Bargaining over Words? Text Analysis of a Model of Monetary Policy by a Committee *

Hie Joo Ahn † Thomas R. Cook‡ Taeyoung Doh§

March 26, 2026

Abstract

Many central banks make policy decision collectively in the form of a committee. With the increased public communications of policy actions, wording over policy statements became an important tool for committee members to express their views. By analyzing sentiments expressed in transcripts of the Federal Open Market Committee (FOMC) meetings and FOMC statements using a large language model, we explore the interaction among FOMC members over statement language for various topics identified by word clustering. Specifically, we examine how statement language aggregates preferences of individual participants expressed during FOMC meetings. The Chair tends to have a dominant role in the aggregation process by setting the agenda but the Chair's power was often constrained by the opinion of a "pivotal" member. We find that the pivotal member was not necessarily the median voter but also a leading member in the hawkish or the dovish camp to be consistent with the FOMC's consensus-seeking culture of lowering the number of dissents. Over time, the Chair's influence in statement language increased with more frequent private communications between the Chair and Reserve Bank president before FOMC meetings, which allowed the Chair the flexibility the baseline proposal to better accommodate members' preferences. An agenda-setting Chair model of the monetary policy committee can explain these empirical findings.

JEL classification: D71; E52; E58
Keywords: Monetary policy committee, Large language model, Monetary policy communication.

*We thank seminar participants at the Federal Reserve Bank of Kansas City, Federal Reserve Board, and Sogang University for helpful comments. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors , the Federal Reserve Bank of Kansas City, the Federal Reserve System or of anyone else associated with the Federal Reserve System. This version: November 2025

†Federal Reserve Board: hiejoo.ahn@frb.gov
‡Federal Reserve Bank of Kansas City:Thomas.Cook@kc.frb.org
§Federal Reserve Bank of Kansas City: Taeyoung.Doh@kc.frb.org

1 Introduction

> "CHAIRMAN BERNANKE. Just to get a sense of the Committee's preferences so that I understand them, if we were to go back to alternative B as written with the forward-leaning language of the last paragraph, without the "financial conditions", how many dissents would I have? Any dissents? [Simultaneous speakers] I'm asking the question, just so I understand Committee preference.
> Mr. EVANS. No reference to mid-2013?
> CHAIRMAN BERNANKE. That's right. Alternative B as written, except we take out the "financial conditions" language and we use the more forward-leaning paragraph 4.
> MR. TARULLO. So no action at all?
> CHAIRMAN BERNANKE. No action at all
> MR. EVANS. Mr. Chairman–
> CHAIRMAN BERNANKE. I'm trying to get preferences.
> MR. EVANS. I could not. I mean, it just seems clear that we would simply be postponing that very discussion until September, when we would have the very same outcome. And so I could not support that.
> CHAIRMAN BERNANKE. Okay. Thank you. We have to get this done by 1:30. Yes?
> MS. YELLEN. You might have other negative votes."
> (August 9, 2011 FOMC Meeting Transcript FOMC (2011b)).

> "The Committee currently anticipates that economic conditions–including low rates of resource utilization and a subdued outlook for inflation over the medium run–are likely to warrant exceptionally low levels for the federal funds rate at least through mid-2013."
> (August 9, 2011 FOMC Post-meeting Statement FOMC (2011a)).

The post-meeting FOMC statement in August 2011 changed the language associated with forward guidance on the future path of interest rates from "exceptionally low levels for an extended period" to "exceptionally low levels at least through mid-2013". As noted by Lunsford (2020), the August 2011 forward guidance language was a clear example of policy inclination toward accommodation. Three members who did not support additional policy accommodation dissented. The staff of the Federal Reserve Board prepared the three alternative versions of the FOMC statement in which alternative A and C are dovish and hawkish versions, respectively while alternative B is typically regarded as a more consensus-oriented reference version. In this case, the alternative B and the alternative C do not change forward guidance language from the previous meeting while the alternative A calls for the change. The above quote from the transcript of the FOMC meeting shows that the change in language was the output of an intense bargaining over words.

While the three dissents as in the August 2011 meeting do not happen frequently, the deliberation over language to communicate policy stance is a main purpose of the FOMC meeting. With the increased use of public communications, voting at an FOMC meeting involves agreeing on the exact language used in policy statements as well as policy actions such as changing the target interest rate. Bargaining over words is an important

part of monetary policy deliberation because the policy committee members discuss their views on the appropriate monetary policy stance through alternative wordings on outlook as discussed in Baerg (2020). As shown by Gáti and Handlan (2024), changes in statement language can shape market expectations of future policy actions and generate immediate responses from financial markets, even if the FOMC does not decide to change any policy instrument. In addition, text analysis of FOMC members' remarks during meetings can reveal their preferences on key macroeconomic variables such as inflation target even when it is not explicitly communicated to the public as studied by Shapiro and Wilson (2022).

In this paper, we combine a text analysis of FOMC discussion on policy options (e.g., Hansen, McMahon and Prat (2018)) with a collective decision making model of the monetary policy committee (e.g.,(Riboni and Ruge-Murcia, 2010, 2020, 2025)) in which the aggregation of individual members' preferences into the committee decision is explicitly modeled.[1] We identify topic scores of sentences in the statement data by fitting a topic model (e.g., Hoffman, Bach and Blei (2010), Griffiths and Steyvers (2004)). Each document (a sentence in the statement data) can contain multiple topics and the fitted topic model provides information on the prevalence of alternative topics in the document. Then we extract sentiment scores from documents using a pre-trained Large Language Model (LLM) in Hartmann, Heitmann, Siebert and Schamp (2023) that is able to capture the context-dependent meaning of a word. By combining a topic model with a pre-trained LLM, we quantify sentiment scores conditional on topics for documents. Next, we apply the same methodology to the transcript data to back out sentiment scores conditional on topics. We isolate remarks by voting FOMC members that are relevant for topics identified from the statement data and calculate sentiment scores using the LLM model. Finally, we consider different forms of regimes through which FOMC members' topic sentiment scores are aggregated into those in the released statement. Specifically, we consider 1) the equal weight regime, 2) the median voter regime, and 3) the chair-centric regime in which the weight is proportional to the closeness of a member's topic sentiment score to the chair's topic sentiment score. The existing literature using text analysis to extract FOMC members' preferences (e.g., Shapiro and Wilson (2022)) is not explicit about the aggregation regime but tends to assume the equal weight regime after controlling the individual fixed effect for the preference.

We find that the chair-centric regime is preferred more than 60% of sample period although the frequency varies a lot across topics. For instance, regarding the real economic activity topic, the chair-centric regime is preferred only 44% while it is preferred 87% for the topic on inflation conditions. In addition, the chair-centric regime is more dominant (preferred 82%) during Yellen's tenure than Greenspan's tenure (preferred 57%) or Bernanke's tenure (preferred 54%). This does not necessarily imply that Yellen was more "dictatorial" inside the FOMC but might be related to the fact that the average length of statements were greater during her tenure than other chairs, giving her more room to incorporate other members' remarks into her view at the end

[1] Riboni, Ruge-Murcia and Tran (2025) use the textual analysis of FOMC transcripts to tease out sentiment measures about policy and economic outlook from members' remarks. While they estimate sentiments for various topics of FOMC members, they consider only the interest rate as the aggregated outcome variable without addressing changes in statement language sentiments, which is out main question.

of the meeting. In addition, she engaged more frequently with Reserve Bank presidents through private calls before FOMC meetings, which allowed her to incorporate their preferences into statement language better. In fact, dissenting votes significantly declined during her tenure (6 in total, 0.19 per meeting) than Bernanke's tenure (48 in total, 0.74 per meeting) and Greenspan's tenure (82 in total, 0.54 per meeting).[2]

Even without any change in the aggregation regime, the statement topic score can shift materially if the distribution of individual members' policy preferences changes over time. In particular, under the equal weight regime which is more likely to influenced by outliers, the statement topic sentiment score can be notably different if there are fat tails in the underlying distribution of individual FOMC members' preferences. We illustrate this pattern by contrasting the aggregation of sentiment scores on the interest rate topic in the October 2008 FOMC meeting with that in the August 2011 FOMC meeting. In the August 2011 FOMC meeting, members with more dovish views successfully pushed the committee to adopt language that is more palatable for their preferences by creating a "winning" coalition. The intense conversations between the Chairman Bernanke and other members described in the previous excerpt highlight this. In contrast, the dovish outlier member in the October 2008 FOMC meeting failed to do so. Interestingly, the equal weight aggregation regime was preferred to the median voter regime in both meetings. This finding suggests that although the median member's preference is often touted as representing the FOMC's view (e.g., focus on the median number in the summary of economic projections), he or she may not be always the pivotal member in the committee's decision. With the FOMC's consensus-driven culture, not only the median voter but also a leading member in the hawkish or the dovish camp can be a pivotal member. We show that an agenda-setting Chair model of the monetary policy committee in Riboni and Ruge-Murcia (2010) modified by multiple proposals made by the Chair and imperfect information on members' preferences can explain this finding.

We find an interesting difference in the prevalence of topics across statements and transcripts. While topics such as the interest rate and risks to the outlook are more prominent in statements than in transcripts, the opposite is true for the topic on inflation and real economic activity conditions. Since Reserve Bank presidents tend to comment on national economic conditions in relation to their interactions with district business people and others, their remarks tend to include more detailed information on economic activities and inflation. Indeed, Bobrov, Kamdar and Ulate (forthcoming) and Fos and Xu (2022) argue that district-level economic conditions have oversized impacts on voting members' policy deliberation. In contrast, discussions on risks to the outlook are more prevalent in statements than transcripts, perhaps reflecting the fact that the FOMC does not want to lock in itself into a single most likely scenario.

Monetary policy affects the real economy through financial market conditions. One challenge in linking our topic sentiment scores in statements with high frequency asset price movements around FOMC announcements is that we do not know statement language expected by financial market participants from statement and transcript data alone. For this purpose, we use a proprietary dataset maintained by a consulting firm who has been sending out the draft language to their paid subscribers including major

[2]Numbers for Greenspan and Bernanke are from Thornton and Wheelock (2014) and the number for Yellen is by the authors' calculation. For Greenspan's tenure, we consider his entire tenure from August 1987 to January 2006, not just the period included in our text analysis.

investment banks a few days before FOMC meetings. By applying the topic model estimated by statement data and the pre-trained LLM to this dataset, we calculate market expectations of topic sentiment scores and isolate the surprise component in the topic sentiment score that we can relate to high-frequency asset price movements.

We connect our surprise component measures for various topics with monetary policy shocks identified by high-frequency bond market responses around FOMC announcements in Bauer and Swanson (2023). They decompose bond market responses into two terms: 1) a news component that is predictable with pre-announcement information on economic and financial variables and 2) a purely unexpected component. While we find that our measure of the surprise component in the sentiment on policy guidance does not have a statistically significant relationship with the raw high-frequency bond market response, it does have a significant relationship with the news component and the purely unexpected component orthogonal to news. Our findings suggest that text analysis is useful for identifying parts of information contained in statement that can inform the FOMC's reaction to inter-meeting economic news.

Related Literature: Our paper is related to a burgeoning literature on applying text analysis to monetary policy analysis (Aruoba and Drechsel (2024), Baerg (2020), Chappell Jr, McGregor, Vermilyea et al. (2005), Cieslak and McMahon (2023), Doh, Song and Yang (2022),Handlan (2022) Hansen et al. (2018), Schonhardt-Bailey (2013), Shapiro and Wilson (2022), Fisher, McCaughrin, Prazad and Vandergon (2023), Bertsch, Hull, Lumsdaine and Zhang (2025), Gáti and Handlan (2024), Bobrov et al. (forthcoming), Fos and Xu (2022), Riboni et al. (2025), Laarits, Matthies, Vasudevan and Yang (2025), Hack, Istrefi and Meier (2025), Kazinnik and Sinclair (2025) for a nonexhaustive list of papers). Aruoba and Drechsel (2024) and Handlan (2022) use text-analysis of FOMC statements or Greenbook/Tealbook documents prepared by the Federal Reserve Board staff to identify a monetary policy shock while Hack et al. (2025) construct a new measure of monetary policy shock using the variations in the hawk-dove balance in the FOMC, which purges out the confounding effect of the time-varying Taylor rule coefficients in identifying a monetary policy shock. Doh et al. (2022) examine the financial market effect of changing statement language to suggest an alternative monetary policy stance. Baerg (2020), Chappell Jr et al. (2005), Hansen et al. (2018), and Schonhardt-Bailey (2013) analyze transcript data to show how monetary policy deliberation process changed over time. Shapiro and Wilson (2022) estimate the FOMC's objectives using text analysis of transcript data. Cieslak and McMahon (2023) create a monetary policy stance measure from the difference between hawkish and dovish remarks during a FOMC meeting and explore how it affects the risk premiums in asset prices subsequently. Fisher et al. (2023) investigate if adding topic sentiment scores from transcripts and tealbook to the futures market data improves the forecasting power of the future interest rate especially during easing cycles. Gáti and Handlan (2024) estimate monetary communication rules that regress interest rate and macroeconomic forecasts by the Fed onto words in the statement, which reveals the value of information contained in statement language. Bertsch et al. (2025) apply a collection of LLM models to the Federal Reserve policymakers' speeches to classify those discussing financial stability concerns.

We build on these papers and connect them with the literature of explicitly modeling the committee decision process through the aggregation of individual policymakers'

preferences (e.g., Hansen, McMahon and Rivera (2014), (Riboni and Ruge-Murcia, 2010, 2014, 2020, 2025)). However, these papers mostly focus on the voting decision regarding changing the policy rate rather than the decision leading to the adoption of statement language used in public communications. Unlike the single-dimension of the rate decision, statement language represents multiple dimensions of monetary policy, which can be related to different asset price responses (e.g., Swanson (2021)) and allows more room for objections from individual policymakers although they would not formally dissent to the committee's decision on the interest rate. Without the textual analysis of transcript data, it is challenging to understand how the heterogeneity in members' preferences affect the committee's decision because explicit dissents on the rate decision are relatively infrequent. Since we allow a different aggregation scheme for the statement sentiment score for each topic, we can address how each aspect of monetary policy can be determined through the FOMC's deliberation process. We represent the heterogeneity in the views of FOMC members in the form of member-specific topic sentiment scores. While we do not investigate the source of this heterogeneity, Bobrov et al. (forthcoming) and Fos and Xu (2022) show that Reserve Bank presidents' frequent contacts with business leaders in their districts and their familiarity with regional economic conditions may lead them to overweight evidence they obtain from district-level contacts. Closely related to our paper is Laarits et al. (2025), who provide prompts to an LLM to construct how the FOMC aggregates members' arguments. However, they emphasize the information aggregation aspect of the committee decision and downplay the preference heterogeneity among members. Kazinnik and Sinclair (2025) also apply LLM prompts to extract FOMC members' preferences from their speeches and relate them to the released statement. While their aggregation methodology is closely related to ours, they fix the aggregation weight parameters in a judgmental way rather than calibrating to match the aggregated outcome from members' remarks with statement language as we do.

Our paper is also related to applying a topic model developed in Blei, Ng and Jordan (2003) to textual data (e.g., Griffiths and Steyvers (2004), Hansen et al. (2018)). We combine topic analysis with sentiment analysis using a pre-trained large language model (Siebert in Hartmann et al. (2023)) to create topic-conditional sentiment scores. The literature on sentiment analysis typically trains a language model for one set of textual data to pin down model parameters and tries to predict the sentiment for a different set of textual data using trained parameters. In our case, we fit an additional layer on topic-sentiment scores of FOMC voting members from transcripts to predict topic-sentiment scores from statements by trying three different weighting schemes. Conditional on a weighting scheme, we do not need any additional model parameter.

Finally, our paper is related to the literature of identifying a monetary policy shock using high-frequency bond market responses during announcements (e.g., Bauer and Swanson (2023)). By regressing high-frequency asset price movements onto multiple measures of surprise components in topic sentiment scores from the statement data, we find which part of statement language is more responsible for a particular asset market movement, which would not be possible in the event-study approach using high-frequency financial market data alone.

Our paper is organized as follows. Section 2 explains the empirical methodology used in the paper. Section 3 describes the text data we use. Section 4 provides empirical results for topic-sentiment analysis and aggregation regimes. Section 5 describe an

agenda-setting Chair model of the monetary policy committee and discusses open issues for modelling the monetary policy committee. Section 6 concludes.

2 Empirical Methodology

2.1 Construction of Topic Sentiment Scores

For textual data we have, topic labels are not provided explicitly, meaning that we have to recover topics implied in documents by using certain features in documents. A widely used feature is the frequency of words used in a document, which represents each document by "bag of words". Essentially, a topic is the cluster of words based on their co-frequency patterns in documents. A document have multiple topics although the prevalence may differ across topics.

To identify topics, we fit a Latent Dirichlet Allocation (LDA) model for text data. The LDA model consists of the three classes of variables: 1) word-frequency data, 2) latent variables related to the topic distribution per document and the word distribution per topic, and 3) hyperparameters determining the prior distributions of latent variables in 2). For our analysis, the main objective is to recover the topic distribution per document. More specifically, we are interested in identifying a K-dimensional vector of probabilities for topic scores $\theta^l = (\theta_{1,l}, \ldots, \theta_{K,l})'$ given a document l (a sentence in our data) where K is the total number of topics. Suppose that there are total N words used in the entire corpora of documents. Each topic (topic k, for example) is represented by an N-dimensional vector of probabilities of words appearing in that topic $\Phi^k = (\Phi_{k,1}, \ldots, \Phi_{k,N})$. Let $W = (W_1, \ldots, W_D)$ represent a word-frequency matrix for D documents that we use to fit a topic model. The size of W is $\times D$ with each column representing the word-frequency vector in a single document. Figure 1 illustrates how a topic model decomposes information from the textual data.

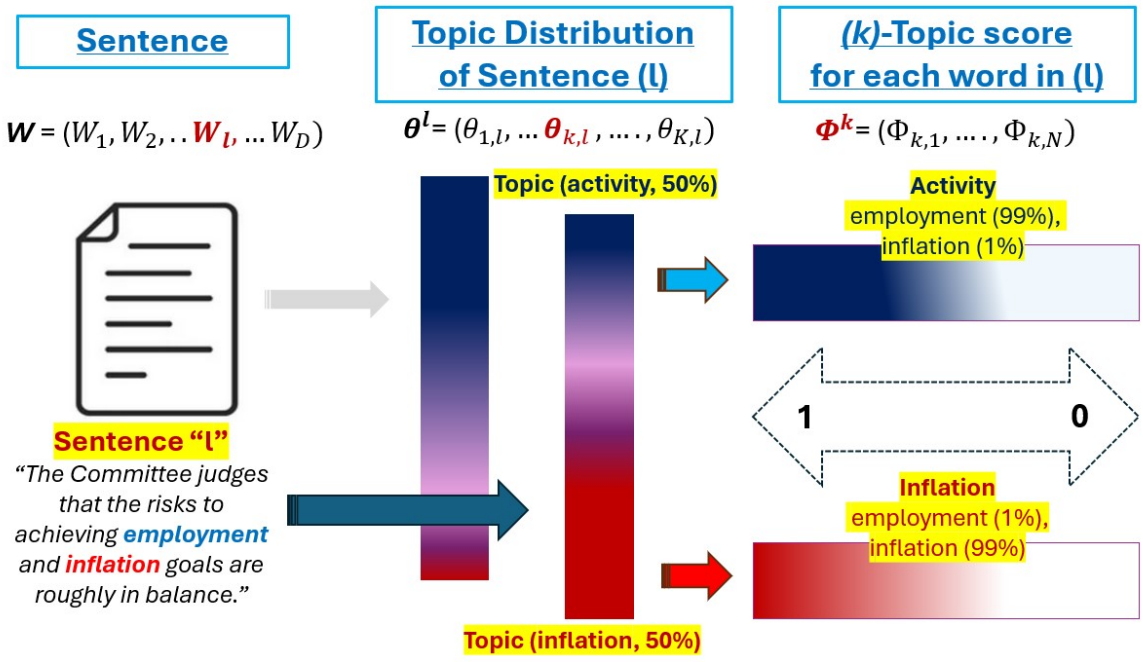
While we observe W, we need to estimate Φ and Θ. Following Wallach, Murray, Salakhutdinov and Mimno (2009), we use Dirichlet priors for Θ and Φ described by hyperparameters α and β.[3] We are interested in the document-specific distribution, which is determined by the posterior estimate of Φ and Θ given W, α, β. To link document-specific topic distributions with the word-frequency matrix, we introduce a latent topic indicator $z_{i,l}$ for each (word, document) pair. Let Z is a matrix of topic indicators for all the (word, document) pairs in the dataset. The plate diagram in Figure 2 illustrates how an LDA model works. The likelihood for W given prior parameters Φ and Θ can be calculated as follows:

$$p(W|\Phi, \Theta) = \int p(W|\Phi)p(\Phi|Z)p(Z|\Theta)dZ. \tag{1}$$

Combining this likelihood with Dirichlet priors $p(\Theta|\alpha)$ and $p(\Phi|\beta)$ leads to the kernel of the posterior density $p(\Theta, \Phi|W, \alpha, \beta) \propto p(\Theta, \Phi|W)p(\Theta|\alpha)p(\Phi|\beta)$. The main challenge in the inference is that integrating the posterior kernel with respect to Θ and Φ is intractable because the posterior kernel depends on the summation of the

[3]The Dirichlet distribution is a multivariate generalization of the beta distribution and conjugate with the multinominal distribution, which is convenient to model uncertainty of the discrete probabilities. For more details, see Murphy (2023).

Figure 1: Topic Model Illustration



Notes: The example sentence is from the post-meeting FOMC statement released on September 18, 2024.

product between elements in Θ and Φ (e.g., $\sum_{k=1}^{K} \theta_{k,l}\Phi_{k,l}$) as noted by Blei et al. (2003).[4] Although the direct sampling from the exact posterior density is not feasible, we can obtain the posterior estimates using approximate inference algorithms. We use the collapsed Gibbs sampling algorithm proposed by Griffiths and Steyvers (2004) that provides the posterior mean estimates of Θ and Φ from posterior draws of Z obtained by Gibbs sampling. With posterior samples of Z, Θ and Φ can be estimated to fit these estimates of Z. Since $z_{i,l}$ follows a multinomial distribution given θ^l, we can obtain the posterior estimate of θ^l using the conjugacy of the Dirichlet prior with the multinomial distribution.[5]

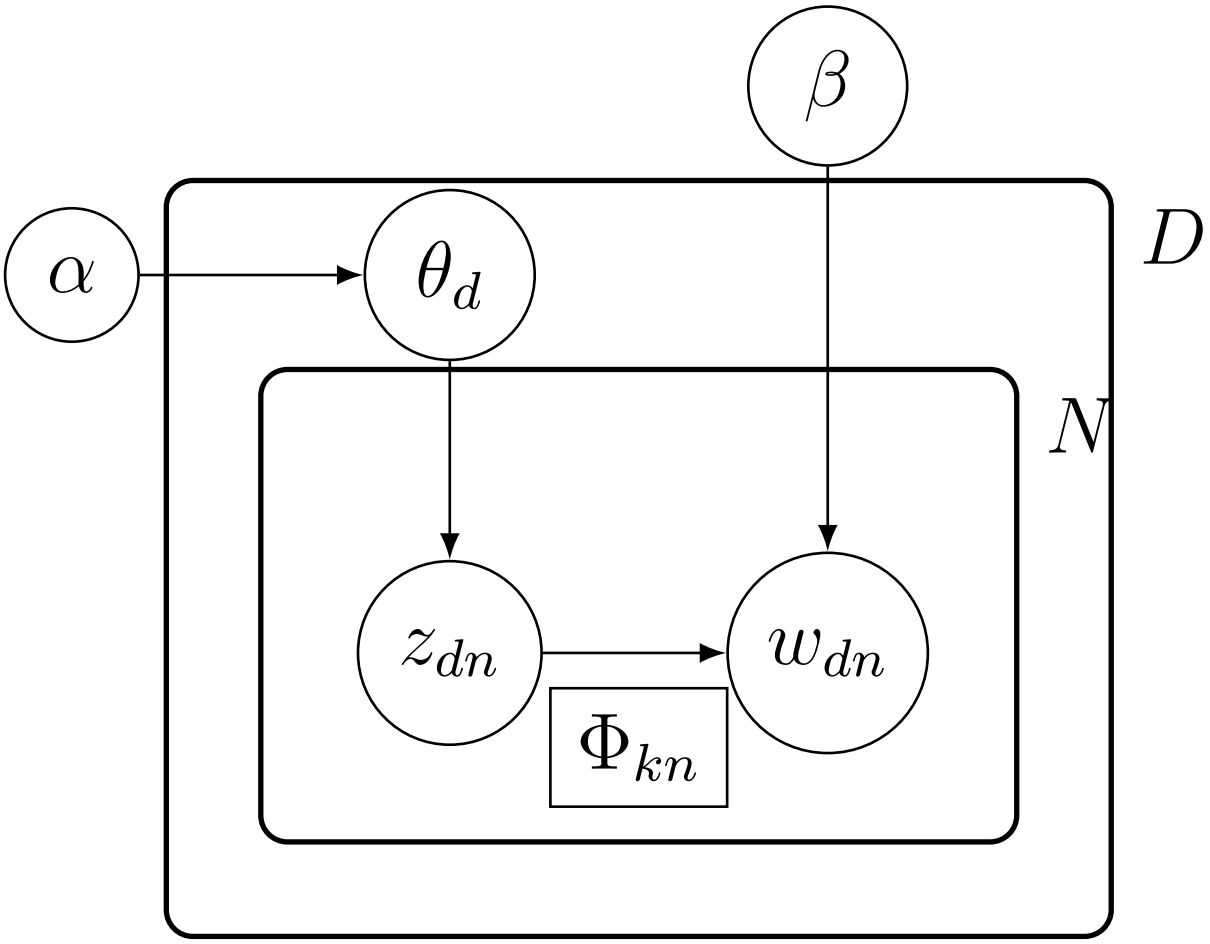
For each sentence, we quantify the sentiment by using the SIEBERT model in Hartmann et al. (2023). SIEBERT is a pre-trained large language model that is a fine-tuned version of RoBERTa in Liu, Ott, Goyal, Du, Joshi, Chen, Levy, Lewis, Zettlemoyer and Stoyanov (2019).[6] The large language model first transforms an input text to a vec-

[4]The appendix provides the analytical expression of the posterior kernel that illustrates this issue.

[5]Another method to obtain the posterior estimates of Θ and Φ is to directly approximate the posterior distribution by minimizing the Kullback-Leibler distance between the approximate posterior distribution chosen from a "variational" family of probability distributions and the exact posterior distribution. See Blei et al. (2003) for more details on the variational inference method.

[6]One alternative to using a pre-trained model for sentiment analysis is Chen and Mankad (2022), which directly put an additional layer onto an existing topic model to extract sentiment scores. They train this topic sentiment model using online review data with pre-labeled sentiment scores. In contrast, we leverage a pre-trained large language model to obtain sentiment scores and calculate topic-sentiment scores by multiplying the output of each model. A structural topic sentiment model can be more internally coherent but we do not have a huge amount of pre-labeled textual data on FOMC monetary policy deliberations and communications to train sentiment scores for statement and transcript data. Relying on the transfer learning of a pre-trained language model is a more tractable alternative in this

Figure 2: Plate Diagram for Latent Dirichlet Allocation

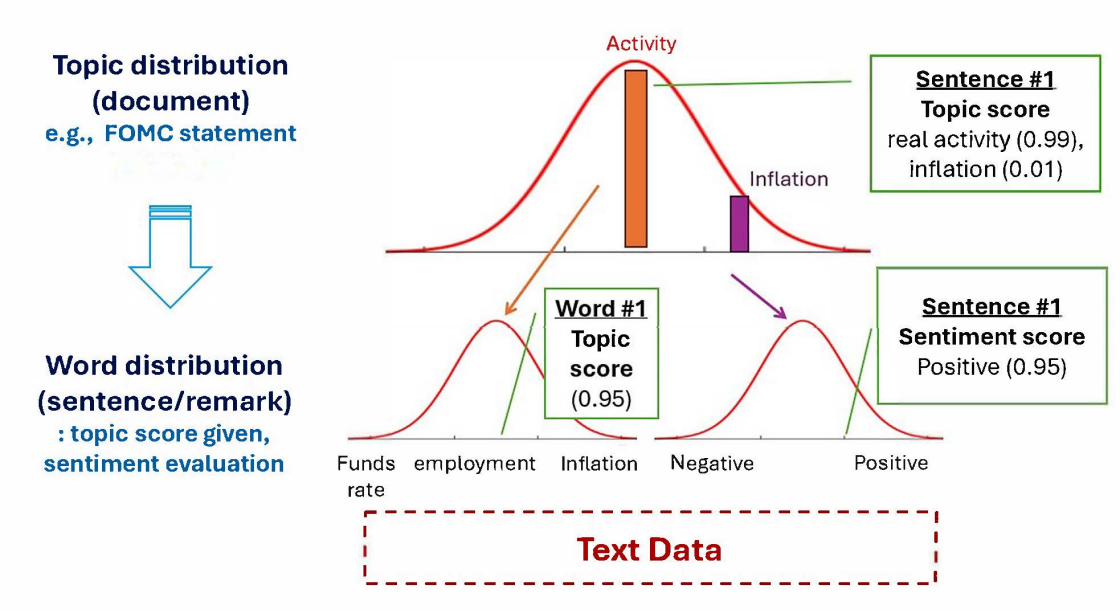


Notes: The outer boundary represents a document out of D documents and the inner boundary describes a word out of N words in a document. Φ_{kn} is estimated from post-meeting statement data and fixed when the model is fit for other datasets. θ_d and z_{dn} are newly estimated for each dataset. α and β are set at default values of 1 and 0.25, respectively.

tor of tokens, which are numeric representations of words. Then, they apply a deep neural network architectures to obtain the numeric representation of the entire input text, which is called the "embedding representation." The model trains parameters that generate the embedding representation of a given text to maximize the performance in standard natural language understanding experiments such as text similarity classification, for example. SIEBERT further optimizes the embedding representation by fine-tuning additional neural network layers on top of the output from RoBERTa with 15 different pre-labeled datasets (e.g., online reviews and tweets) to classify the sentiment of a document as positive or negative. SIEBERT provides the sentiment score ranging from -1 (perfectly negative) to 1 (perfectly positive). In practice, most sentiment scores are very close to -1 or 1 without any meaningful probability around the neutral level (0). We express the sentiment score of the l-th document by s_l. Then

case. Handlan (2022) and Doh et al. (2022) show that a fine-tuned large language model can better identify a monetary policy shock or facilitate the counterfactual language analysis.

Figure 3: Hierarchical Characterization of Text to Construct $s_{l,k}$



Notes: We run the sentiment model for the example sentence in Figure 1.

we construct a topic-sentiment score of the l-th document by multiplying the sentiment score by topic scores. Hence, the k-th topic-sentiment score of the l-th document is given by

$$s_{l,k} = \theta_{l,k} s_l. \tag{2}$$

Each FOMC statement consists of several sentences. To construct topic-sentiment scores at the statement level, we sum the sentiment scores of individual sentences by topic scores.

$$s_k^{\text{FOMC}} = \sum_{p=1}^{P} \frac{\theta_{p,k} s_p}{P}, \tag{3}$$

where p is the index of a sentence in the respective FOMC statement and P is the total number of sentences in the FOMC statement.[7] Figure 3 illustrates how we construct $s_{l,k}$ from a text.

Now, we extend the same method to individual remarks made by FOMC voting members. From transcript data, we can identify the sentiment of the p-th remark made by the i-th member (s_p^i). We can construct the topic-conditional sentiment score by multiplying s_p^i by a topic distribution for the p-th remark by the i-th member. By aggregating the topic conditional sentiment scores for all the remarks made by the i-th member, we can construct the member-specific topic sentiment score $s_k^{\text{Member},i}$. If each FOMC member makes the same number of remarks equal to the number of sentences in the FOMC statement, we can simply aggregate $s_k^{\text{Member},i}$ across different members to match the topic-sentiment score at the FOMC statement. However, the number of remarks by FOMC members varies a lot across members, creating a challenge in the

[7] In practice, the length of the statement varies over time and P can be better represented by P_t.

aggregation because the simple average treats each member's sentiment scores equally without considering the relevance of remarks for policy topics. In addition, not all the remarks by the same members are equally relevant for topic-sentiment scores because the relative frequency of topic key words is wildly different across remarks. To address this issue, we compute the relative frequency of topic key words among total words in each remark ($freq^i_{k,p}$) and apply the softmax function to obtain weights as follows:

$$\hat{freq}^{\,i}_{k,p} = \frac{\exp(freq^i_{k,p})}{\sum_j \exp(freq^i_{k,j})}. \tag{4}$$

Then we can create the topic-sentiment score of the i-th member by the following weighted mean of topic-sentiment scores made by the i-th member.

$$s^{\text{Member},i}_k = \sum_p \hat{freq}^{\,i}_{k,p} \theta^i_{p,k} s^i_p. \tag{5}$$

With this normalization, the magnitude of $s^{\text{Member},i}_k$ matches the one in s^{FOMC}_k. For example, if the i-th member happens to a dictator and gets a full weight while all the other members get zero weights, $s^{\text{Member},i}_k$ would be equal to s^{FOMC}_k assuming that the FOMC statement is the summary of the dictator member's remarks during the FOMC meeting.

2.2 The Statement Topic Sentiment Score through the Aggregation of Statement Topic Scores at the Transcript Level

To connect members' topic sentiment scores with those in statements, we consider the aggregation weight regime. We allow the aggregation weight regime to be different for different topics.

$$s^{\text{FOMC}}_k = \sum_i w_{i,k} s^{\text{Member},i}_k. \tag{6}$$

Without further restrictions, $w_{i,k}$ is not well identified based on only (s^{FOMC}_k, $s^{\text{Member},i}_k$, $\forall i$). We model $w_{i,k}$ as a parametric function of (s^{FOMC}_k, $s^{\text{Member},i}_k$) and consider the three different functional forms of the weight, drawing on the literature from the voting model.[8]

1. **Equal weight model**

$$w_{i,k} = \frac{1}{N}. \tag{7}$$

[8]Our terminology of the aggregation regime reminds the classification in Riboni and Ruge-Murcia (2010). However, the definition is different for two reasons. First, they characterize policy preferences in terms of a change from the status quo. Second, while their simple majority set-up is the same as the median voter model in our study, our equal weight regime is different from their supermajority consensus model. Also, our chair-centric model is similar to their chair-dominance model in spirit but the exact characterization is different because they explicitly constrain the chair's agenda-setting power by the median voter's preference. We sketch out the extension of their agenda-setting chair model in the appendix in which the Chair's proposal is constrained by a pivotal member under the supermajority rule.

2. **Median voter model**

$$w_{i,k} = \begin{cases} 1\,, \text{ if } s_k^{\text{Member},i} = Q_{0.5}(s_k^{\text{Member}}), \\ 0\,, \text{ otherwise.} \end{cases} \tag{8}$$

3. **chair-centric model**

$$w_{i,k}(\theta) = \frac{1 - \frac{(s_k^{\text{Member},i} - s_k^{\text{Chair}})^2}{\max\limits_l (s_k^{\text{Member},l} - s_k^{\text{Chair}})^2}}{N - \frac{\sum_j (s_k^{\text{Member},j} - s_k^{\text{Chair}})^2}{\max\limits_l (s_k^{\text{Member},l} - s_k^{\text{Chair}})^2}}. \tag{9}$$

$Q_\tau(s_k^{\text{Member}})$ represents the τ-th quantile of members' sentiment scores for the k-th topic. The chair-centric model can be defined if there is at least one member whose topic-sentiment scores is different from the Chair. In practice, this condition is always satisfied given the heterogeneity of views among FOMC participants. While weights conditional on the aggregation regime are deterministic because we assume parameters governing the aggregation function are known, we do not know which aggregation regime is supported by the data. Our approach treats the aggregation regime as drawn from a mixture of the above three regimes, in which the mixing probability can be estimated by topic sentiment scores from statements and transcripts. Let's index the aggregation regime by $r = 1, 2, 3$ and represent the topic sentiment score from the statement as the sum of

$$\text{Loss}(s_k^{\text{FOMC}}, s_k^{\text{Member},1}, \ldots, s_k^{\text{Member},M}) = \sum_r P_r (s_k^{\text{FOMC}} - \sum_i w_{i,k}^r s_k^{\text{Member},i})^2\,, \sum_r P_r = 1. \tag{10}$$

To minimize the above loss function, we can select the regime that provides the lowest quadratic distance between s_k^{FOMC} and $\sum_i w_{i,k}^r s_k^{\text{Member},i}$. While our method does not exclude the possibility that multiple regimes might have the same loss function value, in our application, this does not happen. For every FOMC statement topic sentiment score, we are able to select the single best-fitting aggregation regime.[9]

Our method provides a more flexible framework for aggregating FOMC members' preferences (topic sentiment scores at the individual FOMC member level) to the committee's preferences (topic sentiment scores at the statement level). For example, Shapiro and Wilson (2022) measure the sentiment of an individual FOMC member by the person's use of negative words relative to positive words based on the dictionary created by Loughran and McDonald (2011). Their sentiment measure is the net negativity from the overall speech by each individual member, which is assumed to be proportional to the person's loss function value with some measurement error. Let's derive the k-th topic sentiment score as Shapiro and Wilson (2022) do.

[9]Under the current assumption, the preferred aggregation regime has no persistence.

$$s_{i,k}^{SW} = \frac{c_{k,n}^{\text{Member},i} - c_{k,p}^{\text{Member},i}}{c_{k,n}^{\text{Member},i} + c_{k,p}^{\text{Member},i}}, \tag{11}$$

$$\propto L_{i,k}, \tag{12}$$

where $c_{k,n}^{\text{Member},i}$ and $c_{k,p}^{\text{Member},i}$ are the frequency of negative and positive words associated with the k-th topic from the i-th member's remarks at the transcript. The net negativity is assumed to be proportional to the member's loss function of the k-th topic. Suppose that the k-th topic is about inflation. Then we can represent the committee's sentiment score for inflation topic as follows based on Shapiro and Wilson (2022):[10]

$$s_{\pi,t}^{\text{FOMC}} = \sum_i s_{\pi,t}^{\text{Member},i} = \sum_i (\alpha_i + \delta \pi_t^2 + \theta \pi_t). \tag{13}$$

Compared to our framework, they impose two substantial restrictions. First, each member's topic sentiment score receives an equal weight because the committee's topic sentiment score is a simple sum of individual members' topic sentiment scores. Second, after controlling for individual fixed effects, each member's sentiment responds to realized inflation in the same quadratic way. In other words, δ and θ do not vary across individuals. Both assumptions can be quite restrictive. Our framework does not depend on these assumptions because topic-sentiment scores depend on embeddings of remarks, which take into account not only responses to the first and the second moments of realized inflation but also responses to other indicators such as survey data on inflation expectations. By identifying topic sentiment scores from linking a topic model with a large language model for sentiment analysis instead of quantifying the topic sentiment by regressing the overall net negativity in FOMC members' remarks on realized inflation data, our framework allows a more generalized loss function. Also, we decide the preferred aggregation regime based on the topic-sentiment scores contained in the released FOMC statement instead of assuming the equal weight regime for all the FOMC meetings. As a simple check for the validity of the second restriction in Shapiro and Wilson (2022), we calculate the correlation of the time series of sentiment scores on inflation topic of two members (Former Chair Bernanke and Former Governor Kohn), which turns out to be small (around 0.13). Hence, the common variation of inflation sentiment scores across different members is unlikely to hold in our data. This finding suggests that we need to consider a more flexible framework that can take into account a rich pattern of heterogeneity in individual members' preferences.

While we abstract time dimension in the above description of our methodology, we apply our method for documents at each meeting, creating a time series of probability distribution over the aggregation regime. In the subsequent empirical analysis, we estimate topic sentiment scores for FOMC statement and transcript data from March

[10] While $c_{k,n}^{\text{Member},i}$ and $c_{k,p}^{\text{Member},i}$ are indexed by topic k, Shapiro and Wilson (2022) do not construct the net-negativity measure in a topic-specific way. They construct the overall net negative measure and identify a portion of their net negativity measure associated with their topic of interest, which is inflation. One issue with their inflation sentiment measure as pointed out by Ahn, Cook, Doh, Kastrisis and Wedewer (2025) is that most of the time variation in the negativity is explained by non-inflationary control variables in the regression, which complicates the interpretation of "inflation" sentiment.

2004 to December 2017.

2.3 Isolating the Surprise Component in the Statement Topic Sentiment Score

Topic sentiment scores in the statement can provide information on how policymakers interpret macroeconomic conditions. But they do not provide information on what was expected by market participants before announcements, making it difficult to link them with financial market responses to announcements through which monetary policy transmits to the real economy. To isolate surprise components in statement topic sentiment scores, we construct market expectations for statement topic sentiment scores using proprietary data on "proxy" statements written by a consulting firm and circulated among paid subscribers including major investment banks a few days before FOMC meetings.[11] By applying the topic model estimated from the FOMC statement data to Mr. Meyer's proxy statements, we construct market expectations of topic sentiment scores and compute the surprise component as the difference between them.

$$s_k^{\text{Surprise}} = s_k^{\text{FOMC}} - s_k^{\text{Meyer}}. \tag{14}$$

Let y_t be a financial market response during the 30 minute window around a FOMC announcement. We run the following regression of y_t to surprise components in statement topic sentiment scores.

$$y_t = \alpha + \sum_{k=1}^{K} \beta_k s_{k,t}^{\text{Surprise}} + \epsilon_t. \tag{15}$$

3 Data

Our data for FOMC statements start from March 2004 and end in September 2024 while the transcript data covers the period from March 2004 to December 2017.[12] Our sample period covers tenures of thre different Federal Reserve chairs, including the last two years of Greenspan's tenure (March 2004-Jan 2006), the whole tenure of Bernanke (Jan 2006- Jan 2014), and the four years of Yellen's tenure (Jan 2014 - Dec 2017) except for her last meeting as the chair (Jan 2018).

We tokenize words in statement and transcript data and rule out some stop words not relevant for topics of our interest as well as several common contractions such as "Mr." and "Mrs.".[13] Figure 4(a) shows frequently appearing words in FOMC statement when we do not clean up irrelevant words for our purpose. Economic concepts like

[11] Laurence Meyer who was a governor at the Federal Reserve Board from 1996 to 2002 wrote newsletters containing his musings about statement language. We use data from August 2008 to December 2019 because Mr. Meyer started to provide his version of statement language from August 2008 and we have access to high-frequency (intra-daily) financial market data through Bauer and Swanson (2023), who share the data up to December 2019.

[12] The starting point for the statement data follows Doh et al. (2022). While we do not currently use alternative versions of statement prepared by the staff of the Federal Reserve Board, which started from March 2004, we plan to use information from alternative versions of statement in the future to refine our sentiment measure.

[13] We use the same list of stop words in Shapiro and Wilson (2022).

Figure 4: Word Cloud from FOMC Statement Data



(a) Unfiltered



(b) Filtered

Notes: The unfiltered version uses all the vocabularies in the statement while the filter version rules out stop words and numbers. The size of a word corresponds the frequency of the word in the text.

"prices" and "job" appear less frequently than more procedural words like "Committee" and "intermeeting". In contrast, once we clean up the vocabularies, "prices" appear as the most frequently used word shown in Figure 4(b).

For sentiment analysis, we apply the SIEBERT at a sentence level for FOMC statement data while we do that at a remark level from transcript data. Sometimes, FOMC members make long remarks that consist of many sentences. The maximum number of

tokens in one embedding is 512 in SIEBERT. We truncate long remarks whose token size exceeds 512 into several sub-remarks with the token size of less than 512. In total, we obtain topic sentiment scores for 2465 sentences from statement data and 15692 remarks from transcript data. Some of economic concepts like "federal funds rate" or "price stability" contain multiple words. To treat them as a single token, we document a list of economic terms in Ahn et al. (2025) that consist of multiple words and tokenize them as if single words. These terms are created by the human reading of FOMC transcripts.

For transcript data, we consider remarks made by only voting FOMC members and filter out remarks made by the staff of the Federal Reserve System and non-voting FOMC members. Since our focus is on the aggregation regime of individual policymakers' preferences expressed in remarks into the committee's preferences expressed in statements, it would be natural to focus on those who can have practical influences on the outcome, which is statement language.

One technical issue in using the SIEBERT to extract the sentiment of a long remark in the transcript is the truncation of the remark at the fixed location. The BERT model generates an embedding representation of text data only up to 512 tokens. If the text data is longer than that, we truncate the remark into subremarks and calculate sentiments of these subremarks. However, there is no guarantee that the fixed location of truncation reflects the coherency of subremarks because it can truncate the long sentence in the middle. To examine how much the truncation can contaminate our sentiment scores, we employ a generative AI model (Google's Gemma 3) to truncate remarks at the end of a complete thought or argument.[14] While the correlation of individual topic sentiments for the transcripts derived from the default length-based segmentation and the segmentation based on the Gemma 3 model is weakly correlated on average with the correlation of 0.26, the aggregated outcomes are similar for both methods, suggesting that although the truncation error may be important to find out the exact language that can explain an individual member's topic sentiment score at a given meeting, it does not materially affect each member's topic sentiment score aggregated across multiple subremarks.

4 Empirical Analysis

4.1 Topic Analysis from Statement Language

One of key hyperparameter in a topic model is the number of topics. We set the number of topics to 6. We mainly select this number in terms of topic coherence and the sufficient coverage of different topics.[15] Except for the pre-processing of words to be used as inputs, a topic model does not include any restriction based on the economic theory and generates a bundle of words purely based on the co-frequency patterns. Hence, it is often difficult to obtain the economic interpretation of selected topics. We

[14] We provide the detailed description of our procedure as well as the exact prompt language used in the appendix.

[15] While one can use a quantitative measure such the perplexity score capturing the predictive power of a topic model for the next word token, we do not follow this method because our main purpose is to obtain a comprehensive list of internally coherent topics. The internal coherency metric is not necessarily tied to the predictive power of the next word token.

address this issue by picking up the exemplary sentence that can best represent each topic based on the topic probability score of sentences instead of listing a few frequently used words in each topic. For instance, the best sentence that represents the k-th topic, $d^{k,\star}$, is selected by maximizing the k-th topic probability, $\theta_{d,k}$ among all the sentences in the data.

$$d^{k,\star} = \underset{d}{\text{argmax}}\ \theta_{d,k}. \tag{16}$$

Table 1 illustrates these exemplary sentences for six topics from our model. Four out of six topics are associated with describing economic and financial conditions at various levels while the other two topics discuss monetary policies. Interestingly, the topic model separates out topics on the current economic conditions (topic 2, 5, and 6) from the discussion on the future outlook (topic 3) although we do not force the model to recognize this difference. In addition, the model also distinguishes the discussion on policy guidance (topic 1) from that on policy actions (topic 4).[16] Our results suggest that analyzing the co-frequency pattern of words alone can provide sensible classifications of topics discussed in FOMC statements. It is often challenging to obtain such a clean classification result from the topic analysis of a large number of unstructured textual data but we benefit from the fact that FOMC statements are more structured documents incorporating the long deliberation on the appropriate policy path given economic and financial conditions.

Figure 5 shows the average topic distribution for both FOMC statements and transcripts. One notable difference between transcript data and FOMC data is that there are more frequent discussions on inflation and real economic activities but less frequent discussion on outlook risks in transcript data. In contrast, FOMC statements tend to mention risks to the outlook most frequently. The difference can be due to the fact that Reserve Bank presidents often cite their contacts with district business people to report on economic conditions during FOMC meetings, in which words describing inflation and real economic activities are likely to be prevalent.

This pattern largely holds even if we consider sub-sample periods by different chairs of the Federal Reserve as shown in Figure 6. However, there are some variations across different chairs. During the Greenspan period in our sample, statement language mentioned risks to outlook relatively less frequently than in transcript data. In contrast, discussions on outlook risks in statements increased subsequently during the tenure of Bernanke and Yellen. At the same time, real economic activities and inflation, which were prevalent topics during the Greenspan's tenure became relatively less prominent. Overall, the time-variation in topic prevalence suggests that language in FOMC statements became more forward-looking during Bernanke and Yellen's tenure as policymakers relied heavily on forward guidance as the short-term nominal interest rate was constrained by the zero lower bound. Even after the interest rate was lifted off the zero lower bound, the last paragraph in the December 2017 statement provided a lengthy description of conditions on the future adjustment of the policy rate as follows:

"In determining the timing and size of future adjustments to the target range

[16] While the best example for topic 4 mentions asset purchases, it can vary over time. During the Greenspan's tenure, the Federal Reserve did not engage in asset purchases and topic 4 includes actions on interest rate changes.

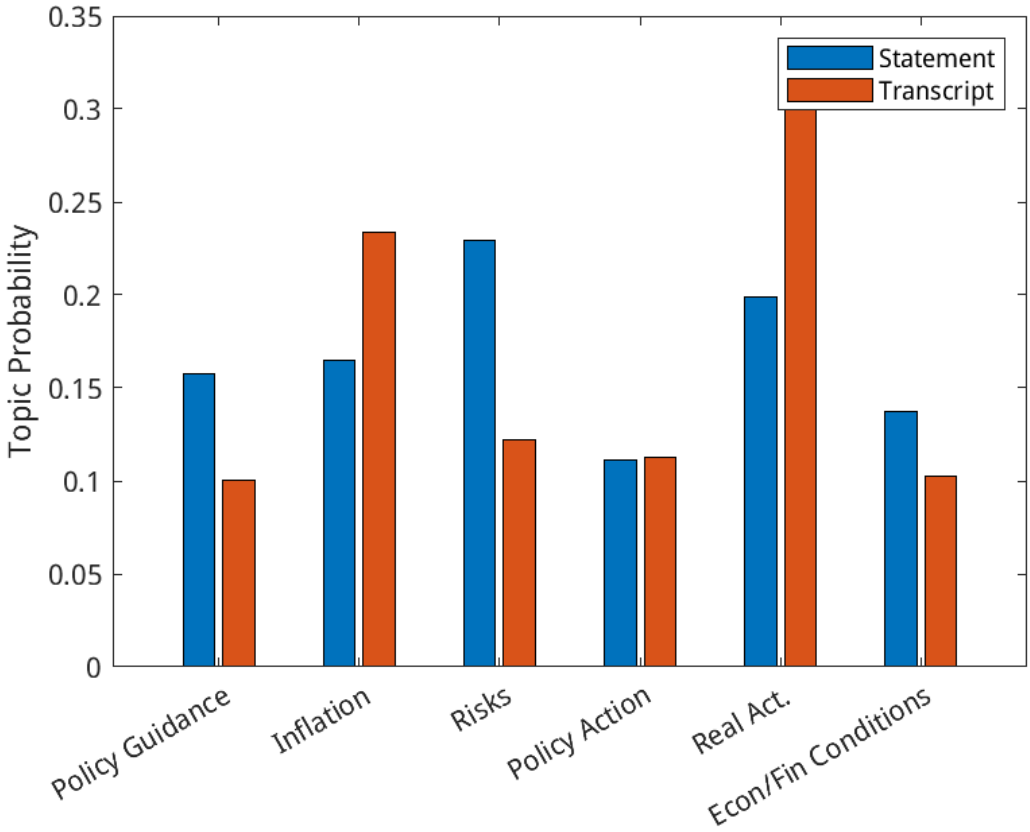
Table 1: Example Sentences for Topics

Topic	Example Sentence
1. Policy Guidance:	"In determining how long to maintain this target range, the Committee will assess progress-both realized and expected-toward its objectives of maximum employment and 2 percent inflation."
2. Inflation Conditions:	"The rise in energy prices, however, has not notably fed through to core consumer prices."
3. Outlook Risks:	"The economic outlook is uncertain, and the Committee is attentive to the risks to both sides of its dual mandate."
4. Policy Actions:	"The Committee is maintaining its existing policies of reinvesting principal payments from its holdings of agency debt and agency mortgage-backed securities in agency mortgage-backed securities and of rolling over maturing Treasury securities at auction."
5. Real Economic Activity:	'Firms have brought inventory stocks into better alignment with sales."
6. Economic and Financial Conditions:	"This assessment will take into account a wide range of information, including measures of labor market conditions, indicators of inflation pressures and inflation expectations, and readings on financial and international developments."

Notes: We select the example sentence for each topic based on the highest probability for the topic across the entire sentences in our FOMC statement data.

for the federal funds rate, the Committee will assess realized and expected economic conditions relative to its objectives of maximum employment and 2 percent inflation. This assessment will take into account a wide range of information, including measures of labor market conditions, indicators of inflation pressures and inflation expectations, and readings on financial and international developments. The Committee will carefully monitor actual and expected inflation developments relative to its symmetric inflation goal. The Committee expects that economic conditions will evolve in a manner that will warrant gradual increases in the federal funds rate; the federal funds rate is likely to remain, for some time, below levels that are expected to prevail in the longer run. However, the actual path of the federal funds rate will depend on the economic outlook as informed by incoming data."

Figure 5: Topic Distribution: Entire Sample Period March 2004- September 2024



Notes: Real Act. and Econ/Fin Conditions represent "Real Economic Activity" and "Economic and Financial Conditions" described in Table 1.

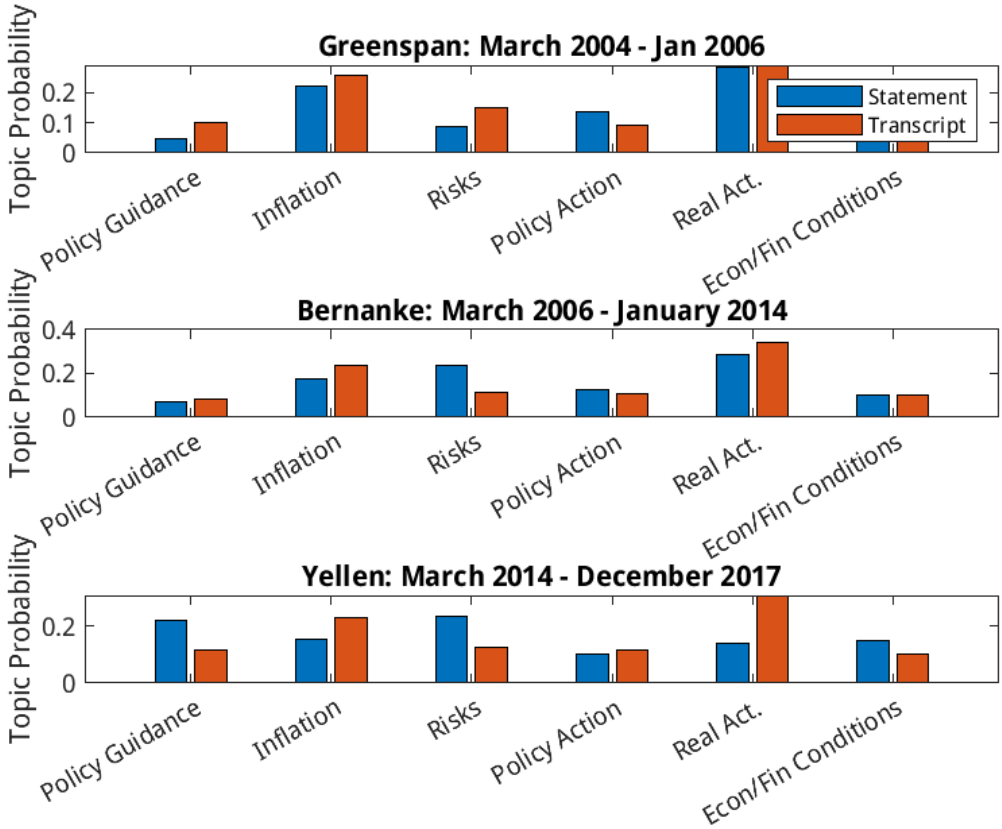
4.2 Combining Sentiment Analysis with a Topic Model

The topic model is useful for uncovering the structure of a large size of textual data. But it does not tell us about the directional tone of sentences or remarks in our data, which is crucial for gauging monetary policy stance. For instance, we may be more interested in whether or not policymakers talk positively or negatively about economic conditions rather than more or less frequently. To quantify the tone of our data, we apply the SIEBERT described in the previous section to our input text data.

Our dataset was not used in training the SIEBERT and there is a possibility that sentiment scores from the SIEBERT may not make sense for our purpose. We check sentiment scores for the SIEBERT matches the human assessment of statement language. For example, a positive sentiment score on policy guidance is associated with keeping the low level of the federal funds rate longer in the following sentence:

> "To support continued progress toward maximum employment and price stability, the Committee expects that a highly accommodative stance of monetary policy will remain appropriate for a considerable time after the asset purchase program ends and the economic recovery strengthens."

Figure 6: Topic Distribution: Subsample Period by Chairs



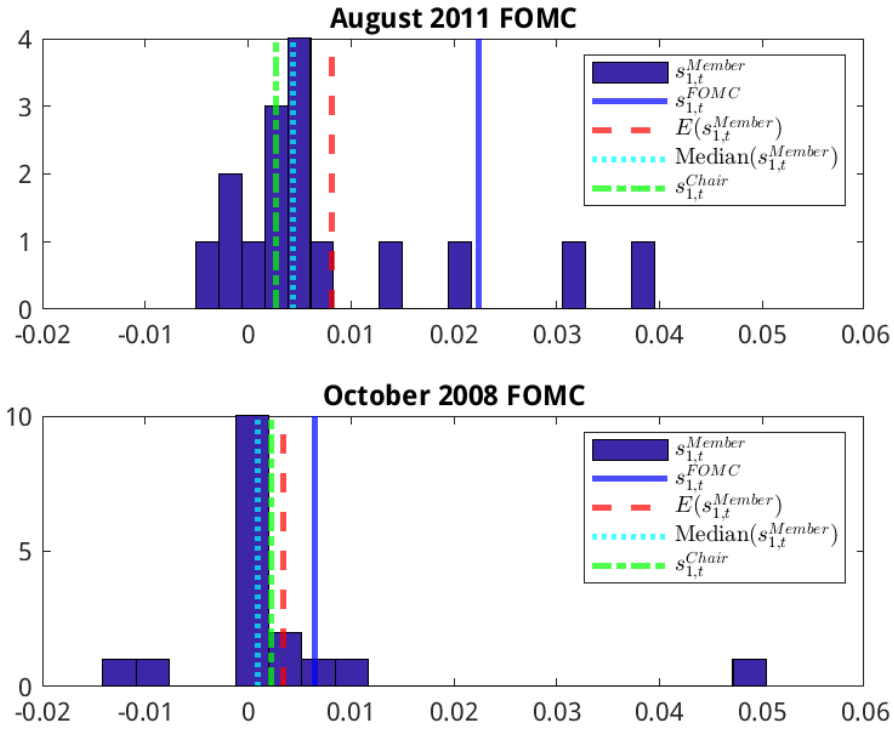
Notes: Real Act. and Econ/Fin Conditions represent "Real Economic Activity" and "Economic and Financial Conditions" described in Table 1.

While topic sentiment scores are generally aligned with the human reading of sentiments in statement language, using information from FOMC statements does not reveal how the committee comes to adopt particular wordings. To analyze it, we need to explore transcript data in which individual members express their opinions about statement language and bargain over statement language in some cases as our quote from the August, 2011 meeting transcript suggests.

4.3 Statement Language through the Aggregation of Individual Policymakers' Preferred Language

A standard model of monetary policy typically assumes a single decision maker. In this case, the systematic response of policy to economic conditions is often constant. However, if we take the committee decision making process seriously, the committee's reaction function is determined by the aggregation of members' preferences. Since the committee's policy stance is an outcome of the negotiation among voting members, we need to look into the distribution of sentiment scores of them. Focusing on the single pivotal member's preference may not fit sentiment scores expressed in FOMC statements all the time. Even the Chair's proposal with the support of simple majority can be modified in the middle of policy deliberations if it faces the intense dissents from

Figure 7: Sentiment Scores on Policy Guidance Topic



Notes: FOMC stands for statement data. In the August 2011 FOMC meeting, two members expressing most positive sentiment scores were Charles Evans (president of the Federal Reserve Bank of Chicago) and Elizabeth Duke (governor at the Federal Reserve Board). In the October 2008 FOMC meeting, the member who expressed the most positive sentiment score was Sandra Pianalto (president of the Federal Reserve Bank of Cleveland).

a group of members as we saw in the except of the August 2011 meeting transcript.

Figure 7 illustrates this point by contrasting sentiment scores from the October 2008 FOMC statement on the interest rate policy topic with the corresponding one from the August 2011 FOMC statement. In both cases, we plot the histogram of individual topic sentiment scores based on transcript data. The chair's sentiment score from transcript data is not much different for both meetings. However, the FOMC statement sentiment score is much lower in the October 2008 meeting. Since the mean transcript sentiment score is closer to the FOMC statement sentiment score than the median transcript sentiment score or the Chair's sentiment score, the preferred aggregation regime at both meetings is the equal weight regime. There are more outliers in the August 2011 meeting and they appeared to play a larger role in determining the statement topic sentiment score. As the except in the introduction suggests, Charles Evans bargained hard to push the FOMC's position to his preferred one. He is one of outlier members who expressed a large positive sentiment score for the discussion related to policy guidance at the meeting. On the opposite side, Jeffrey Lacker (then the president of the Federal Reserve Bank of Richmond) has a negative sentiment score for the same topic at the same time as remarks in Table 2 illustrate.

Table 2: Remarks on the Policy Guidance Topic at the August 2011 FOMC Meeting

Member	**Remark**	**Sentiment Score**
Evans	"Although it might be standard to say that another meeting's worth of data might resolve uncertainty the most recent NIPA revisions have answered enough of these questions for me,Without substantial clarification of our policy intentions in line with our dual mandate responsibilities I think there's a distinctly high probability that we will continually revisit these periods of anemic growth without ever escaping the current malaise,Consequently I favor a strong indication today of additional accommodation,I think alternative A is certainly a plausible demonstration of that,I am quite attracted to the suggestion that you're making today and to Governor Yellen's proposal,Governor Yellen says that she would entertain this accommodative monetary policy to better center our inflation forecast at 2 percent which is our objective,I believe clarifying our policy framework in order to better achieve our dual mandate responsibilities would help provide further policy accommodation at least along the lines suggested in Eggertsson and Woodford and this helps by defining that as extending through mid 2013,Conducting monetary policy in accordance with our dual mandate should mean that as long as high unemployment is generating an extraordinary policy loss the FOMC is willing to provide very large amounts of accommodation and tolerate medium term inflation substantially above 2 percent. I want to thank President Hoenig for listening to me earlier and I certainly take your point,As I mentioned earlier a 9 percent unemployment rate against a very conservative natural rate of 6 percent generates a policy loss that is as large as if inflation were running at 5 percent against a 2 percent objective,But my stronger point is this We talk and act as if 2 percent is an almost unacceptable inflation rate and a ceiling that limits further policy accommodation,A couple of examples First the common view seems to be that we will consider further LSAPs and more only if the risk of deflation is significant,I'm glad there's a floor but the ceiling is too low,Second in an interview the other day."	0.9971
Lacker	"the work I described has bolstered my belief that unemployment disappointingly high and economic growth is disappointingly low for reasons that are related to economic fundamentals rather than insufficient monetary policy stimulus–or insufficient fiscal policy stimulus for that matter."	-0.9635

Notes: Sentiment score for the entire remark of Evans and the subremark of Lacker.

Table 3 shows the preferred aggregation regime by topics and chairs. Overall, the chair-centric regime is preferred in most cases, followed by the equal weight regime, in

Table 3: The Preferred Aggregation Regime Probability by Chairs

Period	**Equal Weight**	**Median**	**Chair Centric**
Greenspan (16 Statements)	0.25	0.18	0.57
Bernanke (64 Statements)	0.21	0.25	0.54
Yellen (31 Statements)	0.14	0.04	0.82
Sum (111 Statements)	0.2	0.18	0.62

Table 4: The Preferred Aggregation Regime Probability by Topics

Topic	**Equal Weight**	**Median**	**Chair Centric**
Policy Guidance	0.27	0.28	0.45
Inflation Conditions	0.08	0.05	0.87
Outlook Risks	0.26	0.21	0.53
Policy Actions	0.12	0.14	0.74
Real Economic Activity	0.33	0.23	0.44
Economic and Financial Conditions	0.12	0.16	0.72

Notes: Topic labels correspond to those in Table 1.

line with Riboni and Ruge-Murcia (2020)'s finding on the agenda-setting power of the chair. Interestingly, the preferred aggregation regime varies significantly across topics as shown in Table 4. the chair-centric regime is preferred less than 50% in topics on real economic activity and interest rate policy. The finding may not be surprising because the description of real economic activity in statement is likely to aggregate information from Reserve Bank presidents' views on district economic conditions. The variation in the preferred aggregation regime over time and across topics implies that the statement sentiment score may shift a lot even if the underlying distribution of individual members' sentiment scores does not.[17]

4.4 High-frequency Financial Market Responses and Surprises in Statement Topic Sentiment Scores

We estimate equation (15) to uncover how financial markets interpret topic sentiment scores in the statement. We use high-frequency bond market responses around FOMC announcements as dependent variables. Table 5 shows regression results for the three measures considered in Bauer and Swanson (2023). The unorthogonalized one is the first principal component (PC) of the changes in the first four quarterly Euro-dollar (ED)

[17]As noted by Ramey (2016), a change in the weight for an FOMC member can be a source of a monetary policy shock.

Table 5: High-frequency Bond Market Responses to Surprises in Topic Sentiment Scores

	BS (Unorthogonalized)	BS (Orthogonalized)	BS (Response to News)
β_1 Policy Guidance	-0.081	-0.187*	0.106*
β_2 Inflation	0.003	-0.054	0.057*
β_3 Outlook Risk	0.041	-0.016	0.057**
β_4 Policy Actions	0.028	0.008	0.020
β_5 Real Activity	-0.007	-0.028	0.021
β_6 Econ and Fin Conditions	0.016	-0.050	0.066**
R^2	0.021	0.092	0.228

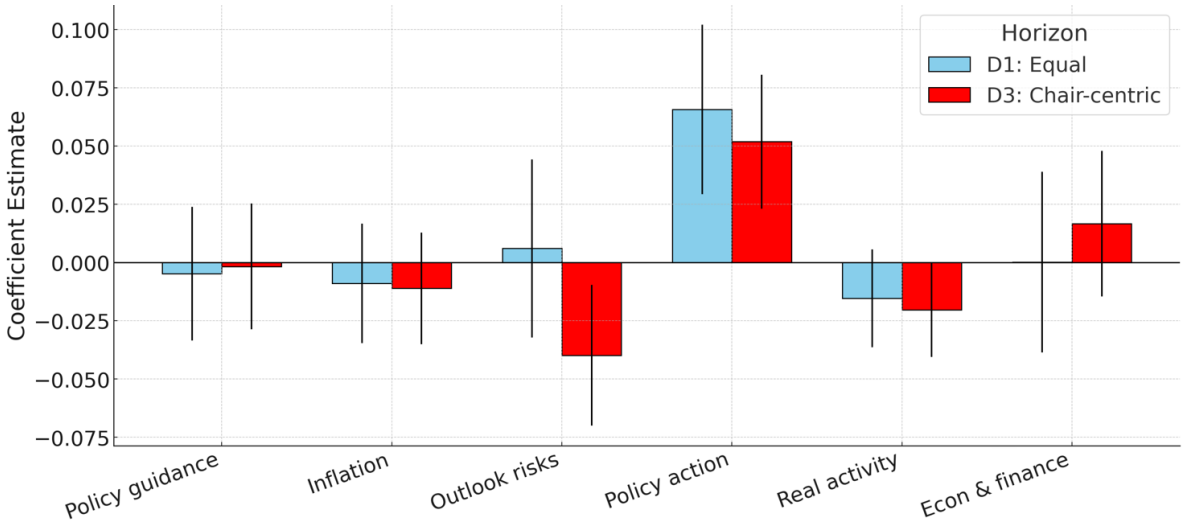
Notes: BS (Unorthogonalized) is the measures of high frequency bond market responses around FOMC announcements in Bauer and Swanson (2023). It can be decomposed into the component predictable by macroeconomic and financial variables before announcements (News) and the component orthogonal to news (Orthognalized). * (**)represents the statistical significance at 10% (5%) level.

futures contracts during 30-minute windows around FOMC announcements, which can be decomposed into the orthogonalized component that cannot be predictable by six macro and financial variables during the inter-meeting period and the news component that is predictable by these variables. According to Bauer and Swanson (2023), the news component may reflect the financial market's imperfect knowledge of the Federal Reserve's reaction function to macroeconomic news while the orthogonalized one may capture the "true" exogenous shock to monetary policy.

We do not find evidence that topic sentiment scores play any statistically significant role in explaining the pure change in ED futures contracts. In addition, for the orthogonalized one, the sentiment score on policy guidance topic plays a statistically significant role. The sign of the regression coefficient (β_1) is negative because a surprising positive sentiment on this topic is associated with using a more accommodative policy to achieve policy goals, which leads to a decline in the interest rate from ED contracts on average.[18]

[18]We should be cautious in interpreting the result because sentiment scores on this topic are more tightly related to the stance on the effectiveness of policy guidance to achieve dual mandate. Hence, guidance on policy tightening in response to stronger-than expected inter-meeting data can have a positive sentiment too if that is perceived as more effective for stabilizing inflation. In fact, β_1 has a positive coefficient to the "news" component in Bauer and Swanson (2023). Nonetheless, as the quote in Section 4.2 illustrates, the positive sentiment on this topic is likely to be associated with the surprising policy easing in our sample period because aggressive policy easing was often advocated as being effective in achieving the dual mandate.

Figure 8: Effects of the aggregation regimes on monetary policy shocks



Notes: Bars depict the estimated coefficients, and vertical lines denote the 95% confidence intervals around those estimates.

4.5 Effects of aggregation regimes on monetary policy shocks

We find a significant time variation of the aggregation regime, which may have a systematic relationship with monetary policy shocks. To examine this channel, we look at how the dominant aggregation regime within the FOMC affects the orthogonalized monetary policy shock. For this, we consider the following regression model:

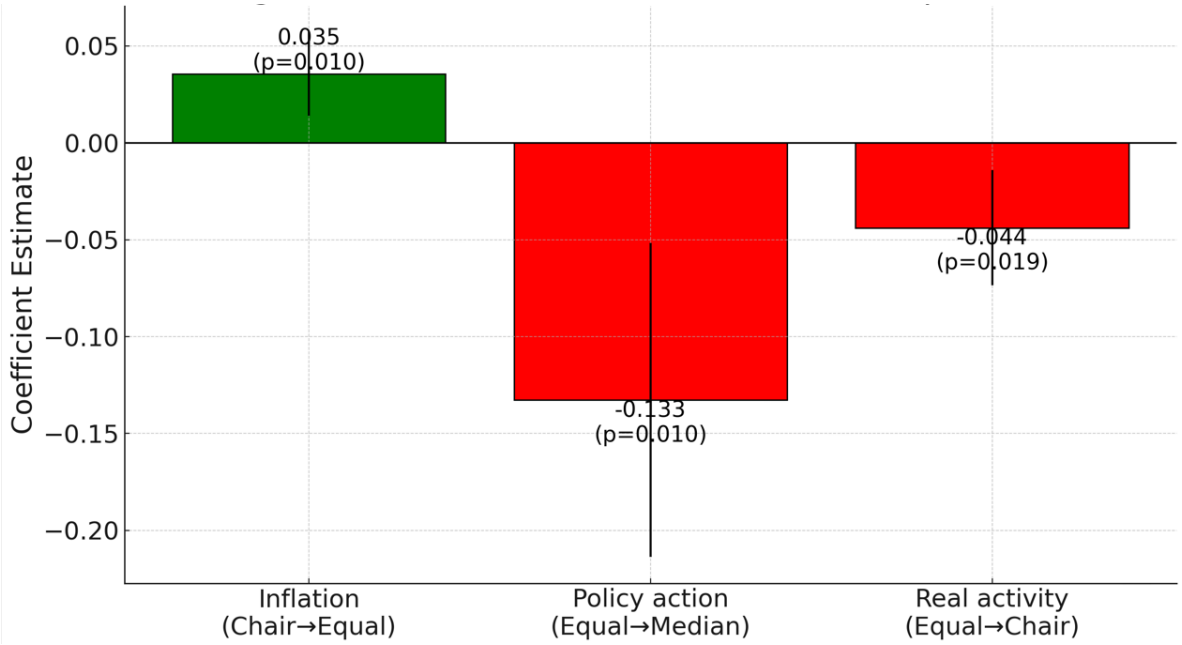
$$y_t^{\text{orth}} = \alpha + \sum_k D_k^1 R_{k,t}^{\text{Equal}} + \sum_k D_k^3 R_{k,t}^{\text{Chair}} + \varepsilon_t, \tag{17}$$

where y_t^{orth} is Bauer-Swanson's orthogonalized monetary policy shock, and $R_{k,t}^{\text{Equal}}$ and $R_{k,t}^{\text{Chair}}$ are regime dummies capturing the equal-weight and chair-centric aggregation modes for topic k, respectively. Note that the median aggregation regime is the base category.

Figure 8 reports the estimated coefficients D_k^1 and D_k^3. Both the equal-weight and chair-centric regimes for policy action are positively correlated with the orthogonalized monetary policy shock. This finding indicates that hawkish voting members tended to be outliers in the sentiment distribution, and that these two regimes capture the preferences of such hawkish members more strongly than the median-voter regime. In contrast, the chair-centric regime for outlook-risk topics is negatively associated with the monetary policy shock, suggesting that when the Chair expressed heightened concern about downside or output risks, the resulting policy stance tended to be more expansionary. The rest of the estimates are not statistically significant.

Next, we focus on changes in the aggregation regimes—specifically, shifts among the equal-weight, median-voter, and chair-centric structures and their effects on monetary

Figure 9: Effects of changes in the aggregation regimes on monetary policy shocks



Notes: Bars depict the estimated coefficients, and vertical lines denote the 95% confidence intervals around those estimates.

policy shocks. The regression specification is modified as follows:

$$y_t^{\text{orth}} = \alpha + \sum_k \sum_{p,q \in e,m,c} D_k^{p \cdot q} R_{k,t}^{p \cdot q} + \varepsilon_t, \tag{18}$$

where $R_{k,t}^{p \cdot q}$ denotes a transition from aggregation regime p to regime q for topic k (equal, median, or chair), and $D_k^{p \cdot q}$ represents the estimated coefficient associated with that transition.

Figure 9 shows that transitions out of the equal-weight regime are associated with expansionary market responses, particularly for *policy-action* and *real-activity* topics. Because the equal-weight regime places relatively greater weight on hawkish outliers in the sentiment distribution, departures from this regime tend to be interpreted as more accommodative. Conversely, transitions into the equal-weight regime from the chair-centric regime—especially for the *inflation* topic—are associated with monetary tightening, as indicated by the positive coefficient estimates.

Taken together, these results suggest that changes in the internal aggregation regime themselves can constitute an independent source of monetary-policy shocks, distinct from conventional target-rate or forward-guidance surprises.

5 Discussion

Our empirical analysis documents frequent shifts in the pivotal member in the committee's decision. Although the Chair and the median voting member appear to be pivotal in most cases, the equal weight regime also receives a non-negligible share in

the preferred aggregation regime. Riboni and Ruge-Murcia (2020) and Riboni and Ruge-Murcia (2025) describe a model in which the agenda-setting power of the Federal Reserve chair is constrained by committee members' preferences. Although this model can explain the dominant role played by the Chair and/or the median voting member, it does not explain the August 2011 FOMC meeting episode that a leading member in the dovish member played a pivotal role in modifying the Chair's baseline proposal for the FOMC statement. In fact, the FOMC appears to favor the consensus and tries to minimize dissents if possible. With the median voter unchanged, having one more dissent should not matter much to the FOMC's decision if the median voter is the pivotal member. But the FOMC can shift the position on statement language to minimize dissents when members have diverse views. Such a consensus driven decision model suggests that the committee decision is actually close to a supermajority rule than a simple majority rule. We describe an extended version of an agenda-setting Chair model under a supermajority rule in Riboni and Ruge-Murcia (2010) with multiple alternatives and imperfect information of the Chair about members' preferences before the meeting, which can explain the August 2011 episode as well as the FOMC's consensus-driven culture.

5.1 A model of the agenda-setting committee Chair

We first recapitulate the original agenda-setting Chair model in Riboni and Ruge-Murcia (2010) and subsequently present the extended version with multiple alternative and the Chair's imperfect information on members' preferences. The Federal Reserve Chair plays a special role in leading the FOMC in the sense that he or she sets the agenda and try to build the consensus view of the entire committee (Goodfriend (1999)). Riboni and Ruge-Murcia (2010) introduce multiple voting models that recognize the Chair's distinct role. In their model, only the Chair can propose an alternative to the status quo that can be represented by the outcome decided in the previous meeting, whose adoption is determined by votes. The voting rule can be a simple majority where the median voter's preference can constrain the Chair's proposal or a supermajority rule in which both hawkish or dovish members can be pivotal members. We assume that there are n (an odd number) members. Except for the Chair who considers the constraint by the pivotal voter's preference, everyone votes sincerely for his or her preferred interest rate. Members care about the deviation of inflation π_t from the target $\pi^\star$ but have different degrees of inflation aversion captured by μ_i. The economic environment described by the backward-looking Phillips curve and the Euler equation as in Svensson (1997), which determine the dynamics of inflation π_t and output y_t.

- Member Preferences

$$\begin{aligned}
\text{Member j's objective} &: \max E_\tau\Big(\sum_{t\geq\tau}\delta^{t-\tau}U_j(\pi_t)\Big),\\
U_j(\pi_t) &= \frac{-e^{(\mu_j(\pi_t-\pi^\star)}+\mu_j(\pi_t-\pi^\star)+1}{\mu_j^2},\\
\text{Preference Ranking} &: \mu_1\leq\cdots\leq\mu_n,\\
\text{Median Member} &: M=\frac{n+1}{2}\to\mu_M=0.
\end{aligned}$$

- Economic Environment

$$\begin{aligned}\pi_{t+1} &= \pi_t + \alpha_1 y_t + \epsilon_{t+1},\\ y_{t+1} &= \beta_1 y_t - \beta_2(i_t - \pi_t - i) + \eta_{t+1},\\ \begin{bmatrix}\epsilon_t\\ \eta_t\end{bmatrix} &= \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}u_t\\ v_t\end{bmatrix} + \begin{bmatrix}\gamma & 0\\ 0 & \zeta\end{bmatrix}\begin{bmatrix}u_{t-1}\\ v_{t-1}\end{bmatrix}\end{aligned}$$

where γ and ζ are inside $(-1, 1)$ and u_t and v_t follow mean-zero white noise processes with standard deviations of σ_u and σ_v, respectively. By combining two equations, we can solve for the dynamics of inflation as follows:

$$\pi_{t+2} = (1 + \alpha_1\beta_2)\pi_t + \alpha_1(1 + \beta_1)y_t + \alpha_1\beta_2 i - \alpha_1\beta_2 i_t + \epsilon_{t+1} + \alpha_1\eta_{t+1} + \epsilon_{t+2}.$$

- Individual Members' Preferred Policies under Discretion (No Commitment for the Future Policy Rate)

$$\begin{aligned} i^\star_{j,t} &= \underset{i_t}{argmax}\ \delta^2 E_t U_j(\pi_{t+2}),\\ \text{First Order Condition}: E_t(e^{\mu_j(\pi_{t+2}-\pi^\star)}) &= 1.\end{aligned}$$

When innovations are normally distributed, π_{t+2} given the information set at time t is also normally distributed with the variance of σ^2_π, making the FOC equal to $E_t\pi_{t+2} = \pi^\star - \mu_j\frac{\sigma^2_\pi}{2}$. After taking the conditional expectations as of time t, member j's preferred interest rate is given by

$$i^\star_{j,t} = a_j + b\pi_t + cy_t + \xi_t.$$

Under our assumption of preference ranking, the preferred interest rate is also ranked as follows:

$$i^\star_{1,t} \le \cdots \le i^\star_{n,t}.$$

where ξ_t is a linear combination of u_t and v_t.

Note that the preference heterogeneity is only reflected in the intercept term and all the members will react to surprises in inflation and output to the same degree.

Let $s_t = (\pi_t, y_t, \xi_t)$ be a state vector describing macroeconomic conditions. ω_t is the vector of the preferred policy rate as a function of s_t and q_t is the status-quo interest rate. If no alternative has more support than q_t under a policy aggregator $\Psi(s_t, q_t, \omega_t)$, q_t stays. Riboni and Ruge-Murcia (2010) consider multiple voting rules.

1. Majority rule: $k \in [0, \frac{n-1}{2}]$

$$i_t = \Psi(s_t, q_t, \omega_t) = \begin{cases} i^\star_{M+k,t}\,, & \text{if } q_t > i^\star_{M+k,t},\\ q_t\,, & \text{if } i^\star_{M-k,t} \le q_t \le i^\star_{M+k,t},\\ i^\star_{M-k,t}\,, & \text{if } q_t < i^\star_{M-k,t}\end{cases}$$

2. Agenda-setting rule: agenda setter = Chair indexed by c. Only the Chair can propose an alternative to the status quo but voting follows the majority rule where the median voter is pivotal.

$$i_t = \Psi(s_t, q_t, \omega_t) = \begin{cases} i^\star_{c,t}\,, \text{ if } q_t > i^\star_{c,t} > i^\star_{M,t}, \\ q_t\,, \text{ if } i^\star_{M,t} \leq q_t \leq i^\star_{c,t}, \\ 2i^\star_{M,t} - q_t\,, \text{ if } 2i^\star_{M,t} - i^\star_{c,t} \leq q_t < i^\star_{M,t}, \\ i^\star_{c,t}\,, \text{ if } q_t < 2i^\star_{M,t} - i^\star_{c,t}. \end{cases}$$

3. Dictator rule: member d decides outcomes unilaterally.
   $i_t = \Psi(s_t, q_t, \omega_t) = i^\star_{d,t}$.

Riboni and Ruge-Murcia (2014) embeds the agenda-setting Chair under a supermajority rule where $k > 0$. We also assume a supermajority rule in our extension described below. But departing from their model, we assume that the Chair can propose multiple alternative proposals.

5.1.1 Extension with multiple alternatives

In Riboni and Ruge-Murcia (2010), members vote between the Chari's proposal and the status-quo. However, in the actual FOMC deliberation, the Chair typically offers three alternatives written by the staff of the Federal Reserve Board. Let's denote them by $q_{A,t} <= q_{B,t} <= q_{C,t}$ with $q_{B,t}$ the baseline proposal while $q_{A,t}(q_{C,t})$ is a dovish (hawkish) alternative. We can incorporate the consensus-based culture of the Federal Reserve by adopting a supermajority rule in the voting with $k > 0$. In this case, we can list two pivotal members on each side of the preference spectrum. The dovish pivotal member's preferred rate is $i^\star_{\frac{n+1}{2}-k,t}$ while the hawkish one is $i^\star_{M+k,t}$. If none of proposals made by the Chair gets the support of the supermajority, the chair can modify the baseline proposal ($q_{B,t}$) to obtain enough support among members.

- The Centrist Chair: $i^\star_{\frac{n+1}{2}-k,t} \leq i^\star_{c,t} \leq i^\star_{M+k,t}$
  In this case, the Chair's baseline proposal is equal to his/her preferred rate ($q_{B,t} = i_{c,t}$) while $q_{A,t}$ and $q_{C,t}$ are constrained by pivotal voters' preferred rates as follows:

$$i^\star_{1,t} \leq q_{A,t} <= 2i^\star_{\frac{n+1}{2}-k,t} - i^\star_{c,t},$$
$$i^\star_{n,t} \geq q_{C,t} >= 2i^\star_{M+k,t} - i^\star_{c,t}.$$

  $i_t = q_{B,t} = i^\star_{c,t}$ as long as the Chair knows all the members' preferred rate and choose both the baseline proposal and alternatives. However, if $q_{A,t} = i^\star_{1,t}$ or $q_{C,t} = i^\star_{n,t}$ does not satisfy the above constraint, the Chair has to modify $q_{B,t}$ to accommodate the preferred rate of either the dovish (hawkish) pivotal member by setting $q_{B,t} = 2i^\star_{\frac{n+1}{2}-k,t} - i^\star_{1,t}$ ($q_{B,t} = 2i^\star_{M+k,t} - i^\star_{n,t}$) with $q_{A,t} = i^\star_{1,t}$ and $q_{C,t} = 2i^\star_{M+k,t} - q_{B,t}$ ($q_{A,t} = 2i^\star_{\frac{n+1}{2}-k,t} - q^\star_{B,t}$ and $q_{C,t} = i^\star_{n,t}$).

- The Dovish Chair constrained by the pivotal member: $i^{\star}_{c,t} < i^{\star}_{\frac{n+1}{2}-k,t}$
  In this case, the Chair proposes $q_{A,t} = q_{B,t} = q_{C,t} = i^{\star}_{\frac{n+1}{2}-k,t}$ and $i_t = i^{\star}_{\frac{n+3}{2}-k,t}$.

- The Hawkish Chair constrained by the pivotal member: $i^{\star}_{c,t} > i^{\star}_{M+k,t}$
  In this case, the Chair proposes $q_{A,t} = q_{B,t} = q_{C,t} = i^{\star}_{M+k,t}$ and $i_t = i^{\star}_{M+k,t}$.

Instead of exogenous status quo, the preferred rate by hawkish members ($[i^{\star}_{M+k,t}, \ldots, i^{\star}_{n,t}]$) or dovish members ($[i^{\star}_{1,t}, \ldots, i^{\star}_{\frac{n+1}{2}-k,t}]$) can constrain the Chair's proposal and the committee's decision in this extended model. This feature makes the model more sympathetic to the equal weight aggregation regime in which the committee's decision is sensitive to outlier members' opinions.

5.1.2 Extension with multiple alternatives and imperfect information

If the Chair knows preferred rates of all the members perfectly, he or she would not propose the baseline proposal that is rejected in equilibrium. However, if the Chair have imperfect information on members' preferences before making proposals, the threat of dissents creates a bargaining process. For example, $q_{A,t}$ and $q_{C,t}$ are based on expected preferred rates of potential members instead of their actual preferred rates. In this case we can rewrite constraints for $q_{A,t}$ and $q_{C,t}$ for the centrist Chair as follows:

$$
\begin{aligned}
i^{\star}_{1,t} \leq q_{A,t} <= 2E(i^{\star}_{\frac{n+1}{2}-k,t} | \Omega_{c,t-\Delta}) - i^{\star}_{c,t}, \\
i^{\star}_{n,t} \geq q_{C,t} >= 2E(i^{\star}_{M+k,t} | \Omega_{c,t-\Delta}) - i^{\star}_{c,t}.
\end{aligned}
$$

where $\Omega_{c,t-\Delta}$ is the information set of the Chair before the FOMC meeting.

Under these assumptions, the Chair can modify proposals to accommodate the pivotal member's preference. For example, suppose $q_{B,t} = i^{\star}_{c,t} > 2i^{\star}_{\frac{n+1}{2}-k,t} - q_{A,t}$. After seeing the realized preferred rate of each member, the Chair provides the following modified baseline proposal while alternative proposals intact.

$$
\tilde{q}_{B,t} = 2i^{\star}_{\frac{n+1}{2}-k,t} - q_{A,t}.
$$

As long as $\tilde{q}_{B,t} \geq 2i^{\star}_{M+k,t} - q_{C,t}$, this proposal is accepted under the supermajority rule. Notice that under the supermajority rule, the pivotal member does not need to be the median voter, which explains the August 2011 episode we highlight in the paper.

This extended model with imperfect information can explain two main empirical findings documented in this paper. First, when the Chair misjudged the dovish pivotal member's preference before the meeting and learned about the true preference during the meeting, the baseline proposal can be modified in the direction of the dovish alternative as we saw in the transcript of the August 2011 FOMC meeting. Second, if the Chair has better information about members' likely stances before FOMC meetings, the modification of the baseline proposal is less likely to happen because learning about pivotal members' preferences during the meeting is the main source of the modification of the baseline proposal in the model. Indeed, our empirical finding for the predominance of the chair-centric aggregation regime during Yellen's tenure might be related to the fact that Yellen started to engage more intensely with reserve bank president right

before the meeting through one-to-one calls. For example, the Chair's calendar records show that Yellen spent more time for phone calls with multiple Reserve Bank presidents on Thursday and Friday of the week before FOMC meetings than Bernanke.[19].

5.2 Other models of the committee decision making

While we focus on the agenda-setting Chair model of the monetary policy committee, there are other models that try to explain the monetary policy committee's decision-making without explicitly acknowledging the Chair's dominant role. Using the Bank of England voting data, Hansen et al. (2014) show that heterogeneous assessments of the economy by policy committee members are as important as their individual policy preferences. Each member makes a binary decision to "tighten" or "accommodate" policy. Policy options are not set by the governor. In their model, individual members receive independent signals about the aggregate economy. They focus on the efficiency of the committee decision-making in terms of aggregating private information by committee members. They study the optimal behavior of individual policymakers with heterogeneous information and preferences given the majority voting structure. In this environment, the median voter is always pivotal in the committee's decision. Therefore, it is difficult to generate the outcome we observed in the August 2011 episode in which a few members who are more dovish than the median voter can push the Chair to revise the baseline proposal. Perhaps, their model is better suited to explain the Bank of England's decision-making process in which frequent dissents are not unusual instead of the FOMC with a more consensus-driven culture.

Recently, Laarits et al. (2025) use prompts to an LLM to obtain sentiment scores for policy tightening from remarks by FOMC members in FOMC meeting transcripts. They perform many sanity checks to make sure that the LLM generates sensible answers for tasks such as topic classification and tone scoring. They focus on the optimal form of information aggregation to minimize prediction errors for macroeconomic variables. There are a couple of issues in taking this model as a realistic description of the FOMC. First, there is no formal voting process in which some members can be pivotal in the committee decision-making. Second, when members differ in unrealized future risk assessments, they do not necessary change their opinions based on the past prediction error of realized data as shown in Cieslak, Hansen, McMahon and Xiao (2023). Third, members may differ not only in private information about the economy but also their preferences on the tradeoff between inflation stabilization and output stabilization.

Kazinnik and Sinclair (2025) apply an LLM model to recover members' preferences on the interest rate and describe a model in which the final interest rate is a weighted average of the Chair's proposed rate and the median preferred rate by committee members. While the aggregation process somewhat similar to our empirical framework, they use fixed weights (30% for the Chair's proposed rate and 70% for the median rate) instead of determining weights meeting by meeting to fit the final outcome. Since the the median voter is the only counterpart to the Chair, this model also has a difficulty in explaining the fact that members who are not the Chair nor the median voter can be pivotal on occasion as in the August 2011 episode.

[19]Records are released typically with two month delay and publicly available on Calendars of Federal Reserve Chairs

5.3 Open issues

There are several open issues related to modeling the monetary policy committee to explain our empirical findings. We find significant variations in the frequency of the dominance of the chair-centric regime over time and across topics. While shifts in the preferred aggregation regime over time can be explained by the change in the pivotal member who constraint the Chair's proposal , it is impossible to explain shifts across topics in our extension of the agenda-setting Chair model because the model still relies on the single-dimensional monetary policy tool. Therefore, this model does not allow different aggregation regimes across different topics. Nonetheless, shifts in the preferred aggregation regime across topics appear to be a robust feature from text analysis of FOMC transcripts. In our analysis, the chair-centric regime was selected less than 50% only for policy guidance topic and real economic activity topic. Our interpretation of this result is grounded on the heterogeneity among policymakers. For instance, Reserve Bank presidents provide different pieces of information on real economic activities that may not be available to governors at the Federal Reserve Board including the Chair because Reserve Bank presidents actively seek information from their district contacts. Bobrov et al. (forthcoming) and Fos and Xu (2022) show that regional economic conditions collected by Reserve Bank presidents may have oversized impacts on their interpretations of national economic conditions. Also, the constraint on the Chair's power might be more prominent in deciding on policy guidance because policymakers may have strong views on the effectiveness of an accommodative policy stance as shown by three dissenting members at the August 2011 FOMC meeting who was skeptical about a more explicit time-based forward guidance. For this reason, the Chair's power is more likely to be constrained by pivotal members in the committee for this topic. A committee voting model regarding the policy stance with multidimensional aspects is ideal to address this issue.

Using alternative versions of statements prepared by the staff at the Federal Reserve Board, Doh et al. (2022) construct a measures of text-based monetary policy stance. The overall sentiment of the released statement is quantified by the distance in embedding representations of different statements in a large language model. For instance, if the released statement is semantically closer to the hawkish alternative ("Alternative C") than the dovish one ("Alternative A"), the sentiment is classified as favoring providing less policy accommodation. With market-expectations of the committee's sentiment fixed, they perform language counterfactual exercises that quantify the counterfactual market responses if one of alternative versions of statements were released instead of the realized post-meeting statements. Ultimately, adopting the alternative version of statement would involve committee members' voting decisions. Hence, the plausibility of the counterfactual exercise depends on the details of the environment in the committee decision-making such as the voting rule and heterogeneity in members' information and preferences. One shortcoming of our current empirical framework is the lack of the dynamic link between successive FOMC meetings. For instance, members can take a dynamic strategy to maximize the discount present value of their payoffs instead of the static payoff at a given meeting. In this environment, members can act strategically and may not always fully reveal their preferences depending on other members' behavior. Career concern in Iaryczower, Lopez-Moctezuma and Moscariello (2025) (e.g., ignore one's private signal not to be seen as wrong when other members get right about the

prediction on the state of economy) and the concern for the inflation expectations of the private sector in Sibert (2003) (e.g., dovish junior committee member disguising as a hawkish member to lower inflation expectations lower which he or she can exploit later to boost output by creating inflation surprise) can incentivize members to act strategically by suppressing their true preferences and private information for the future gain although all the other members act according to their true preferences and information about the state of the economy. A structural voting model that expands the strategy space of each member from the static one to the dynamic one may address this issue.

6 Conclusion

Text analysis is increasingly used in monetary policy research. However, a standard model of monetary policy assumes a single decision maker and often abstracts the committee structure, which is prevalent in the real world. When there are multiple policymakers, bargaining over words to be used in the official communication of the committee is an important aspect of policy-making.

By combining a topic model estimated by FOMC statement data with sentiment analysis using a pre-trained large language model, we obtain topic sentiment scores for both statement and transcript data. We recover the preferred aggregation regime that generates statement topic sentiment scores as weighted averages of transcript topic sentiment scores. We find empirical support for the chair-centric model overall but there are notable variations in the frequency of favoring the chair-centric regime across different Chairs' tenures and topics. The alternations among the chair-centric regime, the median voter regime, and the equal weight regime suggest that the median voter may not always be the pivotal member who constrains the Chair's agenda-setting power. The August 2011 episode in which the Chair's proposal is modified by a group of more dovish members than the median voter highlights this finding. We show that an agenda-setting Chair model with multiple proposals and imperfect information on committee members' preferences can explain this empirical finding.

We quantify surprise components in statement topic sentiment scores by using statement draft language written by a private consultant for major investment banks as the proxy for the market expectation of statement language. By regressing high-frequency changes in bond market variables onto surprise components in statement topic sentiment scores, we uncover different roles played by alternative topic sentiment scores.Some of time-variations in the preferred aggregation regime are found to be correlated with monetary policy shocks identified by high-frequency bond market data around FOMC announcements, suggesting that more research on the monetary policy committee decision-making process may improve our understanding of the aggregate monetary policy shocks typically used in macroeconomic analysis. Our empirical findings as well as the agenda-setting Chair model with multiple proposals and imperfect information on committee members' preferences provide the first-step to dig into the microfoundation of monetary policy decision-making. We believe that further studies in this area will enhance our understanding of not only the actual policy decision-making process but also the counterfactual policy.

References

Ahn, Hie Joo, Thomas Cook, Taeyoung Doh, Elias Kastrisis, and Jesse Wedewer, "Quantifying the FOMC's Monetary Policy Stance based on Text," *Mimeograph*, 2025.

Aruoba, S Borağan and Thomas Drechsel, "Identifying monetary policy shocks: A natural language approach," Technical Report, National Bureau of Economic Research 2024.

Baerg, Nicole, *Crafting consensus: Why central bankers change their speech and how speech changes the economy*, Oxford University Press, 2020.

Bauer, Michael D and Eric T Swanson, "A reassessment of monetary policy surprises and high-frequency identification," *NBER Macroeconomics Annual*, 2023, *37* (1), 87–155.

Bertsch, Christoph, Isaiah Hull, Robin L Lumsdaine, and Xin Zhang, "Central bank mandates and monetary policy stances: Through the lens of federal reserve speeches," *Journal of Econometrics*, 2025, p. 105948.

Blei, David M, Andrew Y Ng, and Michael I Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, 2003, *3* (Jan), 993–1022.

Bobrov, Anton, Rupal Kamdar, and Mauricio Ulate, "Regional Dissent: Do Local Economic Conditions Influence FOMC Votes?," *American Economic Review: Insights*, forthcoming.

Chen, Li and Shawn Mankad, "A structural topic sentiment model for text analysis," *Available at SSRN*, 2022.

Cieslak, Anna and Michael McMahon, "Tough Talk: The Fed and the Risk Premium," *Available at SSRN*, 2023.

_ , Stephen Hansen, Michael McMahon, and Song Xiao, "Policymakers' uncertainty," Technical Report, National Bureau of Economic Research 2023.

Doh, Taeyoung, Dongho Song, and Shu-Kuei Yang, "Deciphering federal reserve communication via text analysis of alternative fomc statements," *Federal Reserve Bank of Kansas City Working Paper*, 2022.

Fisher, Eric, Rebecca McCaughrin, Saketh Prazad, and Mark Vandergon, "Fed Transparency and Policy Expectation Errors: A Text Analysis Approach," 2023.

FOMC, Statement, "FOMC Statement, August 2011," 2011.

FOMC, Transcript, "FOMC Meeting Transcript, August 2011," 2011.

Fos, Vyacheslav and Nancy R Xu, "Do the Voting Rights of Federal Reserve Bank Presidents Matter?," *European Corporate Governance Institute–Finance Working Paper*, 2022, (856).

Gáti, Laura and Amy Handlan, "Monetary Communication Rules," 2024.

Goodfriend, Marvin, "The role of a regional bank in a system of central banks," in "Carnegie-Rochester Conference Series on Public Policy," Vol. 51 Elsevier 1999, pp. 51–71.

Griffiths, Thomas L and Mark Steyvers, "Finding scientific topics," *Proceedings of the National academy of Sciences*, 2004, *101* (suppl_1), 5228–5235.

Hack, Lukas, Klodiana Istrefi, and Matthias Meier, "The Systematic Origins of Monetary Policy Shocks," 2025.

Handlan, Amy, "Text Shocks and Monetary Surprises: Text Analysis of FOMC Statements with Machine Learning," 2022.

Hansen, Stephen, Michael McMahon, and Andrea Prat, "Transparency and deliberation within the FOMC: a computational linguistics approach," *The Quarterly Journal of Economics*, 2018, *133* (2), 801–870.

_ , _ , and Carlos Velasco Rivera, "Preferences or private assessments on a monetary policy committee?," *Journal of monetary Economics*, 2014, *67*, 16–32.

Hartmann, Jochen, Mark Heitmann, Christian Siebert, and Christina Schamp, "More than a feeling: Accuracy and application of sentiment analysis," *International Journal of Research in Marketing*, 2023, *40* (1), 75–87.

Hoffman, Matthew, Francis Bach, and David Blei, "Online learning for latent dirichlet allocation," *advances in neural information processing systems*, 2010, *23*.

Iaryczower, Matias, Gabriel Lopez-Moctezuma, and Paola Moscariello, "Career Concerns in Collective Decision-Making: The Federal Open Market Committee," Technical Report, National Bureau of Economic Research 2025.

Jr, Henry W Chappell, Rob Roy McGregor, Todd A Vermilyea et al., *Committee Decisions on Monetary Policy: Evidence from Historical Records of the Federal Open Market Committee*, The MIT Press, 2005.

Kazinnik, Sophia and Tara M Sinclair, "Fomc in silico: A multi-agent system for monetary policy decision modeling," *Available at SSRN 5424097*, 2025.

Laarits, Toomas, Ben Matthies, Kaushik Vasudevan, and Will Yang, "Policy by Committee," 2025.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

Loughran, Tim and Bill McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *The Journal of finance*, 2011, *66* (1), 35–65.

Lunsford, Kurt G, "Policy language and information effects in the early days of Federal Reserve forward guidance," *American Economic Review*, 2020, *110* (9), 2899–2934.

Murphy, Kevin P, *Probabilistic machine learning: Advanced topics* 2023.

Ramey, Valerie, "Macroeconomic shocks and their propagation," *Handbook of macroeconomics*, 2016, *2*, 71–162.

Riboni, Alessandro and Francisco J Ruge-Murcia, "Monetary policy by committee: consensus, chairman dominance, or simple majority?," *The Quarterly Journal of Economics*, 2010, *125* (1), 363–416.

_ and _ , "The power of the Federal Reserve chair," 2020.

_ and Francisco Ruge-Murcia, "Dissent in monetary policy decisions," *Journal of Monetary Economics*, 2014, *66*, 137–154.

_ and _ , "Membership Turnover and Policy Disagreement at the FOMC," 2025.

_ , _ , and Linh Tran, "Deliberation and Policy Outcomes: Evidence from the Textual Analysis of FOMC Transcripts," 2025.

Schonhardt-Bailey, Cheryl, *Deliberating American monetary policy: A textual analysis*, The MIT Press, 2013.

Shapiro, Adam Hale and Daniel J Wilson, "Taking the fed at its word: A new approach to estimating central bank objectives using text analysis," *The Review of Economic Studies*, 2022, *89* (5), 2768–2805.

Sibert, Anne, "Monetary Policy Committees: Individual and Collective Reputations," *The Review of Economic Studies*, 2003, *70* (3), 649–665.

Svensson, Lars EO, "Inflation forecast targeting: Implementing and monitoring inflation targets," *European economic review*, 1997, *41* (6), 1111–1146.

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

Team, Gemma, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, and Ramona Merhej et al., "Gemma 3 Technical Report," 2025.

Thornton, Daniel L and David C Wheelock, "Making Sense of Dissents: A History of FOMC Dissents.," *Review (00149187)*, 2014, *96* (3).

Wallach, Hanna M, Iain Murray, Ruslan Salakhutdinov, and David Mimno, "Evaluation methods for topic models," in "Proceedings of the 26th annual international conference on machine learning" 2009, pp. 1105–1112.

Appendix

A. The Posterior Inference of the Topic Model

Dirichlet prior distributions of Θ and Φ given hyperparameters α and β are as follows:

$$p(\Theta|\alpha) = (\frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K})^D \prod_{k=1}^{K}\prod_{l=1}^{D}\theta_{k,l}^{\alpha-1} \ , \ p(\Phi|\beta) = (\frac{\Gamma(n\beta)}{\Gamma(\beta)^n})^K \prod_{i=1}^{n}\prod_{k=1}^{K}\Phi_{k,i}^{\beta}, \tag{19}$$

where $\Gamma(\cdot)$ is the gamma function defined by $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$. Then we can derive the posterior kernel by combining the likelihood of W given Θ and Φ, $p(W|\Theta, \Phi)$ with these prior distributions.

$$p(\Theta, \Phi|W, \alpha, \beta) \propto (\frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K})^D (\frac{\Gamma(n\beta)}{\Gamma(\beta)^n})^K [\prod_{k=1}^{K}\prod_{l=1}^{D}\theta_{k,l}^{\alpha-1}][\prod_{i=1}^{n}\prod_{k=1}^{K}\Phi_{k,i}^{\beta-1}] \prod_{l=1}^{D}\sum_{k=1}^{K}\prod_{i=1}^{n}(\theta_{k,l}\Phi_{k,i})^{W_{i,l}}. \tag{20}$$

Because of the summation over topics, it is not feasible to simplify the powers of $\theta_{k,l}$ and $\Phi_{k,i}$ in equation (20). The exact posterior distribution requires us to integrate out $p(\Theta, \Phi|W, \alpha, \beta)$ with respect to Θ and Φ, which is intractable without the simplification of powers of $\theta_{k.l}$ and $\Phi_{k,i}$.

Suppose we have the posterior sample of Z given (W, Θ, Φ), we can rewrite the likelihood to eliminate the summation over topics because we have the topic index for each word.

$$p(W, Z|\Theta, \Phi) = \prod_{l=1}^{D}\prod_{i=1}^{n}[(\prod_{k=1}^{K}\theta_{k,l}^{\mathbb{I}(z_{i,l}=k)})\Phi_{z_{i,l},i}]^{W_{i,l}}. \tag{21}$$

Let the length of document l be $n(l)$ (total number of words in document l). The posterior estimate of the k-th topic score for document l is given by

$$\widehat{\theta}_{k,l} = \frac{\alpha + \sum_{i=1}^{n(l)}\mathbb{I}(z_{i,l}=k)}{K\alpha + \sum_{j=1}^{K}\sum_{i=1}^{n(l)}\mathbb{I}(z_{i,l}=j)}. \tag{22}$$

When the i-th word of document l is v, we can draw $z_{i,l}$ iteratively using the following conditional distribution:

$$p(z_{i,l}=k|Z_{-(i,l)}, W, \alpha, \beta) \propto [\frac{\sum_{(a,b)\neq(i,l)}\mathbb{I}(z_{a,b}=k, x_{a,b}=v)+\beta}{\sum_{(a,b)\neq(i,l)}\mathbb{I}(z_{a,b}=k)+n\beta}][\frac{\sum_{a\neq i}\mathbb{I}(z_{a,l}=k)+\alpha}{n(l)+K\alpha}], \tag{23}$$

where $x_{a,b}$ is the b-th word in document a.

B. Robustness Check for Transcript Remark Segmentation

Transcripts are produced and stored by the board of Governors in PDF files. The PDF file is challenging to work with for natural language processing. To prepare the data for our analysis, we cleaned the transcript documents manually and with additional machine learning and generative AI techniques. The objectives of cleaning were primarily to correctly label statements with the corresponding speaker, remove extraneous formatting (e.g. pagination) and properly handle footnotes. After initial cleaning, the transcript passages were irregularly broken into minimal and uneven paragraphing (i.e. most remarks were not broken into sensible paragraphs)

There are two important reasons to break transcript passages into paragraphs. First, for purposes of sentiment and topic analysis, many transcript passages are read statements from FOMC members that span several pages, discussing several topics and attaching different sentiments to each. Breaking these passages into discrete passages into paragraphs, with each paragraph ideally expressing a complete thought, allows us to more accurately derive the sentiment of the speaker towards specific topics. Second, for the purposes of using a RoBERTa-based sentiment classifier, splitting passages into paragraphs breaks the transcript into units that generally fit into the RoBERTa context window (512 tokens).

We employed generative AI to break remarks into paragraphs. Specifically, we passed each transcript passage into Google's Gemma 3 (Team, Kamath, Ferret, Pathak, Vieillard and et al., 2025) model and asked it to identify sentences after which a new paragraph should begin. We then programatically inserted the paragraph breaks in the location identified by Gemma 3. This approach restricts the LLM so that it's influence on the text is *indirect* (suggesting locations for edits) and *limited* (only paragraph marks are inserted at edit points, no other text is added or removed). By using the LLM this way, we ensure that the resulting transcript is complete and free from hallucinations.

In general, topic-sentiments for the transcripts derived from length-based segmentation (breaking transcript remarks into 512-token chunks) and paragraph-segmentation (breaking remarks as described here) are weakly correlated with an average correlation coefficient of 0.26 across all topics. This suggests that the method used for segmenting the text does have an impact on the topic-sentiment that is ultimately produced. Qualitatively, however, the aggregation regime preferences are similar for both paragraph and length-based segmentation.

Paragraph Segmentation Prompt

[**System Message:**] *Default System Message*
[**User Message:**] Read through the following passage from an FOMC meeting transcript. It is missing paragraph breaks. Find where breaks should be inserted to split the passage into logical paragraphs. Generally, paragraph breaks should be inserted at points such as:

- where the speaker changes topics
- at the end of a complete thought or argument

Return a json dictionary with one field: "paragraph_breaks" which is a list of strings where each string is the last sentence of a paragraph exactly as it appears in the text.
Transcript passage:
{*remark_text*}