The Voice of Monetary Policy

Yuriy Gorodnichenko  Tho Pham  Oleksandr Talavera
University of California, Berkeley, and NBER  University of Reading  University of Birmingham

First draft: October 31, 2020
This draft: January 30, 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 University of Reading, University of Birmingham, 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 Workshop on Empirical Monetary Economics for 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 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 increasingly relying 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 a lot of weight, but little is known about the effects of their non-verbal communication. To shed more 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 in voice tone can move financial markets.

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), press conferences, particularly the Q&A sessions, play a key role in helping financial markets and the general public


2 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).
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 communicate soft information via non-verbal channels, thus potentially influencing investors’ decision making. 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 press conference during the 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 in terms of three emotions: 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 large effects on the stock market. We also find that inflation expectations and exchange rates respond to the variation 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 reading and posting prepared remarks and speeches. It appears that a certain level of acting skill may be helpful to ensure that the public receives the policy message fully and correctly.

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 CEOs’ speeches during

---

3 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.
earnings conference calls, Hobson et al. (2012) find that the vocal markers of cognitive dissonance can predict the likelihood of irregular restatements of earning 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, 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 firms’ fundamentals. In a more recent study, Hu and Ma (2020) find that positivity about a startup, shown through the 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), we differ in 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, room 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 allows us to adjust the number and the class of emotions, which also cannot be done with commercial software and pre-trained algorithms.

Curti and Kazinnik (2021), a concurrent paper 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 with economic players.

The rest of this paper is organized as follows. In the next section, we provide some anecdotal evidence to justify the importance of non-verbal channel 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 results in this section. Finally, Section 5 concludes and discusses implications.
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 look for more information through the aspects which are not explicitly “scripted”, such as the choice of words, the tone of voice or body language of the Fed Chair. Like words, the non-verbal elements of communication can signal the Fed’s perspective on current and future economic outlook and the future course of monetary policy. 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”

Not surprisingly, 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. Indeed, it is not uncommon to have media reports assessing non-verbal elements of press conferences. For example, the Wall Street Journal 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.6

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 Fed Chairs into a spectrum of emotions, and then we study how variations in voice tone (emotions) can affect financial variables.

3  FOMC speeches: voice and linguistic analysis

Our sample runs from 2011 (when the first 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 collected from the Federal Reserve (Fed) website. The press conferences’ videos are downloaded from the Fed’s official Youtube channel. We use only the audio component of these videos.7 Because 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 reader to Appendix B for more details. Conceptually, we need 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 in determining the emotion of a speaker. Building on earlier research on voice recognition (e.g.,

---

7 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 measurement.
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 chromogram 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 improve the accuracy of the model. Note that the number of Mel spectrogram coefficients, MFCCs, and Chroma coefficients can be adjusted to achieve better prediction.

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 and eight emotions (calm, happy, sad, angry, fearful, surprised, disgust, neutral). The second data set is the Toronto Emotional Speech Set (TESS). To create these data, 2 actresses spoke a set of 200 words in 7 emotions (happy, sad, angry, fearful, pleasantly surprised, disgust, 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). Because 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 on 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 on these outputs gives us the final output (i.e., the classification of emotions).

In this study, we build a fully connected network with two hidden layers of which the architecture is specified as follows. The input layer has 180 nodes which are 180 vocal features (128 Mel coefficients, 40 MFCCs, and 12 Chroma coefficients). The first hidden layer has 200 nodes which are connected with 180 nodes in the input layer through the linear activation function. The second hidden layer has 200 nodes which are connected with 200 nodes in the first hidden 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. 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 inputs are randomly set to 0 at each step during training time (hence, only 70% of inputs are retained for training). After training the model, we use the accuracy score to evaluate the model’s performance

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

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 trained model gives us an overall accuracy score of 84%. Applying the same formula for each emotion, we get the accuracy in predicting angry, sad, neutral, surprised, and happy emotion 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 measure sensitivity to various assumptions and data points. In contrast, using humans to classify Fed Chairs’ voice into emotions is not only much more costly but also likely involves judgment, biases, conflicting interpretations, and perhaps too much “reading between the lines”. However, humans may be better at picking up subtle tone variations.
3.1.2 FOMC audio input

FOMC press conferences’ audios are manually processed and split into smaller audio segments where each segment 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 of three speakers (Ben Bernanke, Janet Yellen, and Jerome Powell). The length of audio segments varies from 10 seconds to more than 5 minutes, but most answers are between 1 and 3 minutes. 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 is classified into positive (“happy” or “pleasantly surprised”), negative (“sad” or “angry”), or neutral. We aggregate the tone to the press conference level as follows:

\[
VoiceTone = \frac{Positive answers - Negative answers}{Positive answers + Negative answers}
\]  

(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 1.

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. The serial correlation in the voice tone across meetings is close to zero (\(\rho =0.08\)).
Although we do not have a standard benchmark to validate \textit{VoiceTone}, we can use reports in the media and other external information to get a sense of how \textit{VoiceTone} aligns with other sources. For example, Ben Bernanke tends to get positive score values of \textit{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.\footnote{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?”} He also had to say that the Fed was working with untested policies. In addition, Bernanke had to answer questions about his own future (i.e., Janet Yellen was expected to be officially nominated to chair the Fed) and his regrets. Thus, it is natural that he sounded rather negative.

Consider now the press conference on September 17, 2015 when Janet Yellen has 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”\footnote{https://www.theguardian.com/commentisfree/2014/jun/18/fed-janet-yellen-economy-inflation (Accessed on January 16 2022)}, “[Yellen] expressed scepticism that the official, headline unemployment rate (6.3%) was an adequate reflection of labor market tightness”\footnote{https://www.businessinsider.com/janet-yellen-dove-2014-6 (Accessed on January 16 2022)}, and “Yellen struck a dovish tone”\footnote{https://www.ftadviser.com/2015/10/06/investments/fixed-income/markets-no-longer-listening-to-fed-vGD6k9R3rbm3YQYknBjsTN/article.html (Accessed on January 16 2022)}. 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 has a low score (-1) when according to the Financial Times\footnote{https://www.ft.com/content/1047a2a8-6f7e-11e8-92d3-6c13e5e92914 (Accessed on January 16 2022)}, “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”\textsuperscript{13}.

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 nuances picked 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 cacophony and hence confusion, verbal and non-verbal messages should be congruent and reinforce each other. But 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 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\textsuperscript{14} 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 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 embeddings of a word.\textsuperscript{15}

\textsuperscript{13}https://www.afr.com/markets/equity-markets/asx-advances-on-dovish-fed-20190322-p516gq
\textsuperscript{14}In a nutshell, 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).
\textsuperscript{15}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).
Devlin et al. (2018) report that, as a result of these features, BERT has very high accuracy in interpreting texts.

Although pre-trained BERT models can assign some interpretations to texts (e.g., positive or negative), previous works on textual analysis of policy communication focused on the dovish-hawkish spectrum. To bridge our work to earlier studies, we need to make an extra 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 \( \geq 0.5 \)), hawkish (the average score \( \leq -0.5 \)), or neutral (the average score is between -0.5 and +0.5). Then we use the word embeddings obtained from BERT as the inputs for a neural network\(^{16}\) 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 score in predicting hawkish, neutral, and dovish sentiment is 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 sentiment of the text of an FOMC statement, remarks, and Q&A as follows:

\[
TextSentiment = \frac{Dovish \text{ text} - Hawkish \text{ text}}{Dovish \text{ text} + Hawkish \text{ text}}
\]  

(3)

where \textit{Dovish text} and \textit{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, \textit{TextSentiment} is in the range of [-1;1] and a positive value

\(^{16}\) 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 \(\times\) 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 second hidden layer is a dense layer with 512 nodes and the third hidden layer is a dense layer with 128 nodes. A dropout layer (the dropout rate of 0.1) is added after the second and third hidden layers to prevent overfitting. The output layer has 3 nodes representing 3 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 details.
of TextSentiment indicates that an expansionary monetary policy is expected or is being implemented.

3.2.2 Sentiment analysis output

We find (Panel B of Table 1) that the sentiment of texts during the terms of Yellen and Bernanke was generally more dovish than Powell’s.\textsuperscript{17} 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 of 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 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

How much 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 in 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

\textsuperscript{17} The text sentiment scores for each meeting are listed in Appendix Table A2.
Q&A responses may generate variation in policy communication that is unrelated to what is contained in the texts of policymakers’ statements, remarks, or even 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) identified as 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 (measured as in Wu and Xia, 2016)\(^{18}\) 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 want to isolate the effect of voice tone. But 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 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 people, 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 to detect the effects of policy communication because it is less clouded by potentially confounding factors such as changes in the fed funds rate.\(^{19}\)

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):

\[
Outcome_{t,t+h} = b_0^{(h)} + b_1^{(h)} \text{VoiceTone}_t + b_2^{(h)} \text{TextSentiment}_t + b_3^{(h)} \text{FFRShock}_t + b_4^{(h)} \text{FGShock}_t + b_5^{(h)} \text{APShock}_t + b_6^{(h)} \text{ShadowRate}_t + b_7^{(h)} \mathbb{I}\{\text{NoPressConference}_t\} + \text{error}_t^{(h)}
\]

\(^{18}\)The updated series of the shadow rate are available at https://sites.google.com/view/jingcynthiawu/shadow-rates.

\(^{19}\)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 an intentional move or a slip-up. If the latter is the case, one may be concerned that voice control was inadequate and introducing unnecessary volatility.
where \( t \) dates FOMC meetings. \( \text{VoiceTone}_t \) measures voice tone of the Q&A session at date \( t \). \( \text{TextSentiment}_t \) indicates the sentiment in the policy statement, remarks, and Q&A responses. \( \text{FFRShock}_t, \text{FGShock}_t, \) and \( \text{APShock}_t \) 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). \( \text{ShadowRate}_t \) 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 identify more cleanly the effects of voice tone on outcome variables. Note that we code \( \text{VoiceTone} \) 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 \( \mathbb{I}\{\text{NoPressConference}_t\} \) which equals to one when an FOMC meeting did not have a press conference.

We estimate specification (4) for each horizon \( h \) (\( h \in [0; 15] \)) separately and plot the estimated coefficients, e.g., \( \{\beta_3^{(h)}\}_{h=0}^H \), 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 dramatic volatility for 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 estimates responses to policy shocks \( \{\beta_3^{(h)}\}_{h=0}^H, \{\beta_4^{(h)}\}_{h=0}^H, \{\beta_5^{(h)}\}_{h=0}^H \) and text sentiment \( \{\beta_2^{(h)}\}_{h=0}^H \) 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 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., $Outcome^{SPY}_{t+k} = \log(SPY_{t+k}^{close}) - \log(SPY_{t+k}^{open})$. Hence, the return on the day of an FOMC meeting is the log difference between close and open prices.

Because the sample is relatively small (68 FOMC meetings), we estimate specification (4) using bootstrap methods to correct for possible biases in the estimates as well as to construct confidence intervals with good coverage. Specifically, 90 percent confidence intervals are based on the 5th and 95th percentiles of the estimates generated in bootstrap draws. 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 (Panel B of Figure 4) that a more positive voice tone leads to an increase in share prices. Specifically, the impact response (i.e., $h = 0$) of the stock market is weak and not significant statistically. 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 talking more about 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 changes in the short-term policy rate probably provided a rather 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 reaction of the stock market (see, e.g., Chodorow-Reich, 2004, 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. 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 have 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 larger responsiveness of treasury yields to macro news during weak forms of forward guidance which can be interpreted as evidence of a larger uncertainty on 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

---

20 Appendix Figure A9 plots the contribution by horizon.
measure the interest rate risk with the following spread: $\text{Outcome}_{t,t+h} = \log \left( \frac{P_{t+h,\text{close}}^{\text{LQD}}}{P_{t,\text{close}}^{\text{LQD}}} \right) - \log \left( \frac{P_{t+h,\text{close}}^{\text{LQDH}}}{P_{t,\text{close}}^{\text{LQDH}}} \right)$, where $P_{t,\text{close}}^{\text{LQD}}$ is the price of LQD ETF (investment grade corporate bonds) and $P_{t,\text{close}}^{\text{LQDH}}$ 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 investors’ 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 rate. A unit decrease in voice tone and a one-standard-deviation increase in the FG shock generate similar responses of the spread, again pointing to the economic significance of the tone of voice. In contrast, an AP shock could signal the 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 many 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 document later 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) or 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. As a first stab at this question, we use the LQD ETF (a fund covering investment grade corporate bonds) and find (Figure 9) that policy actions (FFR/FG/AP shocks) tend to move yields in the same direction as they move yields for U.S. government debt. 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 $Outcome_{t,t+h} = \log \left( \frac{P^{GOVT}_{t+h,close}}{P^{GOVT}_{t,open}} \right) - \log \left( \frac{P^{TIP}_{t+h,close}}{P^{TIP}_{t,open}} \right)$ 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 one 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 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 more 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). 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 5 days after the FOMC meetings). After a more dovish text sentiment (opposite to monetary tightening), the dollar appreciates against the euro while the response of 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 statistically nor economically significant response for the dollar/yen exchange rate. While somewhat unexpected, the relatively

---

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

22 Analyses using intraday data (e.g., Swanson, 2020) find that the dollar appreciates against the euro and yen after a FFR shock.
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. Because 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 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 releases over the examined period, there are few 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. Because macroeconomic news can move financial markets (Gurkaynak et al., 2005), we introduce an additional control in specification (4), the Citi Citigroup Economic Surprise Index, which aggregates macroeconomic surprises in released data into one indicator. We find (Panel A of Figure 15) that voice tone continues to move the stock market. 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).23

One may also 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, press releases, etc. 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. Then we use our BERT-based sentiment algorithm to process the news and construct a measure of media sentiment about monetary policy for each day $t$:

$$\text{MediaSentiment}_t = \frac{\#\text{Dovish articles}_t - \#\text{Hawkish articles}_t}{\#\text{Articles}_t}$$

To smooth out noise, we compute 5-day averages of $\text{MediaSentiment}$ before each FOMC meeting. Including this control does materially change our results (Panel C, Figure 15).

As we indicated above, Fed Chairs vary in their voice tones. Because 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 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 have 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 ten or so days, but then the effects 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 bigger batches more data, trained on longer sequence, and dynamically changes the masking pattern applied to the training data while removing the next sentence prediction objective. At the cost of being much more computationally expensive, RoBERTa can yield a modest improvement in accuracy. Second, the pre-trained BERT sentiment classification is based on the 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 hawkish/dovish spectrum. As an alternative to our approach, we use FinBERT (Araci, 2019), a pre-trained BERT model which was fine-tuned for 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, and then compute the frequency of words in these dictionaries (Appendix D provides more details). 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 could be better in picking 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.

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 in 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. What information is communicated exactly is open to further inquiry, but some 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, 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 inflation dynamics and the pace of the economic recovery after the Great Recession. Through a positive voice tone, he or she can communicate that monetary tightening aimed to fight 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.
Which of these accounts—and there could be other explanations—is a better rationalization of financial market reactions to the non-verbal communication is a challenging (separating forward guidance and information effects is hard, see e.g., Bauer and Swanson, 2020) but fruitful avenue for future research. In any case, 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. Indeed, it is not unusual for Fed officials to have speeches and press conferences aimed to clarify 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 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 potential confusion about the central bank’s intentions.

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 then the initial momentum is 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

24 We thank William English for suggesting this channel.
(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 trades. It can take time to filter out that part of policy communication (perhaps rewatch a press conference) and form a narrative in the market that can justify an adjustment in asset prices. To the extent 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 well-understood changes in policy (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 media coverage develops gradually while a consensus view about the message at a press conference is being reached, one then may observe a gradual response of financial markets to variation 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 (Figure 16): after a positive voice tone, media sentiment becomes more dovish on impact, builds up for the next few days, peaks after about a week and then converges back to zero. This pattern, coupled with the evidence of investors’ reactions to the media sentiment documented in the literature, suggests that FOMC-related

25 There is some anecdotal evidence that information revealed at FOMC press conferences dissimates gradually. For example, by 21 March 2021, the FOMC press conference on 17 March 2021 was watched 55,172 times on Yahoo Finance’s Youtube channel and 143,093 times on CNBC’s Youtube channel. By the time of this writing (December 2021), the number of views rose to 59,988 and 174,148, respectively.
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 better measure messages in written policy documents and thus 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. But 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 of policy communication may have a significant effect on the stock market above and beyond what 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 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 work should dig deeper into understanding nuances of using voice to communicate policy, our results clearly have important policy implications. How messages are said appears to be potentially as important as what is in the messages. 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. This does not make the job of central bankers easier and possibly adds another qualification (voice control) to the job requirement. However, this could be a prerequisite for anybody interested in using a public arena for policy communication. Indeed, to paraphrase Ronald Reagan, how can a Fed Chair not be an actor?
References


### Table 1. FOMC meetings’ statistics

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Bernanke</th>
<th>Yellen</th>
<th>Powell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meetings</td>
<td>68</td>
<td>25</td>
<td>32</td>
<td>11</td>
</tr>
<tr>
<td>Press conference</td>
<td>36</td>
<td>12</td>
<td>16</td>
<td>8</td>
</tr>
</tbody>
</table>

**Panel A.** Voice analysis of responses in Q&A during press conferences

<table>
<thead>
<tr>
<th>Answers (count)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>377</td>
<td>200</td>
<td>109</td>
<td>68</td>
</tr>
<tr>
<td>Negative</td>
<td>285</td>
<td>43</td>
<td>131</td>
<td>111</td>
</tr>
<tr>
<td>Neutral</td>
<td>30</td>
<td>0</td>
<td>28</td>
<td>2</td>
</tr>
<tr>
<td>Voice tone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.09</td>
<td>0.64</td>
<td>-0.13</td>
<td>-0.30</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.75</td>
<td>0.58</td>
<td>0.61</td>
<td>0.82</td>
</tr>
</tbody>
</table>

**Panel B.** Textual analysis

<table>
<thead>
<tr>
<th>Statement</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawkish paragraphs</td>
<td>37</td>
<td>8</td>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>Dovish paragraphs</td>
<td>223</td>
<td>105</td>
<td>108</td>
<td>10</td>
</tr>
<tr>
<td>Neutral paragraphs</td>
<td>64</td>
<td>6</td>
<td>37</td>
<td>21</td>
</tr>
<tr>
<td>Text sentiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.66</td>
<td>0.86</td>
<td>0.71</td>
<td>0.09</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.53</td>
<td>0.27</td>
<td>0.34</td>
<td>0.94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Remarks</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawkish paragraphs</td>
<td>83</td>
<td>15</td>
<td>47</td>
<td>21</td>
</tr>
<tr>
<td>Dovish paragraphs</td>
<td>244</td>
<td>97</td>
<td>106</td>
<td>41</td>
</tr>
<tr>
<td>Neutral paragraphs</td>
<td>119</td>
<td>41</td>
<td>50</td>
<td>28</td>
</tr>
<tr>
<td>Text sentiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.50</td>
<td>0.75</td>
<td>0.38</td>
<td>0.35</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.37</td>
<td>0.24</td>
<td>0.33</td>
<td>0.47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q&amp;A</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawkish answers</td>
<td>233</td>
<td>77</td>
<td>95</td>
<td>61</td>
</tr>
<tr>
<td>Dovish answers</td>
<td>339</td>
<td>137</td>
<td>120</td>
<td>82</td>
</tr>
<tr>
<td>Neutral answers</td>
<td>120</td>
<td>29</td>
<td>53</td>
<td>38</td>
</tr>
<tr>
<td>Text sentiment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.18</td>
<td>0.29</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.30</td>
<td>0.30</td>
<td>0.26</td>
<td>0.35</td>
</tr>
</tbody>
</table>

**Statement, Remarks, Q&A**

<table>
<thead>
<tr>
<th>Text sentiment</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.52</td>
<td>0.71</td>
<td>0.50</td>
<td>0.15</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.40</td>
<td>0.30</td>
<td>0.35</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Notes: This table shows the statistics related to 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 ($\text{Voice Tone} = \frac{\text{Positive answers} - \text{Negative answers}}{\text{Positive answers} + \text{Negative answers}}$). Hawkish and Dovish are the number of hawkish and dovish answers/sentences in the text, respectively. The average text sentiment is measured by $\text{Text Sentiment} = \frac{\text{Dovish text} - \text{Hawkish text}}{\text{Dovish text} + \text{Hawkish text}}$. 


Figure 1. Conceptual neural network

Notes: This figure shows a conceptual neural network with 3 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 ($H_N$, $k=[1;4]$) which are the activation functions of the input features $IN_i$ ($i=[1;3]$). A node $H_N$ is connected with the input the through weight ($w_{ik}$) and bias ($b_k$): $\sum_{i=1}^{3} IN_i \times w_{ik} + b_k$. 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 5 output nodes.
Figure 2. Voice tone vs. Text sentiment.

Notes: The figure shows the joint distribution of voice tone and text sentiment across FOMC meetings.
Figure 3. Policy Words vs. Actions.

Notes: The 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: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Figure 5. Response of VIX (CBOE Volatility Index) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Figure 6. Response of VIXY ETF (VIX Short-Term Futures) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Figure 7. Response of LQD ETF (investment grade corporate bond) minus LQDH ETF (interest rate hedged corporate bond) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Figure 8. Response of GOVT ETF (U.S. government debt) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Figure 9. Response of LQD ETF (corporate debt) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Figure 10. Response of IVR ETF (debt for the real estate sector) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Figure 11. Response of GOVT ETF (nominal U.S. government debt) minus TIP ETF (inflation-protected U.S. government debt) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Figure 12. Response of GLD ETF (gold) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% 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: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Figure 14. Response of the Euro to one U.S. Dollar (dollar/euro) exchange rate to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Figure 15. Robustness checks

Notes: the figure reports 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% bootstrap confidence intervals.
Figure 16. Post-FOMC media coverage vs. FOMC voice tone

Notes: The figure plots impulse responses for cumulative media sentiment ($\sum_{t=0}^{h} MediaSentiment_{t+h}$) to a unit increase in voice tone. The specification is given by equation (4). Dashed lines show 90% bootstrap confidence intervals.
Online Appendix

Additional Information, Tables, and Figures
Appendix A

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

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

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

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: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Appendix Figure A2. Response of SHY ETF (1-3 Year Treasury Bond ETF) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Appendix Figure A3. Response of IEI ETF (3-7 Year Treasury Bond ETF) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Appendix Figure A4. Response of IEF ETF (7-10 Year Treasury Bond ETF) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Appendix Figure A5. Response of TLH ETF (10-20 Year Treasury Bond) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Appendix Figure A6. Response of TLT ETF (20+ Year Treasury Bond ETF) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Appendix Figure A7. Response of VIXM ETF (VIX Mid-Term Futures ETF) to policy actions and messages.

Notes: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% 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: the figure reports estimated slope coefficients $b$ (Specification (4)) for policy communication/actions. Dashed lines show 90% bootstrap confidence intervals.
Appendix Figure A9. Variance decomposition (SPY ETF)

Absolute contributions to R-squared

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:

$$\text{frames}_s = \frac{\text{duration}_s \times \text{sample rate}}{512}$$

(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.\(^{26}\)
- A Chromagram is a representation of 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, hence, can capture 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 range. Second, the extracted features also include a vector of 40 MFCCs, which are considered 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.\(^{27}\) All obtained features are then averaged over all frames, so that we get a set of 180 features, or inputs, for each audio file.

\(^{26}\) Mel scale is a log transformation of frequencies which “mimic” humans’ perception of sound. That is, pitches of equal distance on the Mel scale are of equal distance when judged by humans.

\(^{27}\) 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
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 better-performing 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 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 difference between the mean energy in the top and bottom quantiles of the spectrum. One could also compute 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. As a part of fine-tuning exercise, we also experimented with using 5 features (Mel spectrogram coefficients, MFCCs, Chromagram, Tonnetz, and Contrast) as the inputs. However, this combination did not help 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 validated on the remaining 20%.

Network structure

Our neural network is a fully connected network with 2 hidden layers. It 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 $(w_{k,l})$ and bias $(b_k)$: $\sum_{i=1}^{I} I_i \times w_{k,l} + b_k$.

- The input layer has 180 nodes which are 180 features (128 Mel coefficients, 40 MFCCs, and 12 chroma coefficients).
- The first hidden layer has 200 nodes which are connected with 180 nodes in the input layer through the linear activation function.
- The second hidden layer has 200 nodes which are connected with 200 nodes in the first hidden layer through the linear activation function.
- The output layer has 5 nodes representing 5 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 second hidden layer.
- To prevent overfitting, 3 Dropout layers with a dropout rate of 0.3 are added after the input layer and each of the hidden layers. This means that 30% of inputs are randomly set to 0 at each step during training time (hence, only 70% of inputs are retained for training).

**Training parameters**

- The number of training epochs is 2,000. It means that the entire training dataset is passed forward and backward through the network 2,000 times.
- Batch size is 64. It 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 are 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:

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

(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 get the accuracy score 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-cased 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 labelled FOMC statements data, validated on 10% of the sample, and validated on the remaining 10%.

Network structure
Our neural network consists of 5 layers.
- 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 by wrapping a LSTM layer with a Bidirectional layer. The LSTM has 512 units, dropout rate of 0.1, and recurrent dropout rate of 0.1. We use the default activation function (hyperbolic tangent). Following the bidirectional LSTM is a global average pooling 1D layer which is added to flatten the 2-dimensional data into 1-dimensional data.
- The second hidden layer is a dense layer, which has 512 nodes.
- The third hidden layer is a dense layer, which has 128 nodes.
- After the second and third hidden layers, a dropout layer with the dropout rate of 0.1 is added to prevent overfitting.
- The output layer has 3 nodes representing 3 sentiment classes (hawkish, neutral, dovish). Again, we use the softmax activation function for this multi-class classification task.

Training parameters
- The number of training epochs is 200.
- 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:

<table>
<thead>
<tr>
<th>Embeddings</th>
<th>Average</th>
<th>Hawkish</th>
<th>Neutral</th>
<th>Dovish</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT</td>
<td>81%</td>
<td>85%</td>
<td>77%</td>
<td>79%</td>
</tr>
<tr>
<td>RoBERTa</td>
<td>78%</td>
<td>88%</td>
<td>68%</td>
<td>78%</td>
</tr>
</tbody>
</table>
Appendix D. Textual analysis – search and count approach

We build four lists of nouns, adjectives, and verbs (Appendix Table D1) of which combinations 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 2 percent for some time so that inflation averages 2 percent over time and longer-term inflation expectations remain well anchored at 2 percent” contains 2 parts, “With inflation running persistently below this longer-run goal” and “the Committee will aim to…2 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 (2020), 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.*

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

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