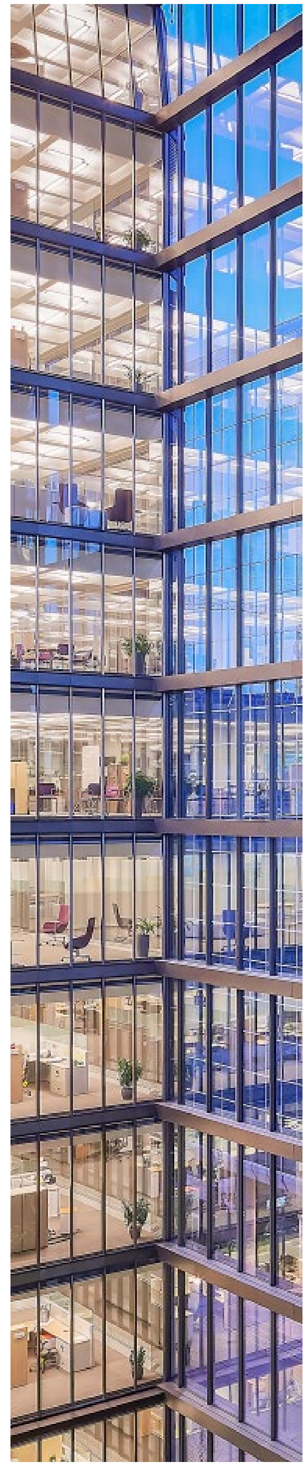
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel—2022-20

Last updated: May 5, 2022



More Than Words: Fed Chairs' Communication During Congressional Testimonies

by Michelle Alexopoulos,[1] Xinfen Han,[2] Oleksiy Kryvtsov[3] and Xu Zhang[2]

[1] University of Toronto

[2] Financial Markets Department
Bank of Canada

[3] Economic and Financial Research Department
Bank of Canada

m.alexopoulos@utoronto.ca, xhan@bankofcanada.ca, okryvtsov@bankofcanada.ca, xzhang@bankofcanada.ca

DOI: https://doi.org/10.34989/swp-2022-20 | **ISSN** 1701-9397

Acknowledgements

We thank participants at the Bank of Canada seminar and Conference on Non-traditional Data, Machine Learning, and Natural Language Processing in Macroeconomics for their comments and suggestions, and the Internet Archive's TV archive group for helping secure video footage. Victoria Bolitho, Aryan Dhar, Jenny Duan, Maryanne Gonzalez, Walter Muiruri, Colleen Spencer, and Sarah Zanette provided superb research assistance.

Abstract

We measure soft information contained in the congressional testimonies of U.S. Federal Reserve Chairs and analyze its effect on financial markets. Our measures of Fed Chairs' emotions expressed in words, voice and facial expressions are created using machine learning. Increases in the Chair's text-, voice-, or face-emotion indices during these testimonies generally raise the S&P500 index and lower the VIX—indicating that these cues help shape market responses to Fed communications. These effects add up and propagate after the testimony, reaching magnitudes comparable to those after a policy rate cut. Markets respond most to the Chair's emotions expressed about issues related to monetary policy.

Topics: Central bank research, Financial markets, Monetary policy communications
JEL codes: E52, E58, E7

"Unlike his predecessor Alan Greenspan, who was famous for convoluted public testimony, Bernanke was clear, concise and brief. He erased any doubts that he, as a longtime academic, could handle the political hot seat. He answered tough questions with confidence, authority and an occasional gentle smile."

– *The Baltimore Sun (February 16, 2006)*[1]

1 Introduction

Central bank leaders have the difficult task of communicating monetary policy to the public. Not only do they need to present complex information in simple and relatable terms, but they also need to be credible and convincing, all the while being at the center of the media's spotlight. The literature has mostly studied *what* central bankers say, analyzing the content and design of central bank press releases, speeches, and policy reports.[2] But *how* central bankers deliver this content to the public, and the impact of the delivery itself, has received less attention. In this paper we address this gap, and instead of the message itself we study how it is delivered by the messenger.

It is well-known in psychology that communication is mainly transmitted via non-verbal cues, such as tone of voice, body language, and facial expressions (Mehrabian 1972). Moreover, humans are less adept at controlling their non-verbal cues than their words (Kahneman 2013). So when central bankers explain their policy during public events, "soft" information contained in their non-verbal or emotional signals may be as meaningful as their words. To study this hypothesis, we measure emotional cues of the Chairs of the U.S. Federal Reserve during congressional testimonies and analyze how they influence financial markets.

Our dataset of emotional cues is constructed using 32 semi-annual congressional testimonies between 2010 and 2017 that were given by two recent Fed Chairs, Ben Bernanke and Janet Yellen. Utilizing audio and video inputs from C-SPAN videos and text inputs from publicly available testimony transcripts, we apply machine learning and big data meth-

[1] `https://www.baltimoresun.com/news/bs-xpm-2006-02-16-0602160182-story.html`

[2] Recent studies include Hansen & McMahon (2016), Bholat et al. (2019), Ehrmann & Talmi (2020), Fraccaroli et al. (2020), Cieslak & Vissing-Jorgensen (2021), Gómez-Cram & Grotteria (2022). Algaba et al. (2020) review econometric methodology for constructing quantitative sentiment variables from qualitative textual, audio, and visual data, and using them in an econometric analysis of the relationships between sentiment and economic variables.

ods to construct three distinct measures of each Fed Chair's emotions expressed via his or her words, voice, and face. To measure stock market prices and their volatility, we use tick-by-tick prices for the S&P500 index and the VIX.

Our results highlight the salience of the soft information contained in the Fed Chair's emotional signals for shaping market responses to Fed communications. In particular, we find that increases in the Chair's text-, voice-, or face-emotion indices during both remarks and Q&A parts of the testimony raise the S&P500 index and lower stock market volatility as measured by the VIX, in most cases. To validate the estimated responses to voice- and face-emotion indices during the remarks, we design an alternative test that exploits a unique feature of semi-annual testimonies whereby the Chair delivers virtually identical results on both days, in front of the House and the Senate. This alternative identification of the effects of Chair's vocal and facial emotions during the remarks corroborates our findings that positive emotional cues work to increase stock prices and decrease stock market volatility.

We provide evidence that market responses during the testimony are economically meaningful. First, we demonstrate that the responses during the testimony add up and propagate in days after the testimony, reaching magnitudes comparable to those after a policy rate cut. Second, during the testimony, market activity is elevated: asset prices are more volatile and trading volumes are higher. Finally, we use changes in the quantity of TV viewership and media coverage to demonstrate that semi-annual testimonies attract public attention, on par with FOMC press conferences.

The magnitudes of the financial market responses vary by the topics discussed during the Q&A rounds and by the Fed's messenger. Discussions of issues directly related to monetary policy (the central bank's reserves, balance sheet management, policy rate, and inflation) are the key drivers of financial asset responses. Markets are more sensitive to Bernanke's emotions, with positive responses to all three of his indices, albeit voice responses are not significant. For Yellen, only the response to her voice-emotion is positive and significant, and it is only half the magnitude of the response to Bernanke's vocal cues. There is no systematic link with other dimensions of the testimony. The responses to congressional members' emotions are weaker, suggesting that markets mostly react to information reflected in the Fed Chair's emotions. We also show that the results are only stronger when we explicitly control for market-wide news during testimonies.

Our paper is closely related to two recent papers that analyze the effects of Fed Chairs' emotional cues on financial markets. Gorodnichenko et al. (2021) find that positive voice tone raises stock prices and lowers their volatility in the days following FOMC press conferences. Curti & Kazinnik (2021) estimate that the Chairs' negative facial expressions during FOMC press conferences lead to lower S&P500 and higher VIX levels. Similar to these papers, ours provides evidence that the Fed Chair's emotions carry meaningful information for stock prices and volatility, both during and after the testimony. While Gorodnichenko et al. (2021) and Curti & Kazinnik (2021) do not find significant evidence that voice or face emotions move interest rate expectations during or after FOMC press conferences, we find that positive textual, vocal, and facial expressions during congressional testimonies raise interest rate expectations, albeit by small magnitudes.

The analysis in our paper complements the analyses in Gorodnichenko et al. (2021) and Curti & Kazinnik (2021) along several new dimensions. While these papers study communications during FOMC press conferences, we study Fed Chair's congressional testimonies, which offer several advantages. Unlike press conferences (Boguth et al. 2019), testimonies do not normally accompany monetary policy announcements, and therefore communications during testimonies are not influenced by the announcement itself. In addition, unlike the tightly scripted and one-directional communication observed during relatively short press conferences, the testimony largely comprises an hours-long Q&A session between the Fed Chair and Congress members in an unscripted, two-directional, and sometimes contentious environment. Such a setting provides more time and scope for the Chair to express her-/himself in more ways than one. While post-FOMC press conferences were introduced in 2010, an examination of testimony coverage can also provide information on the effects of displayed emotions over a much longer time frame. Overall, our evidence suggests that the Fed Chair's emotional cues during testimonies exert similar influence on financial markets as during press conferences.

The second contribution of our analysis is that we consider Fed Chair's emotions jointly. How a person combines his/her words, voice, and face to express themselves, and how these emotions are distilled by others, remains an open research question. Therefore, focusing on only one or two emotions may omit some of soft information that could be inferred from the Chair's delivery. Indeed, in our sample, our three emotion indices are at best weakly correlated, suggesting that the Fed Chair may be using their emotional vehicles separately. We also find that markets are twice as sensitive to a typical (one-standard-deviation) change in the Chair's voice pitch than his/her text sentiment, and roughly five times more sensitive to the change in his/her facial expressions. These rankings are similar whether the Chair delivers the remarks or responds to questions on topics around monetary policy during Q&A. This evidence resonates with the view in psychology that communication is much more than words, and underscores the need for a holistic approach to central bank communication by both academics and practitioners.

Our work also relates to the voluminous finance literature that studies the effects of limited investor attention on financial market outcomes. Variations in investor attention are associated with market-wide news (Yuan 2015), news-searching and news-reading activity (Ben-Rephael et al. 2017), uncertainty about macroeconomic factors affecting the likely path of interest rates (Benamar et al. 2020), and fluctuations in trading volumes (Barber & Odean 2007, Dellavigna & Pollet 2009). We show that congressional testimonies are widely covered in the news and social media, both contemporaneously and in the days following the event. Moreover, they are associated with elevated trading volumes, which the literature would suggest is indicative of heightened investor attention. Emergent literature in behavioral finance uses advanced machine learning techniques to obtain evidence that sentiment indices constructed from audio and photo/video inputs can predict equity returns (Mayew & Venkatachalam 2012, Obaid & Pukthuanthong 2022, Edmans et al. 2022) and detect misreporting (Hobson et al. 2012). Hu & Ma (2021) show that visual, vocal, and verbal persuasiveness is effective during delivery of start-up pitches. We use similar techniques to construct sentiment indices for Fed Chairs.

Finally, our paper also relates to the literature that studies the effects of Fed announcements on financial markets using high-frequency data (Kuttner 2001, Gürkaynak et al. 2005, Nakamura & Steinsson 2018, Cieslak & Schrimpf 2019, Gürkaynak et al. 2021, Swanson 2021).[3] These papers identify the effects of monetary policy surprises by analyzing market behavior within a narrow window around monetary news releases. We build on this approach by analyzing market responses within seconds and minutes after the Fed Chair registers soft information captured in text-, voice-, or face-emotion indices.

The remainder of the paper is organized as follows. Section 2 describes testimony and financial data and explains the construction of three emotion indices. Section 3 lays out the estimation specifications and summarizes the main results. Section 4 discusses the fac-

[3] Andrade & Ferroni (2021) and Altavilla et al. (2019) study the ECB announcements.

tors that drive the results. Section 5 argues that the estimated responses are economically significant. Finally, Section 6 concludes.

2 Data and measurement

2.1 Testimony data

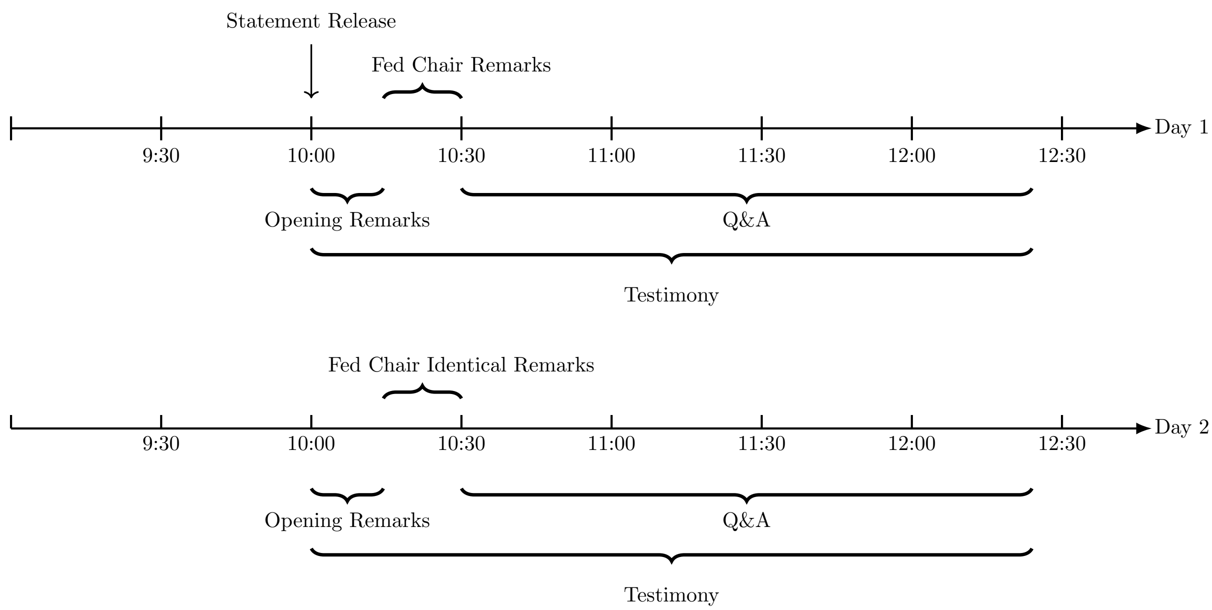


Figure 1. A typical testimony timeline.

Notes: The timeline of events around and during a two-day semi-annual testimony by the Chair of the Federal Reserve for House and Senate Chambers of the U.S. Congress.

To fulfill the requirements of the Humphrey–Hawkins Full Employment Act of 1978, the Federal Reserve issues two Monetary Policy Reports each year. In each, the Federal Reserve summarizes its past policy decisions along with their predicted impacts, as well as recent financial and economic developments for Congress. After each semi-annual report's release, the Chair of the Federal Reserve delivers two congressional testimonies—one in front of the House Financial Services Committee and another in front of the Senate Banking, Housing, and Urban Affairs Committee. The two testimonies normally take place within a day or two of each other, and the order of appearance before the Congress chambers alternates. The timing of events during a typical congressional testimony is depicted in Figure 1. The Fed Chair's remarks are released at the beginning of the first day of these testimonies—usually at 10 a.m. The hearing, often starting at 10:00 a.m., begins with opening remarks by the Committee Chair and other high-ranking committee members and are followed by the prepared remarks of the Fed Chair. The Q&A session then begins upon the conclusion of the Fed Chair's statement. The Q&A session consists of five-minute segments allotted to each committee member, in the order of their seniority, alternating by party affiliation (Congressional Research Service 2010). The testimony lasts several hours and ends with brief concluding remarks by the Committee Chair. The timeline of the second day of the testimony is similar, with the Fed Chair usually delivering precisely the same remarks.

Our data contain textual, vocal, and video inputs for 32 congressional testimonies by Fed Chairs that occurred between February 24, 2010, and July 13, 2017 (see Table 1). The sample covers 16 testimonies by Ben Bernanke (February 24, 2010–July 18, 2013) and 16 testimonies by Janet Yellen (February 11, 2014–July 13, 2017). The testimony transcripts we use were created by CQ transcriptions and obtained from LexisNexis's Nexis-Uni online database.[4] The videos of the C-SPAN broadcasted testimonies are mainly from Internet Archive's TV News collection.[5]

We organize the data along the timeline in semantic blocks. The Fed Chair remarks part of each testimony is divided into blocks of 10 sentences, and the subsequent Q&A part is divided into blocks of Q&A rounds with each congress member. For each block, the text-emotion index reflects the sentiment of the language content in the block. Voice- and face-emotion indices are constructed with audio/video frames recorded within the block of sentences.

We opt to organize our data by blocks rather than by fixed-time windows because it prevents breaking the natural flow of speech and speech-emotions mid-sentence in the remarks

[4]This source, available from https://www.lexisnexis.com/en-us/professional/academic/nexis-uni.page, is used for our analysis since these transcripts capture an unedited version of what was stated during the testimonies and often matched what was heard in the recordings more accurately than the official edited transcripts released with considerable lag.

[5]https://archive.org/. The House and Senate maintain general control over the footage that is recorded and broadcast. The Cable Satellite Public Affairs Network (C-SPAN) is a specialized nation-wide television network that provides de facto exclusive video coverage of Congress proceedings (Eckman 2017).

	All testimonies	Bernanke testimonies	Yellen testimonies
Sample	24 Feb 2010–13 Jul 2017	24 Feb 2010–18 Jul 2013	11 Feb 2014–13 Jul 2017
# of testimonies	32	16	16
Duration (mins)	154	146	161
remarks	15	15	14
Q&A	139	131	148
# of blocks			
remarks	217	127	90
Q&A	661	309	352

Table 1. Summary statistics.

Notes: The full sample includes 32 semi-annual monetary policy testimonies between 2010 and 2017. The remarks part of each testimony is divided into blocks of 10 sentences, and the subsequent Q&A part is divided into blocks of Q&A rounds with each congress member. The number of blocks reflects blocks included in the regression analysis.

or in the middle of the Q&A round. The blocks' lengths are long enough to allow time for accruing speech-emotions and financial market trades, and at the same time short enough to avoid washing out meaningful variation in emotions over the course of the testimony. On average, a sentence lasts 8 seconds, so a block of 10 sentences during prepared remarks lasts slightly more than a minute. A Q&A round, in contrast, is typically 5 minutes. Using this approach, a typical testimony has around 7 remarks blocks and 21 Q&A blocks, and it lasts around 2.5 hours.

2.2 Emotion indices

Based on the audio and video inputs from C-SPAN videos and the text inputs from publicly available testimony transcripts, we construct three distinct measures of each Fed Chair's emotions expressed via his or her words, voice, and face. The details of data processing and construction of indices are provided in the Appendix.

The measure of emotions contained in text or words of the Fed Chair is based on the text-sentiment classifier trained by fine-tuning Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2018), a state-of-the-art natural language processing transformer model, with authors' annotated testimony training data. The process of fine-tuning allowed us to better capture the sentiment expressed during Congressional testimonies.

Experiments in psychology have demonstrated that increases in vocal pitch may reflect a variety of emotions.[6] Therefore, following Dietrich et al. (2019), we utilize the changes of vocal pitch as our measure of vocal emotions. Using the vocal signal processing tool, *Praat* (Boersma & Weenink 2001), we extract the fundamental frequency ($F0$)[7] at 0.015 second intervals. The vocal pitch is measured by calculating the mean $F0$ of each audio sentence.

For the face-emotion measure, we combine the video frame outputs from face recognition and facial expression analysis software[8] to obtain facial muscle action values. Using upper facial actions and the Facial Action Coding System (FACS) created by Ekman & Friesen (1969), we then compute the Chair's face-emotion score as the average of four basic negative emotions—*Sad*, *Angry*, *Fear*, and *Disgust*. We multiply this score by -1 so that high index values indicate less-negative face emotions.[9]

The emotion scores, *Scores*, are calculated at sentence level for text emotions, at 0.015 second intervals for voice, and at video frame level for facial expressions. We define three emotion indices $\text{TEXT}^i_{\tau,b}$, $\text{VOICE}^i_{\tau,b}$, $\text{FACE}^i_{\tau,b}$ for speaker i (or person on screen i), block b, testimony τ as the mean of corresponding emotion scores in that block, standardized by its standard deviation over all blocks in the Q&A:

$$\text{INDEX}^i_{\tau,b} = mean(Scores)/sd_{\text{INDEX}},$$

where $\text{INDEX} \in \{\text{TEXT}, \text{VOICE}, \text{FACE}\}$. Speaker superscript i denotes a chair or a congress member. We define a single index for congress members by pooling all Q&A blocks for different congress members. For the voice-emotion index, raw scores are de-meaned for each speaker to remove differences in individuals' average voice pitch. For the face-emotion

[6]For example, the evidence presented in Kamiloglu et al. (2020) highlights that heightened pitch can be seen coinciding with positive emotions, even though other research commonly explores the relationship between high pitch and stress levels.

[7]$F0$ corresponds to the rate of vocal fold vibrations: high pitch is associated with rapid vibrations and low pitch with slow vibrations

[8]We use Azure Video Indexer for face recognition and identification, and we use FaceReader for facial expression analysis. We process our videos with a frame rate of 29.97, i.e., 29.97 frames per second.

[9]We exclude muscles around the person's mouth because they are activated when the person is speaking, introducing measurement error. We also do not include the basic emotion *Happy* in computation of the face-emotion score because it tends to imply counterintuitive and less significant results, especially for the remarks. We conjecture that because *Happy* is easier to control (e.g., by showing a smile), it is less informative about the speaker than the four negative emotions.

index, we use data for the person on screen instead of the person speaking. By construction, positive index values indicate positive sentiment for text, higher pitch for voice, and less-negative face emotions.

The text-emotion index is different from the stance index used in the literature, which measures the degree of hawkish or dovish sentiment conveyed in the central banks' communications (Ehrmann & Talmi 2020). Therefore, we also construct a stance index for each block of sentences using the dictionary in Gorodnichenko et al. (2021). We use the stance index as a control variable in our empirical analysis.

2.3 Time alignment of emotion and financial data

Synchronizing emotion indices with financial transactions is crucial for accurately identifying the effects of the Fed Chair's communication on financial markets over the course of the congressional testimony. We synchronize the data in three steps. First, we derive corresponding timestamps at the block level for the three emotion indices. Next, we align the block level emotion indices with the time clock on the testimony day. Finally, time-stamped emotion data are matched with tick-by-tick financial market data for analysis. Figure 2 shows the organization of the testimony and financial market data.

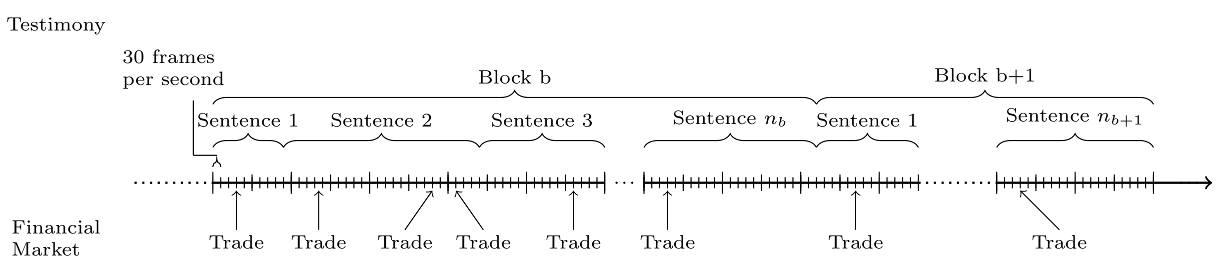


Figure 2. Time alignment of text, voice, video, and financial data.

Notes: The remarks part of each testimony is divided into blocks of 10 sentences, and the subsequent Q&A part is divided into blocks of Q&A rounds with each congress member. The emotion indices summarize the Chair's emotions for each block.

Text sentiment is not correlated with either voice or face indices, whereas voice and face are related (see Appendix). Their relation, however, changes over the testimony. During the remarks, voice and face emotions of the Chair are positively correlated, suggesting that the Chair is using them jointly to support the delivery of his or her remarks. By contrast, voice and face emotions are uncorrelated during the Q&A, suggesting they fulfill a somewhat different role during the Q&A, when the Chair responds to the questions from the congress members. Emotions of Chairs are positively correlated with respective emotions of congress members, suggesting that emotions of the Chair's answers somewhat resonate with emotions of members' questions.

3 Empirical analysis of financial market responses

We estimate financial market responses using high-frequency data for salient financial assets. We use the S&P500 index from TickData to measure stock market price responses, and the VIX from Refinitiv for stock market volatility. To measure U.S. interest rate expectations, we use five-quarter-ahead Eurodollar futures contracts from the Time and Sales database from Chicago Mercantile Exchange. These data are time-stamped by a second.

3.1 Financial market responses: Remarks

The dependent variable $Outcome_{\tau,b+h} - Outcome_{\tau,b}$ is a cumulative change in the outcome for the financial instrument over h minutes starting from the end of block b of testimony τ. For example, for the S&P500 index, $Outcome_{\tau,b+h} - Outcome_{\tau,b}$ denotes the h-minute change in the log price of the S&P500 after the end of block b in testimony τ. We restrict the data to regular trading hours, between 9:35 a.m. and 3:40 p.m.

We use the Jordà (2005) local projections method to estimate the effect of emotions by a Fed Chair in block b during the Chair's remarks for testimony τ on financial market outcomes after h minutes, $h = 1, ..., H$, using the following empirical specification:

$$\begin{aligned} Outcome_{\tau,b+h} - Outcome_{\tau,b} =& \beta^{(h)}_{\text{TEXT}} \text{TEXT}^{\text{CHAIR}}_{\tau,b} + \beta^{(h)}_{\text{VOICE}} \text{VOICE}^{\text{CHAIR}}_{\tau,b} + \beta^{(h)}_{\text{FACE}} \text{FACE}^{\text{CHAIR}}_{\tau,b} \\ &+ \text{controls} + \text{constant} + \varepsilon^{(h)}_{\tau,b}. \end{aligned} \tag{1}$$

The set of controls includes testimony fixed effects and the stance index measuring dovish/hawkish statements. Because congressional testimonies are not accompanied by a policy announcement, we do not need the controls for policy surprise, which were necessary in studies of

FOMC press conferences (Gorodnichenko et al. 2021, Curti & Kazinnik 2021). Specification (1) is estimated by fixed-effects panel regression with Driscoll-Kraay standard errors (Driscoll & Kraay 1998). Estimated coefficients $\hat{\beta}_{\mathrm{M}}^{(h)}$, $M \in \{\mathrm{TEXT}, \mathrm{VOICE}, \mathrm{FACE}\}$, provide the responses of the left-hand variable to the emotion index M at the h-minute horizon. The null hypothesis for this regression is that variations in Fed Chair's emotions captured by three indices are not influencing financial markets, i.e., $\beta_{\mathrm{M}}^{(h)} = 0$.

Figure 3 provides the estimated responses of S&P500 and VIX to one-standard-deviation increases in text-, voice-, and face-emotion indices during the Chair's prepared remarks. Positive changes in all three indices lead to positive S&P500 responses within minutes, and for text and face the responses are significant at 1 bp and 5 bps, respectively. The increases in the Chair's text sentiment and voice pitch lower VIX by 3 and 12 bps, respectively. It is worth noting that the responses to changes in the degree of negative facial expressions are similar to those uncovered in Curti & Kazinnik (2021), who focused on the impact of facial expressions on markets following FOMC press conference communications. However, our responses to a typical face-emotion shock appear somewhat larger even though we include the text-emotion and voice-emotion variables in our regressions. Moreover, at the aggregate level, the results suggest that increased pitch during the remarks section of the testimony is interpreted as a positive emotional cue resulting in an increase in the S&P500 and a corresponding decrease in the VIX.

The response of VIX to the Chair's face emotion is counterintuitive, suggesting decreased market uncertainty when the Chair's face emotion is negative. It is important to note that the measurement of the Chair's face emotions during the remark period is potentially confounded by the fact that for more than half of the time during the remarks the Chair is moving his or her head down to read the remarks. When the Chair's head is tilted down, it is more difficult for both the software and individuals watching the testimony to accurately identify all facial movements and related emotions. An examination of the responses by the Chair suggests that the counterfactual response for the VIX is only seen for Bernanke. Moreover, as indicated by the results from our alternate identification strategy discussed next, the counterintuitive result for the VIX's response to facial emotions during remarks is not robust.

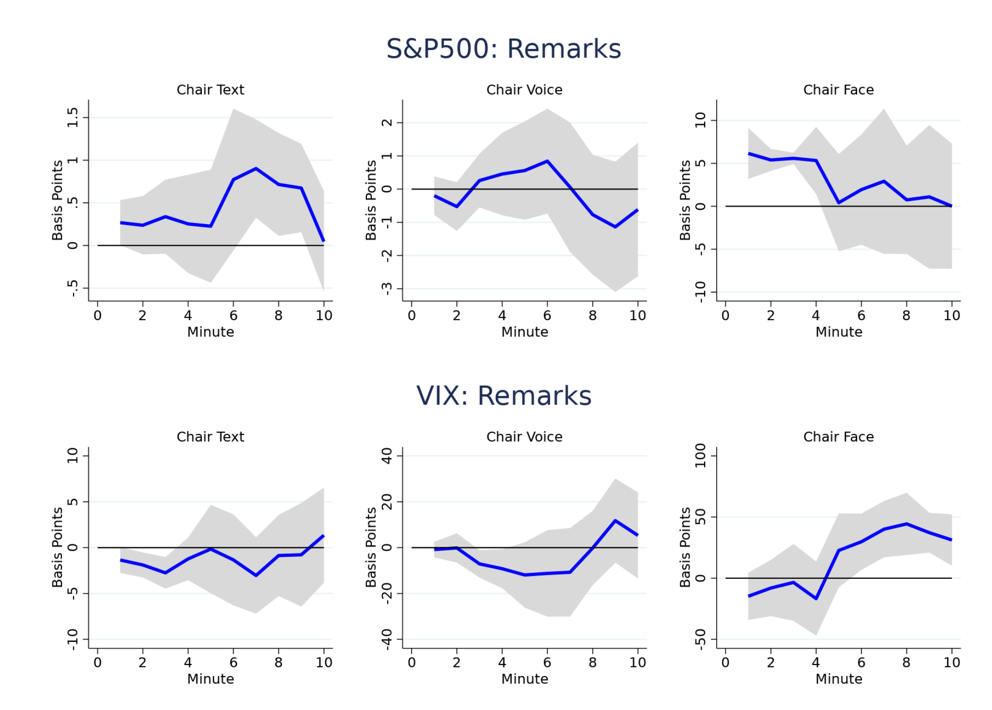


Figure 3. Responses during the remarks.

Notes: The figure provides responses of the change in log S&P500 (top) and the change of log VIX (bottom) to a one-standard-deviation positive impulse in the Fed Chair's text-emotion index (left), voice-emotion index (middle), and face-emotion index (right) during remarks. Responses are estimated using specification (1). The shaded areas represent the 90 percent confidence interval based on Driscoll-Kraay standard errors.

3.2 Day 1 vs. Day 2: Fed Chair's identical remarks

A unique feature of the semi-annual testimonies is that they take place on two separate days (to the House and to the Senate), and on both days the Chair delivers virtually identical remarks.[10] This implies that the text-emotion index is identical between two days, and markets should respond to the Chair's voice and face emotions only insofar as they differ from his or her voice and face emotions during Day 1's remarks. We estimate the effect of these voice- and face-index differentials between Day 2 and Day 1 remarks using the following

[10] There are two exceptions. Bernanke delivered very different remarks on March 2, 2011 (Day 2) than on March 1 (Day 1). Yellen delivered remarks on July 12, 2017 (Day 1) while she delivered no remarks on July 13 (Day 2). We exclude these observations from this analysis. Among the remaining testimonies, three pairs of testimonies (February 2010, February 2013, and February 2014) contained minor differences in several sentences of the remarks. They do not influence the results.

specification:

$$Outcome_{\tau,b+h} - Outcome_{\tau,b} = \beta^{(h)}_{\text{VOICE}} \triangle\text{VOICE}^{\text{Chair}}_{\tau,b} + \beta^{(h)}_{\text{FACE}} \triangle\text{FACE}^{\text{Chair}}_{\tau,b} + \text{controls} + \text{constant} + \varepsilon^{(h)}_{\tau,b}, \quad (2)$$

where the dependent variable is the h-minute change in the outcome variable for the remarks on Day 2, and $\triangle\text{VOICE}^{\text{Chair}}_{\tau,b}$ and $\triangle\text{FACE}^{\text{Chair}}_{\tau,b}$ are the differences in voice- and face-emotion indices for block b of the remarks on Day 2 and same block b on Day 1.

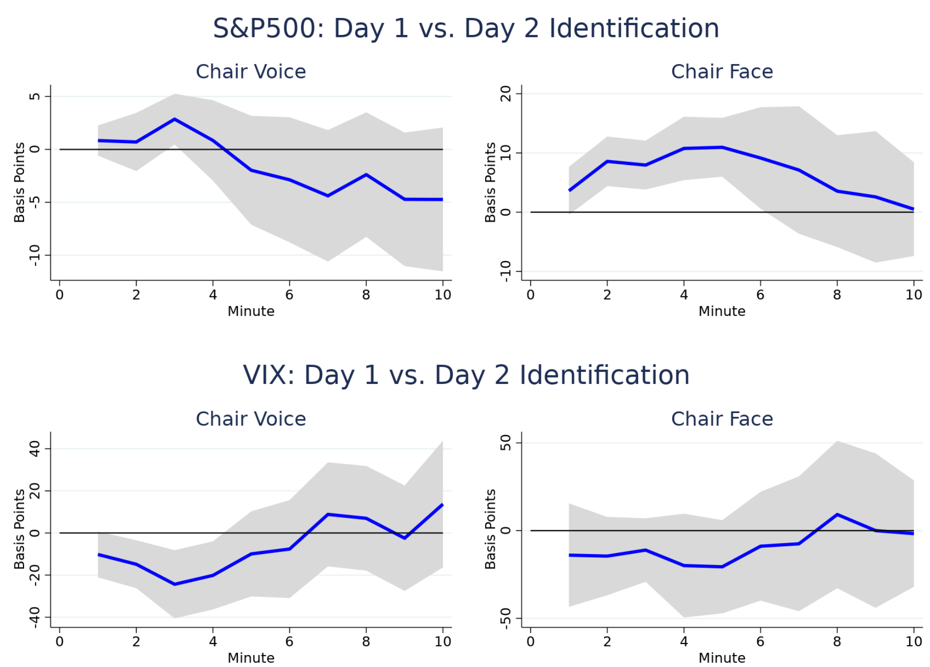


Figure 4. Responses: Day 1 vs. Day 2 testimonies.

Notes: The figure provides the responses of the changes in log S&P500 (top) and changes in log VIX (bottom) to a one-standard-deviation variation in the voice and face emotions of the Fed Chair between the Day 1 and Day 2 testimonies. Responses are estimated using specification (2). The shaded areas represent the 90 percent confidence interval based on Driscoll-Kraay standard errors.

The results are shown in Figure 4. S&P500 voice and face responses are in the same direction as in the first exercise, and they are quantitatively twice as large (and larger than Curti & Kazinnik (2021) find for FOMC press conferences). Both voice and face VIX responses are now negative, although VIX responses to the Chair's face emotion are not statistically significant. However, as we show in Section 5, they become larger and significant in days after the testimony. All in all, this alternative identification of the effects of the Chair's vocal and facial emotions during the remarks corroborates positive responses of stock prices and negative responses of stock volatility obtained using specification (1).

3.3 Financial market responses: Q&A

To set up the analysis of Q&A data and focus on the effects of the Chairs' responses to questions, we discard testimony blocks shorter than 10 sentences, blocks where the Chair speaks less than 20% of sentences, and blocks where the speaker's face is recognized for less than 15% of the frames.[11] Out of total 741 Q&A rounds, we end up with 661 complete Q&A rounds between the Chair and the member, where the Chair has the opportunity to provide full answers.

We estimate the responses to variations in the Fed Chair's emotion indices during Q&A using the baseline specification (1). In addition to controls used for the remarks, we include the following controls: three variables measuring the portion of each block containing the Chair's speech, voice, or face, and three emotion indices and the stance index for congress members.

The estimated financial market's responses to exchanges during the Q&A (Figure 5) are similar to responses we documented for the Chair's prepared remarks. The S&P500 index increases after the Chair's positive text- and voice-emotion indices change, and the responses are statistically significant at the 10% level. The VIX decreases in response to positive changes in all three indices, although not significantly for face emotions.

Quantitatively, stock market returns and stock volatility during Q&A are somewhat less sensitive than during the remarks, especially for face emotions. This is not very surprising. The Chair's remarks are prepared, and the flow of speech—and associated emotions—is uninterrupted and one-sided, from the Chair to the audience. In contrast, during the Q&A section of the testimony, the Chair is responding to questions on a variety of topics, and his or her answers are mostly unscripted and frequently interrupted by a congress member. Therefore, what and how the Chair says during Q&A varies from round to round, which may make it harder for the public and markets to distill. In the next section, we show that when

[11] This could be due to a wide angle of the camera, the speaker's head tilting down, or the camera being set from the side of the speaker so only the left or right part of the face is captured.

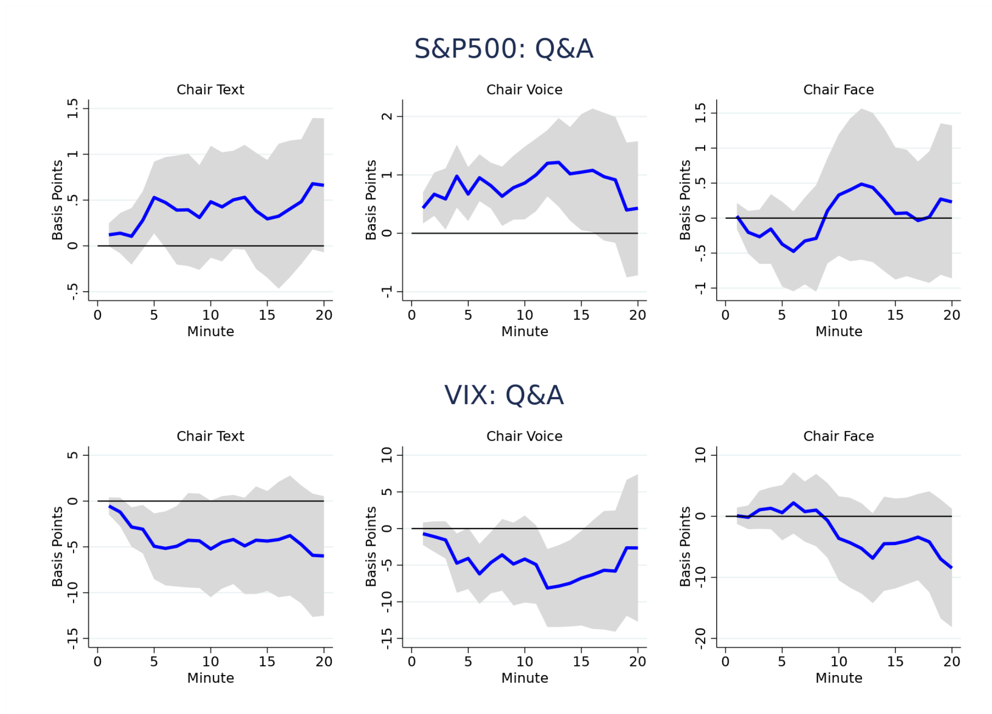


Figure 5. Responses during Q&A.

Notes: The figure provides responses of the changes in log S&P500 (top) and the changes in log VIX (bottom) to a one-standard-deviation positive impulse in the Fed Chair's text-emotion index (left), voice-emotion index (middle), and face-emotion index (right) during Q&A. Responses are estimated using specification (1). The shaded areas represent the 90 percent confidence interval based on Driscoll-Kraay standard errors.

the Chair discusses topics more relevant for financial markets, the responses to the Chair's emotions are as large as the responses during the remarks.

The estimated responses of interest rate expectations, measured by five-quarter-ahead Eurodollar futures, are not as stark as stock market responses (Figure 6). All three of the Chair's emotions have positive effects, albeit magnitudes are economically small. In the Appendix, we show that the responses tend to be positive for the remarks and also for longer yields (e.g., 10-year Treasury Note futures). Such responses suggest that markets associate the Chair's positive emotions with a more hawkish monetary policy stance in the future.[12]

In all, market responses to the Chair's emotions during the delivery of the congressional

[12]The responses are indicative of the information channel of monetary policy, whereby interest rate surprises are interpreted as the Fed's countercyclical responses to changes in economic outlook (Nakamura & Steinsson 2018, Cieslak & Schrimpf 2019, Jarociński & Karadi 2020)

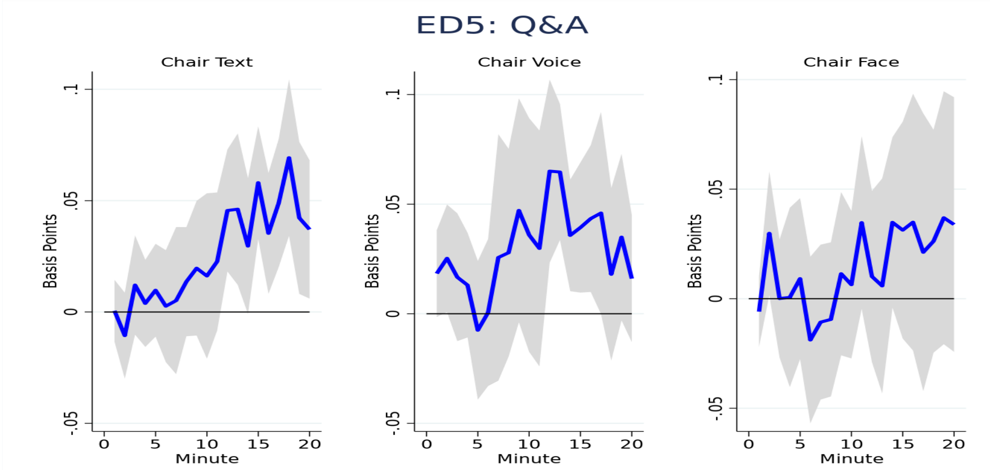


Figure 6. Interest rate expectation responses during Q&A.

Notes: The figure provides responses of the change in ED5 yields to a one-standard-deviation positive impulse in the Fed Chair's text-emotion index (left), voice-emotion index (middle), and face-emotion index (right) during Q&A. Responses are estimated using specification (1). The shaded areas represent the 90 percent confidence interval based on Driscoll-Kraay standard errors.

testimonies have similar effects to those during or shortly after the Chair's FOMC press conferences found elsewhere. Gorodnichenko et al. (2021) find that positive voice tone raises stock prices and lowers their volatility in the days following FOMC press conferences. Curti & Kazinnik (2021) estimate that the Chair's negative facial expressions during FOMC press conferences lead to lower S&P500 and higher VIX levels. We emphasize one difference in our results. We find that positive voice and facial expressions during congressional testimonies raise interest rate expectations. By contrast, Gorodnichenko et al. (2021) and Curti & Kazinnik (2021) do not find significant evidence that voice and face emotions move interest rate expectations during/after FOMC press conferences.

Like these studies, our evidence indicates that soft information expressed by the Fed Chair during a public event influences financial markets. Our paper goes a step further, showing that such soft information is expressed via a combination of text, voice, and face variations. We demonstrate that all of these means of communication tend to move stock returns in the same direction, although their relative impacts may depend on the circumstances of the communication event. In the next section, we conduct additional analyses of these circumstances to shed more light on the determinants of market reaction to the Fed Chair's emotions.

4 Determinants of financial market responses

Our estimates show that financial markets react to soft information contained in the Chair's discourse during the testimony, but do those responses depend on certain contexts or circumstances arising over the course of the testimony? Understanding these contexts or circumstances may help us discern some of the determinants of financial market reaction we document in the preceding section. In particular, we demonstrate that financial markets are somewhat differential to two key elements of the testimony—what was discussed and the Fed Chair person—while other elements seem less relevant.

4.1 Q&A topics

In the 32 testimonies of our study, there are a total of 741 Q&A rounds. Within each round, a congress member and the Fed Chair discuss several questions, six on average. In all testimonies, 4,323 questions and answers are covered. We use Grootendorst (2022)'s BERTopic algorithm to identify topics discussed in this set of question–answers. BERTopic leverages the word and sentence representations derived from the transformer model BERT as inputs, and creates dense clusters by using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm (Campello et al. 2013).

The algorithm identifies 11 fairly narrow topics for two-thirds of 4,323 question–answer pairs in these testimonies (see Appendix). The remaining one-third are general and not associated with a narrow topic. For our analysis, we drop question–answers related to general pleasantries and those lasting less than 15 seconds to eliminate cross-talk, interruptions, platitudes, and introductions. This leaves us with 2,323 question–answer pairs. To estimate the responses conditional on topics discussed, we run specification (1) on a panel of question–answers, where blocks b are now question–answer pairs for each topic instead of the entire Q&A rounds we used above. We recompute emotion indices at a question–answer level, but leave normalization intact (i.e., dividing by standard deviations at the Q&A round level) for ease of comparison.

We find that the Q&A results in Section 3.3 are driven by discussions of issues directly related to monetary policy—the central bank's reserves and balance sheet management, and the central bank's policy rate and inflation. This topic was discussed 7% of time. Figure 7 shows that stock returns respond positively and significantly to positive changes in all three

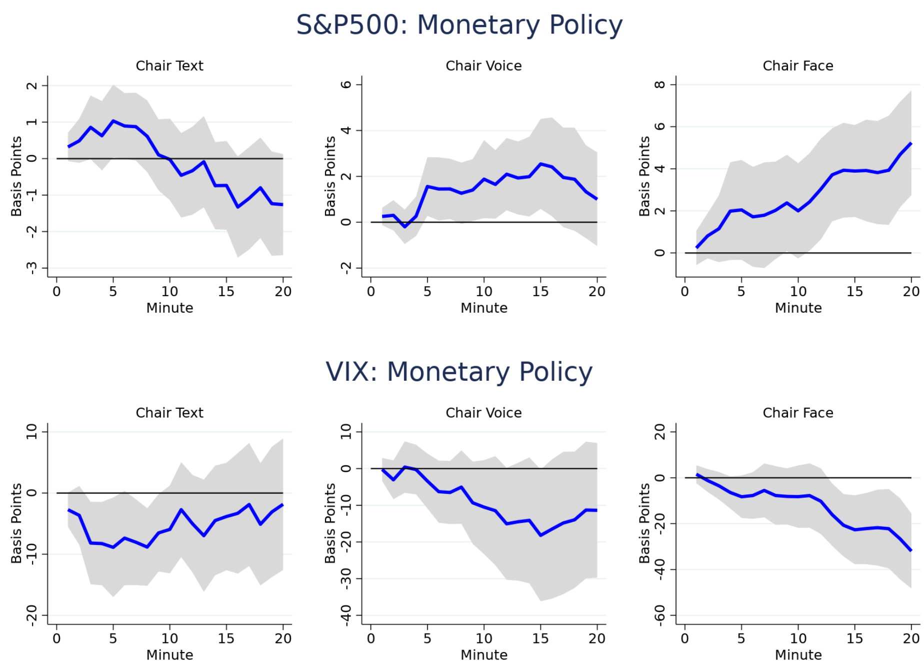


Figure 7. S&P500 and VIX responses to monetary policy topics during Q&A.

Notes: The figure provides the responses of the changes in log S&P500 (top) and log VIX (bottom) to a one-standard-deviation variation in the text-, voice-, and face-emotions of the Fed Chair conditional on discussing topics related to the Fed's monetary policy. Responses are estimated using specification (1) on a panel of testimony blocks, where blocks b are question–answers for the selected topic, instead of entire Q&A rounds. The shaded areas represent the 90 percent confidence interval based on Driscoll-Kraay standard errors.

indices of the Fed Chair's emotions during discussions of monetary policy, and VIX responses are negative and significant. Quantitatively, these responses have similar magnitudes to those we document for the remarks: by 1 bp (text), 2 bps (voice), and 5 bps (face) within 5 to 20 minutes.

These results are not surprising: markets are more likely to tune in to statements regarding the Fed's interest rate and balance sheet policies. By contrast, the responses are either less systematic or less sensitive to discussions of bank regulations related to the Fed's regulatory mandate (35%) or discussions of other economic topics (fiscal policy 8%, housing and mortgage markets 5%, job market and unemployment 7%, trade and China 1%, growth and productivity 0.7%, unidentified topics 33%).

4.2 Bernanke and Yellen

The emotions measured by our indices, of course, reflect many idiosyncrasies of the "messenger": cultural and educational background, previous work experience, demographic features such as age and gender, temperament, and mannerisms. We should not be surprised, therefore, if such differences between Fed Chairs translate into different market responses.

To this end, we repeat the estimations of remarks and Q&A responses separately for Bernanke and Yellen testimonies. In all, markets are more sensitive to Bernanke's emotions, with positive responses of stock prices and negative responses of volatility to his positive cues in most cases. By contrast, the responses to Yellen's emotions are less consistent and insignificant in many cases. For example, Figure 8 shows that the S&P500 response to Yellen's voice emotion is half the magnitude of the response to Bernanke's voice emotions, and the responses to her text sentiment and face emotion are insignificant or of the opposite sign.

These findings suggest that the market's reaction to Fed messages is tightly linked to the messenger. With data for only two of the Fed's Chairs, it is difficult to discern what it is about the messenger that markets are reacting to.[13] Furthermore, different responses could also be associated with different states of the economy during each Chair's respective tenures. Future work can draw firmer conclusions by adding testimony data for other Chairs and expanding the years covered by the analysis.

4.3 Other factors

Other dimensions of congressional testimonies appear less influential. In the Appendix we parse the responses by Day 1 versus Day 2 testimonies, the Senate versus the House testimonies, and the first versus the second halves of the Q&A of the same testimony. Although some of them show significant responses, there is no strong systematic link with the responses reported above. We find the responses to congressional members' emotions are qualitatively similar but quantitatively weaker than the responses to the Chair's emotions, suggesting that financial market activity during the testimony is associated mostly with information

[13]For example, individuals can interpret higher pitch in different ways. Sometimes high pitch is used to stress a point (like stressing a positive outcome in prepared remarks), in other instances high pitch may be interpreted as increased stress levels. Lausen & Schacht (2018) find that in some cases, when the speaker is a woman, men have a harder time accurately identifying emotions.

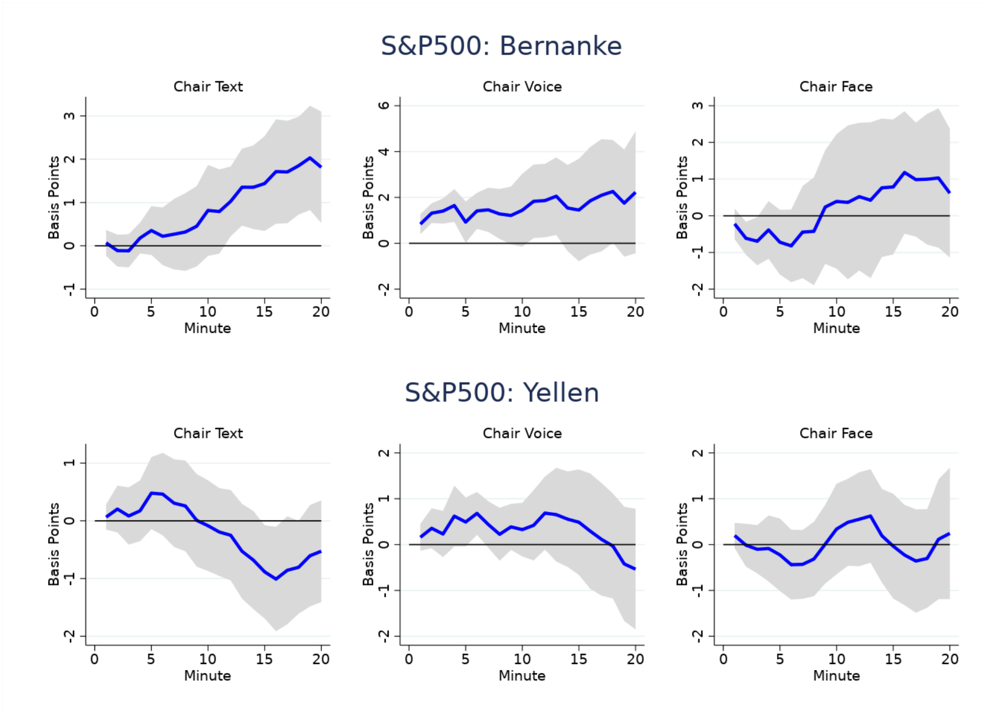


Figure 8. S&P500 responses during Q&A: Bernanke and Yellen.

Notes: The figure provides responses of changes in the log S&P500 to a one-standard-deviation positive impulse in the Fed Chair's text-emotion index (left), voice-emotion index (middle), and face-emotion index (right) for Q&A during Bernanke testimonies (top) and Yellen testimonies (bottom). Responses are estimated using specification (1) and Q&A controls. The shaded areas represent the 90 percent confidence interval based on Driscoll-Kraay standard errors.

reflected in the Fed Chair's emotions.

Furthermore, we also would like to rule out market-wide events that occur during the testimony and may influence the results. To identify such events, we mine information out of contemporaneous business news coverage from CNBC broadcasts, which are archived in the Internet Archives' TV News collection (see Appendix). We assemble the breaking news segments that include macro news releases, political announcements, energy data releases, and other nation-wide news (e.g., hurricanes, terrorist attacks, federal government and regulatory announcements, worldwide events). Out of 661 blocks of testimony data, we identify 86 blocks that overlap with market-wide breaking news not related to the testimony. When we drop these blocks from the regression analysis, the estimated responses become somewhat stronger. Hence, market-wide events do not influence our results.

5 Discussion of the economic significance of estimated responses

In this section we argue that market responses during the testimony are economically meaningful. First, we provide evidence that the effects during the testimony add up and propagate in the days after the testimony, reaching magnitudes comparable to those after a policy rate cut. Second, during the testimony, we find evidence that market activity is elevated: asset prices are more volatile and trading volumes are higher. Finally, we show that semi-annual testimonies attract public attention, reflected in heightened viewership of live broadcasts and increased media coverage, similar to those around the FOMC press conferences. Below we elaborate on each of these points.

Financial market responses to the Chair's three emotional cues are not only statistically significant, as we show above, but also economically significant. A one-standard-deviation change in the text-, voice-, or face-emotion indices during the remarks or relevant parts of the Q&A raises the S&P500 by 1 bp, 2 bps, and roughly 5 bps, respectively. If accumulated over the entire testimony, the effects of soft information from the Fed Chair may reach magnitudes comparable to those after an interest rate cut. For example, an unanticipated 25 bps cut in the Fed funds rate is associated with a roughly 100 bps increase in stock prices (Bernanke & Kuttner 2005).

Indeed, the effects that we document during the testimony appear to add up and persist in the days after the testimony. To determine the magnitudes, we estimate local projections at daily frequency:

$$\begin{aligned} Outcome_{\tau+h} - Outcome_{\tau-1} = & \beta^{(h)}_{\text{TEXT}} \text{TEXT}^{\text{CHAIR}}_{\tau} + \beta^{(h)}_{\text{VOICE}} \text{VOICE}^{\text{CHAIR}}_{\tau} + \beta^{(h)}_{\text{FACE}} \text{FACE}^{\text{CHAIR}}_{\tau} \\ & + \text{controls} + \text{constant} + \varepsilon^{(h)}_{\tau}. \end{aligned} \tag{3}$$

where the dependent variable $Outcome_{\tau+h} - Outcome_{\tau-1}$ is the change in log close price between day $\tau - 1$ and day $\tau + h$, and index values are now the mean of corresponding the Chair's emotion indices over the remarks and Q&A sessions of testimony τ, normalized by its own standard deviation across the 32 days (the data are provided in the Appendix). We include the three emotion indices of the members and the share of each block associated with the Chair's speech as controls.

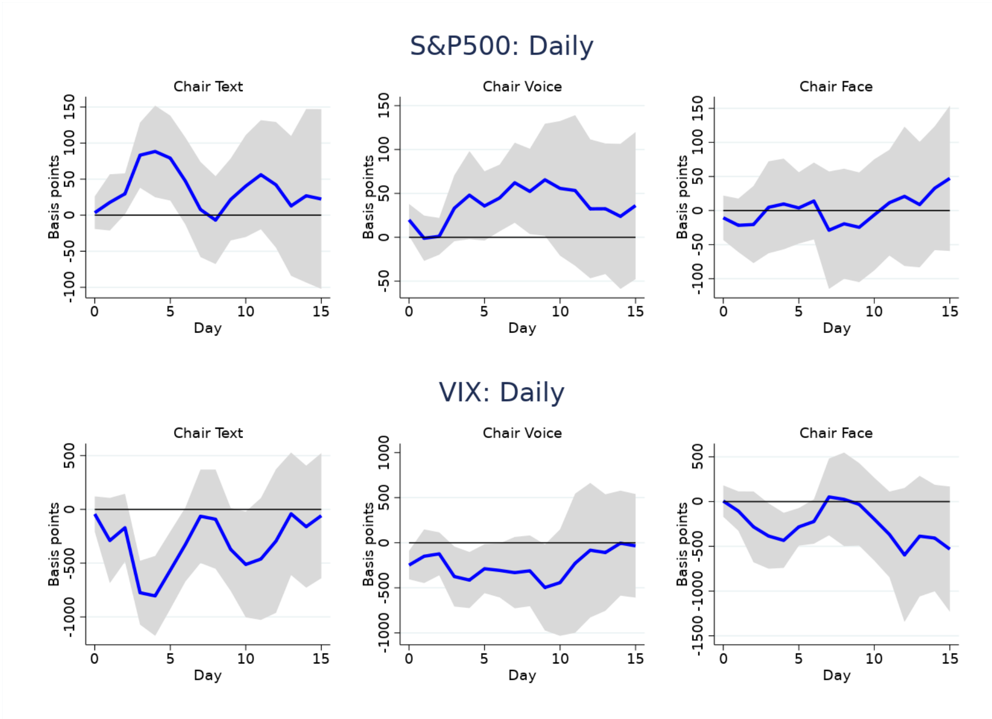


Figure 9. Responses at daily frequency.

Notes: The figure provides responses of the daily change in log S&P500 (top) and the daily change in log VIX (bottom) to a one-standard-deviation positive impulse in the Fed Chair's average daily text-emotion index (left), voice-emotion index (middle), and face-emotion index (right). Responses are estimated using specification (3). The shaded areas represent the 90 percent confidence interval based on Newey-West standard errors.

Figure 9 shows that, for the most part, the S&P500 and VIX responses in days after the testimony have the same direction as responses during the testimony, although due to our small sample of 32 testimonies the responses are not always statistically significant. Note that response magnitudes are of the same order as those after a rate cut. The S&P500 responses to a testimony's cumulative text and voice emotions reach 90 and 60 bps within one or two weeks after the testimony. In particular, responses to vocal cues are similar in magnitude to those reported by Gorodnichenko et al. (2021) for the days after FOMC press conference. The responses to face emotion are around zero, suggesting that the market's interpretation of the Chair's facial emotions for stock prices are short-lived. The VIX responses are all negative and significant.

Second, congressional testimonies attract attention by the financial markets. We observe that, on average, 10,187 SPY trades (an ETF tracking the S&P500) are executed during one block in the Q&A, and 307,159 SPY trades are executed over the course of the whole testimony. Rosa (2018) finds that the Fed Chair's FOMC press conferences and semi-annual testimonies between 2001 and 2012 significantly increase volatility of U.S. asset returns and trading volumes. We conduct our own related exercise and compute standard deviation of log changes and their trading volumes for SPY over 5-minute windows during each of 32 testimonies in our data. We compare these statistics with the day one week prior and one week after the day of each testimony. Even for such a small sample, we find that both price volatility and trade volumes are significantly higher during the testimony than seven days after, and they are also higher than seven days before the testimony, although for price volatility the difference is not statistically significant.

In addition to trading activity, the viewership of the televised or streamed testimonies and their coverage in the print and social media are also elevated. The testimonies are generally live-streamed by C-SPAN and on the Senate and House committees' websites, as well as on the major business news networks, such as CNBC and Bloomberg. Hundreds of thousands of households, investors, and businesses are exposed to these broadcasts contemporaneously on cable TV, through Bloomberg's terminals and TD Ameritrade, and on screens on the floor of the New York Stock Exchange.[14] There is also significant information shared about the testimonies' content in print and social media. We measure media coverage of the testimony by the daily number of related articles in the Dow Jones Factiva database (as a fraction of total daily articles) and the daily number of related Twitter posts. The interest in a testimony builds over the days leading up to it and falls in the days following it, following a fairly standard news cycle pattern. Moreover, on the peak day, which usually corresponds to the first day of testimony, approximately 0.24%–0.75% of news articles and 0.00089%–0.00682% of Twitter posts cover the testimony, which is comparable to coverage of FOMC press conferences.

[14] See, e.g., Comcast (2011-2017) for Nielsen's estimates of CNBC household penetration, Stark (1999) for a discussion of CNBC's large daily audience outside of the home, and https://ctv.kwayisi.org/networks/ for statistics on the viewership of CNBC's programs typically airing the testimonies—Squawk on the Street, Power Lunch and Fast Money Halftime Report. See https://www.bloomberg.com/professional/solution/bloomberg-terminal/ for evidence that there are over 325,000 terminals in use, and Bloomberg Business Wire (2010) and Killam-Williams (2005) for evidence on Bloomberg TV's historical viewers in the United States and Europe based on reported data from Nielsen and the 2010 European Media and Marketing survey.

6 Conclusions

Central bankers are understandably restrained in what and how much they can say about monetary policy. Communications of monetary policy to the public need to be made in non-technical and relatable language (Bholat et al. 2019, Kryvtsov & Petersen 2021), but even simplified communications may not always get through to the audience (Coibion et al. 2020). Furthermore, it is not always desirable to disclose internal information, such as details of internal policy deliberations or staff views on the likely path of future interest rates (Natvik et al. 2020). When words are limited, how can central bank leaders present their institution's policy as credible and be trusted to promote social welfare? Our evidence suggests central bankers do that with more than words.

Even if the sentiment is incorporated in the central bank's written or verbal message, variations in voice pitch and facial expressions of the person delivering the message influence financial markets many times over. Positive emotional cues from the leader tend to be interpreted positively by financial markets. These effects do not disappear when the event is over, but rather they add up and propagate in the days after the event. Markets are more attentive when the central bank leader is speaking and when he or she is discussing monetary policy. These findings suggest that the delivery of central bank communications is potentially as influential for markets and the general public as is the content of these communications.

While the results demonstrate that the impacts of communications are linked to more than words, future research will help study the mechanisms of these effects and further clarify the most important channels. The impacts of soft forms of communication, for example, depend on both the messenger's facial and vocal emotional expressions and the audience's interpretation of these expressions. The messenger may choose to use the expressions in an intentional way, such as to emphasize an important point, or the expressions may unintentionally reveal an emotional state, such as stress, through a wavering voice, nervous gestures, or momentary expressions of shock. The impacts on the audience may depend on the demographic makeup of the messengers (i.e., the Fed Chairs, senators, and the congressional representatives) and the attention levels and characteristics of the audience.

Therefore, future work will focus on three main sets of questions. First, how is the soft information obtained, interpreted, and used by different types of traders (i.e., high-frequency traders vs. others), and which groups are most affected by the emotional signals? Second, what is the role of conventional and social media coverage for disseminating soft information, and how do the emotional signals affect the topics discussed in the news? Finally, are there systematic differences in the interpretation of, and responses to, the communications by messengers that differ by demographic characteristics (including gender, age, cultural background) or political affiliation?

References

Algaba, A., Ardia, D., Bluteau, K., Borms, S. & Boudt, K. (2020), 'Econometrics meets sentiment: An overview of methodology and applications', *Journal of Economic Surveys* **34**(3), 512–547.

Altavilla, C., Brugnolini, L., Gürkaynak, R. S., Motto, R. & Ragusa, G. (2019), 'Measuring euro area monetary policy', *Journal of Monetary Economics* **108**, 162–179.

Andrade, P. & Ferroni, F. (2021), 'Delphic and Odyssean monetary policy shocks: Evidence from the euro area', *Journal of Monetary Economics* **117**, 816–832.

Barber, B. M. & Odean, T. (2007), 'All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors', *The Review of Financial Studies* **21**(2), 785–818.

Ben-Rephael, A., Da, Z. & Israelsen, R. D. (2017), 'It depends on where you search: Institutional investor attention and underreaction to news', *The Review of Financial Studies* **30**(9), 3009–3047.

Benamar, H., Foucault, T. & Vega, C. (2020), 'Demand for information, uncertainty, and the response of U.S. Treasury securities to news', *The Review of Financial Studies* **34**(7), 3403–3455.

Bernanke, B. S. & Kuttner, K. N. (2005), 'What explains the stock market's reaction to Federal Reserve policy?', *The Journal of Finance* **60**(3), 1221–1257.

Bholat, D., Broughton, N., Ter Meer, J. & Walczak, E. (2019), 'Enhancing central bank communications using simple and relatable information', *Journal of Monetary Economics* **108**, 1–15.

Bloomberg Business Wire (2010), 'Bloomberg television delivers largest business day audience of international news channels in Europe'.
URL: *https://www.bloomberg.com/company/press/bloomberg-television-delivers-largest-business-day-audience-of-2/*

Boersma, P. & Weenink, D. (2001), 'Praat, a system for doing phonetics by computer', *Glot International* **5**, 341–345.

Boguth, O., Grégoire, V. & Martineau, C. (2019), 'Shaping expectations and coordinating attention: The unintended consequences of FOMC press conferences', *Journal of Financial and Quantitative Analysis* **54**(6), 2327–2353.

Campello, R. J. G. B., Moulavi, D. & Sander, J. (2013), Density-based clustering based on hierarchical density estimates, *in* J. Pei, V. S. Tseng, L. Cao, H. Motoda & G. Xu, eds, 'Advances in Knowledge Discovery and Data Mining', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 160–172.

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

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

Coibion, O., Gorodnichenko, Y., Knotek, E. S. & Schoenle, R. (2020), Average inflation targeting and household expectations, Working Paper 27836, National Bureau of Economic Research.

Comcast (2011-2017), 'Comcast annual reports on form 10-K'.

Congressional Research Service (2010), Hearings in the U.S. Senate: A guide for preparation and procedure, Technical report.

Curti, F. & Kazinnik, S. (2021), Let's face it: Quantifying the impact of nonverbal communication in FOMC press conferences, SSRN Scholarly Paper ID 3782239, Social Science Research Network.

Dellavigna, S. & Pollet, J. M. (2009), 'Investor inattention and Friday earnings announcements', *The Journal of Finance* **64**(2), 709–749.

Devlin, J., Chang, M., Lee, K. & Toutanova, K. (2018), 'BERT: pre-training of deep bidirectional transformers for language understanding', *CoRR* **abs/1810.04805**.

Dietrich, B. J., Hayes, M. & O'Brien, D. Z. (2019), 'Pitch perfect: Vocal pitch and the emotional intensity of congressional speech', *American Political Science Review* **113**(4), 941–962.

Driscoll, J. C. & Kraay, A. C. (1998), 'Consistent covariance matrix estimation with spatially dependent panel data', *Review of Economics and Statistics* **80**(4), 549–560.

Eckman, S. J. (2017), 'Video broadcasting of congressional proceedings', *CRS reports (Library of Congress. Congressional Research Service)* .

Edmans, A., Fernandez-Perez, A., Garel, A. & Indriawan, I. (2022), 'Music sentiment and stock returns around the world', *forthcoming in Journal of Financial Economics* .

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

Ekman, P. & Friesen, W. V. (1969), 'The repertoire of nonverbal behavior: Categories, origins, usage, and coding', *Semiotica* **1**(1), 49–98.

Fraccaroli, N., Giovannini, A. & Jamet, J.-F. (2020), Central banks in parliaments: A text analysis of the parliamentary hearings of the Bank of England, the European Central Bank and the Federal Reserve, Working Paper 2442, European Central Bank.

Gorodnichenko, Y., Pham, T. & Talavera, O. (2021), The Voice of Monetary Policy, Working Paper 28592, National Bureau of Economic Research.

Grootendorst, M. (2022), 'BERTopic: Neural topic modeling with a class-based TF-IDF procedure', *arXiv preprint arXiv:2203.05794* .

Gürkaynak, R. S., Kara, A. H., Kısacıkoğlu, B. & Lee, S. S. (2021), 'Monetary policy surprises and exchange rate behavior', *Journal of International Economics* **130**, 103443.

Gürkaynak, R. S., Sack, B. & Swanson, E. (2005), 'Do actions speak louder than words? The response of asset prices to monetary policy actions and statements', *International Journal of Central Banking* **1**(1), 55–93.

Gómez-Cram, R. & Grotteria, M. (2022), 'Real-time price discovery via verbal communication: Method and application to Fedspeak', *Journal of Financial Economics* **143**(3), 993–1025.

Hansen, S. & McMahon, M. (2016), 'Shocking language: Understanding the macroeconomic effects of central bank communication', *Journal of International Economics* **99**, S114–S133.

Hobson, J. L., Mayew, W. J. & Venkatachalam, M. (2012), 'Analyzing speech to detect financial misreporting', *Journal of Accounting Research* **50**(2), 349–392.

Hu, A. & Ma, S. (2021), Persuading investors: A video-based study, NBER Working Papers 29048, National Bureau of Economic Research, Inc.

Jarociński, M. & Karadi, P. (2020), 'Deconstructing monetary policy surprises—the role of information shocks', *American Economic Journal: Macroeconomics* **12**(2), 1–43.

Jordà, (2005), 'Estimation and Inference of Impulse Responses by Local Projections', *American Economic Review* **95**(1), 161–182.

Kahneman, D. (2013), *Thinking, Fast and Slow*, 1st edn, Farrar, Straus and Giroux, New York.

Kamiloglu, R., Fischer, A. & Sauter, D. (2020), 'Good vibrations: A review of vocal expressions of positive emotions', *Psychonomic Bulletin & Review* **27**, 237–265.

Killam-Williams, S. (2005), 'Bloomberg TV: Capitalizing on wealthy viewers', *TelevisionWeek* .
URL: *https://www.tvweek.com/in-depth/2005/11/bloomberg-tv-capitalizing-on-w/*

Kryvtsov, O. & Petersen, L. (2021), 'Central bank communication that works: Lessons from lab experiments', *Journal of Monetary Economics* **117**, 760–780.

Kuttner, K. N. (2001), 'Monetary policy surprises and interest rates: Evidence from the Fed funds futures market', *Journal of Monetary Economics* **47**(3), 523–544.

Lausen, A. & Schacht, A. (2018), 'Gender Differences in the Recognition of Vocal Emotions', *Frontiers in Psychology* **9**, 882.

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

Mehrabian, A. (1972), *Nonverbal communication*, Aldine Transaction, New Brunswick, N.J. OCLC: 752953369.

Nakamura, E. & Steinsson, J. (2018), 'High-frequency identification of monetary non-neutrality: The information effect', *The Quarterly Journal of Economics* **133**(3), 1283–1330.

Natvik, G. J., Rime, D. & Syrstad, O. (2020), 'Does publication of interest rate paths provide guidance?', *Journal of International Money and Finance* **103**, 102123.

Obaid, K. & Pukthuanthong, K. (2022), 'A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news', *Journal of Financial Economics* **144**(1), 273–297.

Rosa, C. (2018), 'Fedspeak: Who moves U.S. asset prices?', *46th issue (December 2016) of the International Journal of Central Banking* .

Stark, S. (1999), 'If only Nielsen counted barroom viewers: CNBC execs fight to show advertisers audience bigger than home only numbers.', *InvestmentNews* **(October 18)**. **URL:** *https://www.investmentnews.com/if-only-nielsen-counted-barroom-viewerscnbc-execs-fight-to-show-advertisers-audience-bigger-than-home-only-numbers-2042*

Swanson, E. T. (2021), 'Measuring the effects of Federal Reserve forward guidance and asset purchases on financial markets', *Journal of Monetary Economics* **118**, 32–53.

Yuan, Y. (2015), 'Market-wide attention, trading, and stock returns', *Journal of Financial Economics* **116**(3), 548–564.

More Than Words: Fed Chairs' Communication During Congressional Testimonies*

— *For Online Publication* —

Michelle Alexopoulos   Xinfen Han   Oleksiy Kryvtsov   Xu Zhang†

May 5, 2022

*The views expressed herein are those of the authors and not necessarily those of the Bank of Canada.

†Alexopoulos (University of Toronto): m.alexopoulos@utoronto.ca; Han (Bank of Canada): xhan@bankofcanada.ca; Kryvtsov (Bank of Canada): okryvtsov@bankofcanada.ca; Zhang (Bank of Canada): xzhang@bankofcanada.ca.

Contents

A Construction of text, voice, and face indices

Figure A.1 depicts our data processing procedure to derive text, voice, and face emotion metrics.

A.1 Text-emotion index

Our text-sentiment classifier assigns to each sentence of the testimony an emotion score $T0$, taking values 1 (positive), -1 (negative), or 0 (neutral). The classifier is based on Bidirectional Encoder Representations from Transformers (BERT),[1] a natural language processing transformer model, implemented in the Hugging Face's repository (Wolf et al. 2019). We fine-tune the pre-trained BERT model with testimony sentences classification training data annotated by authors. Two authors annotated 2818 testimony sentences independently, classifying them into 3 groups by positive, negative, and neutral sentiment. The training data is constructed with 2474 sentences for which both authors' classifications are identical. The purpose of augmenting the pre-trained BERT model is to adjust our text classification to better reflect the context of the testimony. We provide examples in subsection A.4.

We mainly use the F1 score to measure our text sentiment classifier's performance (see Table A.1). The F1 score is an accuracy measure for a classification model, which is a useful metric for an imbalanced training data set (our case). F1 is defined as the harmonic mean of a model's precision and recall, where precision measures the share of positive from classifier predicted positive classes, while recall measures the share of positive out of true positive cases.

$$\text{F1(classX)} = 2 * \frac{precision(classX) * recall(classX)}{precision(classX) + recall(classX)},$$

$$precision : \frac{TP}{(TP + FP)},$$

$$recall : \frac{TP}{(TP + FN)},$$

where:

$$TP = TruePositive, FP = FalsePositive, FN = FalseNegative$$

[1] Bidirectional Encoder Representations from Transformers (BERT) is "designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context." (Devlin et al. 2018). The BERT model has been trained on English Wikipedia and BookCorpus (Zhu et al. 2015), and it has displayed state-of-the-art performance on a number of general natural language understanding tasks.

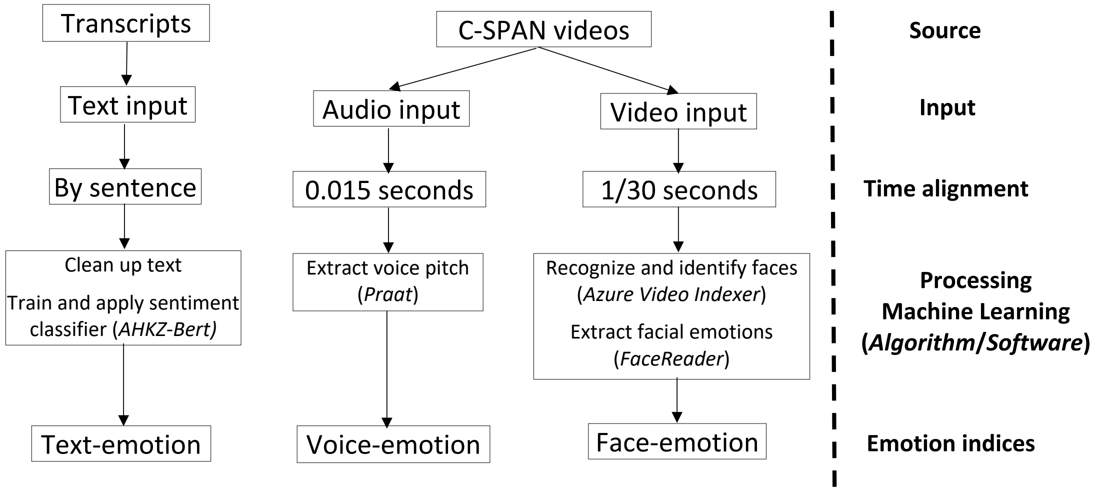


Figure A.1. Text, audio and video data processing procedures

	Precision	Recall	F1 score
Positive	0.85	0.8	0.82
Neutral	0.95	0.97	0.96
Negative	0.84	0.75	0.79

Table A.1. BERT-based fine-tuned sentiment classifier performance

The text-sentiment index for speaker i, block b, testimony τ is defined as the mean of sentence scores in this block by speaker i, normalized by its standard deviation over all blocks in the Q&A:

$$\text{TEXT}^{i}_{\tau,b} = mean(T0)/sd_{\text{TEXT}}.$$

Speaker superscript i denotes a chair or a congress member. We define a single index for congress members by pooling all Q&A blocks for different congress members.

The text-emotion index is different from the stance index used in the literature which measures the degree of hawkish or dovish sentiment conveyed in the central banks' communications (Ehrmann & Talmi 2020). Therefore, we also construct a stance index for each block of sentences using the dictionary in Gorodnichenko et al. (2021) (GPT hereafter). We then parse each testimony sentence using the observed punctuation, and search and count words associated with the GPT dictionary in each part of the sentence.[2] These counts are then aggregated over the entire block to form the stance index we use as a control variable in our empirical analysis.

Specifically, our stance index for testimony τ block b is defined as:

$$\text{STANCE}_{\tau b} = \frac{\#\ \textit{dovish sentences} - \#\ \textit{hawkish sentences}}{\#\ \textit{sentences}},$$

where # *dovish* (*hawkish*) *sentences* is the number of sentences with dovish (hawkish) meaning, and # *sentences* is the number of sentences in the block.

A.2 Voice-emotion index

To create our voice-emotion index, we extract testimony-related audio inputs directly from C-SPAN videos. We first convert the audio file to 48,000 Hz sample rate with mono channel in *wav* format, then we preprocess it to mark every section where voice activities are detected.[3] The output of the process is a list of time intervals in seconds. Using the output, in combination with manual verification, we identify the major pauses in the audio. These are normally major unintentional pauses, for example, when the microphone breaks during the testimony. We then split the transcript into text chunks by excluding the identified pause periods. Each chunk includes sentence-parsed transcript text. We then synchronize text chunks with the testimony audio by applying the forced alignment algorithm implemented using the *aeneas* Python library.[4] The forced alignment process determines the time interval in the audio file that contains the speech text fragment. After *aeneas* produces start and end timestamps for each sentence in a text chunk, we combine the output of all text chunks to produce sentence level transcript timestamps. Using these timestamps, we then split the testimony audio into individual sentence audio files. In order to correct inconsistencies

[2]To maximize the number of identified words, we search and count words from GPT dictionary in three formats: the original word used in the sentence; the stemmed word, which usually refers to a crude process that cuts off the end or the beginning of the word, e.g. *sudies* to *studi*; the lemmatized word, which takes into consideration the morphological analysis of the words, and only remove the inflectional endings to return the base of a word, e.g. *studies* to *study*. We then remove the duplicate findings between each set.

[3]We use python interface to WebRTC Voice Activity Detector (VAD) for this purpose. https://pypi.org/project/webrtcvad/.

[4]https://github.com/readbeyond/aeneas.

between the transcript and the audio speech,[5] we conduct iterative manual verification to ensure that the audio file splits as accurately as possible. These audio segments are the inputs into our main vocal signal processing tool, *Praat* (Boersma & Weenink 2001).

Following Dietrich, Hayes & O'Brien (2019), we utilize the changes in vocal pitch as our measure of vocal emotions. The vocal pitch is measured by calculating the mean fundamental frequency (F0) of each audio sentence. F0 corresponds to the rate of vocal fold vibrations: high pitch is associated with rapid vibrations and low pitch with slow vibrations.[6] Using *Praat*, we extract $F0$ values at 0.015 second intervals.

There are two strands of literature on the link between vocal expressions and the underlying emotions. One focuses on the discrete basic emotions, e.g., happiness, sadness, anger, fear (Laukka 2005, Gelder et al. 1997), while the other studies affective states that represent the broad dimensions of emotions, e.g., activation, valence, potency and emotion intensity (Cowie & Cornelius 2003, Laukka et al. 2005). Our approach differs from Gorodnichenko et al. (2021) who classify Fed Chairs vocal expressions along multiple discrete emotion dimensions. Specifically, our study concentrates on the broad dimensions of vocal emotions; namely, we associate F0 with emotion activation and intensity since high emotion activation and intensity is usually associated with high mean F0 (Laukka et al. 2005, Dietrich, Enos & Sen 2019, Dietrich, Hayes & O'Brien 2019).

Our voice-emotion index $\text{VOICE}_{\tau,b}$ for speaker i in block b testimony τ is defined as the mean of vocal pitch in this block by speaker i, de-meaned by speaker i and normalized by its standard deviation over all blocks:

$$\text{VOICE}^{i}_{\tau,b} = mean(F0 - \overline{F0}_i)/sd_{\text{VOICE}}.$$

A.3 Face-emotion index

To construct the testimony video inputs, we download, cut and merge C-SPAN broadcasting TV recordings that include Fed Chairs' testimonies. To assess facial expressions, we

[5]The transcript often excludes the conversations unrelated to the testimony. For example Bernanke's testimony session on July 17, 2013 experienced some audio problems, and the announcement during the problem periods is not captured in the transcript. The transcript also does not reflect the crosstalks during testimonies. These are potential sources that may cause inconsistencies between the transcript and audio speech.

[6]The range of vibrations is normally between 60 and 180 cycles per second (Hz) for men, and 160 to 300 Hz for women.

process each testimony video with FaceReader software.[7] FaceReader analyzes only one face in each frame. Therefore, to identify the person on screen we proceed in several steps. We first use Azure Video Indexer's functions (*Face detection* and *Celebrity identification*) to detect and identify all faces in each frame.[8] We then match the face locations (derived from FaceReader's facial landmarks) with the locations identified from face detection algorithms. Finally, we query the person's name from the identified-person database for the matched faces and manually verify if the match is correct.

Influential research in psychology, Ekman & Friesen (1969), argues that there exists universal facial emotions across countries and culture, and they can be identified by detecting facial muscles movement. Ekman & Friesen created Facial Action Coding System (FACS) to label different areas of facial muscles and to use as the standard rating scale to rate area muscle movements. These identified muscle areas are defined as action units, and combinations of them produce facial emotions. For example, "Disgust" is associated with action units 9 (Nose wrinkle), 15 (Lip corner depressor), and 16 (Lower lip depressor). We provide examples in appendix A.5 to illustrate how emotions are commonly constructed using these action units.

Based on a frame-by-frame analysis,[9] FaceReader captures not only action units expressed by the face, but also their intensity, which is expressed as a number between 0 (lowest intensity) and 1 (highest intensity). The emotion score for a basic emotion is the average intensity of its corresponding action units.

For each frame f, we compute a raw face-emotion score $FaceScore_f$ as the average of four basic negative emotions, Sad, $Angry$, $Fear$, and $Disgust$:

$$FaceScore_f = -(Sad_f + Fear_f + Anger_f + Disgust_f)/4.$$

This emotion score has values ranging between -1 (highest negative emotions) and 0 (no negative emotions). In constructing these scores, we exclude action units that are associated with speaking (e.g., lips, mouth, and cheeks) since Ekman et al. (2002) explain that these action units make it harder to distinguish emotions when the person of interest is speaking.

[7]FaceReader was originally developed by VicarVision, and currently distributed by Noldus, `https://www.noldus.com/facereader`. It uses Active Appearance Models (AAM) (Cootes & Taylor 1999) for face modelling and over 10,000 manually annotated image data set to train an artificial neural networks for facial emotion classification (Bishop 1995). It also uses a deep artificial neutral network to recognize facial patterns (Gudi 2015), which helps FaceReader to analyze partially hidden faces.

[8]Video Indexer "identifies over 1 million celebrities — like world leaders, actors, actresses, athletes, researchers, business, and tech leaders across the globe." `https://docs.microsoft.com/en-us/azure/azure-video-analyzer/video-analyzer-for-media-docs/`.

[9]Video inputs are collected for each frame, at 29.97 frames per second.

We then define the face-emotion index for person i's face in block b testimony τ as the mean of the face-emotion score for that block, normalized by its standard deviation over all blocks:

$$\mathrm{FACE}^{i}_{\tau,b} = mean(FaceScore)/sd_{\mathrm{FACE}}.$$

A.4 Examples of text sentiment

This Table provides examples of the raw text sentiment score for two testimonies: Bernanke's July 22, 2010, and Yellen's February 10, 2016.

Table A.2. Examples of text sentiment

Speaker	**Sentence**	**score**
	July 22, 2010 Testimony	
Castle	With respect to the Stimulus Act, the recovery bill, whatever one wishes to call it, you know, obviously jobs were saved and jobs were – were created by that to some degree.	1
Castle	The jobs saved are primarily, in my judgment, a lot of the governmental jobs in which state and local governments received funding and saved teachers or whatever it may be.	1
Castle	The jobs created were in many instances patchwork-type things like fixing up highways or whatever it may be.	1
Castle	Have you or has anybody that you know of studied the – the bottom line aspect of those jobs today?	0
Castle	I mean, all that – most of that happened last year at some point or another.	0
Bernanke	Well, as you know, it's intrinsically very difficult to get an exact count.	0
Castle	I know that.	0
Bernanke	Because we don't know what would have happened in the absence of the program.	0
Bernanke	And so, economists use models and other ways of trying to estimate what the effect has been.	0
Bernanke	The CBO gave a very broad range of estimates, between 1 million and 3.5 million jobs, which is a very wide range, you can see.	0
Bernanke	But it encompasses what most private sector economists have estimated.	0
Bernanke	And it encompasses what the Federal Reserve has estimated, which is somewhere in the middle of that – of that range.	0
Bernanke	So there has – there has been some job creation.	1
Frank	In the monetary report, I cited three passages where you cite the events in Europe that began with the Greek debt crisis.	0

Continued on next page

Table A.2 – *Continued from previous page*

Speaker	Sentence	score
Frank	But do you agree, or let me just ask you, what role did the crisis that began with the Greek debt crisis and roiled much of Europe and the euro zone, what effect did it have on what's going on in the economy here and your estimates of that?	0
Bernanke	It certainly did have some negative effects.	-1
Bernanke	The increased financial concerns led to declines in the stock market, increased credit spreads, and was one of the reasons why we marked down our outlook for the U.S. economy.	-1
Bernanke	That's absolutely right.	0
Bernanke	I think that, first, I think that situation is improving.	1
Bernanke	Confidence has been coming back in part because of the Federal Reserve support for the dollar funding markets.	1
Bernanke	There have been a few other things we've seen in the data such as the weakness in the housing market after the end of the tax credit, for example.	-1
Bernanke	And of course the labor market has been disappointing in the last couple of – last couple of months.	-1
Bernanke	But again, our baseline scenario is that as the effects of the European financial crisis pass, that we will continue to see moderate growth in the economy.	1
	February 10, 2016 Testimony	
Luetkemeyer	You know, let's start off first with what happens if we have a downturn and you've already got $4 trillion on your balance sheet.	0
Luetkemeyer	What levers are still allowed or are available to you to do something?	0
Yellen	Well, the Fed has an array of tools.	0
Luetkemeyer	Which are?	0
Yellen	Well, most importantly, the path of the short-term interest rates.	0
Luetkemeyer	I mean, how is lowering the rates going to help when they're almost nothing right now?	0
Yellen	Well, one of the ways in which markets works is that they form expectations about what the likely path of the Fed Funds Rate will be over time.	0
Yellen	Those expectations influence longer-term rates in the market.	0
Yellen	And that shift in expectations moves longer- term rates.	0

Continued on next page

Table A.2 – *Continued from previous page*

Speaker	**Sentence**	**score**
Yellen	I think you can see that just over the last several weeks, as I mentioned longer-term Treasury yields have come down, as market participants have become more fearful about a recession.	-1
Mulvaney	You – by your own testimony, are using traditional tools of monetary policy.	0
Mulvaney	Your written testimony begins by saying that the economy has made further progress towards the Federal Reserve's objective of maximum employment.	1
Mulvaney	You go on to say that inflation is low in the near-term but it will rise to its two percent objective over the median term.	1
Mulvaney	Are we in normal times?	0
Yellen	The economy is in many ways close to normal in the sense that the unemployment rate is declined to levels that most of my colleagues believe are consistent with full employment in the longer run.	1
Yellen	In other words, we have needed for seven years to pull the Federal Funds Rate and – both in nominal and inflation in real terms – inflation adjusted or real terms at exceptionally low levels to achieve growth averaging 2 percent or a little bit above.	0

A.5 Examples of action units and facial emotions

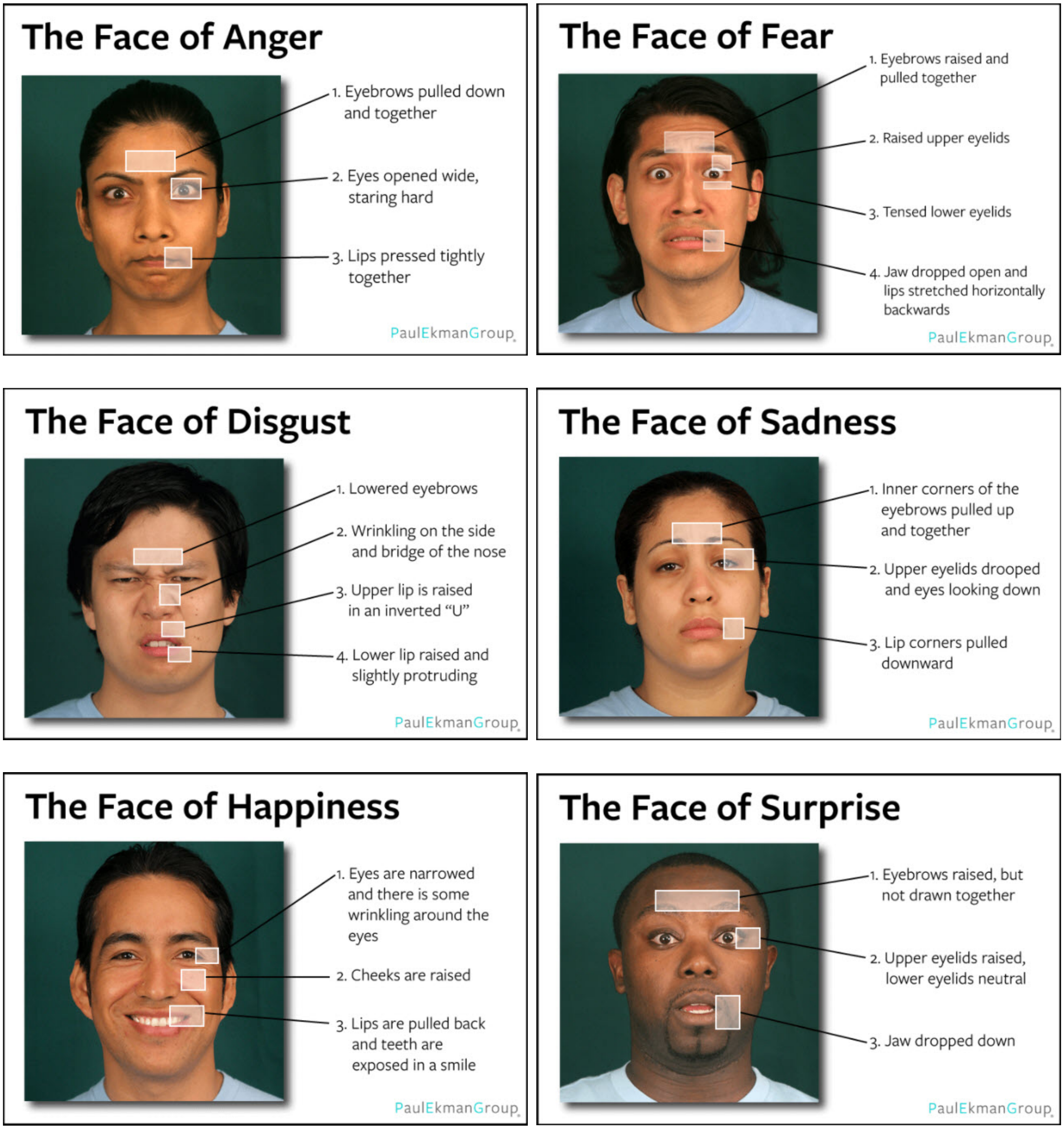


Table A.3. All graphs are retrieved from PaulEkmanGroup website in Feb 2022.

A.6 Snapshots of the Fed Chair and Congress members' face-emotions

	Face: Ben Bernanke Face emotion score: -0.222
	Face: Michael Castle Face emotion score: -0.293

Table A.4. Facial Emotions - July 22, 2010 Testimony

Face: Janet Yellen
Face emotion score: -0.369

Face: Blaine Luetkemeyer
Face emotion score: -0.390

Table A.5. Facial Emotions - February 10, 2016 Testimony

Notes: The Face emotion scores shown next to the pictures are at the frame level and are not standardized. To calculate this score, we first compute the average of four basic negative emotions—*Sad*, *Angry*, *Fear*, and *Disgust*, and then multiply it by -1. High values indicate less-negative face emotions.

B List of semi-annual Humphrey-Hawkins testimonies

In the table we show the index values corresponding Chair's emotion indices over the Remarks and Q&A sessions of the 32 testimony days, normalized by its own standard deviation across the 32 days.

Testimony date	Committee	Chair Text	Chair Voice	Chair Face
2010-February-24	Committee on Financial Services	2.48	2.01	-0.92
2010-February-25	Committee on Banking, Housing, and Urban Affairs	1.72	-0.56	-0.2
2010-July-21	Committee on Banking, Housing, and Urban Affairs	2.35	-0.05	-1.98
2010-July-22	Committee on Financial Services	2.06	-0.94	-3.89
2011-March-01	Committee on Financial Services	2.09	-0.99	-0.87
2011-March-02	Committee on Banking, Housing, and Urban Affairs	0.78	-1.37	-2.66
2011-July-13	Committee on Financial Services	-0.7	1.66	-1.95
2011-July-14	Committee on Banking, Housing, and Urban Affairs	-1.03	0.13	-0.21
2012-February-29	Committee on Financial Services	0.35	0.29	-2.17
2012-March-01	Committee on Banking, Housing, and Urban Affairs	1.05	-0.63	-0.39
2012-July-17	Committee on Banking, Housing, and Urban Affairs	-0.11	1.02	-1.27
2012-July-18	Committee on Financial Services	-1.57	0.37	-1.64
2013-February-26	Committee on Banking, Housing, and Urban Affairs	0.63	0.96	-0.94
2013-February-27	Committee on Financial Services	0.68	-0.04	-1.59
2013-July-17	Committee on Financial Services	1.92	-0.18	-1.29
2013-July-18	Committee on Banking, Housing, and Urban Affairs	1.27	-0.49	-0.09
2014-February-11	Committee on Financial Services	0.28	1.13	-1.61
2014-February-27	Committee on Banking, Housing, and Urban Affairs	0.73	1.11	-1.46
2014-July-15	Committee on Banking, Housing, and Urban Affairs	0.2	1.16	-3.56
2014-July-16	Committee on Financial Services	0.52	0.93	-2.67
2015-February-24	Committee on Banking, Housing, and Urban Affairs	1.16	0.07	-1.88
2015-February-25	Committee on Financial Services	-0.09	-1.41	-1.32
2015-July-15	Committee on Financial Services	1.09	-0.48	-1.23
2015-July-16	Committee on Banking, Housing, and Urban Affairs	1.58	0.97	-0.59
2016-February-10	Committee on Financial Services	0.25	0.95	-2.37
2016-February-11	Committee on Banking, Housing, and Urban Affairs	0.5	-0.01	-1.36
2016-June-21	Committee on Banking, Housing, and Urban Affairs	-0.4	-0.94	-2.5
2016-June-22	Committee on Financial Services	1.15	-0.44	-2.34
2017-February-14	Committee on Banking, Housing, and Urban Affairs	0.97	-1.08	-2.71
2017-February-15	Committee on Financial Services	1.26	-1.3	-2.36
2017-July-12	Committee on Financial Services	2.13	0.35	-3.85
2017-July-13	Committee on Banking, Housing, and Urban Affairs	-0.46	-2.1	-2.14

Table B.6. List of semi-annual Humphrey-Hawkins testimonies.

C Correlations of emotion indices

In this section, we present the block level correlations of different emotion indices in Table C.7 and the correlations between Chair and member's emotions during Q&A in Table C.8.

	Text & Voice	Text & Face	Voice & Face
Remarks, full sample	0.07	0.06	0.47***
Bernanke	0.05	0.09	0.51***
Yellen	0.11	-0.15	0.09
Q&A Chair, full sample	-0.04	-0.03	0.03
Bernanke	-0.05	-0.15***	0.00
Yellen	-0.04	0.08	-0.06
Q&A Member, full sample	0.03	-0.003	0.07*

Table C.7. Correlations between three emotion indices

Notes: The text-, voice-, and face-emotion indices are defined in the text. "Chair" refers to statistics conditional on Chair's emotions; "Members" refers to statistics conditional on Congress members' emotions. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

	Chair & Member Text	Chair & Member Voice	Chair & Member Face
Full sample	0.10**	0.01	0.10**
Bernanke	0.07	0.19***	0.10*
Yellen	0.09*	-0.05	0.17***

Table C.8. Correlations between chair and member for text-, voice- and face-emotions

Notes: The text-, voice-, face-emotion indices are defined in the text. "Chair" refers to statistics conditional on Chair's emotions; "Members" refers to statistics conditional on Congress members' emotions. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

D Additional figures

D.1 Breaking news

These figures provide estimation results after dropping testimony blocks overlapping with the breaking news. The estimated responses become somewhat stronger. Hence, market-wide events do not influence our results.

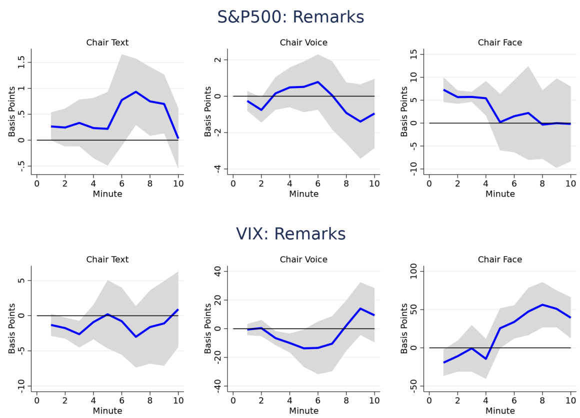


Figure D.2. Responses during the remarks, controlling breaking news.

Notes: The figure provides responses of the change in log S&P500 (top) and the change of log VIX (bottom) to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) during remarks. Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

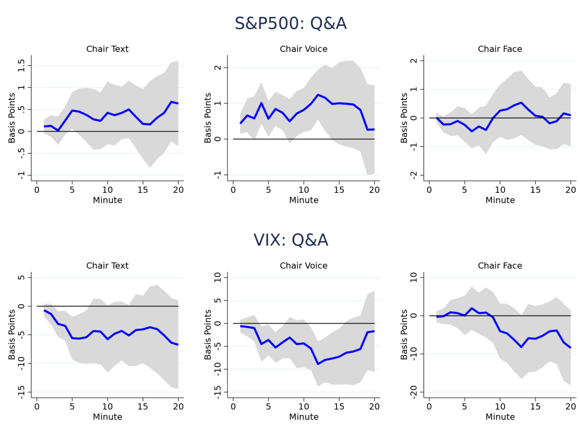


Figure D.3. Responses during Q&A, controlling breaking news.

Notes: The figure provides responses of the change in log S&P500 (top) and the change of log VIX (bottom) to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) during Q&A. Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

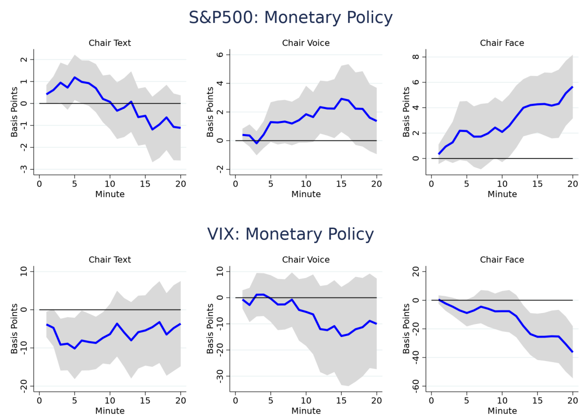


Figure D.4. S&P500 and VIX responses for monetary policy topics, controlling breaking news.

Notes: The figure provides the responses of the changes in log S&P500 (top) and log VIX (bottom) to a one-standard-deviation variation in the text-, voice- and face-emotions of the Fed Chair conditional on discussing topics related to the Fed's monetary policy. Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

D.2 Interest rate expectation responses

This section provides the responses of the yields of ED5 and 10-year Treasury Note futures.[10] The responses are given for the remarks, Q&A, and Monetary topics in Q&A. The responses are positive in most cases.

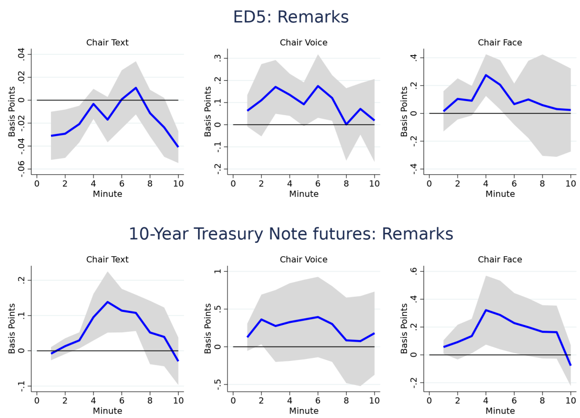


Figure D.5. ED5 and 10-year Treasury Note futures responses.

Notes: The figure provides responses of the ED5 yield (top) and 10-year Treasury Note future yield (bottom) to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) conditional on discussing topics related to the Fed's monetary policy. Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

[10] We obtain the price of the 10-year Treasury Note futures contracts from Chicago Mercantile Exchange Time and Sales database. We follow Cieslak & Schrimpf (2019) and convert futures price changes into yield changes by dividing log futures price changes by the negative of duration. Duration data are obtained from Bloomberg at the daily frequency.

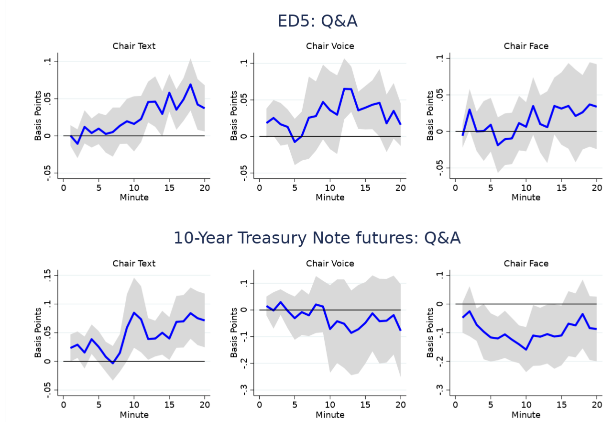


Figure D.6. ED5 and 10-year Treasury Note futures responses.

Notes: The figure provides responses of the ED5 yield (top) and 10-year Treasury Note future yield (bottom) to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) during Q&A. Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

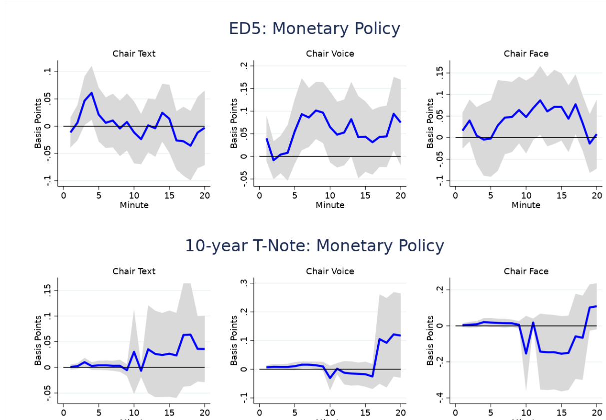


Figure D.7. ED5 and 10-year Treasury Note futures responses.

Notes: The figure provides responses of the ED5 yield (top) and 10-year Treasury Note future yield (bottom) to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) during . Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

D.3 Bernanke and Yellen

This Section provides additional responses for Bernanke and Yellen's testimonies separately: S&P500 responses for the remarks, VIX responses for the remarks and Q&A, and ED5 responses for Q&A. The responses are suggestive that market's reaction to Fed messages is tightly linked to the messenger.

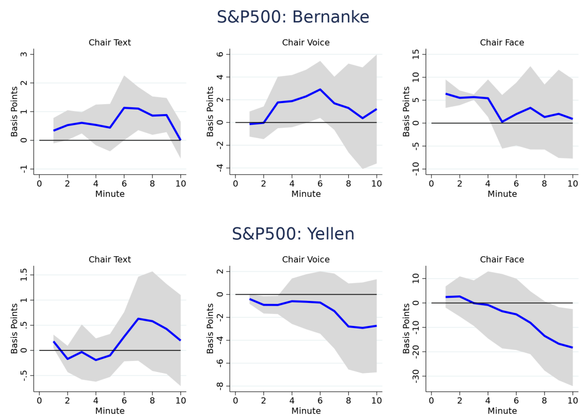


Figure D.8. S&P500 responses during Remarks: Bernanke and Yellen.

Notes: The figure provides responses of the log S&P500 to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) for Remarks during Bernanke testimonies (top) and Yellen testimonies (bottom). Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

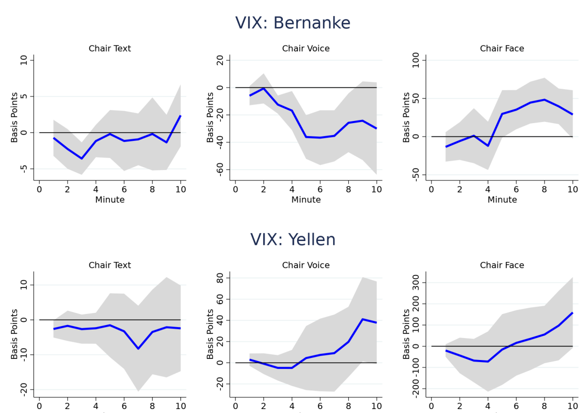


Figure D.9. VIX responses during Remarks: Bernanke and Yellen.

Notes: The figure provides responses of the log VIX to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) for Remarks during Bernanke testimonies (top) and Yellen testimonies (bottom). Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

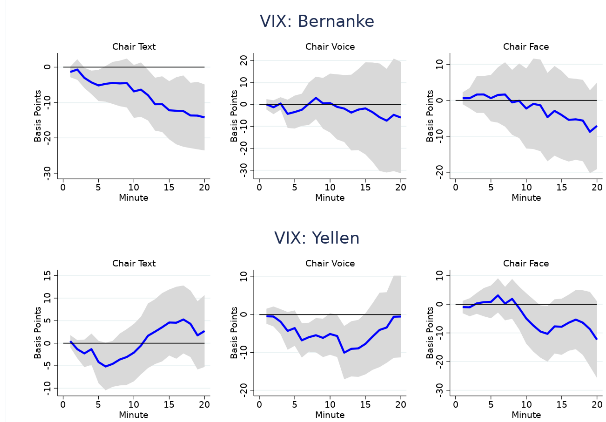


Figure D.10. VIX responses during Q&A: Bernanke and Yellen.

Notes: The figure provides responses of the log VIX to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) for Q&A during Bernanke testimonies (top) and Yellen testimonies (bottom). Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

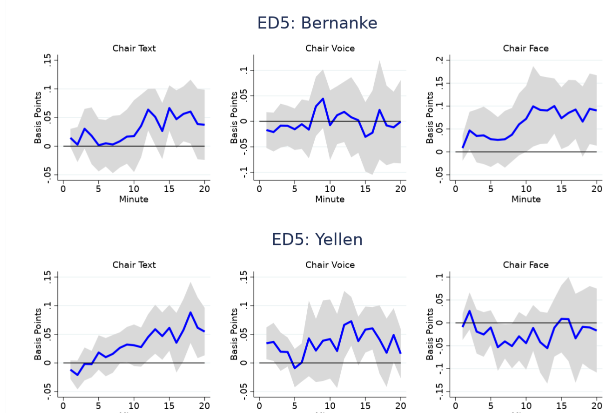


Figure D.11. ED5 responses during Q&A: Bernanke and Yellen.

Notes: The figure provides responses of the ED5 to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) for Q&A during Bernanke testimonies (top) and Yellen testimonies (bottom). Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

D.4 Members

The responses to congressional members' emotions provided here are qualitatively similar but quantitatively weaker than the responses to Chair's emotions, suggesting that financial market activity during the testimony is associated mostly with information reflected in the Fed Chair's emotions.

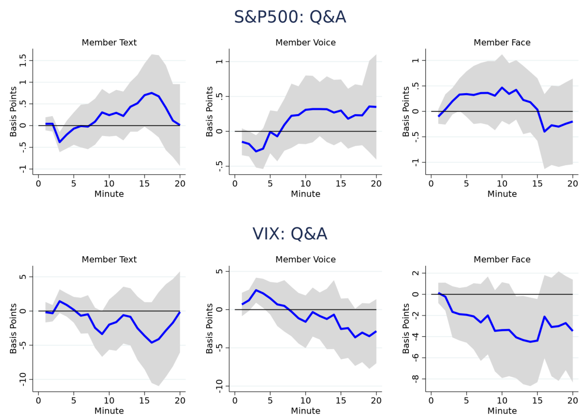


Figure D.12. S&P500 and VIX responses to member's emotions during Q&A.

Notes: The figure provides the responses of the changes in log S&P500 (top) and log VIX (bottom) to a one-standard-deviation variation in the text-, voice- and face-emotions of the Member during Q&A. Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

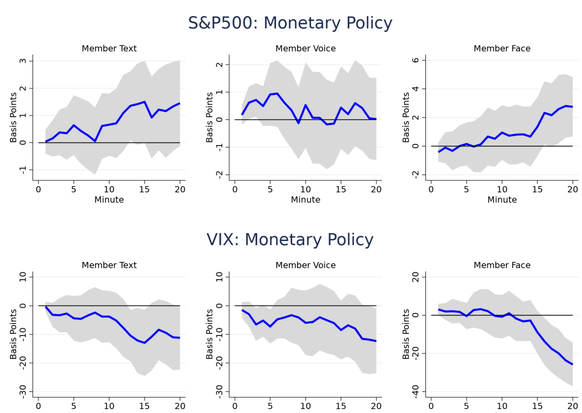


Figure D.13. S&P500 and VIX responses to member's emotions for monetary policy topics during Q&A.

Notes: The figure provides the responses of the changes in log S&P500 (top) and log VIX (bottom) to a one-standard-deviation variation in the text-, voice- and face-emotions of the Member conditional on discussing topics related to the Fed's monetary policy. Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

D.5 Other dimensions of congressional testimonies

We parse the responses the Senate versus the House testimonies, by Day 1 versus Day 2 testimonies, and the first versus the second halves of the Q&A of the same testimony. Although some of them show significant responses, there is no strong systematic link with the responses reported in the main text.

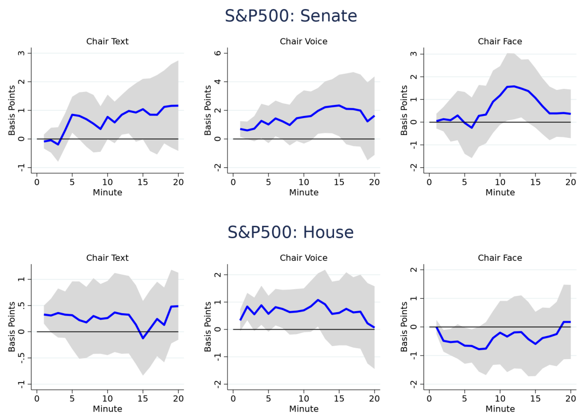


Figure D.14. S&P500 Responses: Senate and house committee

Notes: The figure provides responses of the change in log S&P500 to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) for Q&A in front of the Senate Banking, Housing, and Urban Affairs committee (top) and the House Financial Services Committee (bottom). Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

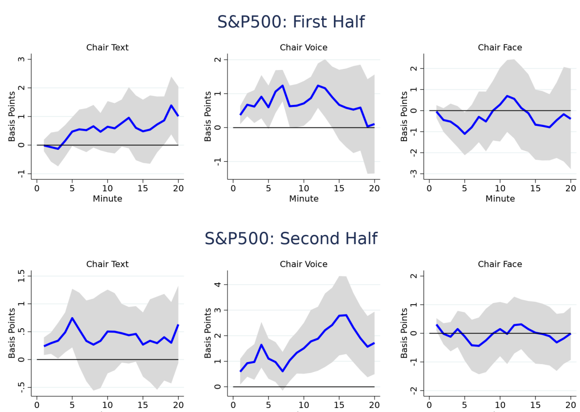


Figure D.15. S&P500 Responses: first and second halves of Q&A

Notes: The figure provides responses of the change in log S&P500 to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) for Q&A during the first day (top) and the second day (bottom). Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

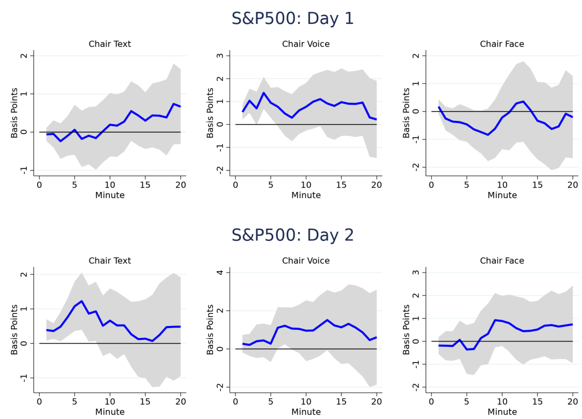


Figure D.16. S&P500 Responses: first and second halves of Q&A

Notes: The figure provides responses of the change in log S&P500 to a one-standard-deviation positive impulse in Fed Chair's text-emotion index (left), voice-emotion index (middle), face-emotion index (right) for Q&A during the first day (top) and the second day (bottom). Shaded area represents the 90 percent confidence interval based on Driscoll-Kraay standard errors.

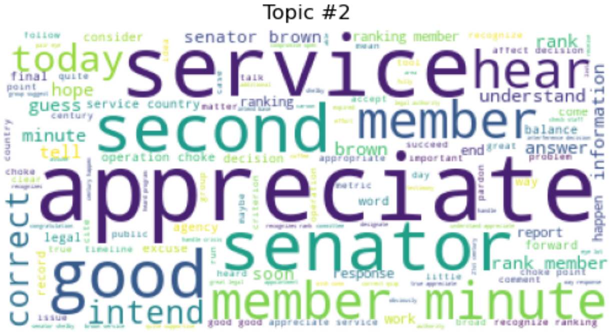
E Q&A round topics

This Section provides topic clusters obtained by using Grootendorst (2022) BERTopic algorithm in the set of 4,323 question-answers across testimonies. BERTopic leverages the word and sentence representations derived from the transformer model BERT as inputs, and creates dense clusters by using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm (Campello et al. 2013).[11] The monetary policy topics are topic #5 and topic #10.

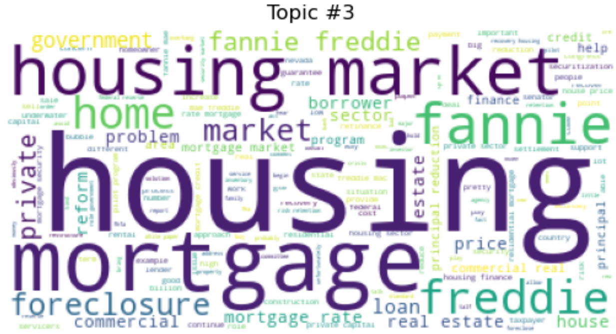
Topic #0

Topic #1

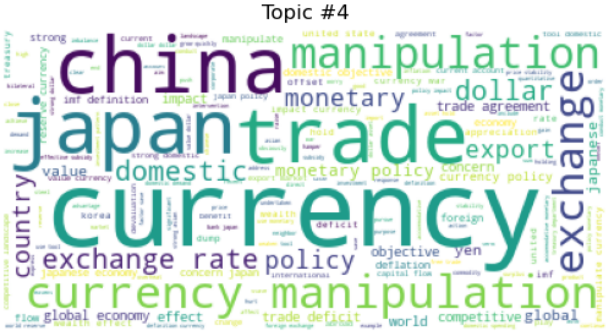
Topic #2



Topic #3



Topic #4



Topic #5

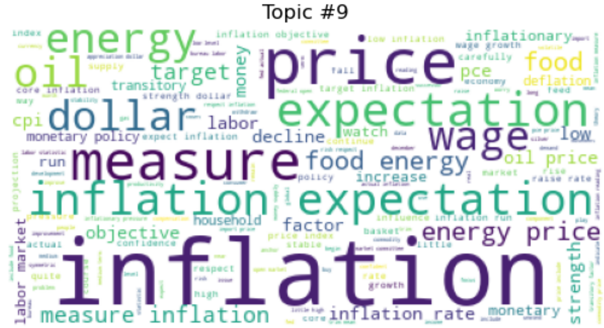
[11] HDBSCAN is a density-based, hierarchical clustering algorithm that constructs a clustering hierarchy tree, and uses a specific stability measure to extract the most significant clusters from the tree.

Topic #6

Topic #7

Topic #8

Topic #9



Topic #10

F Media coverage

Since semi-annual testimonies occur less frequently than FOMC press conferences, it is natural to ask how widely followed is this set of communications, and is the magnitude to the coverage similar to that of the press conferences held following the policy rate announcements?

To examine this, we turn to a familiar archival source for business related news—Dow Jones' Factiva database. For each of our days in question, we examine the number of English language articles that are returned by a keyword search (Table F.9) designed to identify the articles related to the testimony. These searches were also performed for the two days before the testimony and for the two days after the last testimony in the set.[12] The daily counts are then normalized by the number of articles each day that contain the keyword "the" to provide a sense of the magnitude of the coverage.[13]

	Group A		Group B
	Bernanke	Yellen	
Factiva news database	Bernanke, Federal Reserve Chair, Fed Chair, Fed Chairman, Federal Reserve Chairman	Yellen, Federal Reserve Chair, Fed Chair, Fed Chairwoman, Federal Reserve Chairwoman, Fed Chairman, Federal Reserve Chairman	testif*, report*, testim*, deliver*, monetary policy report, humphrey hawkins, humphrey-hawkins, semiannual report, semi-annual report AND congress, senate, congressional, committee, house of representatives, on the hill, Capitol
Twitter	bernanke, fed chair	yellen, fed chair	testimony, testify, testified, testifies, congress, senate, capitol, hill, monetary policy, humphrey hawkins, semi annual, committee

Table F.9. Factiva news database and Twitter search keywords

Notes: We search the Factiva news database and Twitter for testimony related articles and tweets by combining Group A and B keywords set. In particular, we combine Group A - Bernanke (Yellen) with Group B keywords for Bernanke (Yellen) testimony days.

We conduct a similar exercise to examine the number of Twitter posts for the period from the two days before to the two days after a testimony. We use a slightly modified keyword list (Table F.9) to adapt to Twitter's short-text environment. The daily counts are then normalized by the reported average total number of daily Twitter posts.[14]

[12] In the rare cases where the testimony is separated by more than a day, we examine the two days before to two days after each date.

[13] For the purposes of our counts, we do not de-duplicate our article set since we are interested in the magnitude of the coverage.

[14] The average total number of daily tweets has increased from 35 million to 500 million over our study period, https://www.internetlivestats.com/twitter-statistics/#ref-2.

We find similar patterns between the Factiva and Twitter searches.[15] The interest in a testimony builds over the days leading to it, and falls in the days following it, suggesting that the coverage of the event follows a fairly standard news cycle pattern. Moreover, on the peak day, which usually corresponds to the first day of testimony, approximately 0.24%-0.75% of news articles and 0.00089%-0.00682% of Twitter posts cover the testimony. In short, the coverage of the testimonies on the Hill are robustly covered by the print and social media, as well as by major business news networks such as CNBC and Bloomberg.

Next, to ascertain how the coverage compares to that focused on the Federal Reserve's scheduled press conferences, we also created a set of comparable statistics for those dates with a window of +/- two days (i.e., for five days in total). The patterns are similar to those seen in the case of testimony coverage. For the most part, coverage increases over the two days prior, hits a peak on the day of the testimony and generally decreases quickly over the two days post press conference. Second, with the exception of a few dates—the inaugural one and those held in the wake of the taper tantrum, the percent of Factiva's English language documents related to the press conferences range from about 0.24%-0.52%. This would suggest that the testimony is followed in the media at least as much, and often more, than press conferences. Our finding that the media deems the testimonies to be of a similar interest to their audience as the press conference is also consistent with the fact that Economic calendar from Bloomberg ranks Fed press conferences and the Fed testimonies of the same high level of importance.[16]

[15]The correlation between Factiva and Twitter search results is 0.73.

[16]See https://www.bloomberg.com/markets/economic-calendar.

G Breaking news

Since testimonies are 2-3 hours long, it is possible that other major events could occur and affect markets during the period of analysis. To deal with this issue, we mine information out of contemporaneous business news coverage from CNBC broadcasts archived in the Internet Archives' TV News collection. We use CNBC's programming since it remained one of the top business news networks over the time period[17], and it is known to provide accurate and relatively unbiased data and business news to its audience (Vo 2012).

The intuition behind utilizing a business news network's reported breaking news is similar to that behind utilizing print media to capture interest in an event. In short, individuals who are watching the network's programming are doing so to gain insights into events (earnings reports, economic data releases, political events, terror attacks, etc.) that could better inform them and/or impact market activities. On the other side, networks earn profits from subscribership and advertising revenue. To retain their viewership in a competitive environment where events are continuously occurring, the news networks must determine what events and/or breaking news are most in demand by their audience.

An analysis of coverage related to major data releases from the BEA, BLS, and Census, and identified on the Bloomberg Economic Calendar in April 2021 provides some additional insight into reporting and time lags associated with breaking news. The evidence suggests that, first, if CNBC chose to cover the release, its reporting typically occurred within the first few minutes following the official release time. Second, when multiple data releases occurred within a short interval, the data ranked as high importance by Bloomberg's economic calendar, and Dailyfx.com's calendar (e.g., GDP, CPI/Inflation, Michigan confidence survey, initial jobless claims, etc.) were consistently and quickly reported as breaking news, with the medium ranked one being discussed afterwards, if at all.

To create our measures, we collect snapshots of CNBC's onscreen breaking news panels 10 seconds for at least 30 minutes leading up to and testimony, and for an equivalent time after the testimony has finished.[18] The snapshots are then grouped together using a duplicate photo similarity detector, and then OCRed in order to extract the text on screen.[19] The

[17]See, e.g., Comcast (2011-2017) for Nielsen's estimates of household penetration and Stark (1999) for a discussion of CNBC's audience outside of the home.

[18]Ten seconds is used since: (1) this timespan allows us to accurately place the news within our differentiated blocks of time, and (2) the text is usually displayed on screen for at least this long to ensure readability, making the need to switch to smaller time intervals unnecessary for our purposes.

[19]To perform the text extraction, clustering and OCRing, we use a combination of OpenCV python libraries, Tesseract and a visual similarity de-duplicate image finder set at a 90% similarity tolerance to allow for slight variation in pixel coloring and changes in the background seen around the text box.

text was then manually reviewed and corrected, and then categorized into one of 12 topics[20] and one of 5 types of online text[21].

We assign four types of news – macro news release, energy data/commentary, domestic politics, and other market moving news (e.g., extreme weather event, terrorism, Brexit) – as market-wide news that is unrelated to testimonies. Out of 661 blocks of testimony data, we identify 86 blocks that overlap with these news. Our results remain the same if these blocks are excluded.

Below is an example of how CNBC dealt with competing major events on a testimony day on February 15, 2017. In addition to Yellen's second day of testimony, data on Industrial Production and Capacity utilization, business inventories and the housing market index were released, and part of her Q&A session overlapped with President Trump's meeting with Retail CEO's over a proposed border tax, a subsequent press conference held by President Trump and Israeli Prime Minister Netanyahu, and speeches from two Fed Presidents. Three notable observations emerge from our examination of the on-air and breaking news coverage that day. First, while Chair Yellen is onscreen and answering questions, the network reports, usually within 1 to 2 min, what it deems to be important snippets with these "headlines" repeated onscreen multiple times over the course of the day.[22] Second, when the testimony is not onscreen, reports of what is occurring on the Hill are seen via breaking news text or though intermittent recaps - indicating some staff remain tasked with tracking developments. Third, important news related to tax reform, removal of regulation, and policy rate hikes tied to the President's comments or the Regional Fed President's speeches, also tended to appear within 1-2 min of the respective utterances. Overall, the evidence available supports the contention that CNBC reports on the news they deem most relevant for the investors and business communities turning into its programming. Moreover, they do so within a few minutes of information hitting their desks – making it an excellent control for the timing and content of concurrent breaking news that may impact the market during the testimonies.

[20] They are: Testimony related; Other monetary policy (e.g., ECB, BoC, etc.); Company news; Macro news release; Stock index/precious metals/currency/futures movements; Domestic Politics (e.g., regulation, comment from President, etc.); CNBC interview/opinion/analyst-related; Energy data/commentary; Other Survey data; Treasury auction/Treasury department related; Other market moving News (e.g., extreme weather event, terrorism, Brexit); Non-market related news (e.g., sports).

[21] They are: Name/title of person; Data release; General Commentary; Quote; Non-data news release.

[22] The most common that day were "Yellen: Corporate bond market liquidity healthy", "Yellen: No Fed action planned on bond liquidity" and "Yellen: Coming close to achieving Fed mandate".

References

Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press.

Boersma, P. & Weenink, D. (2001), 'Praat, a system for doing phonetics by computer', *Glot International* **5**, 341–345.

Campello, R. J. G. B., Moulavi, D. & Sander, J. (2013), Density-based clustering based on hierarchical density estimates, *in* J. Pei, V. S. Tseng, L. Cao, H. Motoda & G. Xu, eds, 'Advances in Knowledge Discovery and Data Mining', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 160–172.

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

Comcast (2011-2017), 'Comcast annual reports on form 10-K'.

Cootes, T. & Taylor, C. J. (1999), Statistical models of appearance for computer vision.

Cowie, R. & Cornelius, R. (2003), 'Describing the emotional states that are expressed in speech', *Speech Communication* **40**, 5–32.

Devlin, J., Chang, M., Lee, K. & Toutanova, K. (2018), 'BERT: pre-training of deep bidirectional transformers for language understanding', *CoRR* **abs/1810.04805**.

Dietrich, B. J., Enos, R. D. & Sen, M. (2019), 'Emotional arousal predicts voting on the U.S. supreme court', *Political Analysis* **27**(2), 237–243.

Dietrich, B. J., Hayes, M. & O'Brien, D. Z. (2019), 'Pitch perfect: Vocal pitch and the emotional intensity of congressional speech', *American Political Science Review* **113**(4), 941–962.

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

Ekman, P. & Friesen, W. V. (1969), 'The repertoire of nonverbal behavior: Categories, origins, usage, and coding', *Semiotica* **1**(1), 49–98.

Ekman, P., Friesen, W. V. & Hager, J. C. (2002), *Facial Action Coding System: Facial action coding system: the manual: on CD-ROM*, Research Nexus.

Gelder, B., Teunisse, J. & Benson, P. (1997), 'Categorical perception of facial expressions: Categories and their internal structure', *Cognition & Emotion* **11**, 1–23.

Gorodnichenko, Y., Pham, T. & Talavera, O. (2021), The Voice of Monetary Policy, Working Paper 28592, National Bureau of Economic Research.

Grootendorst, M. (2022), 'BERTopic: Neural topic modeling with a class-based TF-IDF procedure', *arXiv preprint arXiv:2203.05794* .

Gudi, A. (2015), 'Recognizing semantic features in faces using deep learning', *CoRR* **abs/1512.00743**.

Laukka, P. (2005), 'Categorical perception of vocal emotion expressions', *Emotion (Washington, D.C.)* **5**, 277–95.

Laukka, P., Juslin, P. & Bresin, R. (2005), 'A dimensional approach to vocal expression of emotion', *Cognition and Emotion* **19(5)**, 633–653.

Stark, S. (1999), 'If only Nielsen counted barroom viewers: CNBC execs fight to show advertisers audience bigger than home only numbers.', *InvestmentNews* **(October 18)**.
URL: *https://www.investmentnews.com/if-only-nielsen-counted-barroom-viewerscnbc-execs-fight-to-show-advertisers-audience-bigger-than-home-only-numbers-2042*

Vo, A. (2012), 'CNBC: Is trust as critical as we believe it is when it comes to news sources?', *Envzone* **(March 4)**.
URL: *https://www.envzone.com/cnbc-is-trust-as-critical-as-we-believe-it-is-when-it-comes-to-news-sources*

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. & Brew, J. (2019), 'Huggingface's transformers: State-of-the-art natural language processing', *CoRR* **abs/1910.03771**.

Zhu, Y., Kiros, R., Zemel, R. S., Salakhutdinov, R., Urtasun, R., Torralba, A. & Fidler, S. (2015), 'Aligning books and movies: Towards story-like visual explanations by watching movies and reading books', *CoRR* **abs/1506.06724**.