting inflation is a lower probability event that reduces rate risk and uncertainty. A positive voice tone can also signal a brighter economic outlook which can reduce uncertainty (uncertainty is countercyclical) and raise expectations about future cash flows. These forces could generate a boom in the stock market. So which effect is at play? If the tone of voice was simply a form of forward guidance, one would expect similar effects of the tone of voice and the forward guidance shock across asset classes. The differences in results for the bond market appear to be inconsistent with this interpretation. However, it has been documented that the bond market reactions to Fed communication could be weaker than, or even different from, the stock market reactions (e.g., Lucca and Moench, 2014; Cieslak and Pang; 2020; Ehrmann and Talmi, 2020). Taken together, while there is suggestive evidence in favour of both the information effect and forward guidance effect, we are not able to provide conclusive evidence on which effect dominates with our data.

Hence, answering the question of which of the above accounts—and there could be other explanations—is a more accurate rationalization of financial market reactions to the non-verbal communication remains a challenging (separating forward guidance and information effects is complex, see e.g., Bauer and Swanson, 2020) but fruitful avenue for future research. Notwithstanding this issue, it is clear that the tone of voice can move multiple financial variables, hence, suggesting that the estimated responses are unlikely to be statistical flukes and are likely to capture some systematic forces in the data.

4.6.2 The shape of the response

The difference in the shape of responses to various forms of policy communication and actions is another area for further research. Specifically, in contrast to step-like contemporaneous responses of financial variables to the Fed's actions, the estimated responses to voice tone (and text sentiment) tend to build gradually over time with weak contemporaneous reactions. We can offer several conjectures to rationalize this pattern of the responses.

First, the tone may be a leading indicator for subsequent policy communication by the Fed chair and other officials. As more information is revealed progressively by the Fed via formal and informal channels, financial variables could take time to respond.[27] Indeed, it is not unusual for Fed officials to organize speeches and press conferences aimed at clarifying the position of the Fed after FOMC meetings. Perhaps the most striking example of such follow-up policy communication happened after the "taper tantrum" episode. At the press conference on June 19, 2013, Bernanke hinted at a reduction of the quantitative easing program, which led to significant movements in the financial markets. To contain these gyrations, a number of Fed officials rushed to clear up any potential confusion about the central bank's intentions.

To assess the plausibility of this channel, we focus on Twitter, a social media platform that central banks have increasingly used to communicate with non-experts (Ehrmann and Wabitsch, 2022). Specifically, we obtain tweets published on the Fed's Twitter accounts (i.e., the Board of Governors' account and the regional Fed accounts) and apply the trained BERT model tailored for central bank communication (as used in the main analysis) on these tweets. The aggregate sentiment of the Fed's tweets on day h = [1,15] after the press conference is measured as: $TweetSentiment_h = \sum_{i=0}^{h} Sentiment_i$. We then use specification (4) to estimate the response of this tweet sentiment measure to the tone of voice measure. The results reported in Panel A of Figure 17 suggest that, after a press conference with a more positive tone of voice, the sentiment of the Fed's tweets is more dovish. This pattern is consistent with the gradual amplification of

[27] We thank William English for suggesting this channel.

policy messages communicated initially via voice tone that are gradually incorporated into asset prices.

Second, the delayed response may capture the dynamics of trading where more informed/attentive investors (e.g., those who are better at reading—perhaps subconsciously—non-verbal communication at press conferences) move first and the initial momentum is then amplified by less informed investors, a mechanism explored in other contexts (e.g., Bikhchandani et al., 1992; Zhou and Lai, 2009). Duffie (2010) and others argue that this can give rise to "slow capital" which can generate predictable movements in stock returns. As discussed extensively in Lucca and Moench (2015), the main challenge of explanations based purely on information or financial frictions is how one can rationalize a delayed or otherwise predictable response to one element of policy events (e.g., the predictable pre-FOMC announcement drift in stock returns) and a sharp, largely unforecastable response to another element of policy events (e.g., investors are not systematically surprised by changes in the Fed funds rate). We conjecture that information frictions are exacerbated for policy messages communicated via the tone of voice and other non-verbal channels. Indeed, market participants cannot rely on readily available objective measures of non-verbal communication that should trigger trading activities. It can take time to filter out that part of policy communication (perhaps by rewatching a press conference)[28] and to form a narrative in the market that can justify an adjustment in asset prices. To the extent that signals in non-verbal communication are decoupled from more standard measures of policy stance (recall that these objects are not perfectly correlated), one can observe rapid responses to changes in policy that are well understood (e.g., changes in the Fed funds rate) and delayed responses to less understood changes in policy (e.g., voice tone).

Third, media coverage may be an important force in financial markets (see Tetlock (2014) for a survey of this literature). Furthermore, financial markets may fail to incorporate information efficiently so that asset prices can respond to dated news. For example, Huberman and Regev (2001), Carvalho et al. (2011), Tetlock (2011), and others document that financial markets react even to "stale news" (i.e., information that has been previously disclosed) covered by the media. To the extent that media coverage develops gradually while a consensus view about the message at a press

[28] There is some anecdotal evidence that information revealed at FOMC press conferences dissiminates gradually. For example, by 21 March 2021, the FOMC press conference held on 17 March 2021 had been watched 55,172 times on Yahoo Finance's YouTube channel and 143,093 times on CNBC's YouTube channel. By the time of writing this (December 2021), the number of views had risen to 59,988 and 174,148, respectively.

conference is being reached, one then may observe a gradual response of the financial markets to variations in voice tone. Building on our exercise in section 4.5.1, we can examine how the sentiment of media coverage for monetary policy evolves after FOMC meetings. Specifically, we use specification (4) with $MediaSentiment_t$ as the dependent variable to construct an impulse response of media sentiment. We find a clear hump-shaped dynamic (Panel B, Figure 17): after a positive voice tone, the media sentiment becomes more dovish on impact, it builds up for the next few days, peaks after around a week and then converges back to zero. This pattern, coupled with the evidence of investor reactions to the media sentiment documented in the literature, suggests that FOMC-related news coverage and sentiment could be a channel through which the tone of voice can move financial markets and continued media coverage can contribute to the delayed responses.

5 Concluding remarks

Press conferences are an important communication tool for delivering and explaining monetary policy decisions to the public. Unlike press releases, transcripts, or minutes, a press conference contains both verbal and non-verbal channels. The latter offers an opportunity to communicate "soft" information. Machine learning applied to text analysis allowed researchers to more accurately measure messages in written policy documents in order to quantify the importance of soft information. Other parts of communication (emotions, moods, tones, body language) could be equally (if not more) important, thus, potentially enriching the policy toolkit. However, these forms of communication have proven to be particularly difficult to quantify. Building on recent advances in voice recognition and deep learning, we attempt to shed new light on the effects of non-verbal policy communication.

Our analysis of variation in the Fed chairs' voice tone during Q&A sessions after FOMC meetings shows that non-verbal communication can have a statistically and economically discernible effect on a variety of financial indicators. For example, our results suggest that the voice tone used in policy communication may have a significant effect on the stock market to a much greater extent than to that which is contained in the Fed's actions or actual words (texts). This reaction is consistent with the Fed communicating a more positive outlook for the economy and a lower probability of monetary tightening in the future. Inflation expectations and exchange rates also respond to voice tone. In contrast, the bond market appears to have more mixed reactions to vocal cues from the Fed chairs.

Although future research should dig deeper into understanding the nuances of using voice to communicate policy, our results clearly have important policy implications. *How* messages are spoken appears to be potentially as important as the content of the messages. That is, non-verbal communication is potentially a new instrument to deliver information to the public. The Fed watchers routinely sieve through policy texts to identify and interpret minute variations in words (e.g., a change from "modest" to "moderate"). With advances in voice/face recognition, one may expect another arms race in policy communication and, hence, one needs to be cautious when using non-verbal cues as a policy communication tool in order to avoid any unintended effects. This does not make the job of central bankers easier and potentially adds another qualification (voice control) to the job requirement. This also may become a prerequisite for any other roles which use a public arena for policy communication. Indeed, to paraphrase Ronald Reagan, how can a Fed chair not be an actor?

References

Andersson, K., 2020. "Speech Emotion Recognition in Neurological Disorders Using Convolutional Neural Network." In *Brain Informatics: 13th International Conference Proceedings* 12241: 287. Italy: Springer, 12241, p. 287.

Apel, M. and Grimaldi, M., 2014. "How Informative Are Central Bank Minutes?" *Review of Economics* 65(1):53-76.

Araci, D., 2019. "FinBERT: Financial sentiment analysis with pre-trained language models." arXiv preprint arXiv:1908.10063.

Bauer, M., and Swanson, E., 2020. "The Fed's Response to Economic News Explains the 'Fed Information Effect'" NBER Working Paper w27013.

Berger, H., Ehrmann, M., and Fratzscher, M., 2011. "Monetary policy in the media." *Journal of Money, Credit and Banking* 43(4):689-709.

Bhavan, A., Chauhan, P., and Shah, R.R., 2019. "Bagged Support Vector Machines for Emotion Recognition from Speech." *Knowledge-Based Systems* 184:104886.

Bikhchandani, S., Hirshleifer, D., and Welch, I., 1992. "A theory of fads, fashion, custom, and cultural change as informational cascades." *Journal of Political Economy* 100(5):992-1026.

Blinder, A.S., Ehrmann, M., Fratzscher, M., De Haan, J., and Jansen, D.J., 2008. "Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence." *Journal of Economic Literature* 46(4):910-945.

Caffi, C. and Janney, R.W., 1994. "Toward a Pragmatics of Emotive Communication." *Journal of Pragmatics* 22(3-4):325-373.

Campbell, J., Evans, C., Fisher, J., and Justiniano, A., 2012. "Macroeconomic Effects of Federal Reserve Forward Guidance." *Brookings Papers on Economic Activity*:1-80.

Carvalho, C., Klagge, N., and Moench, E., 2011. "The persistent effects of a false news shock," *Journal of Empirical Finance* 18:597–615.

Chava, S., Du, W. and Malakar, B., 2021. "Do Managers Walk the Talk on Environmental and Social Issues?". Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3900814.

Chava, S., Du, W. and Paradkar, N., 2020. "Buzzwords?". Available at SSRN: https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3862645.

Chodorow-Reich, G., 2014. "Effects of Unconventional Monetary Policy on Financial Institutions." *Brookings Papers on Economic Activity*:155-227.

Choudhury, A.R., Ghosh, A., Pandey, R., and Barman, S., 2018. "Emotion Recognition from Speech Signals using Excitation Source and Spectral Features." In *2018 IEEE Applied Signal Processing Conference*. IEEE, pp.257-261.

Cieslak, A. and Pang, H., 2020. "Common Shocks in Stocks and Bonds." NBER Working Paper w28184.

Cieslak, A. and Vissing-Jorgensen, A., 2021. "The economics of the Fed put." *The Review of Financial Studies* 34(9):4045-4089.

Cieslak, A., Morse, A. and Vissing-Jorgensen, A., 2019. "Stock Returns over the FOMC Cycle." *Journal of Finance* 74(5):2201-2248.

Coenen, G., Ehrmann, M., Gaballo, G., Hoffmann, P., Nakov, A., Nardelli, S., Persson, E., and Strasser, G., 2017. "Communication of Monetary Policy in Unconventional Times," European Central Bank Working Paper Series 2080.

Coibion, O., Gorodnichenko, Y., and Weber, M., 2019. "Monetary Policy Communications and Their Effects on Household Inflation Expectations." NBER Working Paper w25482.

Coibion, O., Gorodnichenko, Y., Kumar, S., and Pedemonte, M., 2020. "Inflation Expectations as a Policy Tool?" *Journal of International Economics* 124(C):103297.

Curti, F., and Kazinnik, S., 2021. "Let's Face It: Quantifying the Impact of Nonverbal Communication in FOMC Press Conferences". Available at SSRN: https://ssrn.com/abstract=3782239.

Devlin, J., Chang, M.W., Lee, K., and Toutanova, K., 2018. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

Dricu, M., Ceravolo, L., Grandjean, D., and Frühholz, S., 2017. "Biased and Unbiased Perceptual Decision-making on Vocal Emotions." *Scientific Reports* 7(1):1-16.

Duffie, D., 2010. "Presidential Address: Asset Price Dynamics with Slow-Moving Capital," *Journal of Finance* 65(4):1237-1267.

Ehrmann, M. and Fratzscher, M., 2007. "The Timing of Central Bank Communication." *European Journal of Political Economy* 23(1):124-145.

Ehrmann, M. and Talmi, J., 2020. "Starting from a Blank Page? Semantic Similarity in Central Bank Communication and Market Volatility." *Journal of Monetary Economics* 111:48-62.

Ehrmann, M. and Wabitsch, A., 2022. "Central bank communication with non-experts–A road to nowhere?" *Journal of Monetary Economics* 127:69-85.

Ehrmann, M., Gaballo, G., Hoffmann, P., and Strasser, G., 2019. "Can more public information raise uncertainty? The international evidence on forward guidance" *Journal of Monetary Economics* 108(C):93-112.

Enders, Z., Hünnekes, F., and Müller, G.J., 2019. "Monetary Policy Announcements and Expectations: Evidence from German Firms." *Journal of Monetary Economics* 108:45-63.

Fatum, R. and Scholnick, B., 2008. "Monetary Policy News and Exchange Rate Responses: Do Only Surprises Matter?" *Journal of Banking & Finance* 32(6):1076-1086.

Fleckenstein, M., F. Longstaff, and H. Lustig, 2014. "The TIPS-Treasury Bond Puzzle." *Journal of Finance* 69(5):2151-2197.

Gao, Y., Li, B., Wang, N. and Zhu, T., 2017. "Speech Emotion Recognition using Local and Global Features." In *International Conference on Brain Informatics*. Springer, pp.3-13.

Gorodnichenko, Y. and Ray, W., 2017. "The Effects of Quantitative Easing: Taking a Cue from Treasury Auctions." NBER Working Paper w24122.

Gorodnichenko, Y. and Weber, M., 2016. "Are Sticky Prices Costly? Evidence from the Stock Market." *American Economic Review* 106(1):165-199.

Gürkaynak, G., Sack, B., and Swanson, E., 2005a. "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking* 1(1):55-93.

Gürkaynak, R., Sack, B., and Swanson, E., 2005b. "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models." *American Economic Review,* 95 (1):425-436.

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

Hansen, S., McMahon, M., and Prat, A., 2018. "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach." *Quarterly Journal of Economics* 133(2):801-870.

Hansen, S., McMahon, M., and Tong, M., 2019. "The Long-run Information Effect of Central Bank Communication," *Journal of Monetary Economics* 108(C):185-202.

Hattori, M., Schrimpf, A., and Sushko, V., 2016. "The Response of Tail Risk Perceptions to Unconventional Monetary Policy." *American Economic Journal: Macroeconomics* 8(2):111-36.

Hobson, J.L., Mayew, W.J., and Venkatachalam, M., 2012. "Analyzing Speech to Detect Financial Misreporting." *Journal of Accounting Research* 50(2):349-392.

Hu, A. and Ma, S., 2020. "Human Interactions and Financial Investment: A Video-Based Approach." Working Paper.

Huberman, G., and Regev, T., 2001. "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar," *Journal of Finance* 56:387–396.

Husted, L., Rogers, J., and Sun, B., 2020. "Monetary policy uncertainty." *Journal of Monetary Economics* 115(C):20-36.

Jha, M., Liu, H. and Manela, A., 2021. "Natural disaster effects on popular sentiment toward finance." *Journal of Financial and Quantitative Analysis* 56(7):2584-2604.

Jordà, Ò., 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95(1):161-182.

Kozlowski, A.C., Taddy, M. and Evans, J.A., 2019. "The geometry of culture: Analyzing the meanings of class through word embeddings." *American Sociological Review* 84(5):905-949.

Krishnamurthy, A. and Vissing-Jorgensen, A., 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." *Brookings Papers on Economic Activity* 42(2 (Fall)):215-287.

Kuttner, K., 2001. "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market." *Journal of Monetary Economics* 47(3):523–544.

Lamla, M.J. and Vinogradov, D.V., 2019. "Central Bank Announcements: Big News for Little People?" *Journal of Monetary Economics* 108:21-38.

Likitha, M.S., Gupta, S.R.R., Hasitha, K., and Raju, A.U., 2017. "Speech based Human Emotion Recognition using MFCC." In *2017 International Conference on Wireless Communications, Signal Processing and Networking*. IEEE, pp.2257-2260.

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

Lucca, D., and Moench, E., 2015. "The Pre-FOMC Announcement Drift." *Journal of Finance* 70(1):329-371.

Mayew, W.J. and Venkatachalam, M., 2012. "The power of voice: Managerial affective states and future firm performance." *Journal of Finance* 67(1):1-43.

Mehrabian, A., 1971. *Silent Messages: Implicit Communication of Emotions and Attitudes*. Belmont, CA: Wadsworth.

Neuhierl, A. and Weber, M., 2019. "Monetary policy communication, policy slope, and the stock market." *Journal of Monetary Economics* 108(C):140-155.

Pan, Y., P. Shen, and L. Shen, 2012. "Speech emotion recognition using support vector machine." *International Journal of Smart Home* 6(2):101-108.

Romer, C., and Romer, D., 2000. "Federal Reserve Information and the Behavior of Interest Rates." *American Economic Review* 90:429-457.

Rosa, C., 2011a. "The high-frequency response of exchange rates to monetary policy actions and statements." *Journal of Banking & Finance* 35(2):478-489.

Rosa, C., 2011b. “Words that shake traders: The stock market’s reaction to central bank communication in real time.” *Journal of Empirical Finance* 18(5):915-934.

Rozkrut, M., Rybiński, K., Sztaba, L., and Szwaja, R., 2007. “Quest for central bank communication: Does it pay to be “talkative”?” *European Journal of Political Economy* 23(1):176-206.

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.

Tetlock, P.C., 2011. “All the news that’s fit to reprint: Do investors react to stale information?” *Review of Financial Studies* 24:1481–1512.

Tetlock, P.C., 2014. “Information Transmission in Finance” *Annual Review of Financial Economics* 6(1):365-384.

Verma, D. and Mukhopadhyay, D., 2016. “Age driven automatic speech emotion recognition system.” In *2016 International Conference on Computing, Communication and Automation*. IEEE, pp.1005-1010.

Visschedijk, G.C., Lazonder, A.W., van der Hulst, A., Vink, N., and Leemkuil, H., 2013. “Modelling human emotions for tactical decision-making games.” *British Journal of Educational Technology* 44(2):197-207.

Wu, J.C. and Xia, F.D., 2016. “Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound.” *Journal of Money, Credit and Banking* 48(2-3):253-291.

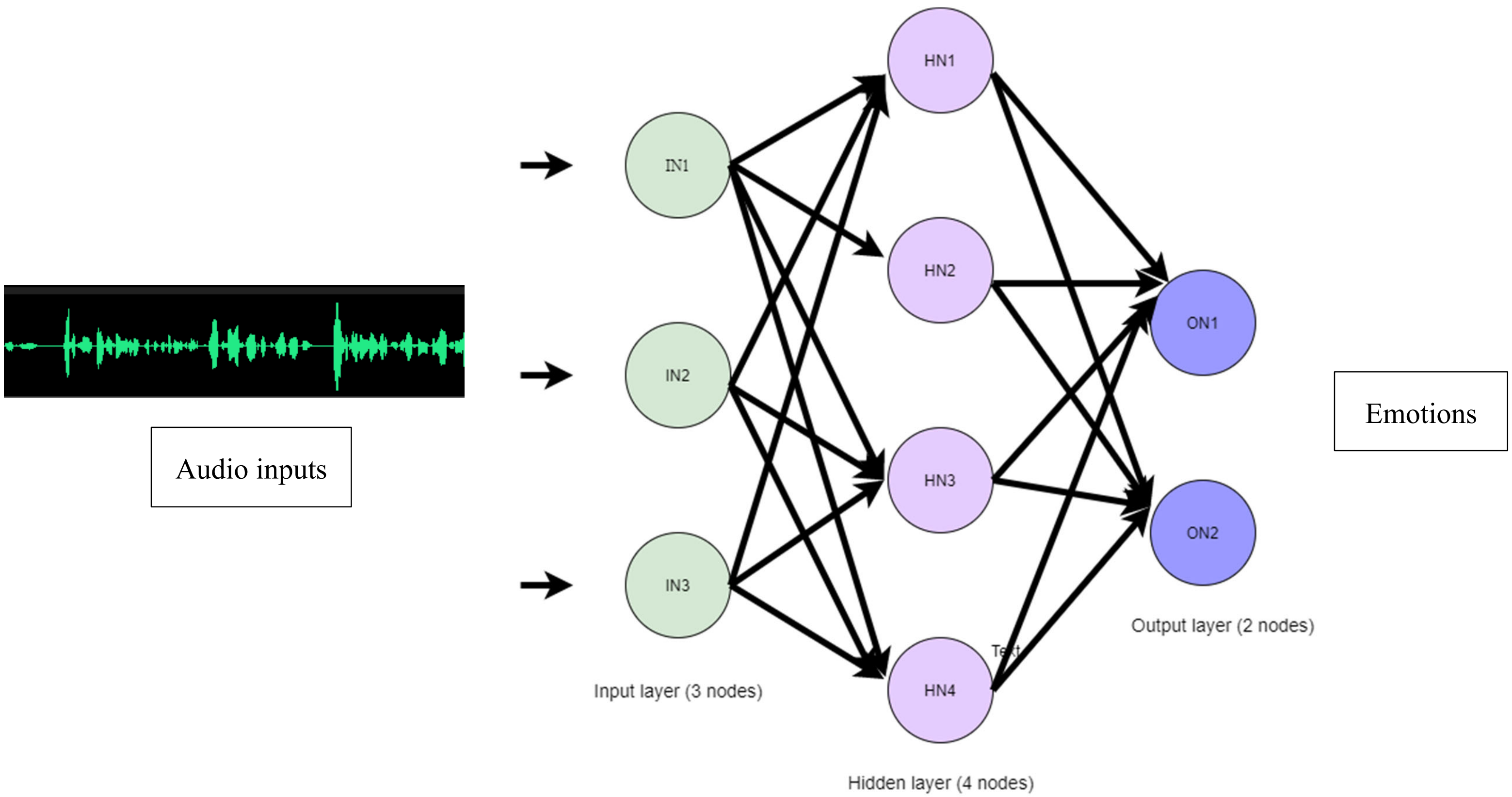
Zhou, R.T. and Lai, R.N., 2009. “Herding and information based trading.” *Journal of Empirical Finance* 16(3):388-393.

Table 1. FOMC Meeting Statistics

	All	Bernanke	Yellen	Powell
	(1)	(2)	(3)	(4)
Meetings	68	25	32	11
Press conferences	36	12	16	8
Panel A. Voice analysis of responses in Q&A during press conferences				
Answers (count)				
Positive	377	200	109	68
Negative	285	43	131	111
Neutral	30	0	28	2
Voice tone				
mean	0.09	0.64	-0.13	-0.30
standard deviation	0.75	0.58	0.61	0.82
Panel B. Textual analysis				
Statement				
Hawkish paragraphs	37	8	20	9
Dovish paragraphs	223	105	108	10
Neutral paragraphs	64	6	37	21
Text sentiment				
mean	0.66	0.86	0.71	0.09
standard deviation	0.53	0.27	0.34	0.94
Remarks				
Hawkish paragraphs	83	15	47	21
Dovish paragraphs	244	97	106	41
Neutral paragraphs	119	41	50	28
Text sentiment				
mean	0.50	0.75	0.38	0.35
standard deviation	0.37	0.24	0.33	0.47
Q&A				
Hawkish answers	233	77	95	61
Dovish answers	339	137	120	82
Neutral answers	120	29	53	38
Text sentiment				
mean	0.18	0.29	0.11	0.14
standard deviation	0.30	0.30	0.26	0.35
Statement, Remarks, Q&A				
Text sentiment				
mean	0.52	0.71	0.50	0.15
standard deviation	0.40	0.30	0.35	0.52

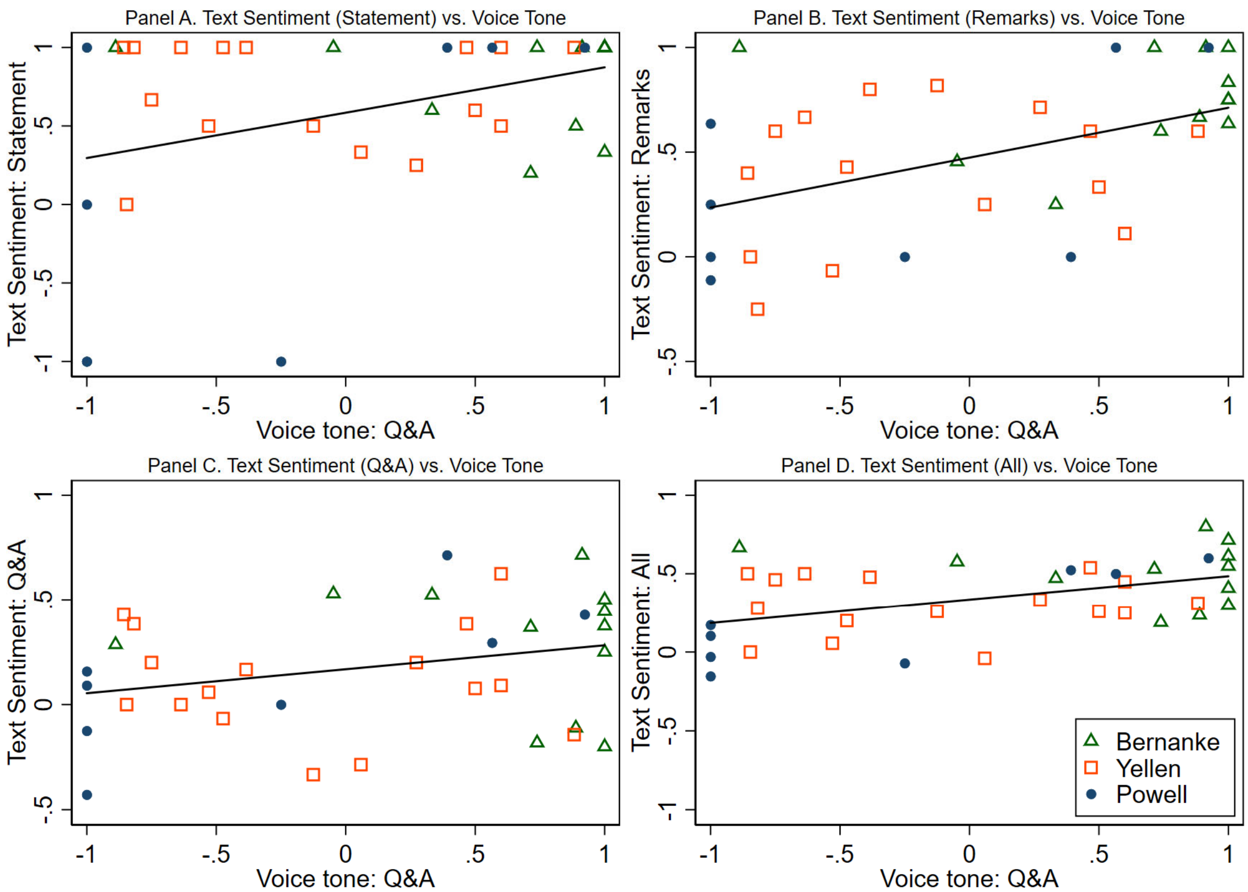
Notes: This table shows the statistics related to the text and voice data of FOMC meetings and press conferences. Column (1) shows statistics for all FOMC meetings during the 2011 – July 2019 period. Columns (2)-(4) show the statistics for the FOMC meetings chaired by Ben Bernanke, Janet Yellen, and Jerome Powell, respectively. Positive, Negative, and Neutral indicate the number of answers expressed in the positive, negative, and neutral emotion, respectively. Voice tone is the average emotion for a given FOMC press conference ($VoiceTone = \frac{Positive\ answers - Negative\ answers}{Positive\ answers + Negative\ answers}$). Hawkish and Dovish are the number of hawkish and dovish answers/sentences in the text, respectively. The average text sentiment is measured by $TextSentiment = \frac{Dovish\ text - Hawkish\ text}{Dovish\ text + Hawkish\ text}$.

Figure 1. Conceptual neural network



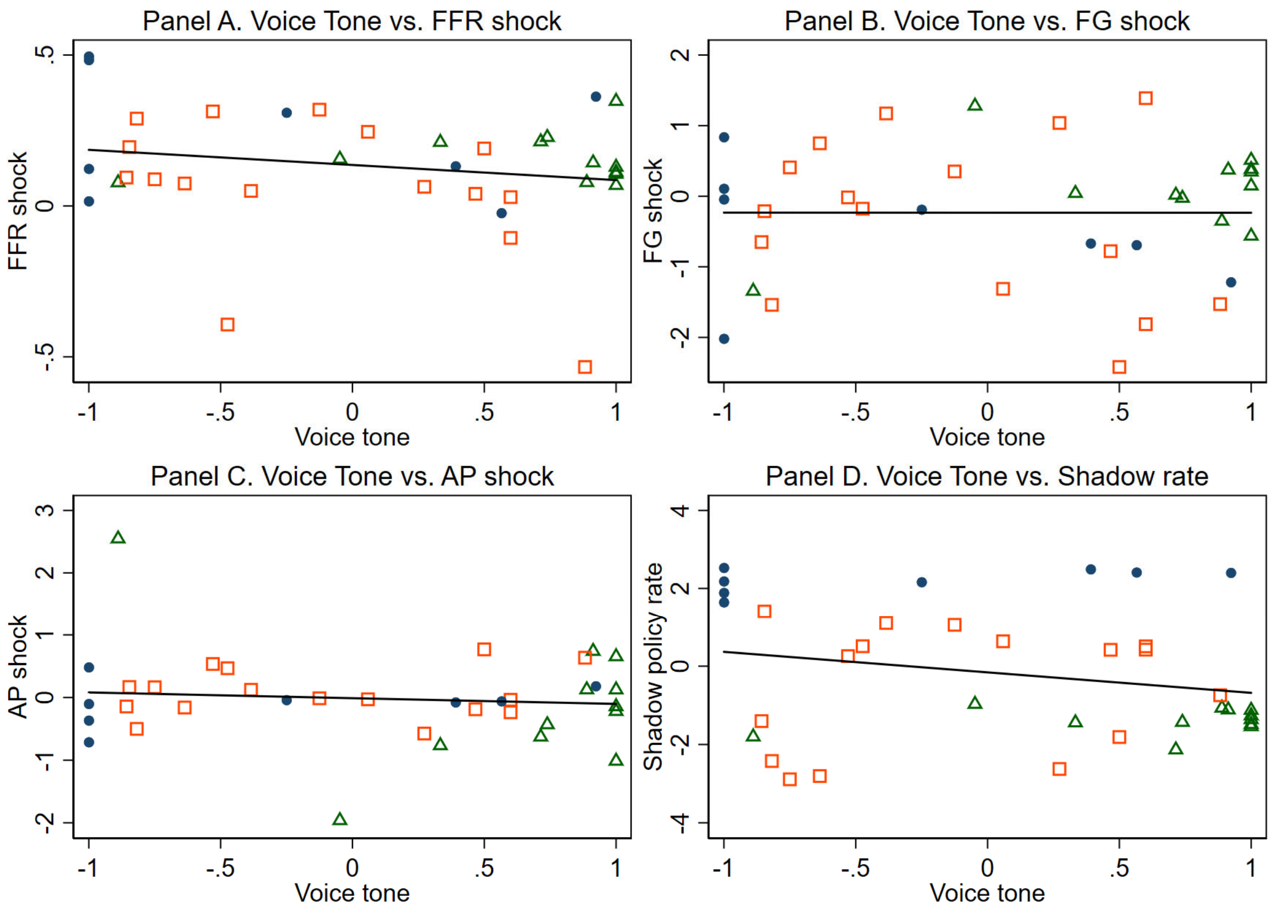
Notes: This figure shows a conceptual neural network with three layers for classifying voice features into emotions. The left layer is the input layer with three nodes and each node is a feature of the training audio data. The middle layer is a hidden layer consisting of four nodes (HK_k, k=[1;4]) which are the activation functions of the input features IN_i (i=[1;3]). A node HK_k is connected with the input, the through weight ($w_{k,i}$) and bias (b_k): $\sum_{i=1}^{3} IN_i \times w_{k,i} + b_k$. The weighted summations are passed on the softmax activation function to obtain the outputs O_k (k=[1;2]). The right layer is the output layer. The actual network has two hidden layers, 180 input nodes and five output nodes.

Figure 2. Voice tone vs. Text sentiment



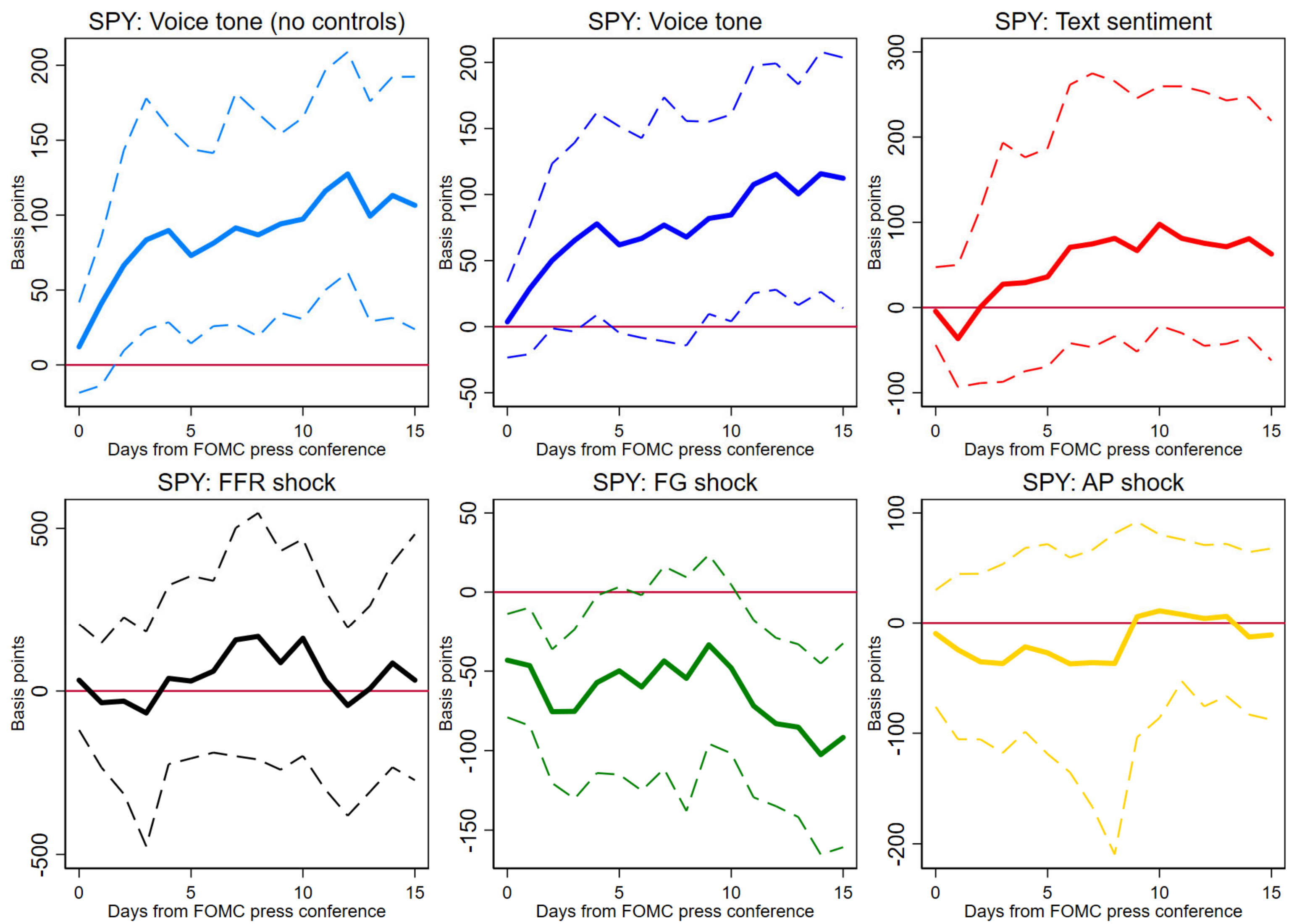
Notes: This figure shows the joint distribution of voice tone and text sentiment across FOMC meetings.

Figure 3. Policy Words vs. Actions



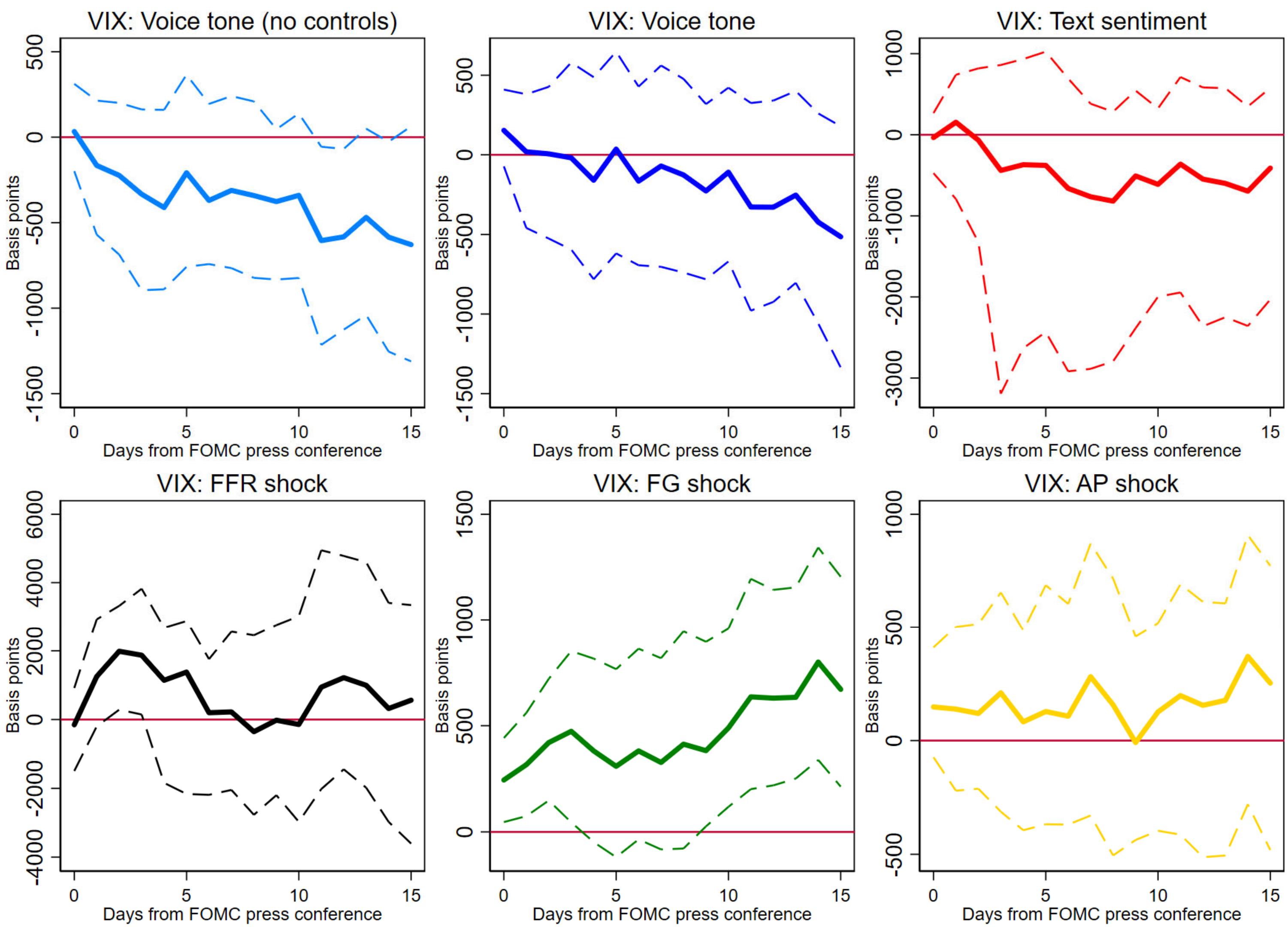
Notes: This figure shows the joint distribution of voice tone and policy actions/stance. Federal Funds Rate (FFR), forward guidance (FG), and asset purchase (AP) shocks are from Swanson (2021). The shadow policy rate is from Wu and Xia (2016).

Figure 4. Response of SPY ETF (S&P 500) to policy actions and messages



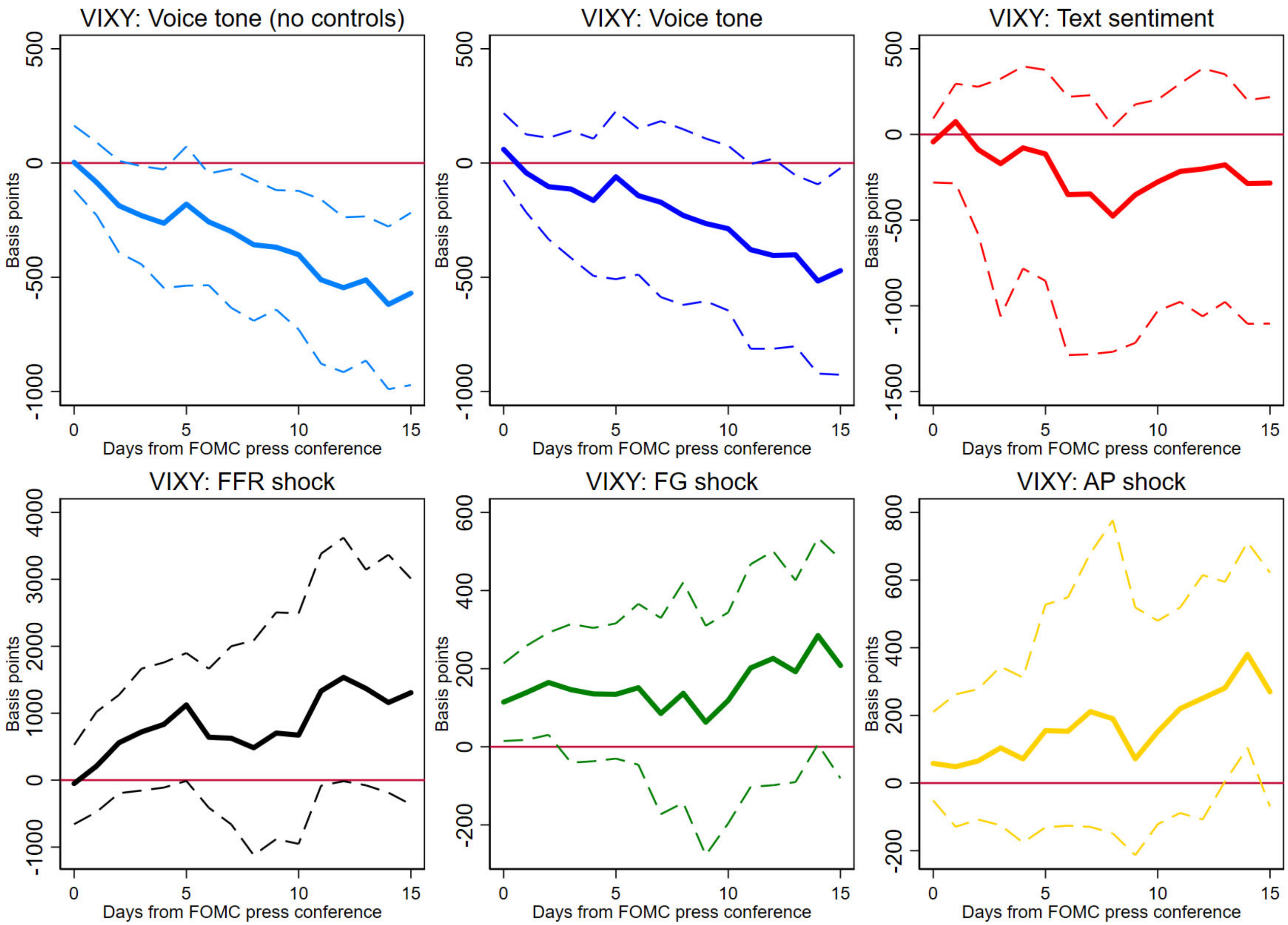
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 5. Response of VIX (CBOE Volatility Index) to policy actions and messages



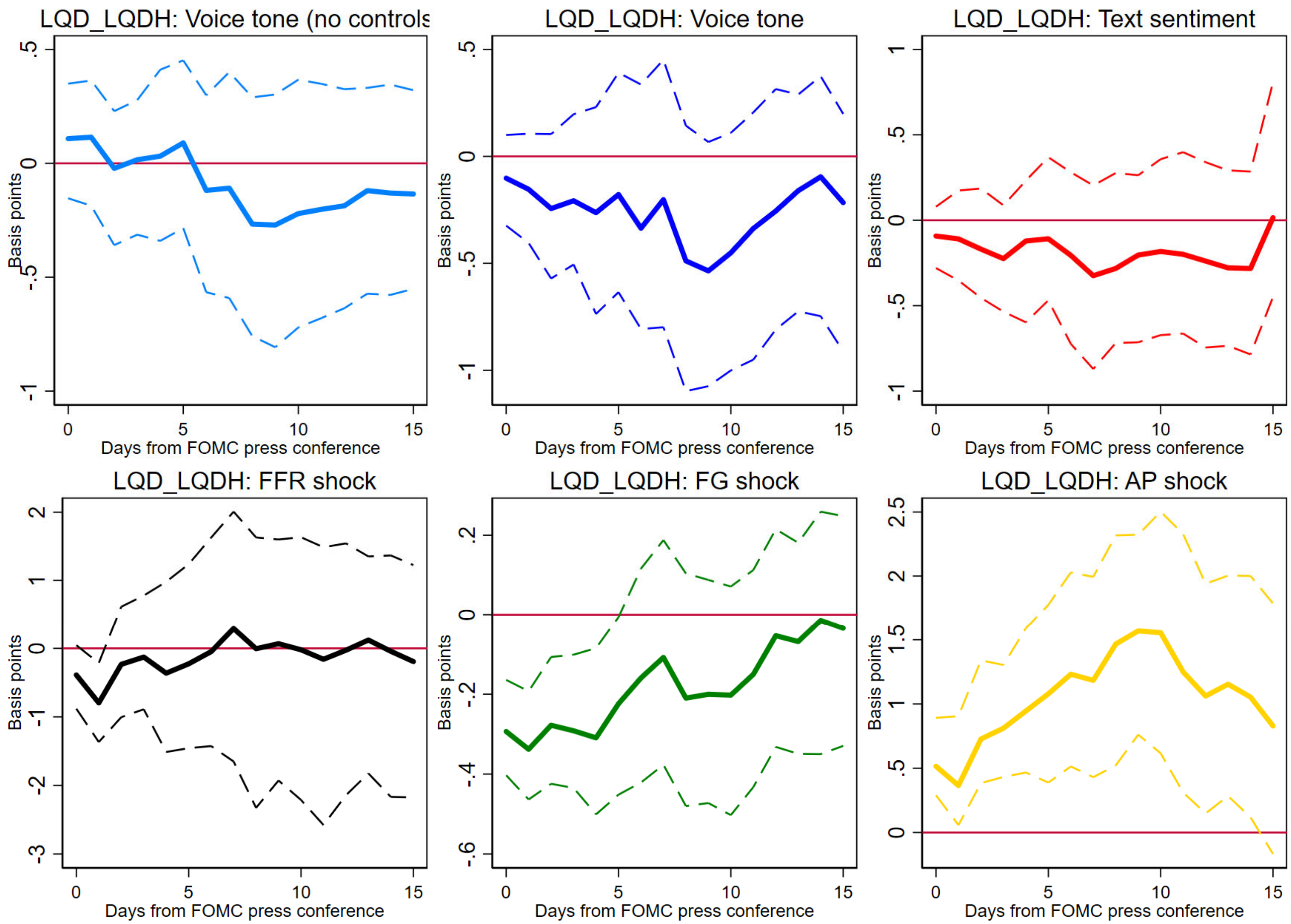
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 6. Response of VIXY ETF (VIX Short-Term Futures) to policy actions and messages



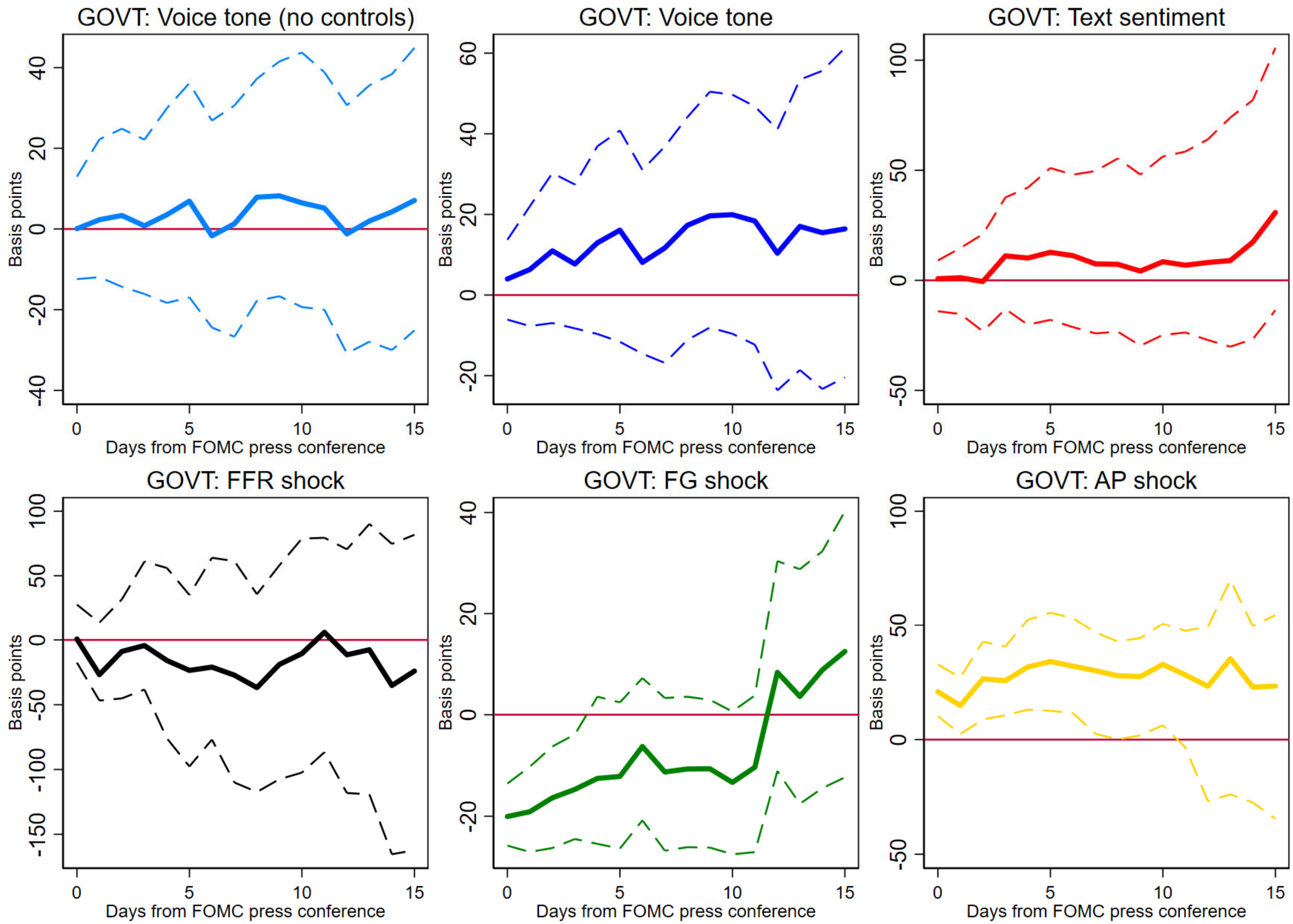
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 7. Response of LQD ETF (investment grade corporate bond) minus LQDH EFT (interest rate hedged corporate bond) to policy actions and messages



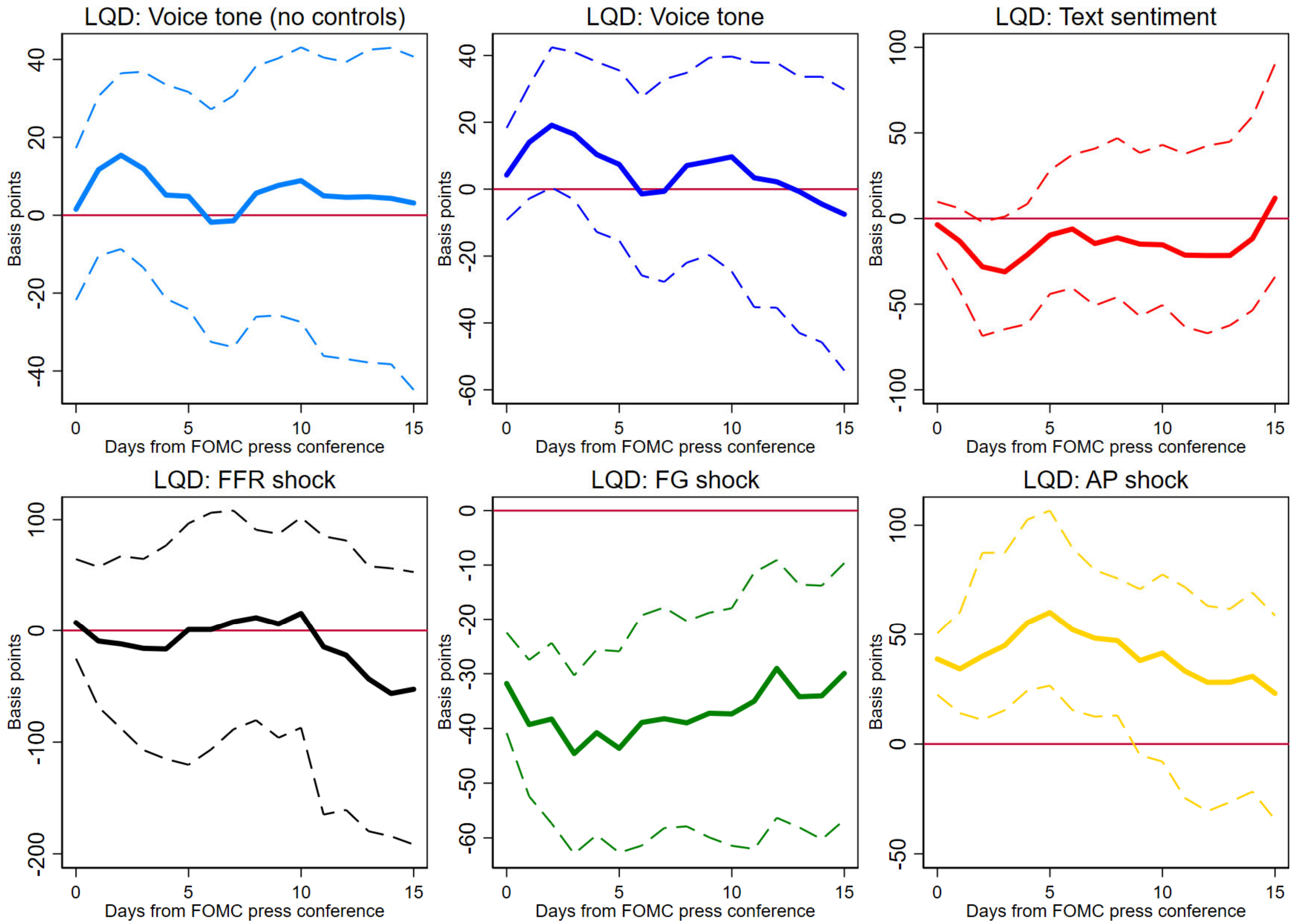
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 8. Response of GOVT ETF (U.S. government debt) to policy actions and messages



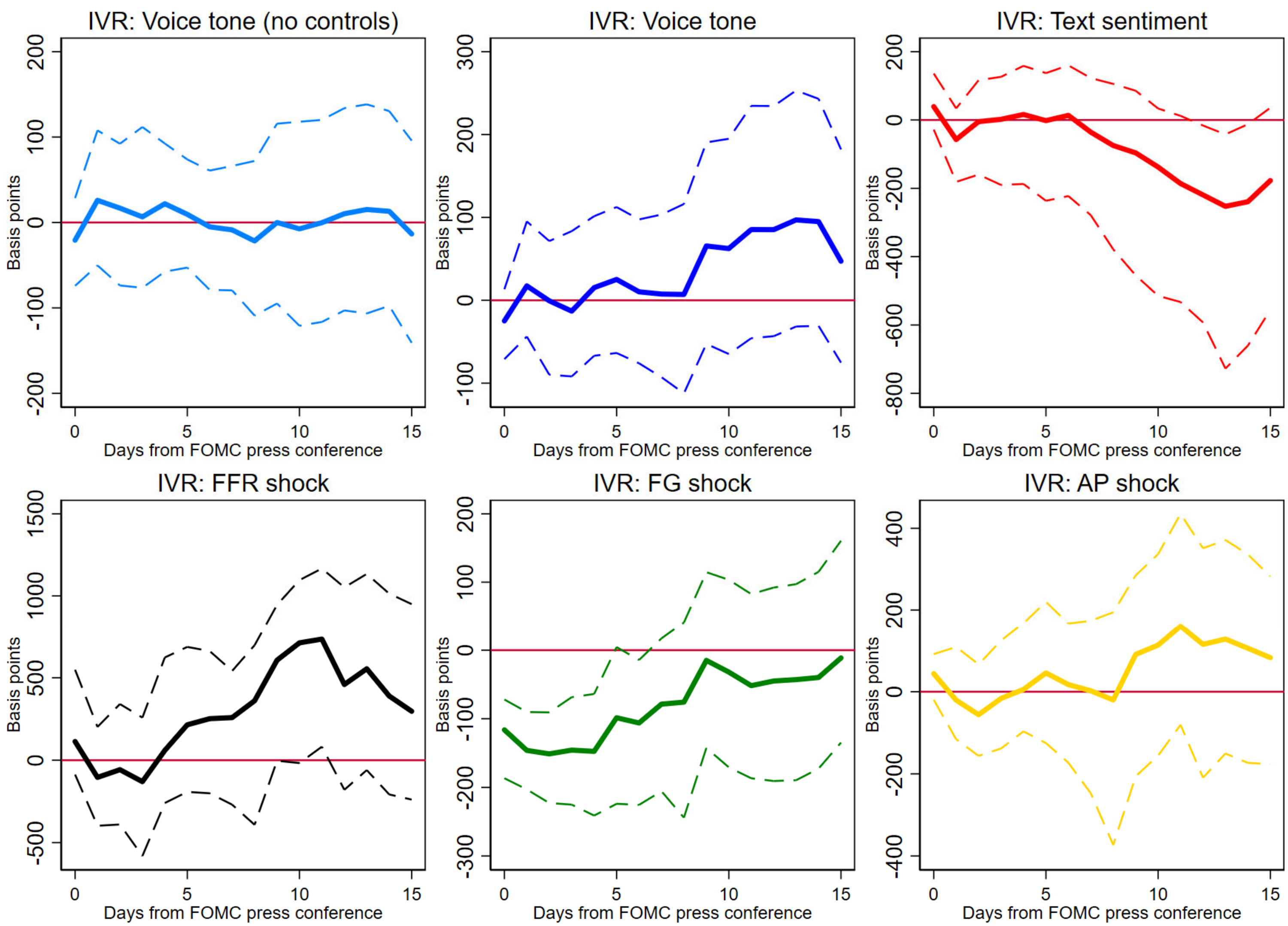
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 9. Response of LQD ETF (corporate debt) to policy actions and messages



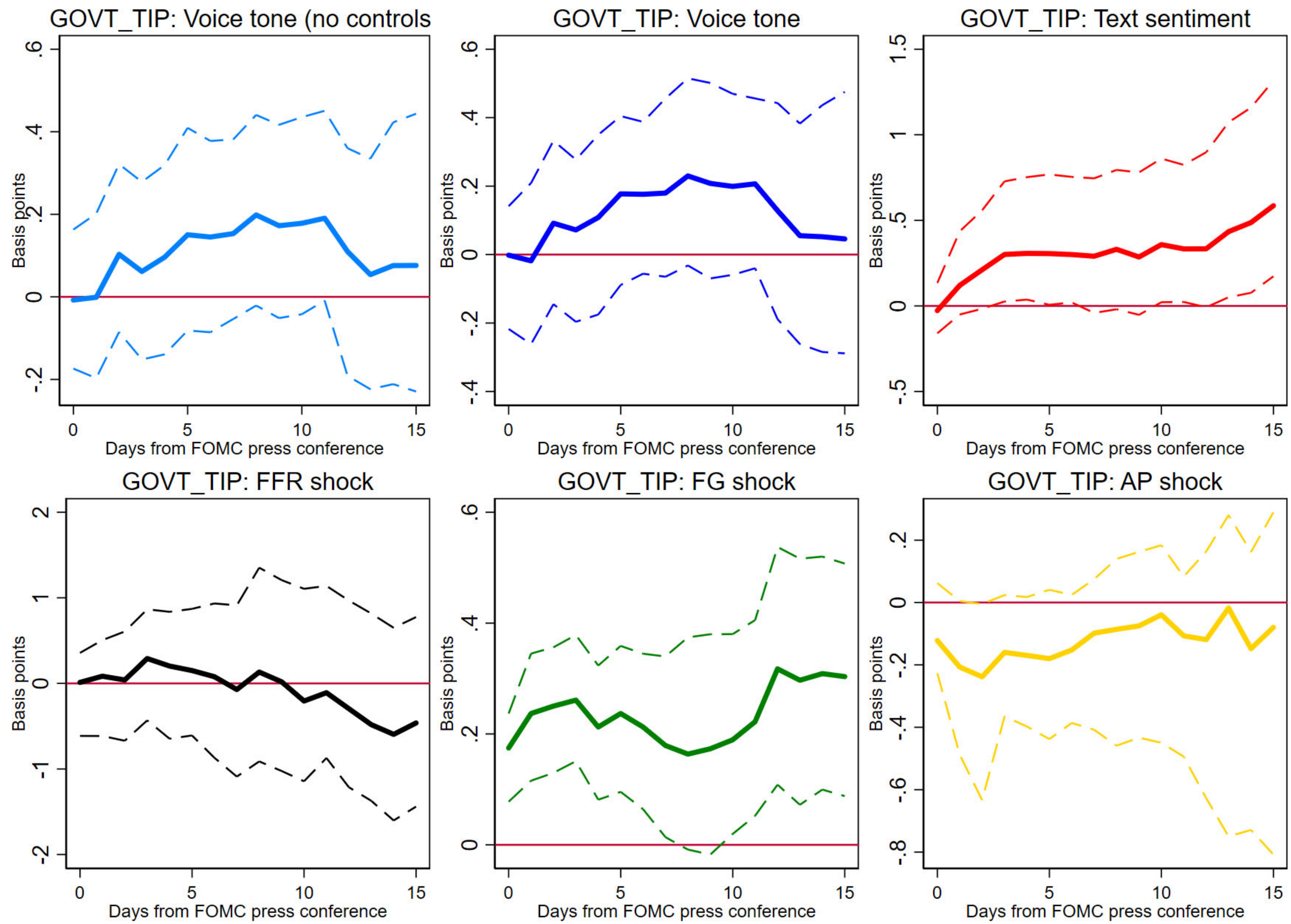
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 10. Response of IVR ETF (debt for the real estate sector) to policy actions and messages



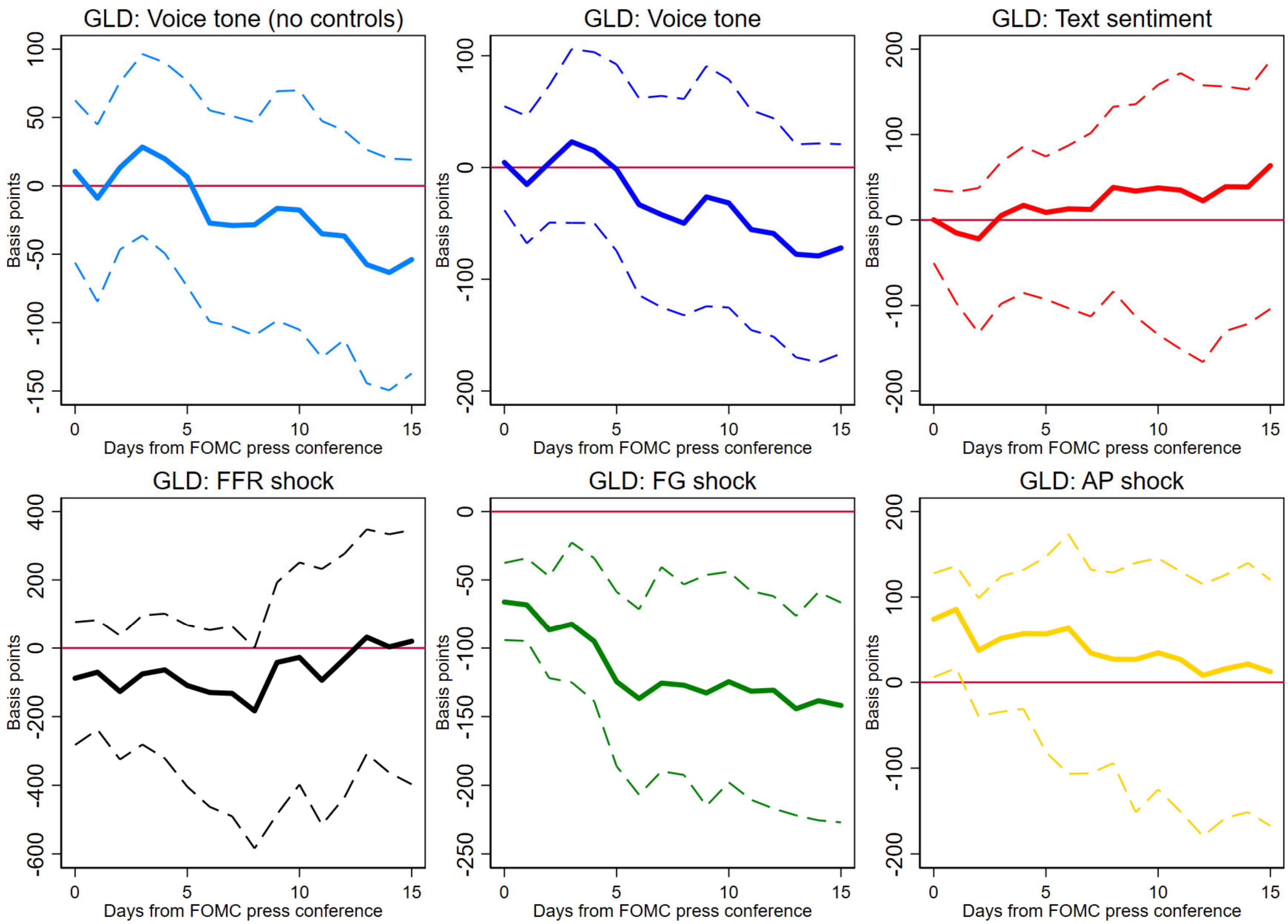
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 11. Response of GOVT ETF (nominal U.S. government debt) minus TIP EFT (inflation-protected U.S. government debt) to policy actions and messages



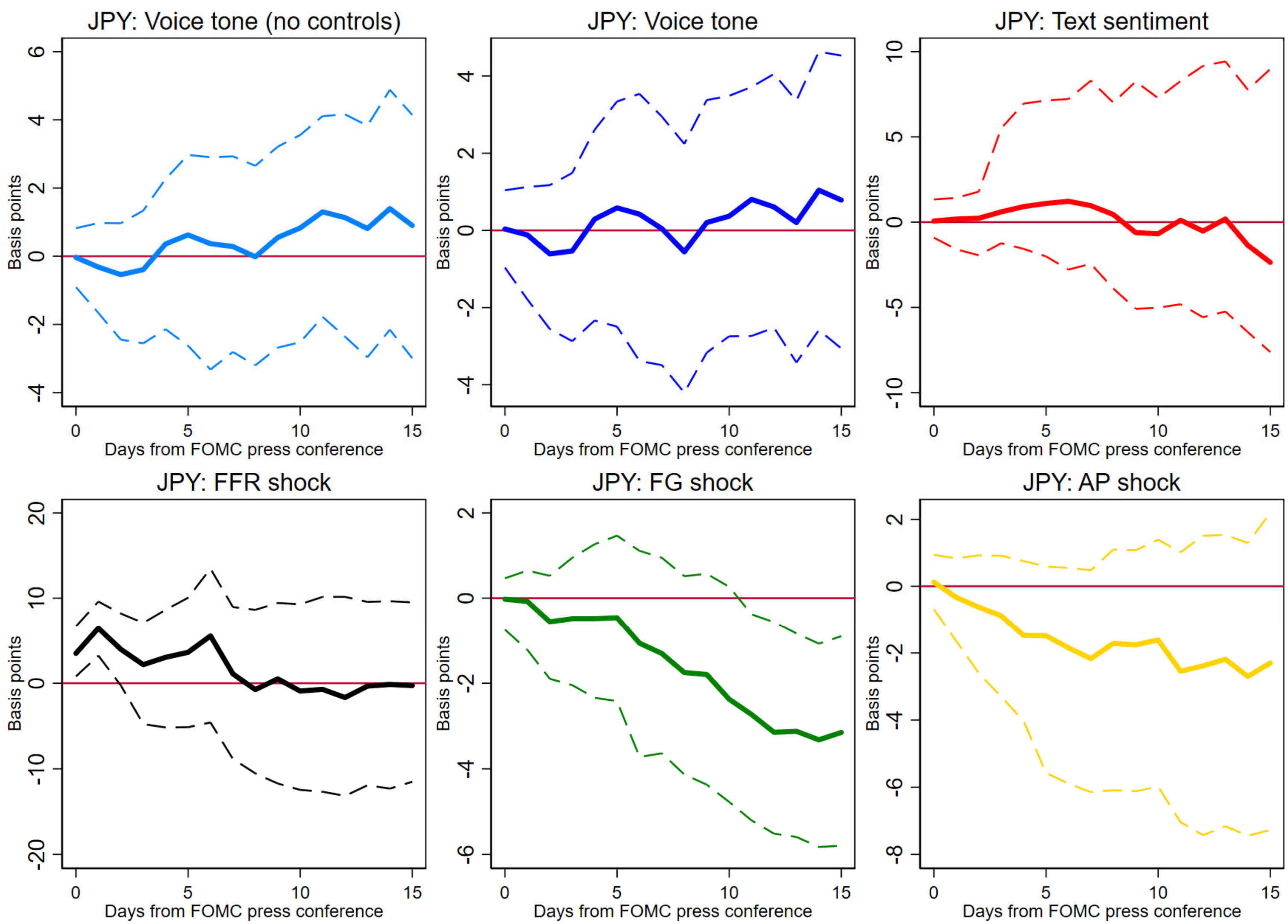
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 12. Response of GLD ETF (gold) to policy actions and messages



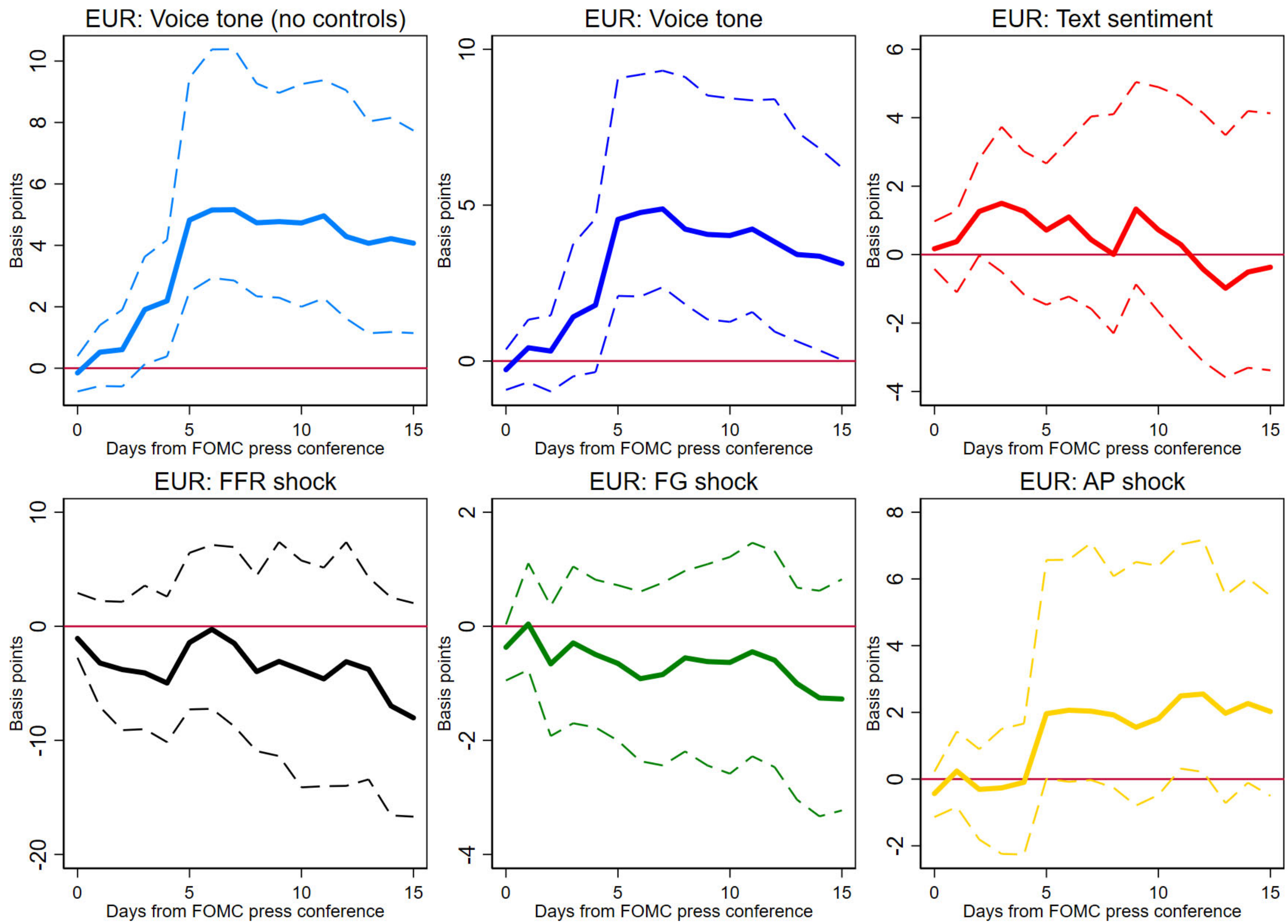
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 13. Response of the Japanese Yen to one U.S. Dollar (dollar/yen) exchange rate to policy actions and messages



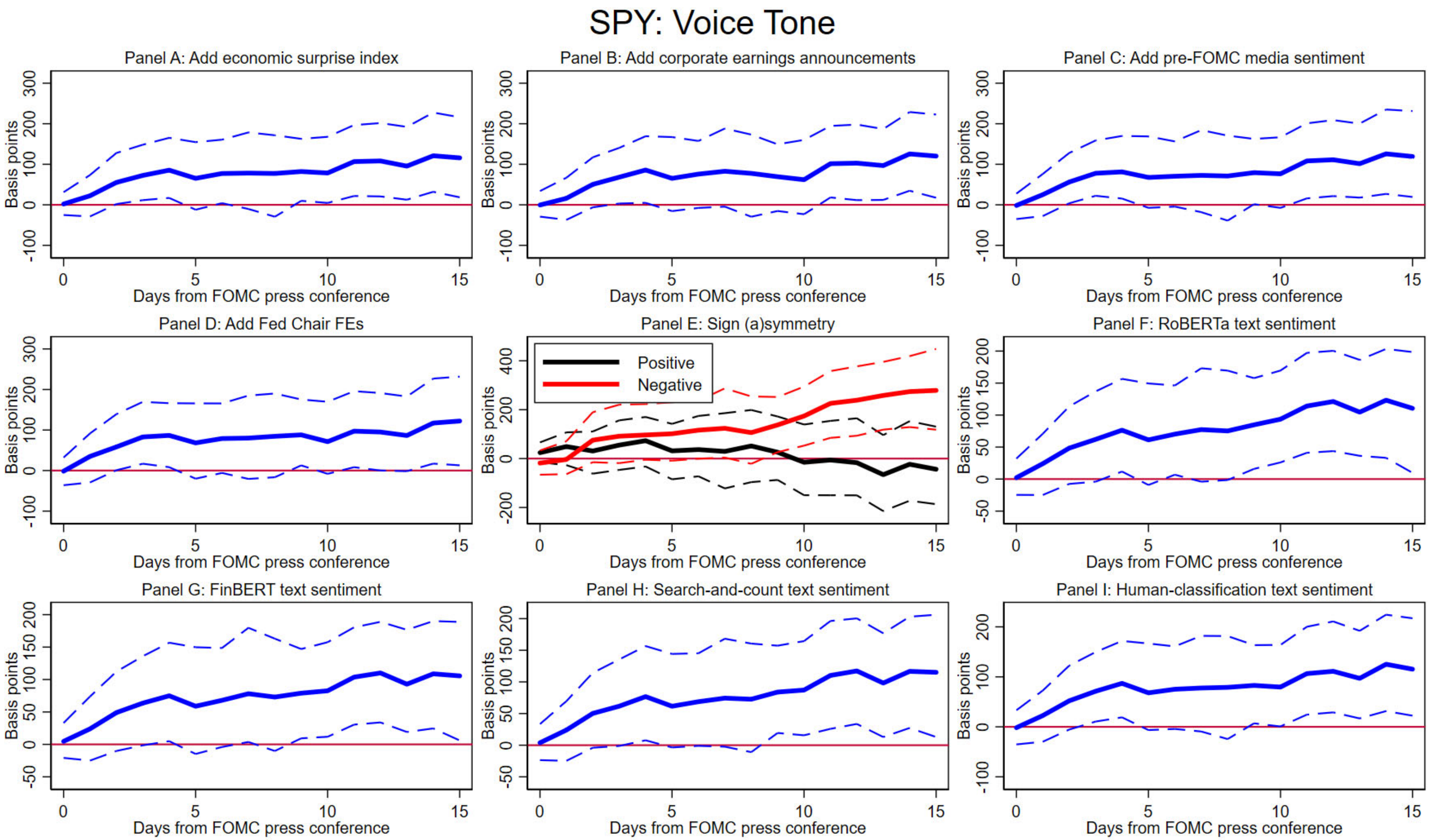
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 14. Response of the Euro to one U.S. Dollar (dollar/euro) exchange rate to policy actions and messages



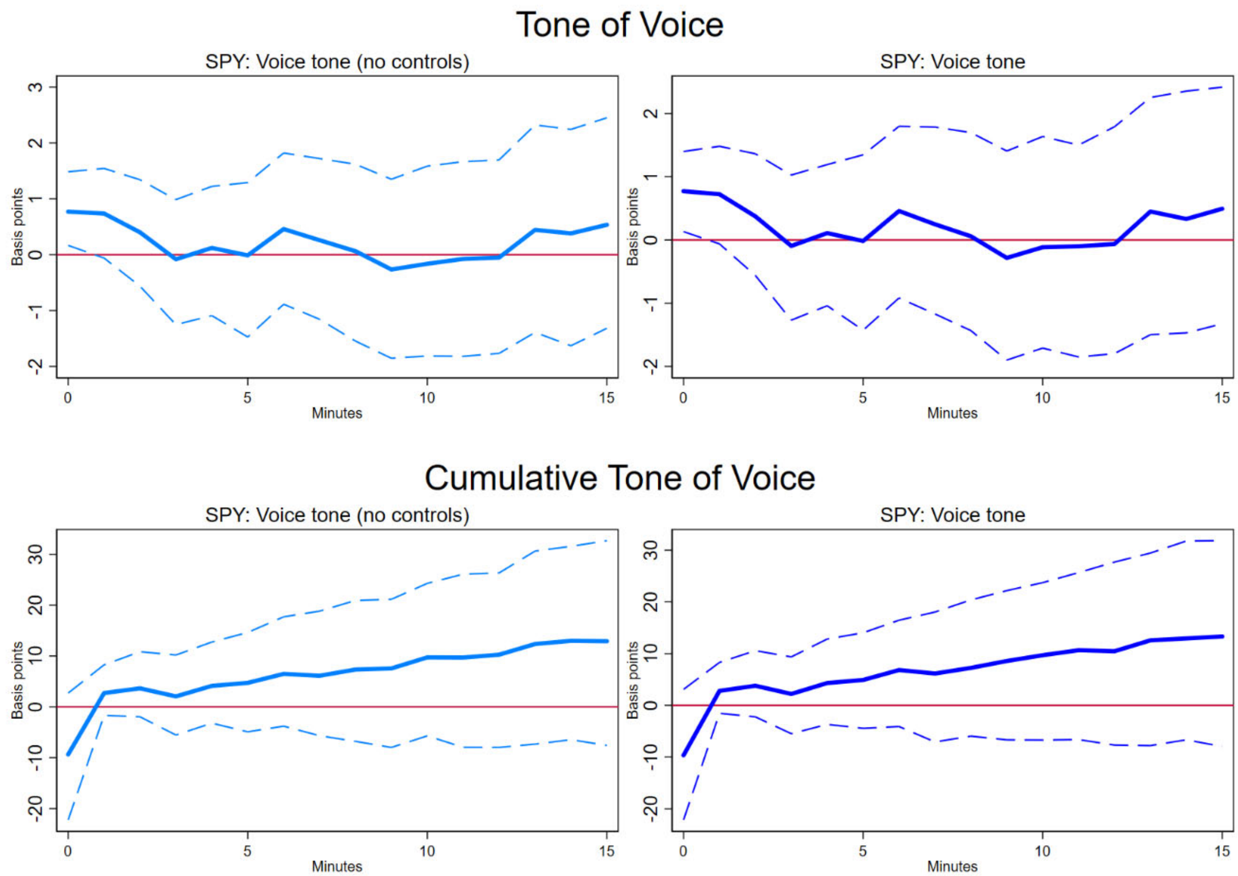
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 15. Robustness checks



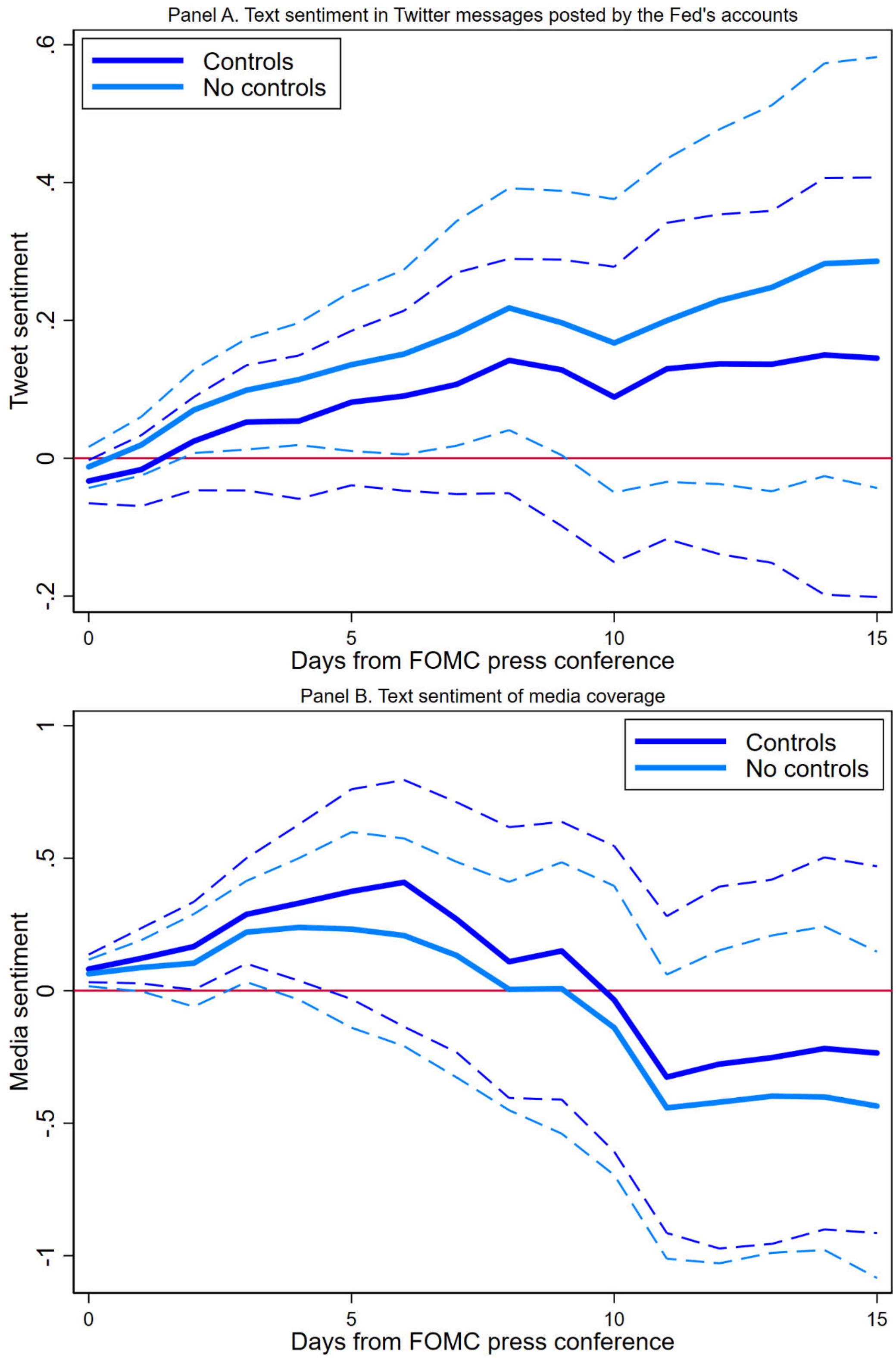
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for voice tone. The outcome variable is SPY, the ETF that tracks the S&P500 index. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 16. Intra-day responses of stock prices (SPY ETF) to voice tone



Notes: The top panels show the impulse responses of SPY ETF prices to voice tone variation answer by answer. These impulse responses are estimated using specification (5'). The bottom panels show the impulse responses of cumulative changes in SPY ETF prices from the start of a press conference to cumulative voice tone variation from the start of a press conference. These impulse responses are estimated using specification (5''). Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Figure 17. Post-FOMC policy commnucation and media coverage



Notes: Panel A plots impulse responses for cumulative text sentiment in tweets posted by the Fed's Twitter accounts ($TweetSentiment_h = \sum_{i=0}^{h} Sentiment_i$) to a unit increase in voice tone. The specification is given by equation (4). Panel B plots the impulse responses for cumulative media sentiment ($\sum_{s=0}^{h} MediaSentiment_{t+s}$) to a unit increase in voice tone. The specification is given by equation (4). Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Online Appendix

Additional Information, Tables, and Figures

Appendix A

Appendix Table A1. Voice tone for responses during Q&A sessions

Press conference date	Speaker	Positive responses	Neutral responses	Negative responses	Tone
April 27, 2011	Bernanke	17	0	1	0.89
June 22, 2011	Bernanke	19	0	0	1.00
November 2, 2011	Bernanke	19	0	0	1.00
January 25, 2012	Bernanke	18	0	0	1.00
April 25, 2012	Bernanke	19	0	0	1.00
June 20, 2012	Bernanke	22	0	1	0.91
September 13, 2012	Bernanke	23	0	0	1.00
December 12, 2012	Bernanke	20	0	3	0.74
March 20, 2013	Bernanke	14	0	7	0.33
June 19, 2013	Bernanke	10	0	11	-0.05
September 18, 2013	Bernanke	1	0	17	-0.89
December 18, 2013	Bernanke	18	0	3	0.71
March 19, 2014	Yellen	7	5	4	0.19
June 18, 2014	Yellen	2	0	14	-0.75
September 17, 2014	Yellen	2	1	9	-0.58
December 17, 2014	Yellen	1	4	10	-0.60
March 18, 2015	Yellen	15	0	5	0.50
June 17, 2015	Yellen	1	3	13	-0.71
September 17, 2015	Yellen	16	1	1	0.83
December 16, 2015	Yellen	4	1	13	-0.50
March 16, 2016	Yellen	12	1	3	0.56
June 15, 2016	Yellen	11	0	4	0.47
September 21, 2016	Yellen	5	0	14	-0.47
December 14, 2016	Yellen	12	5	3	0.45
March 15, 2017	Yellen	9	1	8	0.06
June 14, 2017	Yellen	7	0	9	-0.13
September 20, 2017	Yellen	4	1	9	-0.36
December 13, 2017	Yellen	1	5	12	-0.61
March 21, 2018	Powell	0	0	20	-1.00
June 13, 2018	Powell	0	0	22	-1.00
September 26, 2018	Powell	9	0	15	-0.25
December 19, 2018	Powell	0	0	21	-1.00
January 30, 2019	Powell	16	2	7	0.36
March 20, 2019	Powell	25	0	1	0.92
May 1, 2019	Powell	18	0	5	0.57
June 19, 2019	Powell	0	0	20	-1.00
July 31, 2019	Powell	0	0	24	-1.00

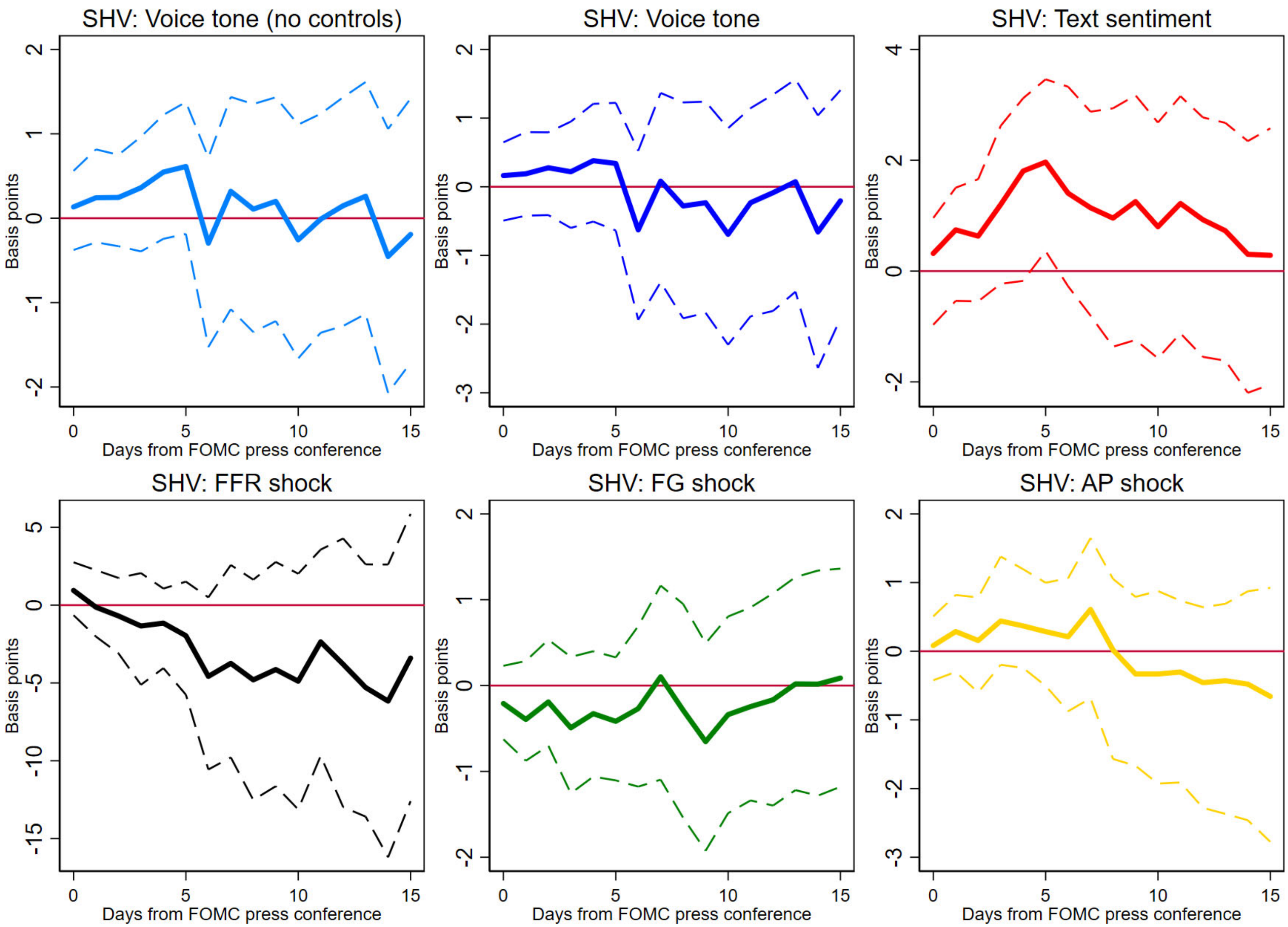
Notes: This table shows the number of positive, negative, and neutral responses as well as the aggregate voice tone for each press conference in the sample.

Appendix Table A2. Text sentiment for statement, remarks and Q&A

Press conference date	Speaker	Text Sentiment	Press conference date	Speaker	Text Sentiment
26/01/2011	Bernanke	1.00	29/04/2015	Yellen	0.60
15/03/2011	Bernanke	1.00	17/06/2015	Yellen	0.50
27/04/2011	Bernanke	0.24	29/07/2015	Yellen	1.00
22/06/2011	Bernanke	0.41	17/09/2015	Yellen	0.31
09/08/2011	Bernanke	1.00	28/10/2015	Yellen	0.60
21/09/2011	Bernanke	1.00	16/12/2015	Yellen	0.06
02/11/2011	Bernanke	0.61	27/01/2016	Yellen	1.00
13/12/2011	Bernanke	1.00	16/03/2016	Yellen	0.25
25/01/2012	Bernanke	0.55	27/04/2016	Yellen	0.33
13/03/2012	Bernanke	1.00	15/06/2016	Yellen	0.54
25/04/2012	Bernanke	0.71	27/07/2016	Yellen	1.00
20/06/2012	Bernanke	0.80	21/09/2016	Yellen	0.20
01/08/2012	Bernanke	1.00	02/11/2016	Yellen	1.00
13/09/2012	Bernanke	0.30	14/12/2016	Yellen	0.45
24/10/2012	Bernanke	1.00	01/02/2017	Yellen	1.00
12/12/2012	Bernanke	0.19	15/03/2017	Yellen	-0.04
30/01/2013	Bernanke	0.60	03/05/2017	Yellen	1.00
20/03/2013	Bernanke	0.47	14/06/2017	Yellen	0.26
01/05/2013	Bernanke	1.00	26/07/2017	Yellen	1.00
19/06/2013	Bernanke	0.58	20/09/2017	Yellen	0.48
31/07/2013	Bernanke	1.00	01/11/2017	Yellen	1.00
18/09/2013	Bernanke	0.67	13/12/2017	Yellen	0.00
30/10/2013	Bernanke	1.00	31/01/2018	Yellen	0.00
18/12/2013	Bernanke	0.53	21/03/2018	Powell	0.10
29/01/2014	Bernanke	0.20	02/05/2018	Powell	1.00
19/03/2014	Yellen	0.33	13/06/2018	Powell	-0.03
30/04/2014	Yellen	0.67	01/08/2018	Powell	-1.00
18/06/2014	Yellen	0.46	26/09/2018	Powell	-0.07
30/07/2014	Yellen	0.00	08/11/2018	Powell	0.00
17/09/2014	Yellen	0.50	19/12/2018	Powell	0.17
29/10/2014	Yellen	0.50	30/01/2019	Powell	0.52
17/12/2014	Yellen	0.28	20/03/2019	Powell	0.60
28/01/2015	Yellen	0.60	01/05/2019	Powell	0.50
18/03/2015	Yellen	0.26	19/06/2019	Powell	-0.15

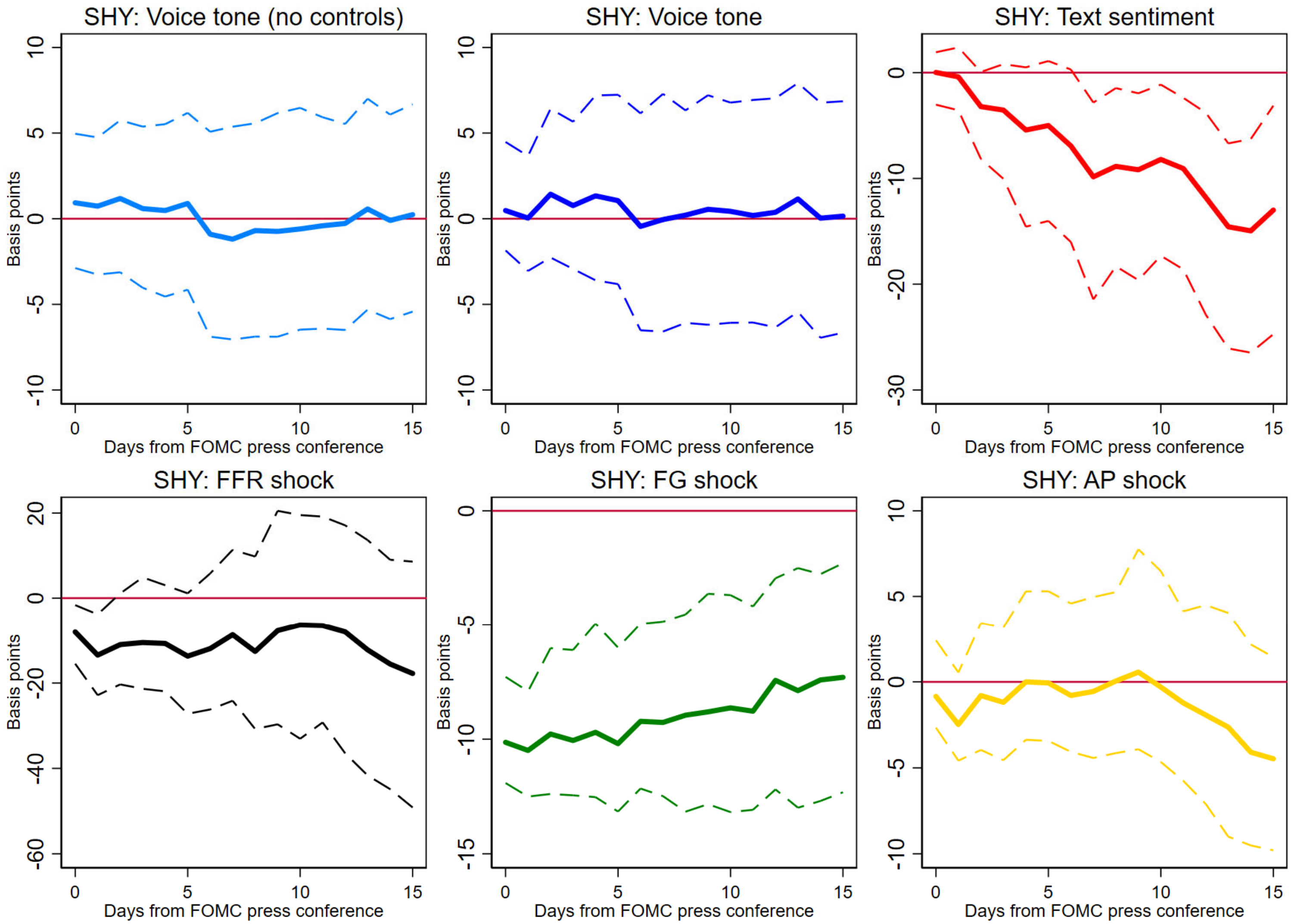
Notes: This table shows the aggregate text sentiment for each FOMC meeting in the sample.

Appendix Figure A1. Response of SHV ETF (Short Treasury Bond ETF; maturities one year or less) to policy actions and messages



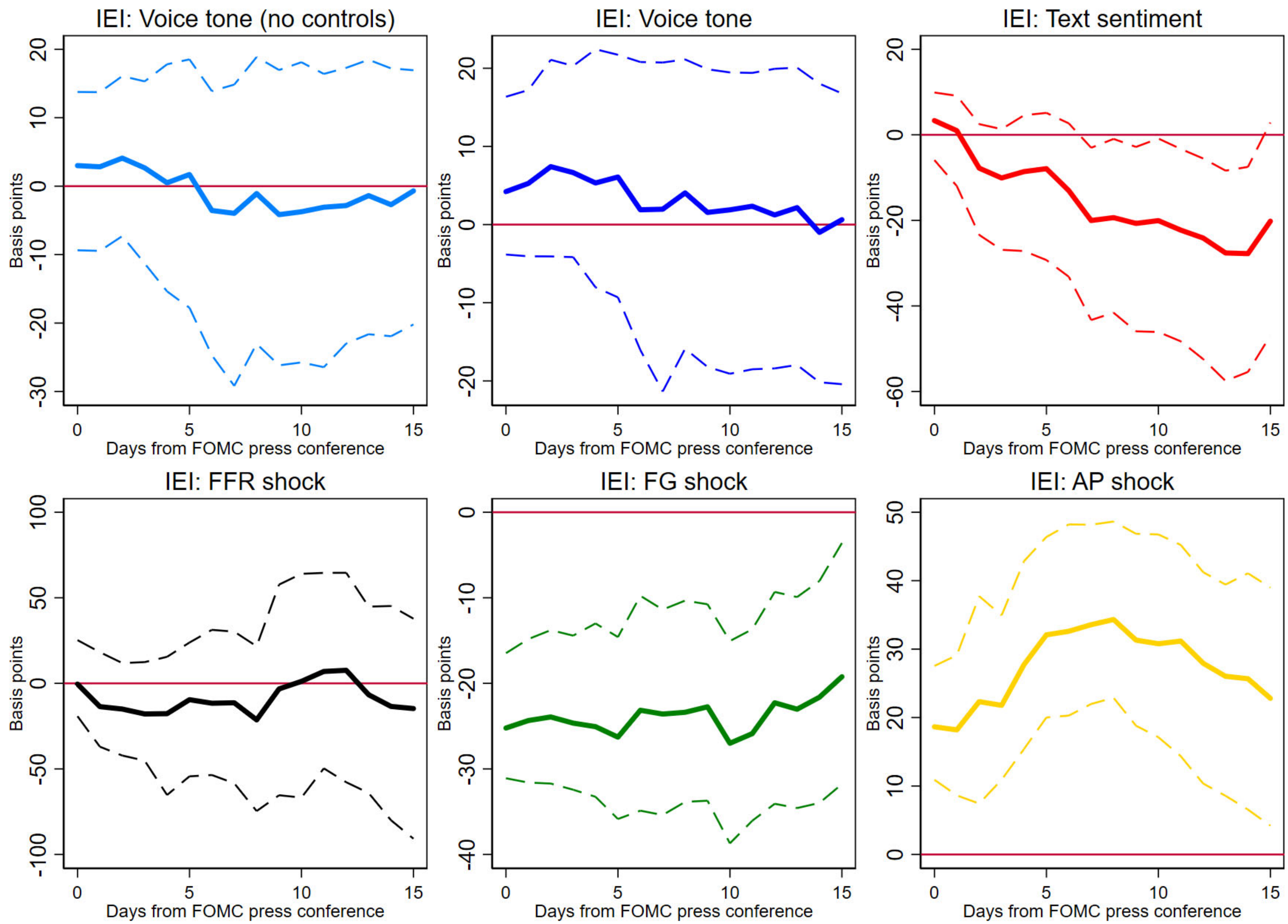
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Appendix Figure A2. Response of SHY ETF (1-3 Year Treasury Bond ETF) to policy actions and messages



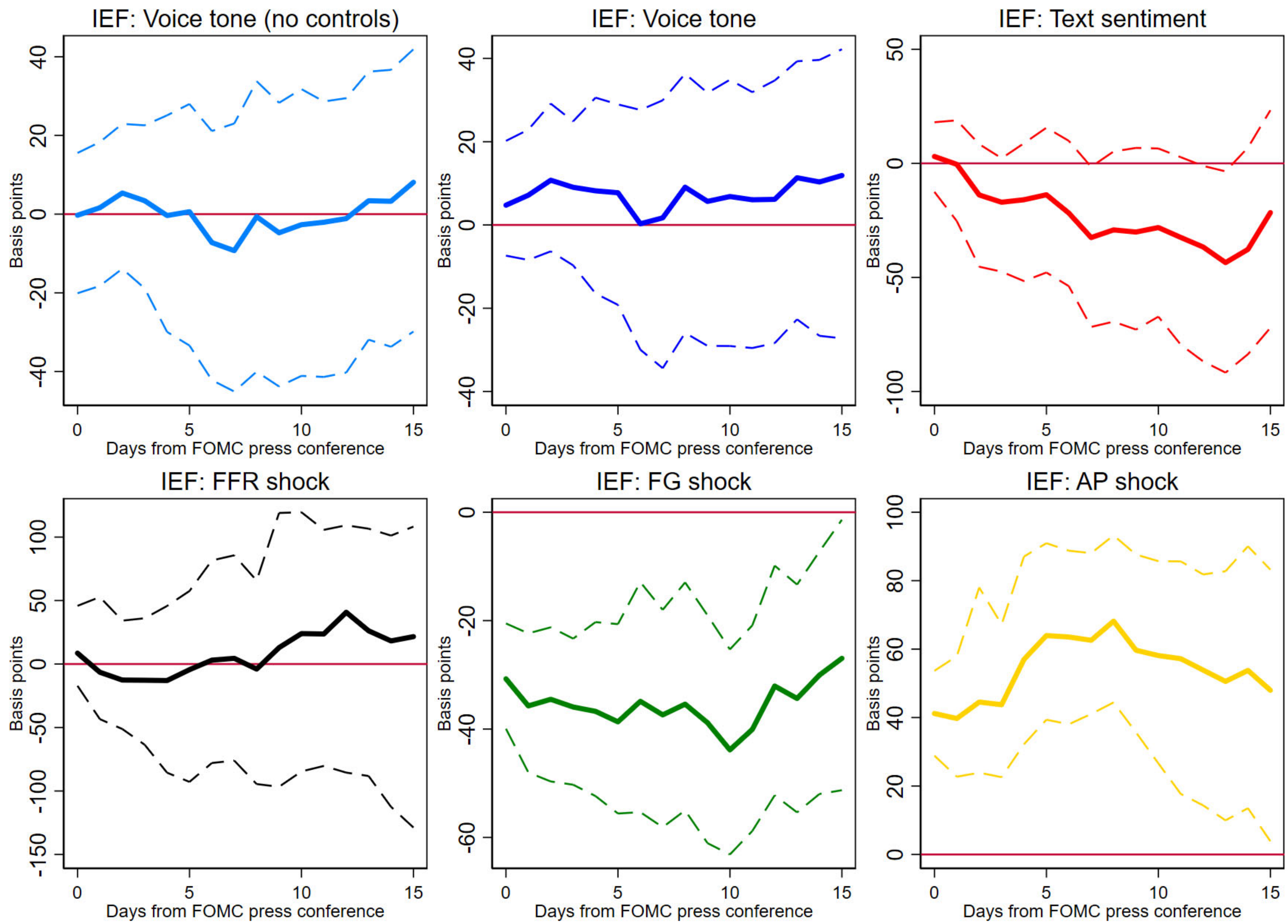
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Appendix Figure A3. Response of IEI ETF (3-7 Year Treasury Bond ETF) to policy actions and messages



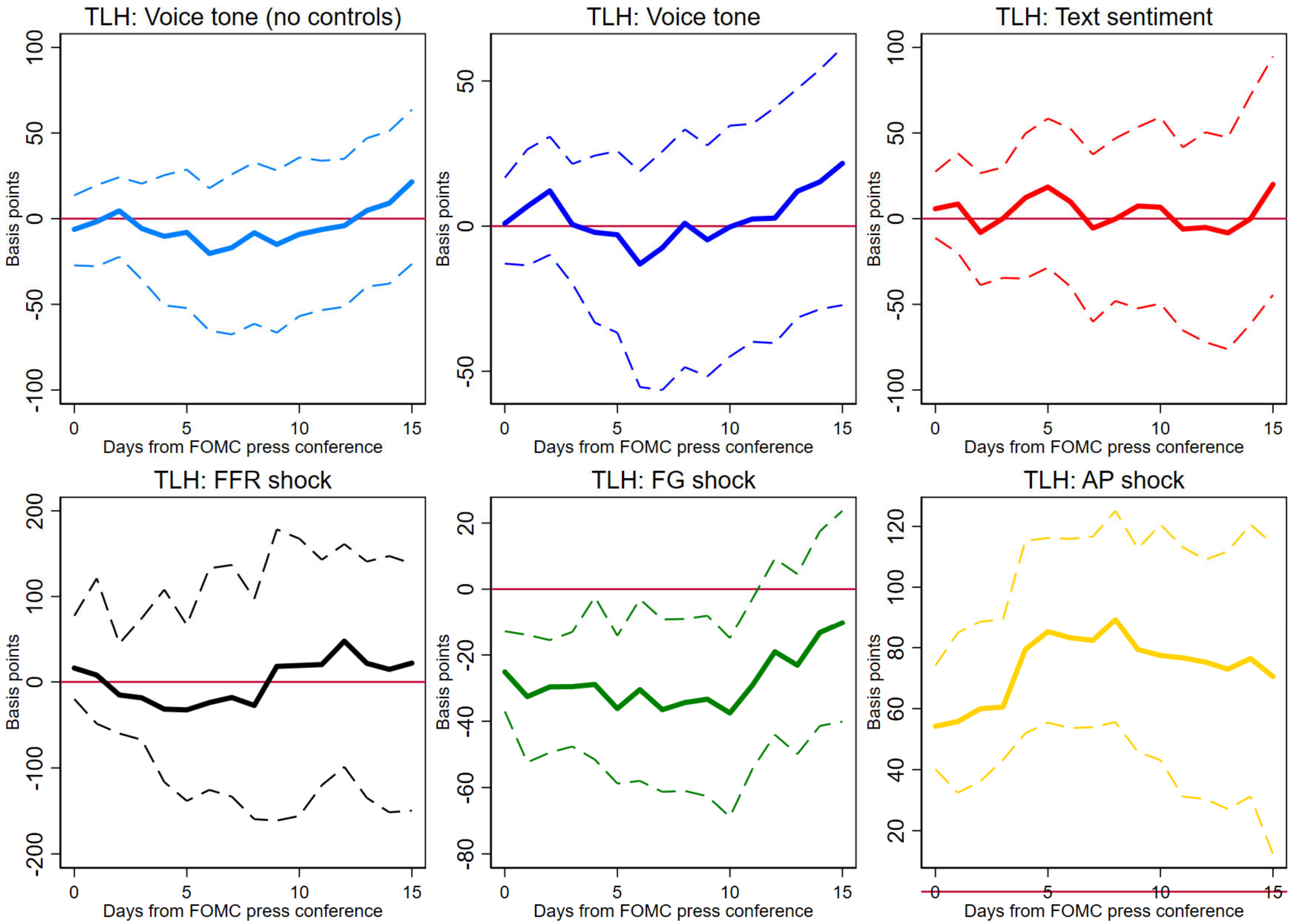
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Appendix Figure A4. Response of IEF ETF (7-10 Year Treasury Bond ETF) to policy actions and messages



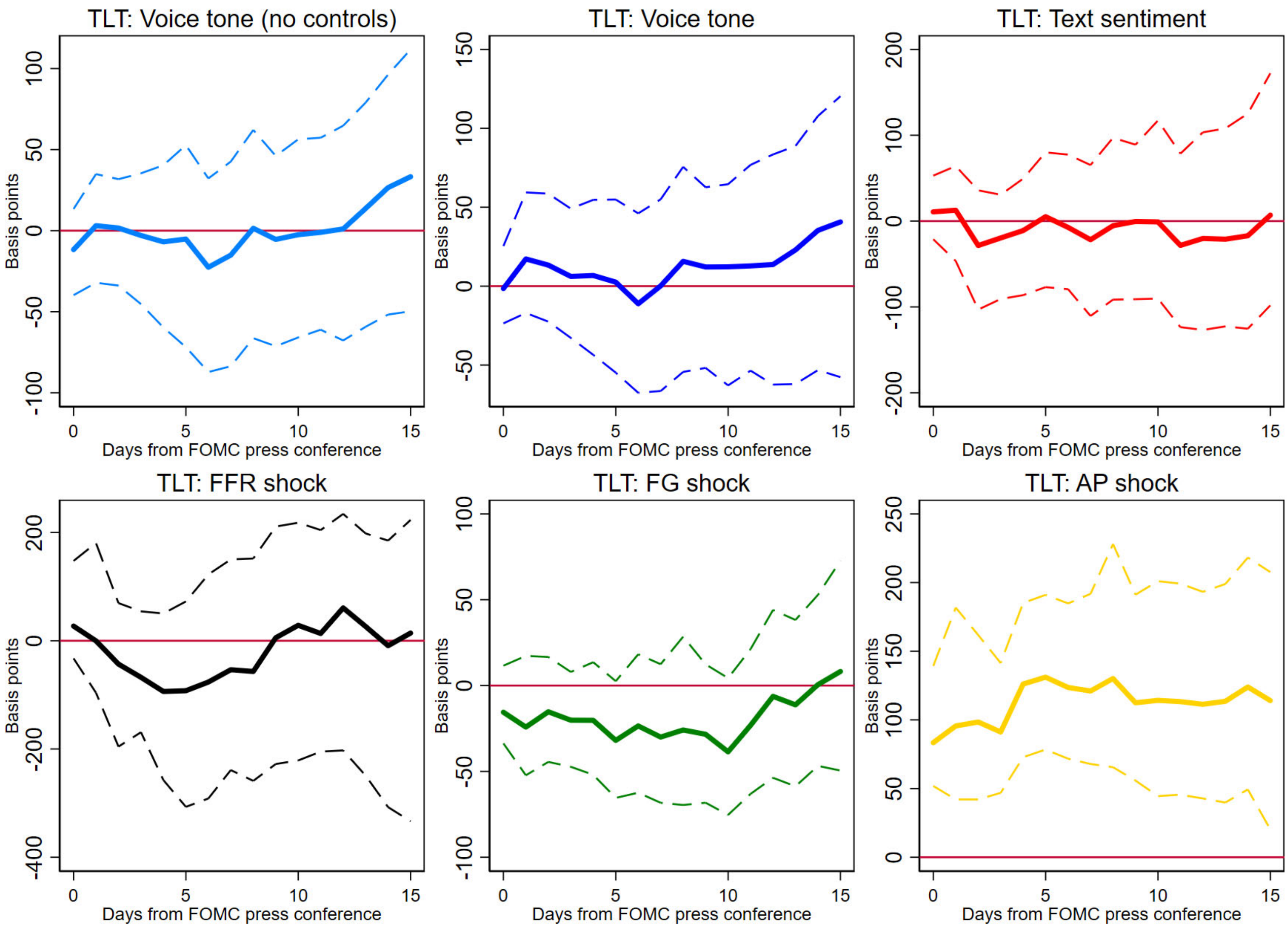
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Appendix Figure A5. Response of TLH ETF (10-20 Year Treasury Bond) to policy actions and messages



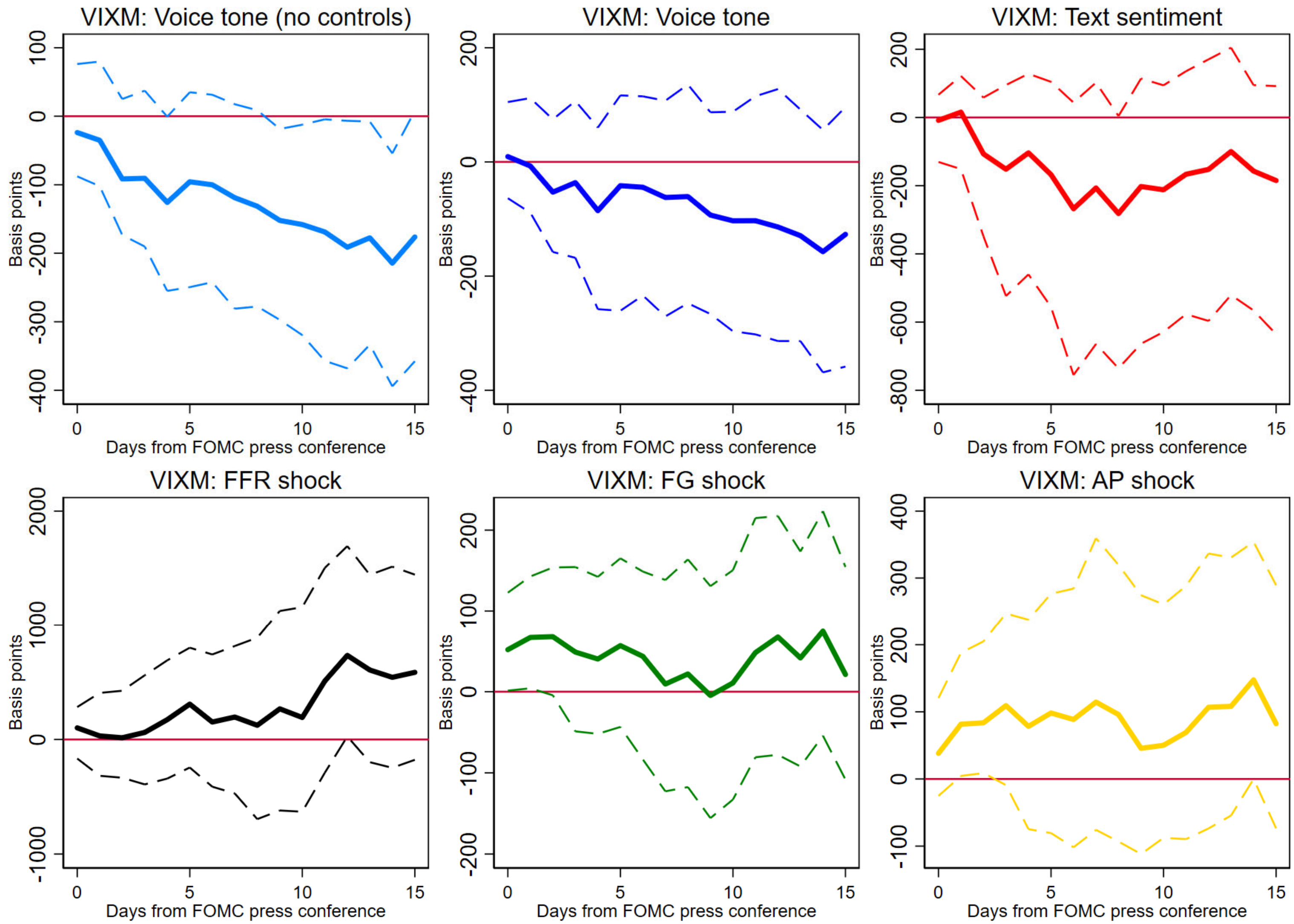
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Appendix Figure A6. Response of TLT ETF (20+ Year Treasury Bond ETF) to policy actions and messages



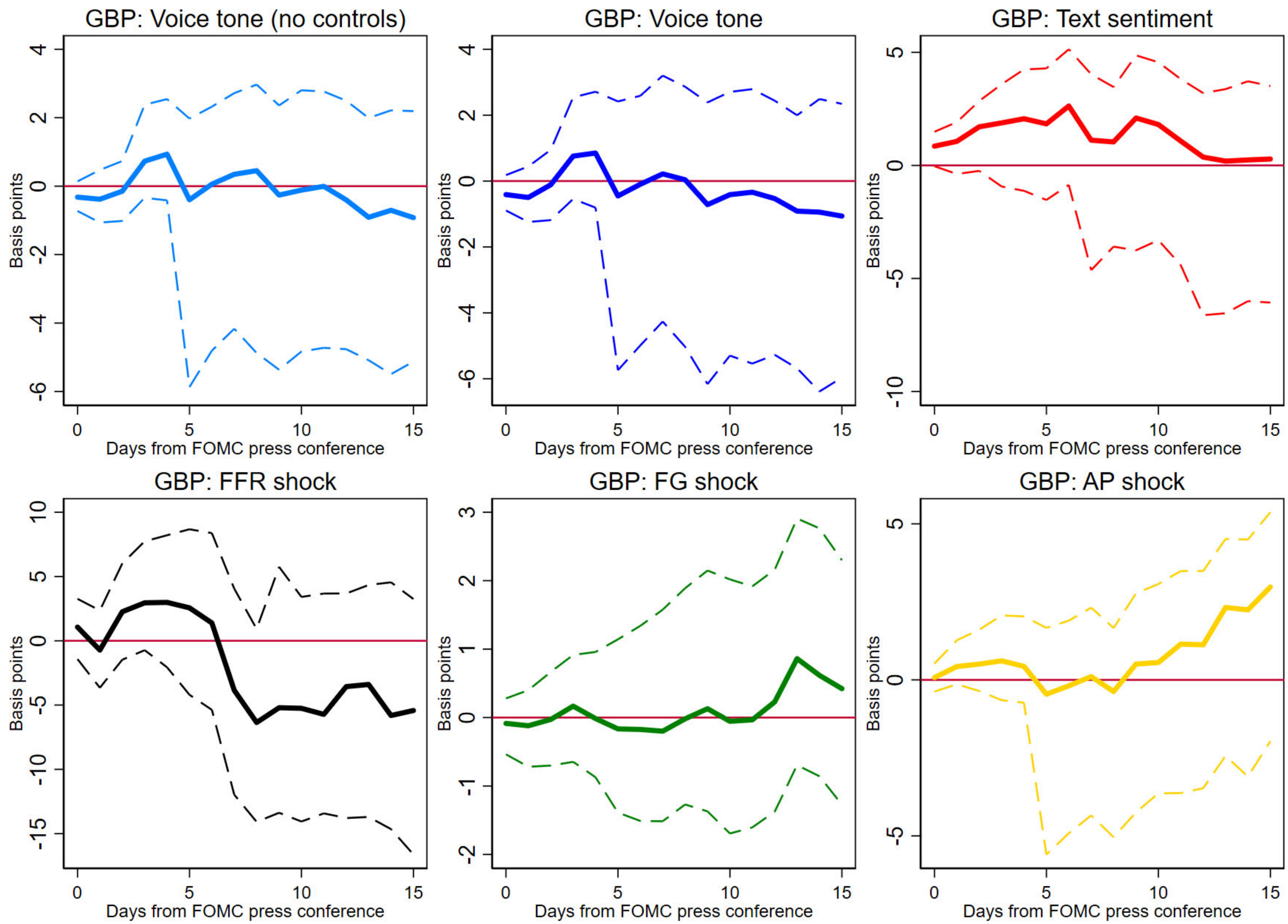
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Appendix Figure A7. Response of VIXM ETF (VIX Mid-Term Futures ETF) to policy actions and messages



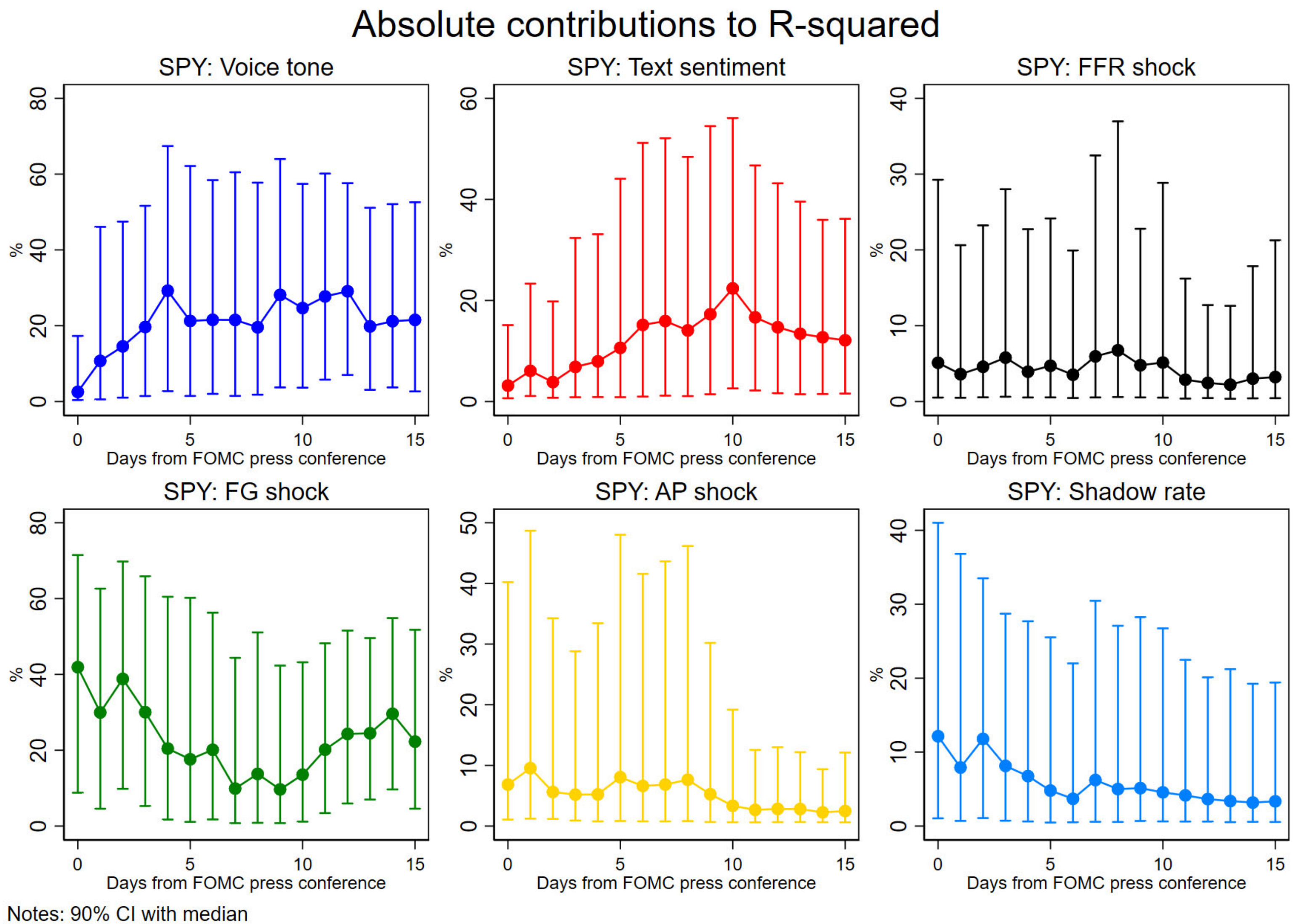
Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Appendix Figure A8. Response of the British Pound to one U.S. Dollar (pound/dollar) exchange rate to policy actions and messages



Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for policy communication/actions. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Appendix Figure A9. Variance decomposition (SPY ETF)



Notes: This figure reports the Shapley decomposition for the estimations with SPY ETF.

Appendix B. Neural network to classify audio tracks into emotions

Audio data

All audio files are converted to 16,000 Hz sample rate and mono channel. When passing the audios into the Librosa package for feature extraction, we use the default frame length (the number of samples in a frame) and the hop length (the number of samples between successive frames) of 2,048 and 512, respectively. Thus, for each audio, the number of frames (or "slices") used for feature extraction is calculated as:

$$frames_s = \frac{duration_s(in\ seconds) \times 16{,}000}{512} \qquad \text{(B1)}$$

Feature extraction

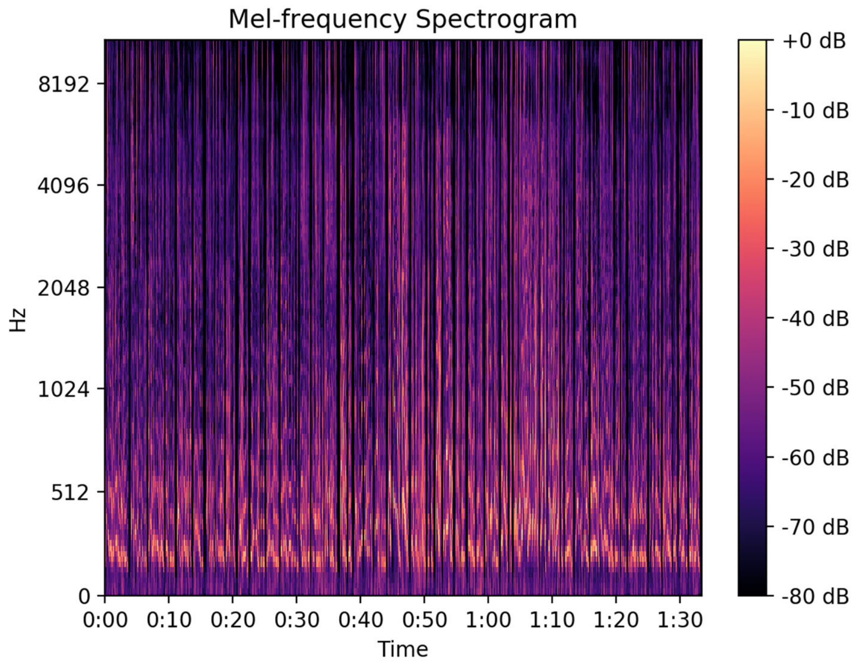
The inputs of our neural network algorithm are essentially the representations of two important vocal aspects, namely frequency (or pitch/highness) and amplitude (or volume/loudness). For an audio signal we can extract the following characteristics:

- A Mel frequency spectrogram is the spectrum of frequencies of an audio mapped onto a Mel scale (instead of the frequency scale) time. It allows us to determine the level of loudness of a particular frequency at a particular time.[29]
- A Chromagram is a representation of an audio signal in which the spectrum of frequencies is projected onto 12 equal-tempered pitch classes or 12 chroma bands (i.e., C, C#, D, D#, E, F, F#, G, G#, A, A#, and B). The Chromagram reflects the distribution of energy along 12 chroma bands over time and, hence, it can capture the melodic and harmonic characteristics of an audio.
- A Mel-frequency cepstral coefficients (MFCC) is a discrete cosine transformation of the Mel frequency spectrogram.

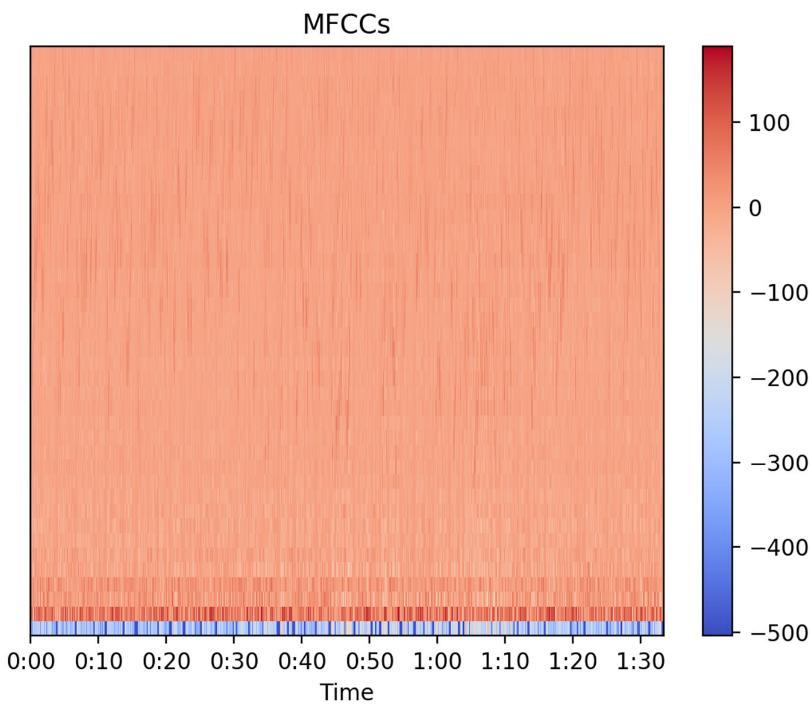
As mentioned in Section 2.1.1, we first extracted a vector of 128 Mel spectrogram coefficients. Appendix Figure B1 presents an example of a Mel spectrogram: the brighter colors around the frequency range of 256 – 512 Hz suggest the stronger (or "louder") amplitude of such a range. Second, the extracted features also include a vector of 40 MFCCs, which are considered to be the decorrelated versions of the Mel spectrogram. The negative MFCCs indicate that the spectral energy is concentrated at the high frequencies, while the positive MFCCs represent the concentration around the low frequencies. This is illustrated in Appendix Figure B2: the majority of cepstral coefficients are positive, corresponding to the stronger amplitude of the 256-512 Hz range suggested in Appendix Figure B1. Finally, the Chromagrams with 12 chroma coefficients are extracted from the audios. In the example shown in Appendix Figure B3, the pitches are scattered and distributed over all pitch classes, which reflects the fact that the examined audio is a sample audio book.[30] All obtained features are then averaged over all frames, meaning that we obtain a set of 180 features, or inputs, for each audio file.

[29] Mel scale is a log transformation of frequencies which "mimic" the human perception of sound. That is, pitches of equal distance on the Mel scale are of equal distance when judged by humans.
[30] This audio can be found at https://secure.toolkitfiles.co.uk/clients/24554/sitedata/files/AudioBook-Tanya-S-Bartlett.mp3.

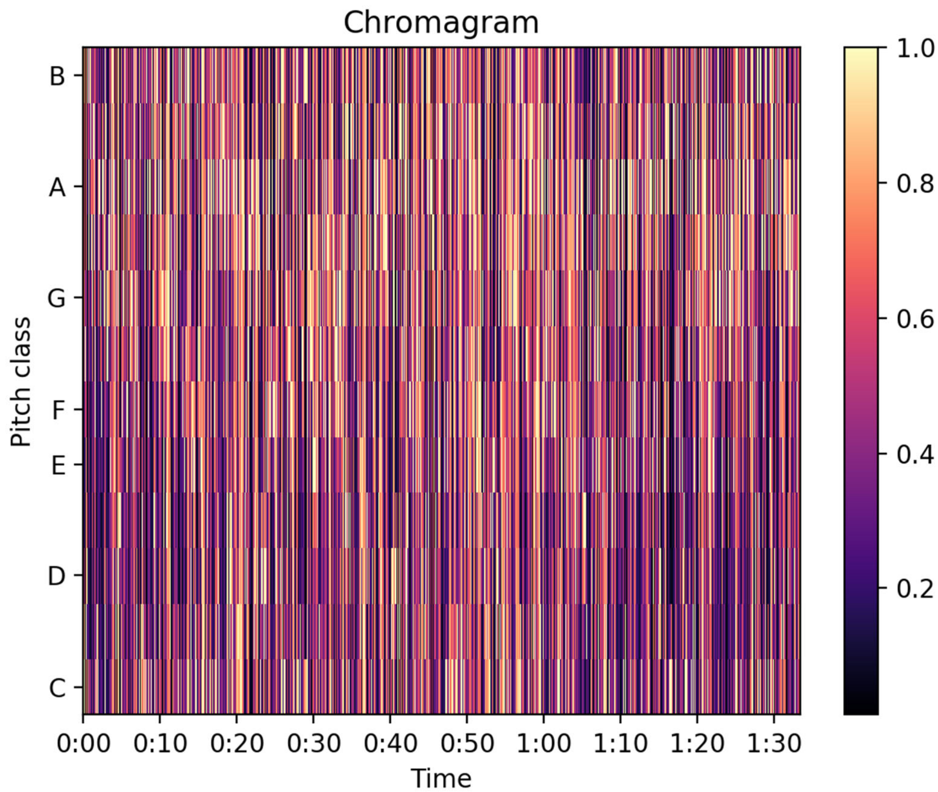
Appendix Figure B1. Example of Mel-frequency Spectrogram



Appendix Figure B2. Example of MFCCs



Appendix Figure B3. Example of Chromagram



It should be noted that the number of Mel spectrogram coefficients, the number of MFCCs, and the number of chroma coefficients are hyperparameters which can be adjusted to achieve a more effective algorithm. Similarly, one might ask whether it is necessary to use both Mel Spectrogram Frequencies and MFCCs as the inputs since the latter is essentially derived from the former. Within the scope of our fine-tuning exercise, we find that using both types of features helps to improve the accuracy of the model.

In addition, there are other spectral features that could be extracted and used in the neural network. For example, a spectral contrast (Contrast) is the level of difference between the mean energy in the top and bottom quantiles of the spectrum. One could also compute the tonal centroid of a chroma vector (Tonnetz), in which the chroma features are projected onto a 6-dimensional basis representing the perfect fifth, minor third, and major third.[31] As part of the fine-tuning exercise, we also experimented with using all five spectral features (Mel spectrogram coefficients, MFCCs, Chromagram, Tonnetz, and Contrast) as the inputs. However, this combination did not improve the accuracy rate.

The neural network

We use Keras, a deep learning API run on top of Google's machine learning platform TensorFlow, to build our neural network. In what follows, we will describe the specific model and training parameters of our network. This network is trained on 80% of TESS and RAVDESS data and tested on the remaining 20%.

Network structure

[31] See https://librosa.org/doc/main/feature.html for more information on various spectral features.

Our neural network is a fully connected network with four layers. This means that every node in one layer is connected to every node in the next layer through an activation function. Particularly, a node in the next layer is connected with all inputs *I* in the previous layer through weight ($w_{k,i}$) and bias (b_k): $\sum_{i=1}^{j} I_i \times w_{k,i} + b_k$.

- The first layer is a dense layer which takes 180 features (128 Mel coefficients, 40 MFCCs, and 12 chroma coefficients) as inputs and passes them through the linear activation function to produce 200 nodes as outputs.
- The second layer has 200 nodes which are connected with 200 nodes in the first layer through the linear activation function.
- The third layer has 200 nodes which are connected with 200 nodes in the second layer through the linear activation function.
- The output layer has five nodes representing five emotions (happy, pleasantly surprised, neutral, sad, and angry). Given that our task is a multi-class classification task, we use the softmax activation function (normalized exponential function), a logistic function, to connect the nodes in this layer with 200 nodes in the previous layer.
- To prevent overfitting, three Dropout layers with a dropout rate of 0.3 are added after each layer before the output layer. This means that 30% of inputs are randomly set to 0 at each step during the training time (hence, only 70% of inputs are retained for training).

Training parameters

- The number of training epochs is 2,000. This means that the entire training dataset is passed forward and backward through the network 2,000 times.
- The batch size is 64. This means that 64 training audio files are propagated through the network (i.e., processed) before the model's weights are updated.
- How the weights are updated is determined by an optimization algorithm. In this study, we use the adaptive moment estimation (Adam) with the default learning rate of 0.001 as the optimizer.
- The loss function, or the error function, is used to optimize the parameter values. Given the multi-class classification task, we use the categorical cross-entropy function which minimizes the distance between the distribution over pre-defined emotions and the "model" distribution over predicted emotions.

To evaluate the model, we use the following formula to calculate the accuracy rate:

$$Accuracy(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n} 1(\hat{y}_i = y_i) \qquad \text{(B2)}$$

where y and $\hat{y}$ are the true emotion and the predicted emotion, respectively. *n* is the number of audio files in the testing dataset.

The accuracy rate of the model used for analysis is 84%. When applying this formula for each of the emotion classes, we obtain accuracy scores of 87%, 84%, 74%, 87%, and 80% for angry, sad, neutral, pleasantly surprised, and happy, respectively.

Appendix C. Neural network to classify (central bank) text sentiment

Text embeddings

We used two different BERT models to extract the word embeddings from texts. The first model is the base uncased model, which has 12 layers, 768 hidden states, 12 heads, 110M parameters, and was trained on lower-case English text. The second model is the RoBERTa model, which has 12 layers, 768 hidden states, 12 heads, and 125M parameters.

The neural network

The sequence of the hidden states at the output of the last layer of the BERT model is used as inputs for the text classification model, which is specified below. This network is trained on 80% of our unique (balanced) labelled FOMC statements data, validated on 10% of the sample, and tested on the remaining 10%.

Network structure

Our neural network's structure is as follows.

- Input layer is the sequence of the hidden states at the output of the last layer of the BERT model.
- The first hidden layer is a bidirectional long short term memory (LSTM) layer created by wrapping a LSTM layer with a Bidirectional layer. The LSTM has 512 units, a dropout rate of 0.1, and a recurrent dropout rate of 0.1. We use the default activation function (hyperbolic tangent). Following the bidirectional LSTM is a dropout layer with a dropout rate of 0.1
- The second hidden layer is a global average pooling 1D layer which is added to flatten the 2-dimensional data into 1-dimensional data, followed by a dropout layer (dropout rate is 0.1).
- The third hidden layer is a dense layer, which has 512 nodes (we use the rectified linear unit activation function). This hidden layer is followed by a dropout layer with a dropout rate of 0.1.
- The fourth hidden layer is a dense layer, which has 128 nodes. The rectified linear unit activation function is used.
- The output layer has three nodes representing three sentiment classes (hawkish, neutral, dovish). We use the softmax activation function for this multi-class classification task.

Training parameters

- The number of training epochs is 200.
- The batch size is 10.
- The optimization algorithm is Adam with the default learning rate of 0.001.
- The loss function (categorical cross-entropy function) is used to optimize the parameter values.

Evaluation

We use formula (B2) to calculate the accuracy score when applying the trained text sentiment model on the testing data. The performance of the model is as follows:

Appendix Table C1

Embeddings	Accuracy score			
	Average	Hawkish	Neutral	Dovish
BERT	81%	85%	77%	79%
RoBERTa	78%	88%	68%	78%

Appendix D. Textual analysis

Appendix D1. Search and count approach

We build four lists of nouns, adjectives, and verbs (Appendix Table D1), combinations of which will indicate either tight monetary policy/strong economic outlook (i.e., hawkish) or expansionary monetary policy/weak economic outlook (i.e., dovish). A phrase combined of (1) A1 and A2 or (2) B1 and B2 is classified as "dovish" while a phrase combined of (1) A1 and B2 or (2) B1 and A2 is classified as "hawkish". To increase the accuracy of our classification, the search and count approach is performed on each part of a sentence then aggregated over the whole document. For example, the sentence "With inflation running persistently below this longer-run goal, the Committee will aim to achieve inflation moderately above two percent for some time so that inflation averages two percent over time and longer-term inflation expectations remain well anchored at two percent" contains two parts: "With inflation running persistently below this longer-run goal" and "the Committee will aim to…two percent". The search and count approach is performed on each part separately, then aggregated over the whole sentence, and then aggregated over the whole document. Since negations such as "won't" or "aren't" can alter the meaning of the text, for each part of a text, a hawkish (dovish) phrase is only counted as hawkish (dovish) if the text does not contain a negation word/phrase. In contrast, if a hawkish phrase is accompanied by a negation word/phrase, then it is counted as dovish and vice versa. A similar approach was applied in Cieslak and Vissing-Jorgensen (2021), where a negative word accompanied by "not" is considered positive. The aggregate sentiment of the text of an FOMC statement/remarks/Q&A is measured as:

$$TextSentiment = \frac{Dovish\ phrases - Hawkish\ phrases}{Dovish\ phrases + Hawkish\ phrases}$$

where $Dovish\ phrases$ and $Hawkish\ phrases$ are the counts of respective phrases in the FOMC statements as well as transcripts when a press conference is held.

Appendix Table D1. Dictionary for hawkish and dovish words

Panel A1	Panel A2
inflation expectation, interest rate, bank rate, fund rate, price, economic activity, inflation, employment	anchor, cut, subdue, declin, decrease, reduc, low, drop, fall, fell, decelarat, slow, pause, pausing, stable, non-accelerating, downward, tighten
Panel B1	**Panel B2**
unemployment, growth, exchange rate, productivity, deficit, demand, job market, monetary policy	ease, easing, rise, rising, increase, expand, improv, strong, upward, raise, high, rapid
Panel C	
weren't, were not, wasn't, was not, did not, didn't, do not, don't, will not, won't	

Notes: This table shows the words/phrases used to classify text into dovish/hawkish.

Appendix D2. Measuring the intensity of text sentiment

As an additional robustness check, we adopt the approach used in Kozlowski et al. (2019) and Jha et al. (2021) to measure the text sentiment's intensity. The steps of this approach can be summarized as follows. First, based on the dictionary in Neuhierl and Weber (2019), we build a dovishness-hawkishness dataset of sentence/phrase pairs with opposite monetary policy stances (see Appendix Table D2 below). Second, we use the BERT algorithm to extract embeddings for the text in this dataset and the policy texts. Third, for each pair of embedding vectors in the dovishness-hawkishness dataset, we calculate the embedding difference between the dovish sentence/phrase and the hawkish counterpart. The average of these dovish-minus-hawkish vectors is considered a dovishness dimension. Finally, the degree of dovishness (or hawkishness) of a given policy text is the cosine similarity score between the policy text's embedding vector and the vector of the dovishness dimension. By construction, this continuous score ranges from -1 to 1 where a positive score indicates a dovish connotation and a negative score represents a hawkish connotation. A higher absolute value of positive (negative) score means a higher degree of dovishness (hawkishness).

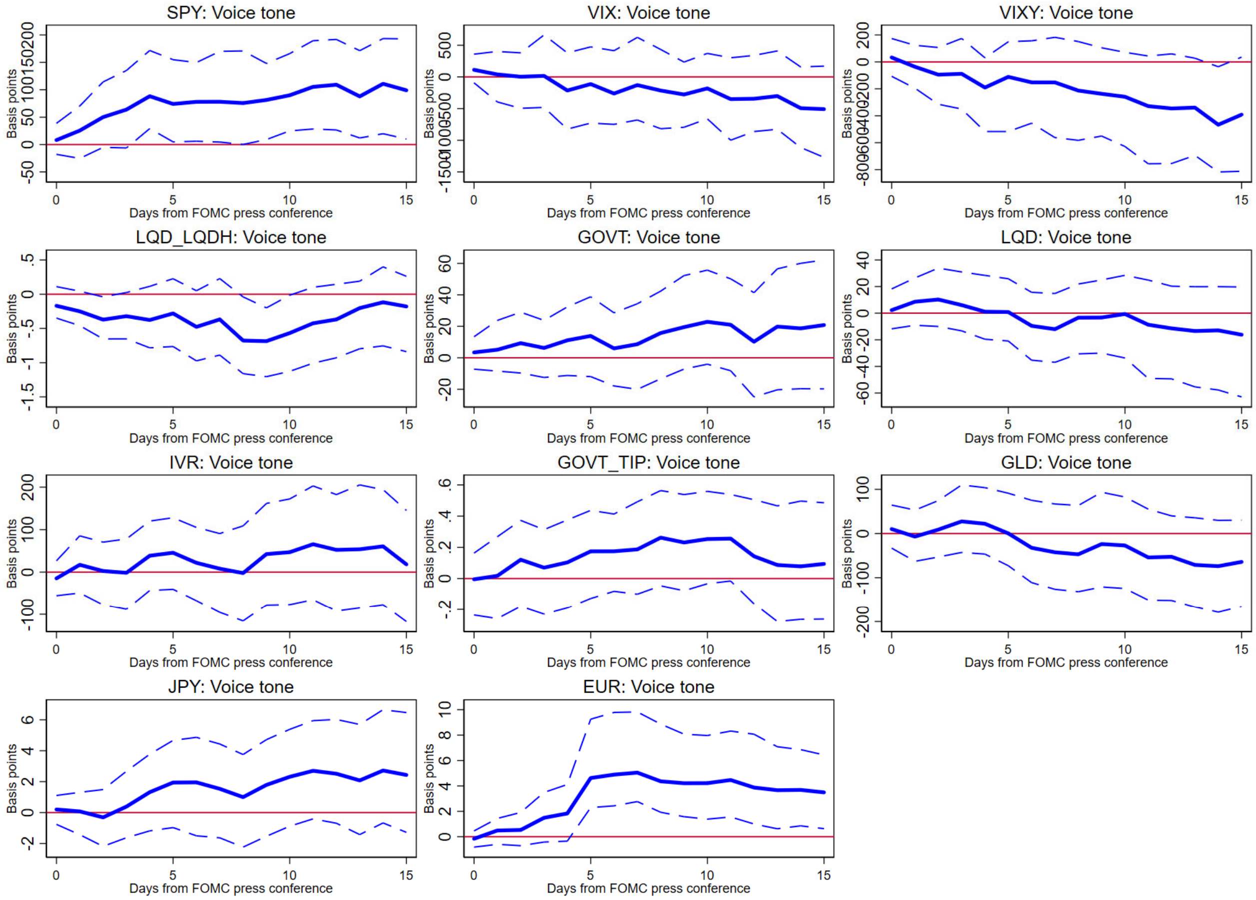
As shown in Appendix Figure D1, the results for the tone of voice measure are consistent when controlling for the degree of dovishness/hawkishness of the policy texts. The consistent findings are also observed when we allow for the non-linear terms in text sentiment (Appendix Figure D2).

Appendix Table D2. Policy stance pairs to measure dovish dimension

Dovish phrases	Hawkish phrases
inflation expectations anchor	inflation expectations increase
anchor inflation expectations	rise inflation expectations
inflation expectations decline	inflation expectations increase
inflation expectations remain stable	inflation expectations higher
inflation expectations stable	inflation expectations higher
stable inflation expectations	rise inflation expectations
lower inflation expectations	higher inflation expectations
reduction inflation expectations	increase inflation expectations
cut federal funds rate	raise federal funds rate
lower federal funds rate	higher federal funds rate
reduce federal funds rate	raise federal funds rate
decrease federal funds rate	increase federal funds rate
reduction federal funds rate	rise federal funds rate
cut interest rate	raise interest rate
lower interest rate	higher interest rate
reduce interest rate	raise interest rate
decrease interest rate	increase interest rate
reduction interest rate	rise interest rate
decline economic activity	increase economic activity
stable inflation	rise inflation
downward pressure inflation	upward pressure inflation
decrease inflation	increase inflation
declined employment	higher employment
employment fallen	employment increased
employment fell	employment increased
unemployment rate rising	unemployment rate lower
increases unemployment rate	declines unemployment rate
rise unemployment rate	drop unemployment rate
higher unemployment rate	lower unemployment rate
dovish monetary policy	hawkish monetary policy
easing monetary policy	tightening monetary policy

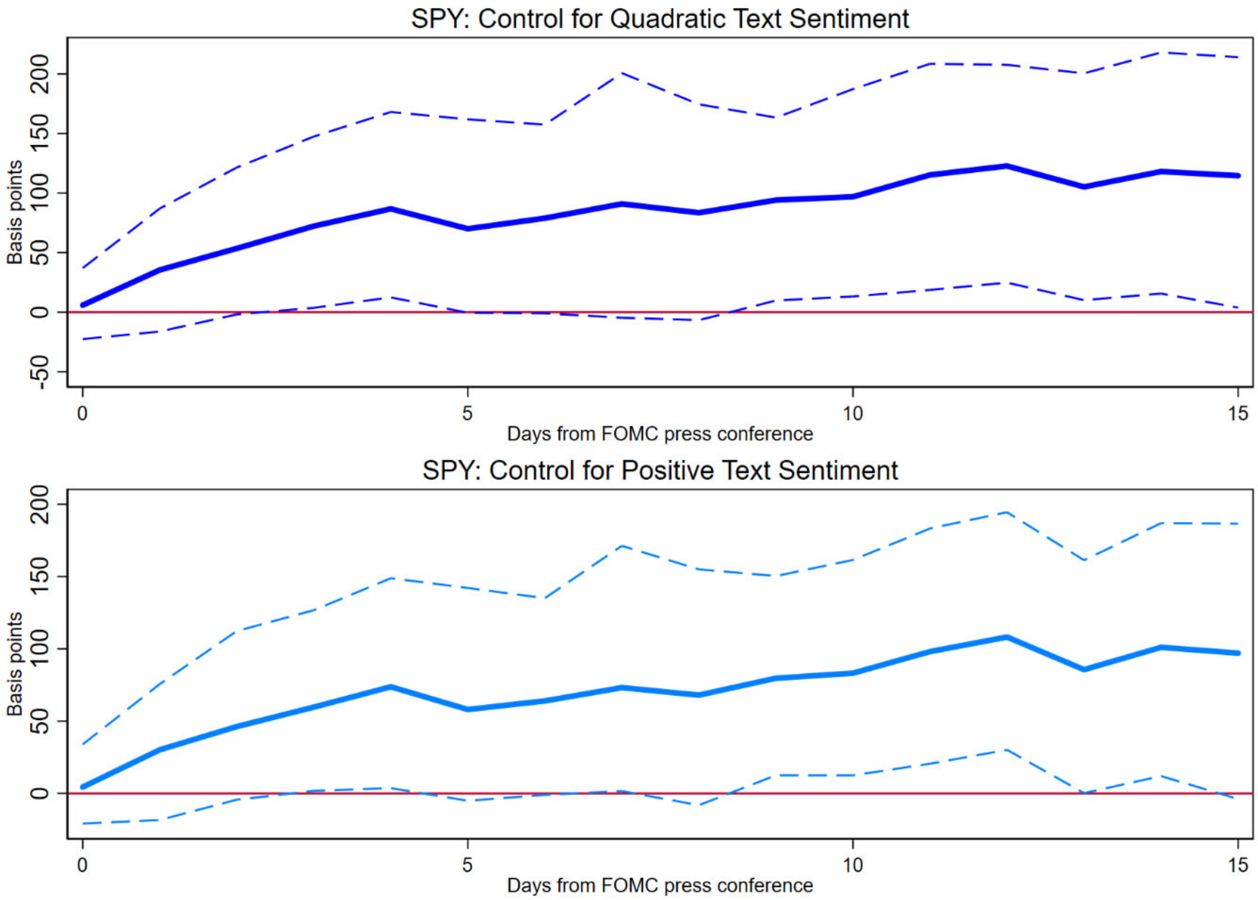
Notes: This table shows the words/phrases used to classify text into dovish/hawkish.

Appendix Figure D1. Control for the intensity of text sentiment



Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for the tone of voice while controlling for the intensity of text sentiment. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.

Appendix Figure D2. Control for the non-linear terms of text sentiment



Notes: This figure reports the estimated slope coefficients *b* (Specification (4)) for the tone of voice while adding the non-linear terms of text sentiment. Dashed lines show 90% bias-corrected and accelerated bootstrap confidence intervals.