Policy by Committee

Toomas Laarits, Ben Matthies, Kaushik Vasudevan, and Will Yang*

August 20, 2025

Abstract

Committees shape decisions across numerous settings, but empirical examination of group decision-making remains limited due to measurement challenges. We address these challenges by constructing a novel dataset of Federal Open Market Committee members' arguments and decisions from over forty years of meetings. Contrary to prevailing theories emphasizing private information aggregation, we find that members employ distinct models to interpret public information, with committee decisions favoring models that better fit recent data. We develop a theoretical framework showing that committees can improve decision-making but must balance against the risk of excess sensitivity to recent data.

*Toomas Laarits is with New York University, email: tlaarits@stern.nyu.edu. Ben Matthies is with the University of Notre Dame, email: bmatthie@nd.edu. Kaushik Vasudevan is with Purdue University, email: kvasude@purdue.edu. Will Yang is with Yale University, email: will.j.yang@yale.edu. For helpful comments and conversations, we thank Nick Barberis, Chris Clayton, Zhi Da, Bill English, Joel Flynn, Paul Fontanier, Stefano Giglio, Peter Hansen, Sid Kankanala, Aleksei Oskolkov, Stefano Pegoraro, John Shim, Alp Simsek, Adi Sunderam (discussant), and Johannes Wohlfart; seminar participants at Yale University, the University of Notre Dame, and New York University; and conference participants at the NBER Behavioral Finance Spring 2025 meeting. The initial draft of this paper was written while Vasudevan was visiting Yale University, whose hospitality he gratefully acknowledges.

1 Introduction

Group decision-making shapes economic outcomes across many contexts. While a long-standing literature following Condorcet (1785) emphasizes that groups can improve decision quality by pooling diverse information, decisions are often made in groups even when information-sharing gains may be limited. For example, the vast majority of central banks make monetary policy decisions by committee, yet, as Blinder (2008) argues, "you can mostly ignore [differential information] in the monetary policy context because virtually all the data that matter are common knowledge." This observation raises questions about why committees persist in such settings and how their deliberations shape the decision-making process and its outcomes.

Although there is a rich theoretical literature on group decision-making, empirical evidence remains scarce, as systematic tests require comprehensive data on groups' decisions and individual members' views, along with sufficient variation to identify how those views shape the decisions. We address these challenges by analyzing a setting that is uniquely well-suited to study group decision-making: the Federal Open Market Committee (FOMC), which sets U.S. monetary policy. We obtain over forty years of meeting transcripts to assemble a comprehensive dataset of committee decisions, as well as members' arguments and policy preferences, allowing us to study the deliberations of experts in a high-stakes setting.

We find that, in contrast to prevailing theories emphasizing private information, committee members employ distinct and persistent models, defined as frameworks used to interpret public information and translate it into forecasts of key variables. Examining how the committee aggregates its members' views, we find that its decisions give greater weight to the preferences of members whose models better fit recent economic data. Guided by our findings, we develop a theoretical framework of group decision-making under model uncertainty in which committees aggregate distinct models held by constituent members. Our framework suggests that group decision-making may enhance decision quality when members are sufficiently willing to consider others' models, but it also carries the risk of excess sensitivity to incoming data.

Empirical Analysis. To conduct our study, we construct a novel dataset of the specific arguments and preferences expressed by each member at each scheduled FOMC meeting from April 1976 to December 2019. The base unit of our analysis is an argument—a specific claim or assessment—made by a committee member about key macroeconomic and financial variables: inflation, growth, employment, credit markets, and the stock market.[1]

[1]We focus on these variables for their policy relevance, as FOMC members make explicit forecasts for the

We employ a large language model (LLM) to parse meeting transcripts, extracting all arguments made by each member at each meeting, alongside the data that members cite. We score each argument based on the support it provides for looser or tighter monetary policy. The score's sign captures the stance supported by the argument—either accommodative or restrictive—and its magnitude captures the strength of support for that stance. From these argument-level data, we construct aggregated member-meeting scores for each variable by averaging each member's argument scores within each meeting. These aggregated scores represent a member's overall stance based on their arguments about a given variable (e.g., inflation or growth). We implement an extensive set of validation tests on these measures, including assessing the comprehensiveness of our argument extraction through a sentence-by-sentence analysis, confirming that aggregate scores track business cycles, and verifying a strong relationship between members' argument-based scores and their quantitative forecasts.

We begin our empirical analysis by studying the drivers of variation in policy stances. We document that members express views that are largely fixed over time, with each member making highly correlated arguments in their discussions of different variables. Member fixed effects explain 20% to 30% of the variation in member-meeting scores, indicating the persistent nature of members' views. Some members consistently view the economy through a more dovish lens, perceiving weak growth and arguing for more accommodative policy; others consistently perceive strong inflation pressures and argue for more restrictive policy. Moreover, members' scores exhibit a strong factor structure across variables: members who are relatively more dovish about inflation also tend to be relatively more dovish when discussing growth and unemployment. The first principal component explains 61% of the variation in member-meeting observations.

We argue that these patterns, with substantial variation being explained by member fixed effects and expressed views being correlated across variables, are consistent with members filtering economic data through heterogeneous and largely fixed models. In contrast, traditionally emphasized private information-based theories would struggle to account for these facts, especially given the structured format of each FOMC meeting, where members explicitly share their information and views.

We provide further evidence in support of model heterogeneity versus private information by analyzing the data cited by members in their arguments. We find that 74% of arguments reflect public information, such as government statistics and Federal Reserve Board staff analysis, while only 26% draw on potentially private sources like discussions with business contacts. Constructing scores using only arguments

macroeconomic variables, and prior work has documented the policy importance of the financial variables, e.g., Cieslak and Vissing-Jorgensen (2021).

based on public information, we find that these exhibit a 0.95 correlation with scores using all arguments and show similar fixed effects and cross-variable correlations. This evidence indicates that heterogeneity in views likely stems from how members interpret public information rather than private information.

To better understand the source of members' disagreements, we examine the content of their arguments. We classify arguments by the economic mechanisms discussed, creating categories specific enough to distinguish between concepts (e.g., *Inflation Expectations and Anchoring* versus *Inflation Dynamics and Persistence*) while grouping together differing views on the same concept (e.g., *inflation is persistent* and *inflation is transitory* both fall under *Inflation Dynamics and Persistence*). Applying this classification of arguments organized by economic concept, we find that about 90% of the variation in members' scores is explained by heterogeneous interpretations of the same topic, rather than discussions of different topics. For example, members' scores primarily diverge not because they emphasize different aspects of inflation, but because they interpret the same aspect of inflation differently, with hawks arguing that inflation is persistent and doves arguing that it is transitory. This result provides further evidence that model heterogeneity drives disagreement.

Given members' persistent and heterogeneous models, our next set of analyses considers how the committee aggregates these models in its decision-making. A central challenge is measurement: previously studied measures include dissents, which are rare for institutional reasons (6.9% of member-meeting observations), and deviations of members' preferred rates from the target rate decision, a single dimension with little variation in the later part of our sample. We address this challenge by using the transcripts to build a new dataset of FOMC decisions, along with novel measures that capture each member's position on each decision and their influence in shaping it. Beyond target rates, we capture decisions on targets for monetary aggregates, unconventional policies such as quantitative easing, forward guidance, and the exact wording of the committee's statement. We use the LLM to extract each decision and code members' alignment with each decision on a seven-point scale from complete opposition to full support, and their influence from zero (no impact) to three (a pivotal role). We validate our measures against monetary policy surprises measured using high-frequency asset price responses to policy announcements, committee voting behavior, and members' stated rate preferences.

Analyzing the FOMC's decisions, we find that the committee systematically tilts its decisions toward the preferences of members whose models best fit recent economic data. To measure model fit in a given meeting, we compute the absolute deviation between a vector of each member's scores by variable and a corresponding vector of recent macroeconomic and financial data. In our main analysis, members with one

standard deviation better model fit show 0.12–0.30 standard deviations higher alignment with committee decisions and 0.04–0.11 standard deviations greater influence, depending on the specification, including meeting, member, and committee role fixed effects. The probability of dissent decreases by 3.4–6.8 percentage points per standard deviation improvement in fit. Moreover, we find that the results hold when we measure members' models using the fixed effects of their variable scores, indicating that the shifting influence of members stems from how well the recent data match a constant, time-invariant aspect of their views.

Theoretical Framework. Our empirical findings suggest a new perspective on committees. Rather than aggregating private information, as in standard theories, committees aggregate different models for interpreting public information. To further explore its implications, we formalize this perspective in a simple theoretical framework of group decision-making under model uncertainty. Each member of a committee observes the same data but has a different, partially misspecified model. The committee's objective is to agree on a model to make its policy decision. Each member advocates for a combination of all members' models, trading off a desire for a committee model that fits the data and a cost for supporting a model that deviates from their own. The committee's choice arises as a Nash equilibrium outcome. Given members' desire to match the observed data well, the committee tilts toward better-fitting models, matching our empirical evidence.

We evaluate the committee's performance based on how closely its chosen model matches the unknown true model in expectation. We find that the committee may outperform its best member's model under three basic conditions. First, the data must be sufficiently informative to distinguish among models. Second, members' models must be "diverse," in that they capture different parts of the data-generating process. Third, members must be sufficiently willing to consider each other's perspectives. If the data are too noisy or members are too stubborn in advocating for their own perspectives, the committee may settle upon a sub-optimal model. Intriguingly, some stubbornness may improve the committee's performance by preventing overreaction to noisy data. This creates a fundamental trade-off in committee design—leveraging model diversity while tempering excess sensitivity to incoming information.

Our results carry practical implications for committee design. Committee performance depends on model diversification, so it is important to select members who bring different perspectives. Structured deliberation processes—like the FOMC's systematic discussion of macroeconomic conditions and policy—can facilitate the aggregation of models and mitigate stubbornness on the part of members. These findings are also more broadly applicable to settings beyond monetary policy where groups must make decisions under model uncertainty.

Literature Review. Our paper contributes to the literature on group decision-making. Prior work takes an informational perspective—that group decision-making may improve decisions by aggregating private information (Gerling et al. (2005) and Sibert (2006) provide literature reviews). In contrast, we document that committee members operate with different models to interpret the same data, indicating that information aggregation is an insufficient explanation for the function of committees, as argued by Blinder (1999, 2008). We argue that a key function of committees is to bring together members with different models; in this regard, we take a related but distinct approach to Schwartzstein and Sunderam (2021, 2024). Given that members' models may be misspecified, our work relates to a growing literature considering how group decision-making affects decision quality in light of behavioral biases (Charness and Sutter, 2012; Enke, Graeber and Oprea, 2023; Barahona et al., 2024).

Our work relates to a long tradition in monetary policy and macroeconomics on decision-making in the face of model uncertainty. Prior work has presented numerous approaches—for example, conservatism (Brainard, 1967), learning (Sargent, 1999; Cogley and Sargent, 2005; Primiceri, 2006), model averaging (Blinder, 1999), and robust control (Hansen and Sargent, 2008). Our work suggests that decision-making by committee is another way policymakers address model uncertainty, and we evaluate the associated trade-offs.

Additionally, our paper relates to recent work on how the FOMC's policy decisions are affected by the composition of its voting members in terms of geographic region (Bobrov, Kamdar and Ulate, 2025; Fos, Tamburelli and Xu, 2025; Fos and Xu, 2025) and policy preferences (Malmendier, Nagel and Yan, 2021; Bordo and Istrefi, 2023; Hack, Istrefi and Meier, 2023). We take a different focus, examining how members' perspectives are weighted in the group decision-making process. We document that the committee tilts its decisions toward members with models that are better aligned with recent data. Given this behavior, our evidence is consistent with Hack, Istrefi and Meier (2025), who find that the FOMC appears to use a time-varying policy rule in making its policy decisions, and provides a micro-foundation for such behavior.

More broadly, our paper relates to empirical work studying central bankers' policy preferences and their determinants (Belden, 1989; Chappell Jr, McGregor and Vermilyea, 2004; Schonhardt-Bailey, 2013; Hansen, McMahon and Rivera, 2014; Istrefi, 2019; López-Moctezuma, 2023), and to work analyzing FOMC transcripts and minutes (Boukus and Rosenberg, 2006; Hansen, McMahon and Prat, 2018; Cieslak and Vissing-Jorgensen, 2021; Shapiro and Wilson, 2022; Cieslak and McMahon, 2023; Cieslak et al., 2023; Bordo, Istrefi and Martínez, 2024). Relative to this literature, we construct a new, granular dataset of the FOMC's decisions and individual members' beliefs, preferences, and economic arguments. In this regard, our paper is related to a small

but growing literature that seeks to identify the mental models used by economic agents in macroeconomic and financial settings (Andre et al., 2022; Andre, Schirmer and Wohlfart, 2023; Andre et al., 2025; Bybee, 2025; Binetti, Nuzzi and Stantcheva, 2024; Ke, 2024; Bastianello, Décaire and Guenzel, 2025). Importantly, we characterize decision-makers' arguments and the resulting outcomes in a high-stakes setting.

2 Background and Data

We construct several measures from FOMC transcripts. We summarize relevant institutional details and data sources and outline the construction of our measures.

2.1 Background

Our sample consists of 365 regularly scheduled FOMC meetings from April 1976 through December 2019. The FOMC meets eight times per year to formulate monetary policy. The committee normally consists of 12 voting members: seven Governors of the Federal Reserve Board in Washington, DC, including a Chairperson; the President of the New York Fed; and four members selected on a rotating basis from the presidents of the eleven other regional Federal Reserve banks.[2] Members' tenures on the committee are staggered, so that individual members do not always make decisions with the same sets of other members. In addition to the voters, the other non-voting regional Fed presidents and numerous Federal Reserve staff members attend and actively participate in the meetings.

Since 1993, transcripts of FOMC meetings have been released to the public on a five-year lag and are available from the Federal Reserve website. These transcripts provide verbatim discussions from every FOMC meeting from 1976 onward. We describe the processing of the transcript data in Internet Appendix IA.A.1. Prior to each meeting, the Federal Reserve Board staff prepares Tealbook A (formerly the Greenbook), which provides discussions and forecasts of macroeconomic and financial conditions. We obtain the reports from the Federal Reserve Board website and the associated forecasts from the Philadelphia Federal Reserve website. The staff also prepares and distributes Tealbook B (formerly the Bluebook), which includes a set of monetary policy alternatives that serves as a baseline for the committee's policy discussion. The committee is not restricted to choosing one of the alternatives and often combines elements across alternatives. We obtain Tealbook B from the Federal Reserve website.

[2]In the sample, there are meetings with fewer than twelve voting members due to empty seats on the Board of Governors not being immediately filled.

FOMC meetings have two parts related to the committee's monetary policy decision: a discussion of the economic and financial situation and a monetary policy discussion.[3] The discussion of the economic and financial situation begins with a presentation by members of the Federal Reserve Board staff followed by questions, prepared statements by members of the committee, and a general discussion. The subsequent monetary policy discussion typically begins with the director of the Division of Monetary Affairs presenting monetary policy alternatives. The presentation is followed by a policy go-around where both voting and non-voting members discuss their policy preferences. Following this discussion, the committee votes on the proposed policy directive. In 1994, the committee began announcing the outcome of its meetings at the end of each meeting via a public statement; a substantial portion of the policy discussion segment thereafter concerns the wording of this statement.

2.2 Extracting Information from Transcripts using LLM

We generate a number of our empirical measures of interest by analyzing the transcripts using the 20250219 release of the Anthropic Claude 3.7 Sonnet large language model ("the LLM"). We discuss our empirical measures and prompts below and present the full text of the prompts in Internet Appendix IA.A.2.

We take a number of steps to ensure the reliability and reproducibility of the LLM's output. First, we use 'chain-of-thought' prompting, wherein the LLM is prompted to engage in stages of intermediate reasoning before answering a query.[4] Prior work has found chain-of-thought prompting to significantly improve performance in complex reasoning tasks (Wei et al., 2022), which has led it to serve as the basis for the most recent generation of complex reasoning models.[5]

Second, we require that the LLM provide explicit transcript quotes to justify its reasoning, ensuring that its analysis is grounded in the text.[6]

Third, we run multiple prompts for each variable of interest. The first identifies the relevant passages for the variable. The second analyzes the relevant subset of text. This multi-step procedure ensures comprehensiveness in capturing all arguments and improves the stability of the LLM's output (Bastianello, Décaire and Guenzel, 2025). We further validate comprehensiveness with a sentence-by-sentence

[3] Meetings typically also feature presentations about open market operations and include discussions of special topics, such as long-run policy implementation frameworks.

[4] Dunn et al. (2024) find strong performance of the Claude 3.5 Sonnet model in identifying topics in FOMC meeting minutes. For chain-of-thought prompts, we use the Anthropic prompt improver tool, discussed at https://www.anthropic.com/news/prompt-improver.

[5] e.g., OpenAI's o1 and o3 models, see https://openai.com/index/learning-to-reason-with-llms/.

[6] Sarkar and Vafa (2024) highlight the potential for look-ahead bias in predictive tasks, which is not a primary concern for our work, given that we assign descriptive tasks to the LLM, e.g., labeling data.

analysis discussed below. To ensure the LLM produces deterministic responses, we also set the 'temperature' parameter of the LLM to zero.

Finally, we verify the LLM's outputs against observable data, as discussed throughout our analysis, ensuring that the variables we construct are consistent with corresponding numerical data, whenever available.

2.3 Measuring Committee Members' Arguments and Beliefs

For each meeting, we construct scores for each committee member based on their arguments with respect to policy-relevant macroeconomic and financial variables: inflation, growth, employment, credit markets, and the stock market. We first score each argument from -3 to 3 based on the support it provides for looser or tighter monetary policy. The score's sign captures the stance supported by the argument—either accommodative (negative) or restrictive (positive)—and its magnitude captures the strength of support for that stance. We construct aggregated member-meeting scores for each variable by averaging scores across each member's arguments in each meeting. In principle, the measures capture a combination of the relative weight the member gives to a variable in their policy stance (e.g., a coefficient in a forward-looking Taylor (1993)-style rule), as well as the member's beliefs about that variable. In practice, as we further detail, we find beliefs and preferences to be tightly linked in members' arguments. For example, members expressing high forecasts of inflation also tend to stress the policy importance of controlling inflation.

To construct member-meeting scores for a variable, we first compile all statements by the member in the meeting and prompt the LLM to flag those discussing the variable. We then prompt the LLM to extract distinct arguments within these statements, providing verbatim quotes for each. The LLM assigns a score from –3 (strongly dovish) to +3 (strongly hawkish) to each argument. Arguments may represent (i) beliefs—descriptions of the variable's current or expected level; (ii) preferences—the appropriate policy response to deviations from the desired level; or (iii) both. We prompt the LLM to label each argument accordingly. For each speaker-meeting-variable, we average the argument scores, assigning zero to speakers who make no arguments about the variable.

Table 1 provides examples from the August 9, 2011 meeting, which we focus on in other examples in the paper to showcase our data. The meeting took place against a backdrop of zero interest rates, with the FOMC viewing the recovery from the Great Recession as slow and weak, and with financial markets performing poorly in light of developments regarding the fiscal situations in the U.S. and Europe. The committee did not move the target Fed Funds rate, which all members tacitly agreed upon. However, it did provide new guidance in its statement that it viewed the economic outlook

Speaker	Variable	Belief or Preference	Score	Quotes
Charles Plosser	Growth	Belief	0	"...it's too soon to tell the extent to which...the current weakness is temporary... or ... that the economy has entered a more persistent regime of slower growth..."
Charles Plosser	Stock Market	Preference	+3	"Our credibility would surely suffer if markets perceived us as directly responding to stock market prices."
Narayana Kocherlakota	Employment	Both	+2	"We adopted our current level of accommodation ... when ... unemployment was 9.8 percent ... Unemployment is lower than it was..."
Daniel Tarullo	Credit Markets	Both	-2	"...tightening financial conditions, along with the blow to confidence they bring, could surely also be more than a modest external shock."
Charles Evans	Inflation	Belief	-3	"Recent developments also suggest that inflation risks over the medium term continue to recede."
Charles Evans	Growth	Belief	-3	"The weak incoming data indicate that it's highly unlikely that the U.S. economy, in the foreseeable future, will achieve anything like a launch velocity."

TABLE 1: COMMITTEE MEMBERS' ARGUMENTS IDENTIFIED BY LLM: EXAMPLES

Note: The table presents examples of arguments extracted for each speaker in each meeting for inflation, growth, employment, credit markets, and the stock market. Each argument is assigned a score from -3 to +3, with -3 indicating that the argument strongly supports dovish policy and +3 indicating that the argument strongly supports hawkish policy. The scores are constructed by providing an LLM with a committee member's speeches and prompting it to identify arguments made by the speaker with respect to the variable of interest, to score the arguments, to provide assessments as to whether the argument reflects the speaker's belief about the current level or future level of the variable, their preference about how policy should address potential deviations of the variable from their desired level, or a combination of the two.

as deteriorating and foresaw that low interest rates were likely to remain warranted through mid-2013.

The first two rows present arguments by Philadelphia Fed President Charles Plosser, who dissented in the meeting out of concern that the committee's actions could be viewed as providing additional accommodation. The table shows his arguments about growth, expressing uncertainty whether weak growth is temporary or structural (score of 0); and about the stock market, that the committee should not appear to react to stock prices (score of +3). The next row lists an argument by Minneapolis Fed President Narayana Kocherlakota, who also dissented, that the additional accommodation is not appropriate given falling unemployment (score of +2). The next row lists Governor Daniel Tarullo's argument highlighting deteriorating financial conditions and the blow they may deal to confidence (score of -2). The final two rows present arguments from Chicago Fed President Charles Evans, indicating a view of receding inflation risks (score of -3) and a weak outlook for growth (score of -3).

Alongside the classifications, the table indicates whether each argument reflects beliefs, preferences, or both. Statements that simply describe forecasts of the variables—for example, Evans' views of receding inflation and weak growth—are coded as beliefs. Statements that discuss how a variable's deviation from its desired level should influence policy—for example, Plosser's comment on the stock market—are coded as preferences. Some statements reflect both, such as Kocherlakota's view about continued accommodation being inappropriate given falling unemployment.

Table 2 reports summary statistics for arguments, split between voting and non-voting members. Among voters, almost every member discusses growth at each meeting, and more than 80% discuss inflation, employment, and credit markets; about half mention the stock market. On average, a voter makes roughly six arguments about growth, four about inflation, three about employment and credit markets, and one about the stock market. Inflation arguments lean slightly hawkish (mean score 0.2), whereas those on growth, employment, and credit markets are slightly dovish (mean scores of –0.2, –0.1, and –0.3, respectively). The vast majority of arguments are coded as reflecting beliefs, and about half of these also reflect preferences. Non-voters show similar patterns to voters, though their arguments are somewhat more hawkish.[7] Because their arguments appear similar, we pool voters and non-voters when analyzing arguments, but focus on voters when examining the committee's decisions.

To validate comprehensiveness, for twenty randomly sampled meetings, we take all of the sentences spoken and prompt the LLM *sentence-by-sentence* to identify arguments with respect to the variables of interest. We then check if each argument is

[7]This pattern accords with evidence that bank presidents—who rotate for voting seats—tend to be more hawkish than governors; see Malmendier, Nagel and Yan (2021) and Bordo and Istrefi (2023).

Voting Members	Inflation	Growth	Employment	Credit	Stocks
Members per meeting			10.9		
Avg pct making argument per meeting	89%	99%	83%	83%	51%
Arguments per member-meeting	4.2	5.9	2.9	2.7	1.0
Avg argument score	0.2	-0.2	-0.1	-0.3	0.0
Pct of arguments reflecting					
Beliefs	36%	55%	45%	33%	35%
Preferences	18%	5%	10%	16%	13%
Both Beliefs & Preferences	46%	40%	45%	51%	52%
Non-voting Members	Inflation	Growth	Employment	Credit	Stocks
Members per meeting			6.7		
Avg pct making argument per meeting	86%	97%	82%	77%	36%
Arguments per member-meeting	4.0	5.7	3.3	2.1	0.6
Avg argument score	0.5	0.1	0.3	0.0	0.2
Pct of arguments reflecting					
Beliefs	42%	60%	56%	37%	33%
Preferences	16%	4%	8%	14%	18%
Both Beliefs & Preferences	42%	36%	37%	49%	49%

TABLE 2: ARGUMENT SUMMARY STATISTICS

Note: The table displays summary statistics for the arguments made related to each variable of interest. The top panel displays statistics for voting members of the committee and the bottom panel displays statistics for the non-voting members. Within each panel, the first row displays the average number of members per meeting. For a given variable, the second row displays the average number of members per meeting making an argument, the third row displays the average number of arguments per member, and the fourth row displays the average score across all arguments. The last three rows of each panel display the proportion of arguments reflecting the speaker's beliefs, the speaker's preferences, or a combination of both.

captured by our dataset. Sentence-by-sentence prompting has the disadvantages of not capturing context, double-counting arguments, and prohibitive cost, but provides a conservative benchmark for comprehensiveness. Our dataset captures 91.4% of arguments identified in sentence-by-sentence prompting. Analyzing the unmatched sentences, they tend to be ambiguous or not policy-relevant, and do not introduce systematic biases. We present the analysis in the Internet Appendix IA.A.3.

Figure 1 plots the average of committee members' scores in each meeting for each variable. The scores demonstrate intuitive behavior, showing cyclical patterns, and attaining troughs during NBER recessions (identified with dark gray bands). Concerns about inflation, all else equal, push toward more hawkish preferences; these are elevated in the 1970s and late 1990s, and were low in the years following the Global Financial Crisis. On the financial side, concerns about credit markets are cyclical, with elevated concerns during the crisis. The stock market is not important in the

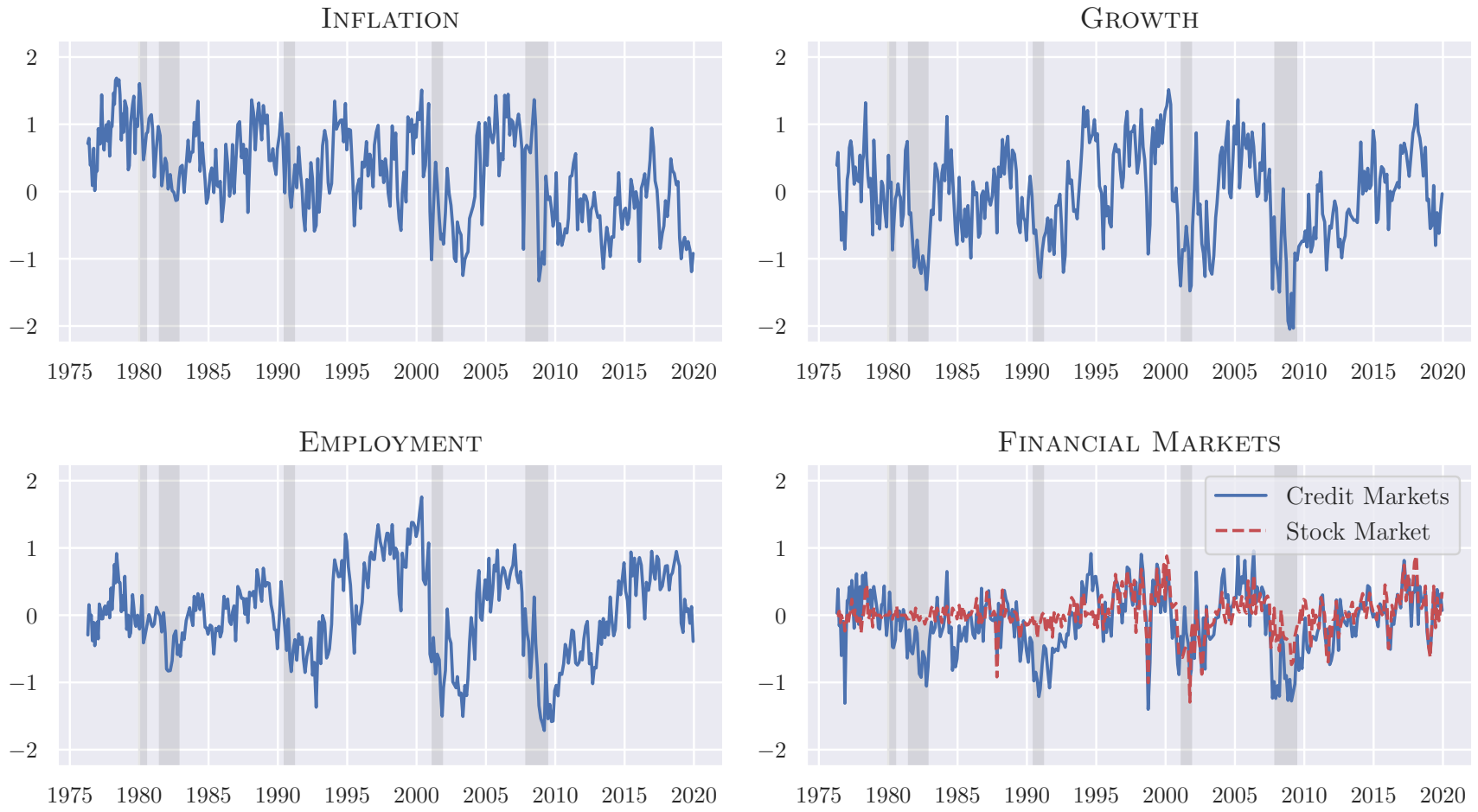


FIGURE 1: COMMITTEE MEMBERS' ARGUMENT SCORES OVER TIME

Note: The figure plots the average of scores assigned to committee members for inflation, growth, employment, credit markets, and the stock market for each FOMC meeting. For each meeting, the arguments made by members are assigned scores from -3 to +3 for each variable, with -3 indicating that an argument strongly supports more dovish policy and +3 indicating that an argument strongly supports more hawkish policy. The dark gray bands indicate NBER recession periods.

early part of the sample, with the exception of the market crash of 1987. Starting in the mid-1990s, the stock market increases in importance, with hawkishness arising in the late 1990s (corresponding to Alan Greenspan's statements about irrational exuberance in December 1996), and dovishness arising in periods of poor stock market performance, corresponding to the "Fed Put" (Cieslak and Vissing-Jorgensen, 2021).

We validate the cross-sectional explanatory power of the meeting-variable scores for each member—constructed by taking averages across argument scores for each argument made—by comparing them with the member's forecasts of inflation, GDP growth, and unemployment. These forecasts are available on a semi-annual basis from 1992 to 2007 from the Federal Reserve's Monetary Policy Report (MPR) to Congress (Romer, 2010); and on a quarterly frequency from the Summary of Economic Projections (SEP), which has been conducted prior to four FOMC meetings per year since October 2007.[8]

Figure 2 sorts committee members into deciles based on their cross-sectionally demeaned GDP, inflation, and unemployment forecasts for a given meeting, and plots the average of these forecasts versus the average cross-sectionally demeaned transcript-measured growth, inflation, and employment member-meeting scores within each decile. The combined panel in the figure standardizes observations within a variable to have zero mean and unit standard deviation, cross-sectionally demeans observations within a meeting, then plots the average (standardized) score versus the average (standardized) forecast, pooled across variables.

The scores constructed from members' arguments exhibit a strong cross-sectional relationship with forecasts, even though the comparison is not strictly apples-to-apples. The scores combine *forecasts* of a variable with the *weight* the variable receives in a member's policy stance. In the Internet Appendix, we construct alternative versions of Figure 2 using scores based only on belief arguments and, separately, only on preference arguments; we find very similar patterns for both. These similarities arise, in part, because members' preference and belief statements tend to point in the same direction for a given variable (member-meeting preference and belief scores have an average correlation of 0.57 across variables). For instance, members who forecast higher inflation also stress inflation as a core focus of policy.

2.4 Influence and Alignment with Respect to Policy Decisions

For each policy decision in each meeting, we construct two measures of how well a member's preferences are reflected in that decision: $Alignment_{i,t}$, a score from -3 to 3,

[8]The Summary of Economic Projections data are available for each year from 2007 through 2019 except for the year 2015; the forecasts for 2015 are expected to be released in 2026, see https://www.federalreserve.gov/monetarypolicy/files/FOMC20160111memo01.pdf.

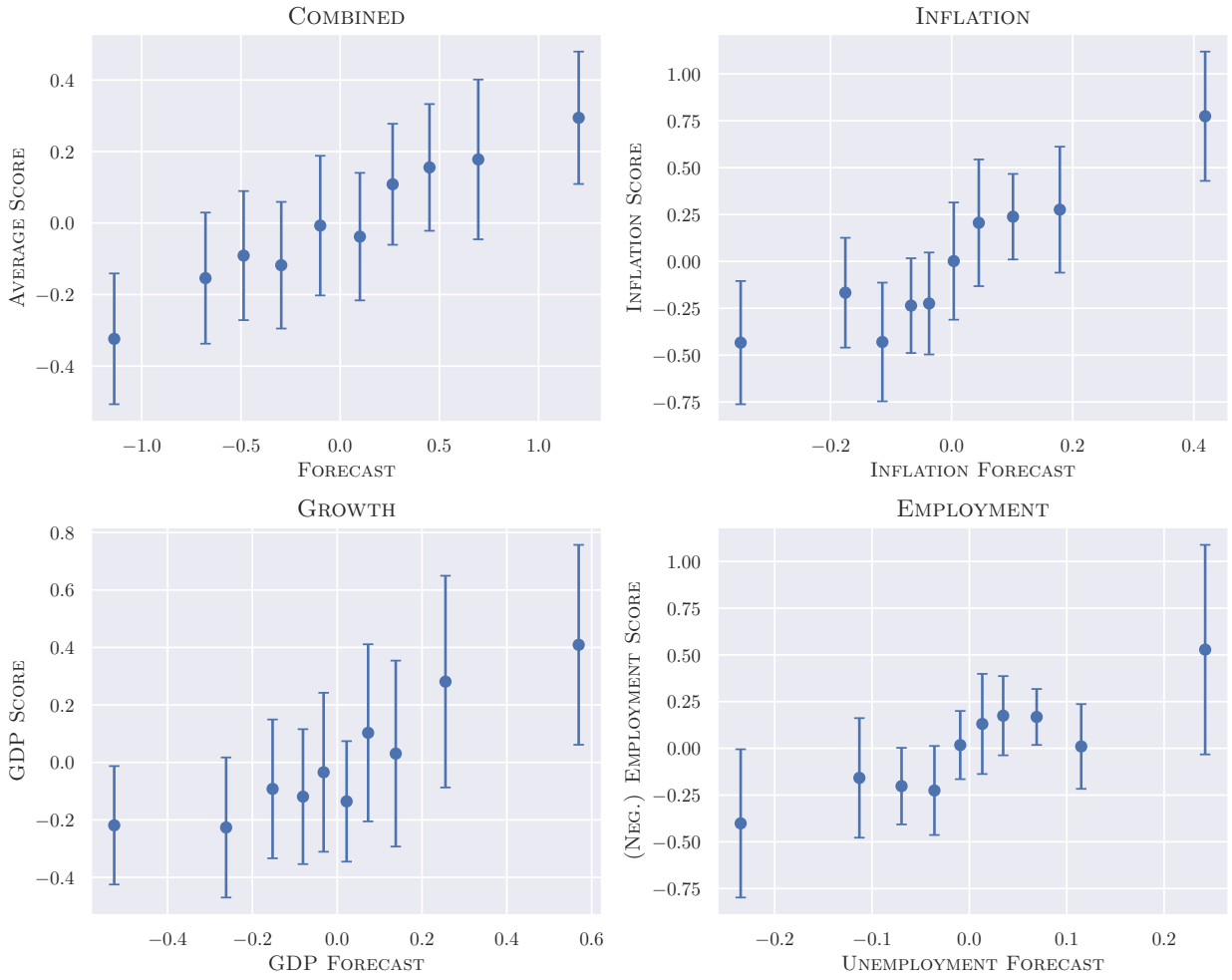


FIGURE 2: COMMITTEE MEMBERS' FORECASTS AND ARGUMENT SCORES

Note: The figure sorts committee members into deciles based on their cross-sectionally demeaned GDP, inflation, and unemployment forecasts, and plots the average of these forecasts versus the average cross-sectionally demeaned transcript-measured growth, inflation, and employment scores for the closest meeting. In the combined panel, observations are standardized to zero mean, unit variance within each variable, cross-sectionally demeaned by meeting, and then sorted into deciles; the panel plots the mean of each decile. Forecasts are from the Federal Reserve's Monetary Policy Reports to Congress, available semi-annually from 1992 to 2007 from Romer (2010); and from the Summary of Economic Projections, conducted prior to four meetings per year since October 2007. The figure also plots plus and minus two standard error bars for each decile. Standard errors are computed by clustering by speaker and meeting within each decile.

where 3 corresponds to perfect alignment and -3 corresponds to complete opposition; and $Influence_{i,t}$, which ranges from 0 to 3, where 3 indicates that the member played a pivotal role in the decision's adoption (for example, by proposing it), and 0 indicates no influence (for example, they opposed it, or did not explicitly address it in their speeches despite voting for it). These measures capture alignment and influence with

respect to target interest rates, as well as with respect to other decisions, for example, targets for monetary aggregates and forward guidance.

To construct the measures, we first assemble a dataset of the policy decisions made in each meeting. To do so for a given meeting, we provide the LLM with the policy discussion section of the meeting and the policy alternatives distributed prior to the meeting. We prompt the LLM to identify all adopted decisions, those presented in the policy alternatives and those proposed and adopted in an *ad hoc* manner alike. We also prompt the LLM to evaluate the hawkish and dovish tenor of each of the decisions by assigning them a score from -3 (highly accommodative) to +3 (highly restrictive).

For each member in each meeting, we then provide the LLM with the policy discussion, the alternatives, and the identified decisions for each meeting. We first prompt the LLM to provide a score from -3 to 3 to capture each member's alignment with each of the decisions. We then prompt the LLM to assign a score from 0 to 3 to capture the influence that the speaker exerted in the decision's adoption. Using the resulting output, we construct $Alignment_{i,t}$ and $Influence_{i,t}$ as the average alignment and influence scores for each of the decisions made.

Table 3 provides illustrative examples. Panel A presents the policy decisions and associated scores from the August 9, 2011 FOMC meeting. The LLM identifies five decisions made during the meeting, with the assigned scores in parentheses: maintaining a target Fed Funds rate of 0–25 bps (0); forward guidance about maintaining the low target rate through mid-2013 (-2); the addition of language to the committee statement that downside risks have increased (-1); the addition of language that the committee has discussed and is prepared to employ the range of policy tools as appropriate (-1); and the decision to continue reinvesting principal payments from domestic securities holdings into Treasury securities (-1). The first decision is assigned a score of zero, with the justification that it maintains status quo policy. The remaining scores are negative, as they introduce easier monetary policy conditions.

Panel B of Table 3 shows examples of alignment and influence scores for the August 2011 policy decisions for four voting members: Narayana Kocherlakota, Daniel Tarullo, Charles Evans, and Governor Janet Yellen.

Narayana Kocherlakota receives alignment and influence scores of 3 and 0 for the Fed Funds rate decision. Each of the alternatives for the meeting included the same target Fed Funds rate, with no discussion or contention, and accordingly, no scope for exerting influence. Kocherlakota receives alignment and influence scores of -3 and 0 for the mid-2013 guidance, which he explicitly argued against, seeing it as too accommodative and responsive to short-term fluctuations.

Daniel Tarullo is assigned alignment and influence scores of 1 and 2 for the mid-2013 forward guidance. The LLM identifies that he expressed concerns about the

PANEL A: EXAMPLE POLICY DECISIONS

Date	Change	Score	Justifications
Aug-11	Target range for the Federal Funds rate at 0 to 1/4 %.	0	Maintains the status quo.
Aug-11	Guidance that "exceptionally low levels for the Federal Funds rate at least through mid-2013" likely warranted	-2	Explicit timeframe for low rates more accommodative than previous "extended period" language.
Aug-11	Added language that "downside risks to the economic outlook have increased."	-1	Signals potential for future easing and reflects a more dovish assessment of economic conditions.
Aug-11	Added language that the committee "is prepared to employ [the range of policy tools] as appropriate."	-1	Signals the Committee's readiness to take further action if needed, which provides a slight easing bias.
Aug-11	Maintain policy of reinvesting principal payments on all domestic securities in Treasury securities.	-1	Maintains the size of the balance sheet rather than allowing it to shrink.

PANEL B: EXAMPLE SPEAKER ALIGNMENT SCORES

Date	Decision	Speaker	Alignment	Influence	Justification
Aug-11	Fed Funds Target	Kocherlakota	3	0	Supported, but not discussed.
Aug-11	Mid-2013 Guidance	Kocherlakota	-3	0	Saw as too much commitment to accommodation
Aug-11	Mid-2013 Guidance	Tarullo	1	2	Expressed concern, but helped shape the discussion.
Aug-11	Policy Tools Language	Tarullo	3	3	Explicitly proposed adopted language.
Aug-11	Mid-2013 Guidance	Evans	3	3	Indicated he would not vote for decision without it.
Aug-11	Downside Risks Language	Evans	3	3	Explicitly identified as missing from Alternative B.
Aug-11	Downside Risks Language	Yellen	3	1	Expressed consistent views but no direct advocacy
Aug-11	Policy Tools Language	Yellen	3	3	Explicitly proposed adopted language

TABLE 3: POLICY CHANGES AND POLICY ALIGNMENT: EXAMPLES

Note: Panel A of the table displays examples of identified policy decisions and the associated scores from -3 (highly accommodative/dovish) to +3 (highly restrictive/hawkish) from the August 2011 meeting. Panel B of the table displays examples of committee members' measured alignment and influence with the proposed changes, which are averaged to construct the $Alignment_{i,t}$ and $Influence_{i,t}$ measures for each speaker. For alignment, scores of -3 indicate complete opposition, and scores of +3 indicate complete alignment. For influence, scores of +3 indicate strong, explicit influence on the policy's adoption and scores of 0 indicate no influence. All measures are constructed by prompting an LLM after providing the policy discussion section of the FOMC meetings and the Tealbook B (Bluebook). The Justifications columns report paraphrased output from the LLM justifying the scores.

guidance being too rigid, which launched a substantial discussion on how best to articulate in the committee's statement that the guidance is based on the committee's current projections. Tarullo receives alignment and influence scores of 3 and 3 for the language about consideration of all available policy tools, which he explicitly introduced and advocated.

Charles Evans receives alignment and influence scores of 3 and 3 for the mid-2013 forward guidance, which he strongly advocated and indicated he would vote against the committee's decision if it were not included. Evans is assigned alignment and influence scores of 3 and 3 for the downside risks language, which he explicitly advocated for as missing from the policy alternative being discussed.

Janet Yellen receives alignment and influence scores of 3 and 1 for the downside risks language. She explicitly noted in the meeting that she "considers the case for policy action compelling" and argued that the "downside risk to growth is particularly serious now," though she did not explicitly discuss the inclusion of such language in the statement. Yellen receives alignment and influence scores of 3 and 3 for the language about the range of policy tools, explicitly proposing the adopted language that the committee is "prepared to" employ these tools.

We validate our LLM-based measure of policy decisions by linking it with monetary policy surprises measured by high-frequency asset price responses to announcements, e.g., as studied by Gürkaynak, Sack and Swanson (2005) and Nakamura and Steinsson (2018). We find a strong relationship between the average scores of decisions made in a meeting—capturing their hawkish or dovish tenor—and surprises to the target rate and future path of rates. This analysis validates our measure as capturing relevant dimensions of policy decisions. We report the results in Internet Appendix IA.C.1.

We validate our measures of policy alignment and influence with two pre-existing measures: dissenting votes, and the magnitude of deviations of members' preferred Fed Funds rate from the committee's choice. For voting data, we obtain the dataset from Thornton, Wheelock et al. (2014), continuously updated by the Federal Reserve Bank of St. Louis. We obtain data on members' preferred target Fed Funds rate for a sub-sample of the data from Chappell Jr, McGregor and Vermilyea (2004), who code these data by reading the transcripts; their dataset contains two samples, one that runs from the beginning of our sample through February 1978, and one that runs from August 1987 through December 1996.

Table 4 sorts member-meeting observations into bins using $Alignment_{i,t}$ scores, and reports the average $Alignment_{i,t}$, $Influence_{i,t}$, and target rate deviation in each bin. The table also reports the percentage of dissents and number of voter-meeting observations for each bin. The table reveals a strong relationship between the align-

	$Alignment_{i,t}$ range					
	[-3, -2]	(-2, -1]	(-1, 0]	(0, 1]	(1, 2]	(2, 3]
Average $Alignment_{i,t}$	-2.16	-1.29	-0.28	0.72	1.73	2.81
Average $Influence_{i,t}$	0.38	0.33	0.38	0.68	0.90	1.30
Percentage dissenting votes	86.4%	74.4%	41.8%	13.5%	1.1%	0.0%
Target Rate Deviation* (bps)	30	23	14	9	4	1
Number of voter-meetings	37	94	239	467	882	2263

TABLE 4: DECISION ALIGNMENT, INFLUENCE, AND DISSENT

Note: The table groups each voting committee member in a meeting based on their $Alignment_{i,t}$ score, where $Alignment_{i,t}$ ranges from -3 to 3, with -3 indicating strong opposition to the policy decisions made by the committee and +3 indicating perfect agreement with decisions. For each range of $Alignment_{i,t}$, the table reports the average $Alignment_{i,t}$; the average $Influence_{i,t}$, a score from 0 to 3 capturing the influence a member exerted in the decisions; the percentage of dissenting voters; and the average Target Rate Deviation (the absolute value of the difference between the chosen Fed Funds target and a member's preferred target rate, coded by Chappell Jr, McGregor and Vermilyea (2004)). The asterisk indicates that the data are only available for sub-samples from April 1976 through February 1978 and from August 1987 through December 1996. The table also reports the total number of voting members within each of the ranges.

ment and influence measures, dissents, and target rate deviations. Of voters with $Alignment_{i,t}$ scores in the range [-3,-2], (-2, -1], and (-1, 0] indicating disagreement with the committee's decision, 86%, 74%, and 42% of them dissent, respectively, with dissent decreasing in likelihood by bin. The average target rate deviation similarly displays a monotonically decreasing relationship with $Alignment_{i,t}$, ranging from 30 basis points for alignment scores between -3 and -2, to 1 basis point for alignment scores between +2 and +3. That is, members who are more misaligned with the committee's policy decision prefer different target Fed Funds rates than those chosen.

The evidence indicates that the LLM-based measures of alignment and influence match up well with other data: low alignment with the committee decision captures dissenting behavior well, and (mis)alignment is strongly related to deviations of members' preferences from the chosen target rate. Given the rarity of dissenting votes, and the narrow scope of the target rate deviations measures, our measures provide more granular and continuous measurement of alignment with policy decisions. Moreover, our influence measure is unique in assessing the role each member plays in shaping the decisions.

3 Members' Arguments and Fixed Models

We next examine members' arguments and their variable-meeting scores, with a focus on the disagreements expressed within them. We present several analyses indicating that members' disagreements are primarily driven by heterogeneous models—ways of interpreting public data and forming forecasts across variables—rather than by members having different private information. We document that members' scores exhibit strong member fixed effects and cross-variable correlations, which, through the lens of a simple framework, are consistent with model heterogeneity. Further, categorizing the data cited by members in their arguments, we show that the majority of data cited are public information (as opposed to private). Finally, categorizing the content of members' arguments, we document that disagreements across members are largely explained by them discussing the same topics, but drawing differing interpretations, rather than by a focus on different aspects of the economy.

3.1 Overview and Framework

Before proceeding, we present a simple framework to guide our analysis in this section, and describe our results through this lens. Consider an economy with state ω_t in period t. Committee members forecast k variables, which shape their policy preferences. All members observe public data D_t, a noisy signal about ω_t. Member i interprets these data through a probability model m_i that induces a map, f_i, of D_t into signals for each variable:

$$f_i(D_t) = (s_{i,t,1}, \ldots, s_{i,t,k})^\top.$$

In addition, member i observes private signals about each variable, $U_{i,t} = (u_{i,t,1}, \ldots, u_{i,t,k})^\top$. Member i's score for variable k is

$$\gamma_{i,t,k} = \beta_{i,k}\big(w\, s_{i,t,k} + (1-w)\, u_{i,t,k}\big), \tag{1}$$

where $\beta_{i,k}$ captures the importance of variable k to member i (e.g., a Taylor-rule coefficient) and w is the weight on public versus private information in making forecasts.

Our first result in this section is that member fixed effects explain substantial variation in $\gamma_{i,t,k}$. Members who are hawkish on a given variable remain so over time. Our second result is that, for a given member, $\gamma_{i,t,k}$ is correlated across variables. For both results, Equation (1) highlights two potential sources: (i) heterogeneous models, that is, members having a persistent map that translates public information into forecasts in a consistent manner across variables; or (ii) members receive persistent private in-

formation that is correlated across dimensions.[9] Given the structure of the FOMC where members publicly share their views each meeting, we interpret the results as evidence that members employ different models. We also provide further support for the heterogeneous model interpretation in additional tests using data on whether the data cited in each argument are public or private and on the topical composition of members' arguments.

3.2 Fixed Effects in Policy Perspectives

With our conceptual framework in hand, we proceed with the analysis. Table 5 reports R^2 values from regressions of member-meeting scores for inflation, growth, employment, credit markets, and the stock market on (i) meeting fixed effects and (ii) member fixed effects. Unsurprisingly, meeting effects explain a sizeable share of the variation—R^2 values of 0.26–0.29 for the macro variables and 0.16–0.18 for the financial variables—capturing the common tilt in policy perspectives that accompanies macroeconomic fluctuations (e.g., greater accommodation when growth is weak).

Member fixed effects are almost as important, with R^2 values of 0.20–0.28 for the macro variables and 0.08–0.17 for the financial variables. Averaging across measures, meeting effects account for 28 percent of the variation and member effects for 28 percent. This echoes evidence from other macro-financial settings, such as Giglio et al. (2021), on persistence of individual heterogeneity in views.

To gauge whether the estimated R^2 values for member effects exceed what could arise by chance, we conduct 1,000 bootstrap simulations. Each simulation resamples scores with replacement inside every meeting—keeping the cross-variable correlations intact while randomizing their assignment to members—and regresses the resampled scores on member fixed effects. The 99th-percentile R^2 values from these simulated regressions are reported in the last row and provide a conservative benchmark under the null of no member fixed effects. The R^2 values far exceed this benchmark, indicating the strong presence of member fixed effects.

Next, we explore how members' scores are related across different variables. Table 6 reports the principal components of members' scores. The first three columns report principal components estimated from member-meeting observations. The first principal component loads positively on each of the variables and explains 61% of the variance. These positive loadings are consistent with the first principal component representing a hawk-versus-dove dimension, where committee members and

[9] For member fixed effects, a third possibility is that members differ in their desired levels of the variables (e.g., distinct inflation targets). Focusing on belief-based arguments suggests this is not the primary driver. Additionally, the persistent nature of members' models may also be consistent with members slowly learning the parameters governing the data-generating process (Johannes, Lochstoer and Mou, 2016; Farmer, Nakamura and Steinsson, 2024; Li, Van Nieuwerburgh and Renxuan, 2025).

	Regression R^2					
	Inflation	Growth	Employment	Credit	Stocks	Average
Meeting FE	0.26	0.29	0.29	0.18	0.16	0.28
Member FE	0.28	0.20	0.21	0.17	0.08	0.28
Bootstrap null (99th pct.)	0.10	0.05	0.06	0.04	0.03	0.05
N	6402	6402	6402	6402	6402	6402

TABLE 5: COMMITTEE MEMBERS' SCORES: FIXED EFFECTS

Note: The table reports R^2 values from regressions of the member-by-meeting inflation, growth, employment, credit market, and stock market scores—and the average of scores across variables—on meeting fixed effects and member fixed effects. The scores range from –3 to 3, where the sign indicates whether arguments made with respect to the variable by the member support a hawkish or dovish stance, and the magnitude captures their relative strength. The last row, "Bootstrap null (99th pct.)," gives the 99th-percentile R^2 from 1,000 bootstrap simulations that reassign member identities by sampling member-meeting observations with replacement within each meeting, a benchmark under the null of no persistent member fixed effects.

meetings that are comparatively hawkish for one variable (e.g., inflation) also tend to be hawkish for other variables (e.g., growth and unemployment). The second and third principal components capture differential loadings on various variables, such as hawkishness based on inflation versus dovishness based on growth and credit market conditions (the second principal component).

The last three columns of the table report principal components from a matrix of member-level fixed effects for each variable, estimated after controlling for meeting fixed effects.[10] Once again, the first principal component loads positively on all variables, capturing a hawk-versus-dove dimension and explaining 84% of the variation. Because these principal components are estimated after controlling for meeting fixed effects, the evidence indicates that a committee member who tends to be persistently hawkish about inflation also tends to be persistently hawkish about growth, employment, credit market conditions, and the stock market. The subsequent components capture differential loadings on variables, such as dovishness based on credit conditions versus hawkishness based on unemployment (the second component).

The principal component analysis reveals that most of the variation in members' scores is explained by a hawk-versus-dove dimension. Without controlling for meeting fixed effects, committee members' scores reflect the dominant macroeconomic or financial variables at the time; for example, committee members may be more concerned with inflation during inflationary periods or credit conditions more during

[10]Due to the inclusion of meeting fixed effects, and the fact that members' tenures on the FOMC are staggered, member fixed effects are estimated controlling for the fact that a given member's tenures may happen to coincide with poor or strong economic conditions on average.

Comparison of Principal Components						
	Full Data			Member FE		
Principal Component Loadings	PC1	PC2	PC3	PC1	PC2	PC3
Inflation	0.54	0.77	-0.34	0.59	-0.28	0.75
Growth	0.56	-0.26	0.13	0.53	-0.09	-0.34
Employment	0.49	-0.07	0.71	0.44	0.84	-0.04
Credit Markets	0.37	-0.55	-0.59	0.40	-0.43	-0.55
Stock Market	0.15	-0.17	-0.13	0.13	0.16	-0.15
Proportion of Variance Explained	0.61	0.14	0.12	0.84	0.08	0.05

TABLE 6: PRINCIPAL COMPONENTS OF COMMITTEE MEMBERS' SCORES

Note: The table reports a comparison of the first three principal components of the inflation, growth, employment, credit, and stock market policy scores. The first three columns (Full Data) show the results where observations are member-meeting scores. The last three columns (Member FE) displays results where observations are member fixed effects for each variable, estimated from regressions of the member-meeting scores for the variables on member and meeting fixed effects.

periods of financial turmoil. However, after controlling for meeting fixed effects, the lion's share of the cross-sectional variation in scores is explained by heterogeneity across members along the hawk-versus-dove dimension.[11]

The dominance of a single hawk–dove factor in explaining cross-sectional variation in members' scores and the close link between those scores and committee members' macroeconomic forecasts (Figure 2) are consistent with each member having a different model to filter information and generate forecasts across variables. Differences in members' models may lead some to have persistently higher or lower forecasts across variables, giving rise to strong member fixed effects.

3.3 Public versus Private Information

We provide further evidence that differences in members' views reflect heterogeneous models—as opposed to differential private information—by showing that the vast majority of data cited in members' arguments reflect public information.

We pass the quotations for each argument to the LLM and prompt it to identify whether the cited data are qualitative or quantitative; whether they describe regional, national, or international quantities; and whether they reflect private or specialized information that other members would not reasonably know. We also pass one thousand randomly sampled arguments to the LLM to construct categories of cited data,

[11] Related, Herbst and Winkler (2021), Scanlan (2023), and Wu (2023) document evidence of a factor structure in macroeconomic forecasters' predictions across different variables.

Summary	
Arguments citing data	77,539
Arguments *not* citing data	27,229
Total arguments	104,768
Information source	
Public information	74%
Private / Specialized info.	26%
Evidence type	
Qualitative	79%
Quantitative	22%
Geographical scope	
National	79%
Regional	14%
International	7%

TABLE 7: INFORMATION CITED IN MEMBERS' ARGUMENTS

Note: The table reports summary statistics on the data cited in arguments made by committee members during FOMC meetings. The categories are constructed by prompting the LLM to identify the data cited in each identified argument made by a committee member during a meeting. Categories may sum to greater than 100% as some arguments cite multiple pieces of data.

and then prompt the LLM to categorize the data cited in each argument. Internet Appendix Table IA.B.1 lists each of the categories and their descriptions.

Table 7 displays details on the information cited by committee members in their arguments. We identify that 74% of arguments reflect public information, while 26% reflect specialized or private information. About 79% of arguments are qualitative in nature, while 22% include some type of quantitative figures. About 79% of arguments reference data at the national level; 14% reference data that are specific to a region of the United States, and 7% reference international data.[12]

To better understand the nature of information cited by members, Figure 3 plots the proportion of arguments in different data categories, broken down by the proportion that reflect public versus private information; an argument may reflect data from multiple categories, so the proportions sum to more than 100%. The most common category is *FOMC Member Analysis*, capturing arguments where members express personal observations without citing any specific sources. An example in this category is the quote from Charles Plosser in Table 1 expressing uncertainty about whether economic weakness is structural.[13] The vast majority (more than 97%) of *FOMC Mem-*

[12]The percentages may not sum to 100%, as a single argument may reference multiple pieces of data.

[13]Similarly, about 25% of arguments are identified as not citing any data, with the speaker simply asserting their preferences or economic relationships.

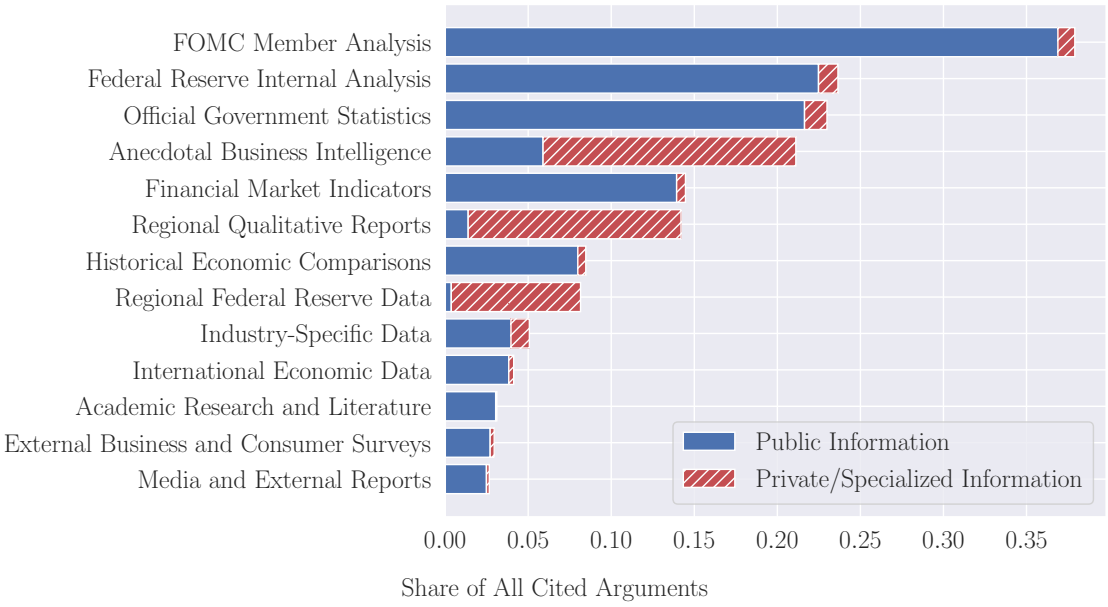


FIGURE 3: DATA CITED IN ARGUMENTS BY CATEGORY

Note: The figure displays the share of arguments made, split by category and whether the argument uses public or private information. Detailed descriptions of the categories are provided in Internet Appendix Table IA.B.1.

ber Analysis only reflects public information. The next two categories of data most commonly cited are *Federal Reserve Internal Analysis* (24% of arguments, e.g., Greenbook/Tealbook forecasts), and *Official Government Statistics* (23% of arguments; e.g., GDP and CPI numbers). Both overwhelmingly reflect public information.

Considering arguments with private information, we identify that 21% of arguments reflect *Anecdotal Business Intelligence* (e.g., conversations with business contacts); about three quarters of *Anecdotal Business Intelligence* reflect private information. The other types of arguments reflecting private information are *Regional Qualitative Reports* and *Regional Federal Reserve Data*, which are qualitative and quantitative descriptions of regional conditions.

We compute member-meeting scores using only arguments that do not reflect private information and compare these with the member-meeting scores used in our main analysis. The correlations of the scores across observations for each variable exceed 0.9; and the average correlation is 0.95. That is, the scores largely reflect public information. Additionally, in Internet Appendix Tables IA.C.2 and IA.C.3, we repeat the fixed effects and principal components analyses from Tables 5 and 6 and obtain very similar results. The strong similarities between the overall scores and those constructed using public information are consistent with members' scores being driven by public information.

3.4 Members' Arguments and Model Heterogeneity

To further explore members' model heterogeneity, we examine the contents of arguments made by members. We classify arguments into categories based on the topic discussed—for example, inflation categories include *Inflation Expectations and Anchoring* and *Inflation Dynamics and Persistence*. We then consider whether members' within-meeting disagreements are explained by heterogeneous interpretations of the same mechanisms, for example, whether inflation is persistent or transitory, or by members focusing on different aspects of the economy. We find that the vast majority of score differences arise from differing views on the same topic, for example, disagreements on the persistence of inflation, with a limited role for topic selection.

3.4.1 Argument Classification

We classify each of the arguments made into different categories. We construct categories by randomly sampling five sets of 500 arguments, prompting the LLM to identify the categories of arguments within each set, and then prompting it to consolidate the resulting five sets of categories. We seek to construct categories that are sufficiently narrow to distinguish between different mechanisms, for example, inflation expectations versus persistence; but sufficiently broad to accommodate differing viewpoints, for example, for the viewpoints *inflation is persistent* and *inflation is transitory* to be identified as discussing the same mechanism and to be classified in the same category.

We report the categories and their descriptions in Appendix IA.B. For each variable, we identify economically intuitive categories capturing differing viewpoints on economic mechanisms. For example, growth categories encompass recent growth and the outlook for growth, monetary policy transmission, and financial conditions. Inflation categories cover inflation expectations, policy tradeoffs, and transmission mechanisms. Employment arguments cover labor market impacts of monetary policy, skill mismatches, and sectoral variations. We identify a total of 53 categories across the five variables.

Figure 4 provides examples of the categories, plotting the average number of arguments per meeting by category and their average scores (accommodative versus neutral versus restrictive). The first panel displays the most frequently discussed categories. These include *Growth Outlook and Risks*, *Employment Growth Patterns and Business Investment*, *Sectoral Growth Patterns and Capacity Constraints*, and *Interest Rates and Credit Market Pricing*. Interestingly, the most discussed topics tend to have neutral scores on average, though this masks substantial across-member heterogeneity, as we show below. The second and third panels plot the categories of arguments with

the most negative average scores (most dovish) and most positive averages (most hawkish). The most accommodative arguments are about financial conditions, and the most hawkish about inflation, for example with respect to policy credibility and the strength of policy transmission.[14]

3.4.2 Topic Selection versus Heterogeneous Interpretations

We next explore whether differences in scores are driven by members focusing on the same aspects of the economy but drawing different interpretations, or by members focusing on different aspects of the economy altogether.

For member i in meeting t, we decompose i's score for variable k, $\gamma_{i,t,k}$, as

$$\gamma_{i,t,k} = \underbrace{\overline{\gamma}_{-i,t,k}}_{\text{Selection}} + \underbrace{(\widehat{\gamma}_{i,t,k} - \overline{\gamma}_{-i,t,k})}_{\text{Interpretation}} + \underbrace{(\gamma_{i,t,k} - \widehat{\gamma}_{i,t,k})}_{\text{Residual}}, \tag{2}$$

where, constructing category weights proportional to the number of arguments made by i in each category, $\overline{\gamma}_{-i,t,k}$ is the weighted average of the average within-category scores of arguments by members other than i; and $\widehat{\gamma}_{i,t,k}$ is the weighted average of the within-category average scores of i's arguments in categories shared with at least one member in meeting t.[15] The selection effect, $\overline{\gamma}_{-i,t,k}$, captures how much of i's score is driven by their choice of topics. The interpretation effect, $\widehat{\gamma}_{i,t,k} - \overline{\gamma}_{-i,t,k}$, captures the extent to which i's scores are driven by them discussing the same topics differently than the average member. The residual is the component of i's score coming from making arguments on topics not discussed by other members.

To illustrate these concepts concretely, we study Charles Plosser's and Daniel Tarullo's arguments from the August 2011 meeting. With meeting fixed effects, Plosser's cross-sectionally demeaned score of +1.94 decomposes into a selection effect of +0.14 and an interpretation effect of +1.80. Tarullo's cross-sectionally demeaned score of -1.12 decomposes into a selection effect of -0.26 and an interpretation effect of -0.86.

The 0.40-point difference in selection effects reflects their choices of topics. Plosser focused on a set of more hawkishly oriented topics than the average member. For example, he argued that some of the weakness in growth might be structural and that "We may be headed for that world as well [like Europe with persistent high unemployment], and monetary policy is not the solution to that problem," which we classify under the category *Structural Factors and Long-term Growth.* Meanwhile, Tarullo

[14]Section IA.C.2 in the Internet Appendix presents time series of argument category proportions by variable. The figures reveal that while some categories of argument receive fairly constant discussion, such as *Growth Outlook and Risks*, others vary cyclically and exhibit secular trends. For example, discussions of the *Phillips Curve and Economic Slack*, *Inflation Target Interpretation*, and *Inflation Measurement and Methodology* become increasingly important starting in the 1990s.

[15]If member i only makes arguments in categories not discussed by other members, $\overline{\gamma}_{-i,t,k}$ is set to zero.

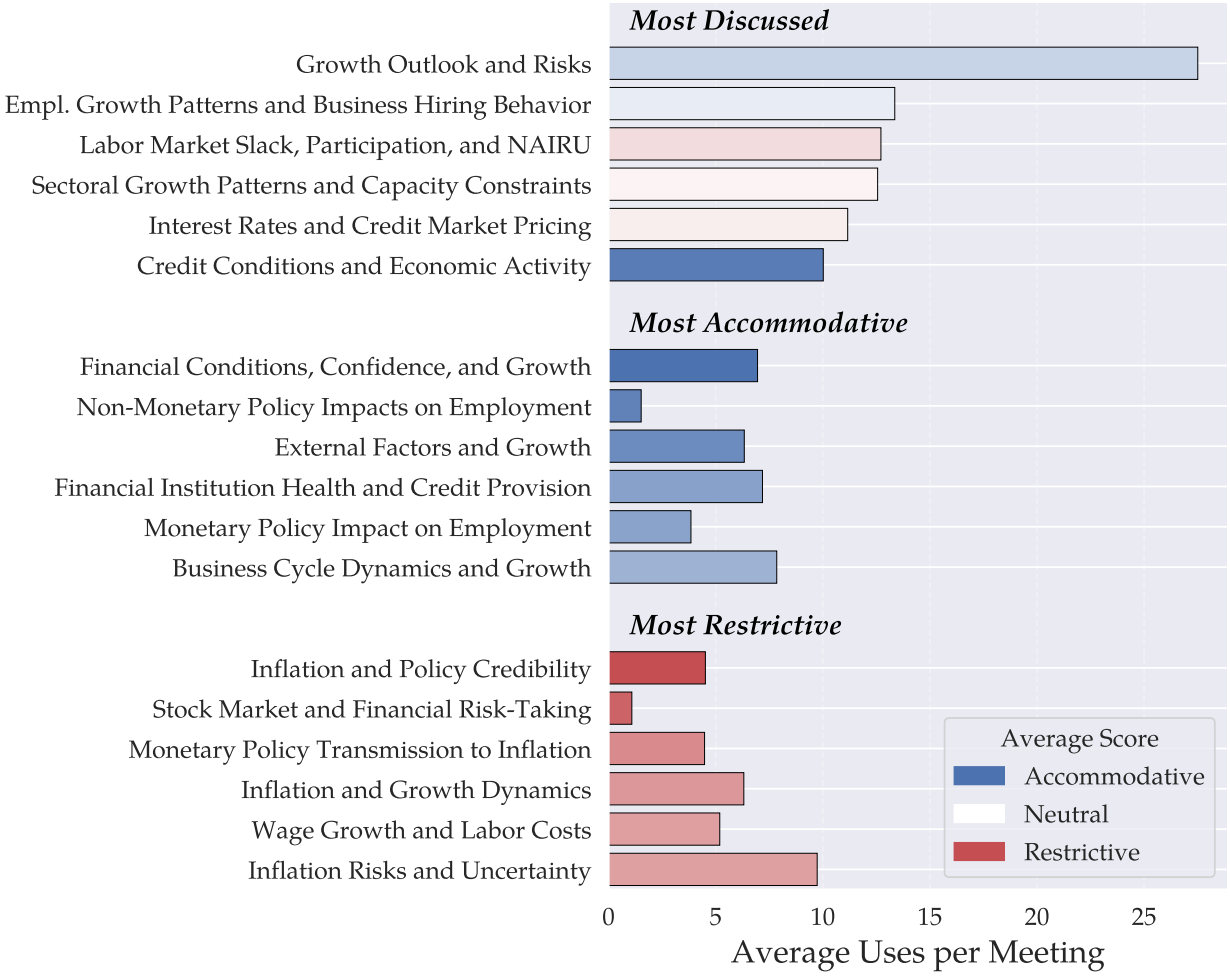


FIGURE 4: ARGUMENT TOPIC CATEGORIES

Note: The figure displays the average number of arguments made per meeting for selected categories. Bar color indicates average policy stance score, ranging from blue (accommodative) through white (neutral) to red (restrictive). Categories are grouped into most frequently discussed, most accommodative (average score < -0.5), and most restrictive (average score > 0.5).

	Selection (%)	Interpretation (%)	Residual (%)
All	8.2	91.1	0.7
Inflation	12.9	85.2	2.0
Growth	9.7	89.7	0.6
Employment	13.4	82.5	4.1
Credit Markets	13.2	85.6	1.3
Stock Market	19.1	70.2	10.7

TABLE 8: WITHIN-MEETING VARIANCE DECOMPOSITION OF MEMBER SCORES

Note: The table displays variance decompositions of members' scores for different variables into (i) a selection effect capturing the average score of arguments made by other members of the same category composition; (ii) an interpretation effect capturing the average difference between a member's score and arguments made by other members of the same category; and (iii) a residual effect that is the difference between the average score of all of a member's arguments versus the average score of those in categories also discussed by other members. The decomposition ignores covariances and normalizes the sum of variances to 100%. The first row reports decompositions constructing a member's score by averaging across all arguments and the last five rows average across arguments made with respect to a given variable.

devoted nearly one-third of his arguments to arguments classified under *Stock Market Valuation and Fundamentals*, viewing market turmoil as a signal of economic fragility.

Interpretation effects are best reflected in the discussion of *Growth Outlook and Risks*. Tarullo warned that "the economy is now sufficiently vulnerable that a modest shock could send us back into recession," receiving a score of -3.0. Plosser, discussing the same topic, saw "signs the economy is gradually improving," receiving a score of +1.0—a four-point interpretation gap. The difference of 3.07 points between their overall scores decomposes into 13.1% from selection and 86.9% from interpretation.

To understand the importance of selection versus interpretation, Table 8 presents a variance decomposition of the contributions of the selection, interpretation, and residual components to cross-sectional differences in members' scores.[16] The first row reports decompositions pooling arguments across variables and averaging scores by member-meeting. The next five rows present decompositions for each variable.

Focusing on the first row, which pools arguments across variables, interpretation effects account for 91.1% of total variance in member scores, while selection effects contribute 8.2%. Residual effects from exclusive categories explain only 0.7%, consistent with high topic overlap across members within meetings. The last five rows demonstrate that this pattern is consistent across variables, with interpretation effects ranging from 70.2% (stock market) to 89.7% (growth).[17]

[16]The decomposition omits covariances and normalizes the variance contributions to sum to 100%. Internet Appendix Table IA.C.4 reports results including covariance terms, which yield identical conclusions.

[17]Internet Appendix Table IA.C.5 presents the variance decomposition without meeting fixed effects,

The results indicate that the vast majority of expressed disagreement arises from heterogeneous interpretation of the same topics, consistent with model heterogeneity. Rather than attending to different parts of the economy, members largely attend to the same aspects of the economy but draw different conclusions.

4 Monetary Policy by Committee

Given the evidence that members hold largely fixed models to interpret the same economic data, we next analyze how the FOMC aggregates these models in its decision-making. Our transcript-based data include a complete set of committee decisions and track each member's position and influence across these decisions, providing additional detailed insight into the aggregation process relative to previously studied measures. Using these novel measures, we find that the committee's decision tilts towards the preferences of members whose models better align with recent data.

4.1 Overview and Framework

Before proceeding, we outline a basic framework to guide and interpret our analysis. In period t, given economic data D_t commonly observed by committee members, we represent how well a member i's model matches the data as $\mathcal{L}(D_t|m_i)$. We use the notation $\mathcal{L}$ to stand in for the likelihood of the observed data given a member's model, as in the theoretical analysis presented in Section 5. Each member i has a policy preference in t, denoted by $\gamma_{i,t}$. The committee makes a decision

$$Decision_t = \sum_i w_{i,t}\gamma_{i,t}, \tag{3}$$

where $w_{i,t}$ represents the (implicit) weight given to member i in the period t decision.

We test whether committee members whose models better align with recent data have more influence on the committee's decisions: that is, whether an increase in $\mathcal{L}(D_t|m_i)$ increases $w_{i,t}$. This prediction stands in contrast to the assumption in prior work, that the committee equally weights its members' preferences, that is, $w_{i,t}$ is equal across members (e.g., Malmendier, Nagel and Yan (2021)).

We estimate regressions of the form

$$y_{i,t} = \alpha_t + \eta_i + \beta\mathcal{L}(D_t|m_i) + controls, \tag{4}$$

where $y_{i,t} \in \{Alignment_{i,t}, Influence_{i,t}, Dissent_{i,t}\}$, α_t captures meeting fixed effects,

where interpretation explains 67.6% of score variation and selection explains 31.9%. The increased selection variance is consistent with business cycle variation in topics explaining across-meeting score variation.

η_i captures member fixed effects, and β captures the relationship between member i's model fit and their alignment with the committee's decision.

The prediction of canonical homogeneous prior models of differential information is that $\beta \approx 0$. Assuming that members fully share their information in the economic situation discussion, members should share common beliefs, providing no scope for information-based differences in alignment with the committee's decision. If some members tend to have less precise information and the committee weights by precision, or if some members tend to have flexible, data-driven models and dynamically drive the decision, such effects, as well as any time-invariant preferences, should be captured by member fixed effects.

A coefficient of $\beta > 0$, which we find, indicates that members whose models better match the data have preferences that are better reflected in the committee's decision. Note that because the regressions always include time fixed effects, a positive β cannot be explained by the fact that all members update their beliefs and policy preferences in the direction of the incoming data.

To interpret the positive β, we can approximate our dependent variables of interest as $y_{i,t} \propto cov(\gamma_{i,t}, \sum_j w_{j,t}\gamma_{j,t})$, i.e., if member i's policy preferences are well aligned with the weighted sum of all members' preferences, member i receives a high alignment score. Our interpretation of $\beta > 0$ is that a higher model fit for member i (higher $\mathcal{L}(D_t|m_i)$) coincides with more influence on the decision (higher $w_{i,t}$) and hence, higher alignment (higher $cov(\gamma_i, \sum_j w_j\gamma_{j,t})$). Regressions with $Influence_{i,t}$ as the dependent variable further provide direct evidence that members with better-fitting models have a higher weight, $w_{i,t}$, by more actively shaping the decision.

4.2 Committee Decisions and Model Aggregation

To proceed, we represent the output of member i's model in period t as a vector of their scores by variable, i.e., as $(\gamma_{i,t,k})_{k\in\mathcal{K}}$ for $\mathcal{K}$ ={Inflation, Growth, Employment, Credit, Stocks}, where $\gamma_{i,t,k}$ is member i's score in meeting t for variable k.

We then estimate how well member i's model matches incoming data. Representing the incoming data as the vector $D_t = (d_{t,k})_{k\in\mathcal{K}}$ where each $d_{t,k}$ is signed so that a positive sign corresponds to hawkish policy, we define our variable of interest capturing member i's model fit as

$$\mathcal{L}(D_t|m_{i,t}) = -\sum_k |\gamma_{i,t,k} - d_{t,k}|. \tag{5}$$

We standardize $\gamma_{i,t,k}$ and $d_{t,k}$ to have zero mean and unit standard deviation to ensure comparable scales. More positive values of $\mathcal{L}(D_t|m_{i,t})$ indicate member i's model better fits recent data.

In our main analysis, we focus on two measures of member i's model in period t: a vector of scores across different variables from meeting t, and a vector of member fixed effects for scores for each variable, estimated over the full sample after controlling for meeting fixed effects.

To operationalize our model fit variable, we measure $d_{t,\text{Growth}}$ and $d_{t,\text{Employment}}$ using Tealbook forecasts of real GDP growth and the unemployment rate for the quarter following t. We measure $d_{t,\text{Inflation}}$ using forecasts of the following quarter's headline CPI growth (from 1983 onward) and GDP deflator growth (pre-1983).[18] The Tealbook is distributed days before the FOMC meeting, reflecting up-to-date forecasts of the variables of interest. It reflects real-time data in the shared information set of members and has been commonly used to estimate the Fed's policy rule (Orphanides, 2001; Coibion and Gorodnichenko, 2012). We measure $d_{t,\text{Stocks}}$ and $d_{t,\text{Credit}}$ as the 1-month stock market return (from Kenneth French's website) and the (negative of the) BAA–$10Y$ credit spread (from the St. Louis Federal Reserve Bank website). We express each $d_{t,k}$ as an expanding window Z-score, to ensure similar scale across dimensions.

We next estimate regressions of members' alignment and influence with respect to the decision on their model fit, as described in Section 4.1. The dependent variables are $Alignment_{i,t}$, $Influence_{i,t}$, and $Dissent_{i,t}$, an indicator variable equal to one if member i dissented. The independent variable of interest is the fit of member i's model to the incoming data in period t, $\mathcal{L}(D_t|m_{i,t})$. We also include meeting fixed effects, member fixed effects, and role fixed effects (Chair, Regional Fed President, or Governor), the last motivated by the potential strength of the committee chair's influence, and by work emphasizing differential dissenting behavior by governors and regional presidents (Belden, 1989; Chappell Jr, McGregor and Vermilyea, 2004).

With the exception of $Dissent_{i,t}$, all variables are standardized to have zero mean and unit standard deviation. Consequently, coefficients can be loosely interpreted as the standard deviation change in a member's alignment with the committee's decision corresponding to a one standard deviation change in the independent variable. For $Dissent_{i,t}$, the coefficients represent the percentage point change in the probability of dissent associated with a one standard deviation change in the independent variable.

Table 9 reports the regression results. Focusing on the first six columns, where a member's model in meeting t reflects arguments made in t, with $Alignment_{i,t}$ as the dependent variable, the regression coefficients range from 0.12 (with meeting and member fixed effects) to 0.23 (without member fixed effects), with t-statistics ranging from 4.00 to 6.23. With $Influence_{i,t}$ as the dependent variable, the coefficients are 0.07

[18]The sample of Greenbook forecasts for headline CPI begin with the October 1979 meeting, whereas the history of forecasts of the GDP deflator goes back to 1967. We use the GDP deflator forecasts for an additional 3 years before the CPI forecast sample begins due to excess volatility of the expanding Z-scores in small samples. Our results are similar using GDP deflator forecasts for the full sample.

(with only meeting fixed effects) and 0.04 (with meeting and member fixed effects), with t-statistics of 3.48 and 2.55. These results indicate that in a given meeting, the committee's decision is more heavily aligned with and influenced by members with better-fitting models. The decline in coefficients with the inclusion of member fixed effects suggests that part of the relationships between the fit, alignment, and influence are captured by members whose models persistently match recent data tending to influence the committee's decision on average; but a substantial part comes from members being more influential in periods in which their models better fit the data.

With $Dissent_{i,t}$ as the dependent variable, coefficients range from -3.43 (with meeting and member fixed effects) to -5.71 (without member fixed effects), indicating that a one standard deviation decrease in model fit corresponds to a 3.4% to 5.7% increase in dissent probability. The unconditional dissent probability is 6.9%, indicating a quantitatively substantial increase in dissent probability corresponding to models that poorly fit recent data.

The last six columns of Table 9 report results where models are measured using member fixed effects for each variable. Across specifications, the effect sizes in the regressions are of a similar or slightly larger magnitude than those in the first six columns. That is, the match between macroeconomic data and a time-invariant component of members' models largely explains the relationship between how well their models fit recent data and their alignment with the committee's decisions.

Appendix IA.C.3 reports variations of the results in Table 9, which all yield very similar conclusions. Table IA.C.6 presents results without controls for member role. The results are generally stronger, driven by the chair being influential and having a well-fitting model, though this relationship does not explain our main results due to the inclusion of role fixed effects. Table IA.C.7 presents results where models are measured using arguments from the $t-1$ meeting, temporally separating the measurement of models from decision influence and alignment. Table IA.C.8 includes non-voting members in the regressions. Tables IA.C.9 and IA.C.10 separately report results where models are measured using only traditional macroeconomic and financial variables, respectively, showing that both help explain the results. Table IA.C.11 presents results using trailing macroeconomic realizations rather than Tealbook forecasts, and Table IA.C.12 presents results where incoming data are constructed by a textual analysis of the Tealbook, based on the idea in Aruoba and Drechsel (2024) that the text contain information not captured by quantitative Tealbook forecasts. Table IA.C.13 reports results excluding observations from 2007 through 2009 and shows that the results are not driven by the Global Financial Crisis.

	$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$		$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model Fit: $\mathcal{L}(D_t\|m_{i,t})$	0.23	0.12	0.07	0.04	-5.71	-3.43						
	(6.23)	(4.00)	(3.48)	(2.55)	(4.87)	(3.51)						
Model Fit: $\mathcal{L}(D_t\|m_{i,FE})$							0.30	0.16	0.11	0.09	-6.83	-3.74
							(7.72)	(2.98)	(4.26)	(3.46)	(5.31)	(2.00)
N	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972
Meeting Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Member Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Role Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

TABLE 9: MODEL FIT AND POLICY DECISION ALIGNMENT

Note: The table reports results from regressions of member i's alignment with the committee's policy decision in period t on $\mathcal{L}(D_t|m_i)$ variables, which capture how well committee member i's model fits the data in period t. The $\mathcal{L}(D_t|m_i)$ variables are constructed as the absolute value of the difference between committee member i's scores across macroeconomic and financial variables ('their model') and macroeconomic and financial data for meeting t. The model $m_{i,t}$ is constructed using arguments by member i in meeting t and the model $m_{i,FE}$ is constructed from member fixed effect in regressions of variable scores on meeting and member fixed effects. The macroeconomic data are forecasts of inflation, growth, and unemployment from the Greenbook/Tealbook A and the financial data are trailing 1-month stock market returns and the $BAA - 10Y$ credit spread. The dependent variables are $Alignment_{i,t}$, a score from -3 (complete opposition) to 3 (perfect alignment) to the policy decision; $Influence_{i,t}$, a score from 0 (no influence) to 3 (pivotal influence) in the decisions adopted by the committee; and a 0/1 indicator variable that equals 1 if member i dissented in meeting t. The sample consists of voting members during FOMC meetings. Standard errors are clustered by meeting and committee member. t-statistics are reported in parentheses.

4.3 Link to Monetary Policy Surprises and Shocks

Our results show that the committee's decisions tilt toward the preferences of members whose models better fit recent data. In Appendix IA.C.1, we link our policy decisions data to monetary policy surprises, measured as the high-frequency asset price response of interest rate futures in the 30-minute window following policy announcements (Gürkaynak, Sack and Swanson, 2005; Nakamura and Steinsson, 2018). This relationship underscores that the committee's tilting behavior has a material effect on policy, with the resulting decisions surprising financial markets. The result is also consistent with recent evidence of the FOMC using a time-varying policy rule (Hack, Istrefi and Meier, 2023), and suggests that monetary policy shocks may reflect "change in the political power of individuals on the FOMC," (Ramey, 2016), with such change driven by the fit of members' models to recent data.

5 Theoretical Framework

Our evidence indicates that committee members have distinct and largely fixed models of the macroeconomy, with the committee tilting its decision toward the preferences of members whose models better fit the recent data. These results indicate that the traditional explanation for committee decision-making—to aggregate diverse information across members—is an insufficient explanation in the case of the FOMC.

Guided by the evidence and to explore its implications, we present a theory of group decision-making where groups provide a way to make decisions in the face of model uncertainty by bringing together members with different models.

5.1 Setup

There are two committee members indexed by $i \in \{A, B\}$. Let μ be the true probability measure on a measurable space $(\Omega, \mathcal{F})$, which is unknown to committee members. A *model* is a probability distribution over the state space Ω, where Ω can be taken to represent all possible macroeconomic realizations. The committee's objective is to select a model, which, in turn, maps to its policy decision.

Each member i has their own personal, fixed model, m_i, which is the density of a probability measure, M_i, on $(\Omega, \mathcal{F})$. We use lowercase to refer to probability densities and uppercase to refer to the corresponding measures throughout. We make three assumptions about members' models.[19]

[19]For expositional ease, we assume that members' models are not updated upon observing data. The results hold with updating if members have heterogeneous prior models and the assumptions are applied to the models they possess upon observing data.

Assumption 1 (Absolute Continuity) *Members' models assign positive probability to all possible states, and only to those states (models are mutually absolutely continuous with respect to μ).*

Assumption 2 (Partial Misspecification) *Members' models are partially misspecified ($\forall i$, $\exists S_i \subseteq \Omega$ such that $\mu(S_i) > 0$ and $M_i(S_i) \neq \mu(S_i)$).*

Assumption 3 (Distinctness) *Member i's model is distinct from member j's ($\exists S \subseteq \Omega$ such that $\mu(S) > 0$ and $M_i(S) \neq M_j(S)$).*

The committee makes its decision as follows. Both members observe data D_t from a prior period, which is a realization of a random variable—for example, realized CPI and GDP growth—whose distribution depends on the state of the world. For any state ω, the probability of observing D_t is $\pi(D_t|\omega)$. Given a model p, the likelihood of observing D_t, $\mathcal{L}(D_t|p) = \int_{\omega\in\Omega} \pi(D_t|\omega)dP(\omega)$, is a weighted average of the conditional probabilities $\pi(D_t|\omega)$ over the state space, where the weights are given by p. Given the observed data, likelihood calculations give higher scores to models that assign high probability to states that appear more likely in light of the data.

After observing the data, each member i advocates for a convex combination of committee members' models,

$$v_i = \alpha_i m_i + (1-\alpha_i)m_j, \text{ where } j \neq i \text{ and } \alpha_i \in [0,1]. \tag{6}$$

The weight $\alpha_i \in [0,1]$ reflects how strongly member i advocates for their own model, e.g., how actively they influence policy. The committee's model m, which maps to the decision of the committee, is given by averaging the members' advocated models:

$$m = \frac{1}{2}(v_A + v_B) = w_A m_A + (1-w_A)m_B, \tag{7}$$

where $w_A \equiv \frac{\alpha_A + (1-\alpha_B)}{2}$ is the weight that member A's model receives in equilibrium. Member i chooses α_i to maximize the objective function

$$U_i(\alpha_i;\alpha_j) = \log \mathcal{L}(D_t|m) - \lambda KL(m_i\|v_i), \tag{8}$$

where $KL(p\|q) = \int_{\omega\in\Omega} \log\left(\frac{p(\omega)}{q(\omega)}\right) dP(\omega)$ is the Kullback-Leibler divergence, and the coefficient λ is positive.

The first term, $\log \mathcal{L}(D_t|m)$, captures that members want a chosen committee model m that explains the data well, all else equal. This term may reflect that members are aware that their models are misspecified, and want to use data in the face of model uncertainty. Even if members are dogmatic that their own models are correct, such a term may still arise. For example, it may reflect difficulty advocating for one's own

model when it is misaligned with the observed data, or it may reflect that the committee must justify its decision to constituencies that also observe D_t. This term leads the committee to tilt its decision toward models that better fit D_t, consistent with the empirical evidence.

The second term, $-\lambda KL(m_i \| v_i)$, captures a penalty for the deviation of member i's advocacy from their personal model. The coefficient λ captures how this penalty scales with deviations, $KL(m_i \| v_i)$. A high λ may emerge from a combination of behavioral factors—such as stubbornness or overconfidence—as well as institutional factors—such as not being sufficiently exposed to other models' merits.

5.2 Equilibrium

We study pure strategy Nash equilibria.

Definition 5.1 (Equilibrium) *A Nash equilibrium is a collection of weights, $\{\alpha_A, \alpha_B\}$, such that*

$$\alpha_i \in \arg\max_{\alpha^i} U_i(\alpha^i; \alpha_j),$$

where $U_i(\alpha^i; \alpha_j)$ denotes the utility of committee member i given their own chosen weight α^i and the other member's chosen weight, α_j.

Proposition 1 (Equilibrium Existence and Uniqueness) *There is a unique equilibrium.*

Proof. All proofs are provided in Appendix A. □

5.3 Committee Performance

We next explore the conditions under which the committee improves decision-making. We evaluate the performance of a model using the model's (ex-ante) expected log likelihood under the true probability measure, $\mathbb{E}_\mu(\log m)$. This is a natural measure, as choosing m to maximize $\mathbb{E}_\mu(\log m)$ is equivalent to minimizing the KL divergence between the true model and m, i.e., minimizing the loss of information from using m to approximate the true probability distribution.

Proposition 2 (Model Diversification) *If each member i's model better explains a non-negligible part of the state space than the other member's model does $\left(\int_\Omega \frac{m_i(\omega)}{m_j(\omega)} d\mu(\omega) > 1, j \neq i\right)$, then there is a set of optimal weights on members' models, $\mathbf{w}^* = (w_A^*, w_B^*)$, such that the performance of a weighted average of the members' models—its expected log likelihood under the true distribution—exceeds that of even the best committee member. That is, there is a set of optimal weights $\mathbf{w}^*$ that satisfy*

$$\mathbb{E}_\mu(\log(w_A^* m_A + w_B^* m_B)) > \max\{\mathbb{E}_\mu(\log m_A), \mathbb{E}_\mu(\log m_B)\}.$$

Proposition 2 provides a necessary condition for committees to outperform their best member, namely that each member's model explains part of the state space better than the other's. This result is independent of the relative performance of the models. For example, a member B may have a model that is substantially more misspecified than member A's model, but as long as the models explain different parts of the state space, there can be benefits to combining them.[20]

However, even with "diversifying" models, in equilibrium, the committee does not, in general, choose the optimal weights $\mathbf{w}^*$. First, because the committee settles upon its model using data—which has an element of randomness—its chosen weights, $\mathbf{w} \equiv (w_A, 1 - w_A)$, will vary. Second, there may be bias, stemming from committee members' stubbornness in advocating for their own models, λ.

To understand the relationships between stubbornness, bias, and variance in the committee's choice, it is useful to consider how the committee's model changes with λ. When $\lambda = 0$, the committee chooses the model that best fits the observed data, where the weights on models are unbiased relative to the optimal w^*, but are sensitive to noise in the data. As $\lambda \to \infty$, each member advocates for their own model, and in equilibrium, the chosen model is always $m = \frac{1}{2}m_A + \frac{1}{2}m_B$. That is, there is no variability, but the weights on members' models are biased relative to the optimum (except when $w_A^* = \frac{1}{2}$). For intermediate values, as λ increases, the equilibrium model weights exhibit more bias and less variance relative to the optimum.

Whether the benefits of committees are realized or not depends upon the bias and variability of the committee's chosen weights relative to the optimal weights. The committee outperforms its best member in expectation when the benefits from model diversification are sufficiently large to offset potential bias from members' stubbornness and random variability in the committee's decision.

Proposition 3 (Committee Outperformance) *Denote $w_A(\lambda)$ as the equilibrium weight on model A. Define*

$$\Delta \equiv \mathbb{E}_\mu(\log(w_A^* m_A + (1 - w_A^*)m_B)) - \max\{\mathbb{E}_\mu(\log m_A), \mathbb{E}_\mu(\log m_B)\}, \tag{9}$$

$$\beta(\lambda) \equiv \mathbb{E}_\mu(w_A(\lambda) - w_A^*), \tag{10}$$

$$\sigma^2(\lambda) \equiv \mathbb{V}_\mu(w_A(\lambda)), \text{ and} \tag{11}$$

$$c \equiv -\mathbb{E}_\mu\left(\frac{\partial^2}{\partial w_A^2}\log \mathcal{L}(D_t|w_A m_A + (1 - w_A)m_B)|_{w_A = w_A^*}\right), \tag{12}$$

where Δ is the expected outperformance of the optimal model relative to the best member's model, $\beta(\lambda)$ is the expected bias of the committee's weight on model A relative to the optimal

[20] The model diversification assumption is not difficult to satisfy. For example, with three states, $(\omega_1, \omega_2, \omega_3)$, each with probability $\frac{1}{3}$ under μ, the assumption is satisfied $\forall \epsilon \in (0, \frac{1}{3}]$ if member A assigns probabilities $(\frac{1}{3} - \epsilon, \frac{1}{3}, \frac{1}{3} + \epsilon)$ and member B assigns probabilities $(\frac{1}{3} + \epsilon, \frac{1}{3}, \frac{1}{3} - \epsilon)$.

weight, $\sigma^2(\lambda)$ is the variance of the committee's weight on model A, and c is the curvature of the expected log-likelihood evaluated at the optimal model weights. Under a second-order approximation, if

$$\Delta > \frac{c}{2}(\beta(\lambda)^2 + \sigma^2(\lambda)), \tag{13}$$

then $\mathbb{E}_\mu(\log(w_A(\lambda)m_A + (1 - w_A(\lambda))m_B)) > \max\{\mathbb{E}_\mu(\log m_A), \mathbb{E}_\mu(\log m_B)\}$.

Proposition 3 provides conditions under which the committee may outperform its best member. First, there must be sufficient gains from model diversification (Δ). These gains are higher when members' models are more complementary in explaining the state space. Second, members must be sufficiently willing to entertain each other's models (λ must not be too high). Third, the data must be sufficiently informative as to not introduce too much random variation into the committee's choice of model when the committee tilts toward models that better fit the data.

Interestingly, while too much stubbornness hinders the committee's performance, it may not be optimal for members to be fully data-driven ($\lambda = 0$). When the maximum likelihood weights on members' models (those chosen with $\lambda = 0$) are sufficiently variable, the committee's performance is maximized with some stubbornness.

Corollary 1 (Stubbornness Can Improve Committee Performance) *Denoting $w_A(\lambda)$ as the committee's equilibrium weight on model A, under a second-order approximation, when*

$$\mathbb{V}_\mu(w_A(0)) > \left(\frac{1}{2} - w_A^*\right)^2, \tag{14}$$

the expected log likelihood of the committee model, $\mathbb{E}_\mu(\log(w_A(\lambda)m_A + (1 - w_A(\lambda))m_B))$, is maximized at $\lambda^* > 0$.

The logic of Corollary 1 is that, to a second-order, maximizing the expected performance of the committee's model, $\mathbb{E}_\mu(\log(w_A M_A + (1 - w_A)M_B))$, is equivalent to minimizing the expected mean-squared error of the chosen weights with respect to w^*,

$$\mathbb{E}_\mu\left[(w_A - w_A^*)^2\right] = \underbrace{(\mathbb{E}_\mu(w_A - w_A^*))^2}_{\text{Bias}} + \underbrace{\mathbb{V}_\mu(w_A)}_{\text{Variance}}. \tag{15}$$

Some stubbornness can help by reducing excess sensitivity to noisy data in model selection. This can be understood in Bayesian terms as putting some weight on a prior of equal weighting models and not too much posterior weight on data when the data are noisy. Stubbornness comes at the cost of pushing the committee towards the biased, equally-weighted model, and is particularly valuable when the benefit of reducing the variability of w_A is high relative to the bias of equal weighting.

Remark 1 (Model Variance and Communication) *Here, the cost of variation in the weights assigned to each member's model from randomness in the data is purely statistical. There may*

also be other costs, for example, related to time-inconsistency and the committee's credibility in the eyes of market participants, as emphasized in the rules versus discretion debate (Kydland and Prescott, 1977; Barro and Gordon, 1983). Variation in members' weights may also make it difficult for the committee to communicate its future stance, consistent with our empirics relating the committee's tilting behavior to monetary policy surprises, potentially stemming from market participants not understanding the FOMC's reaction function.[21]

5.4 Transparency and Dissent Costs

We consider a simple extension to include costs for dissenting from the committee's decision. This analysis is motivated by work suggesting that making its deliberations public through the release of transcripts increased conformity on the committee (Meade and Stasavage, 2008; Hansen, McMahon and Prat, 2018; Iaryczower, López-Moctezuma and Moscariello, 2025).

We modify committee members' objective functions in Equation (8) to be

$$U_i(\alpha_i; \alpha_j) = \log \mathcal{L}(D_t|m) - \lambda KL(m_i \| v_i) - \gamma KL(m \| v_i). \tag{16}$$

The newly introduced term, $\gamma KL(m\|v_i)$, captures a dissent cost parametrized by γ for advocating for a model that deviates from the committee's choice.

Proposition 4 (Equilibrium Multiplicity) *There is a $\bar{\gamma}$ such that $\gamma < \bar{\gamma}$ implies that the equilibrium is unique, and $\gamma > \bar{\gamma}$ implies that there are multiple equilibria.*

All else equal, higher dissent costs can help push the committee toward an optimal model by offsetting members' costs of advocating for models different from their own. However, sufficiently high dissent costs introduce multiple equilibria, and the committee may settle for a sub-optimal model due to conformity. This outcome is particularly likely when the data provide little guidance on the optimal model. In addition to poorly matching the true model, multiplicity may also impose real economic costs by making the committee's decision less predictable.

5.5 Implications for Committee Design

Our framework illustrates novel tradeoffs and practical implications with committee decision-making of the form we document at the FOMC, where the committee's decisions tilt toward the preferences of members whose models better fit recent data. Committees may improve decision-making, but member selection is important, as

[21] Prior work has recognized that committees may give rise to a 'cacophony of voices' on policy, rather than a single unified voice, e.g., Blinder (2007, 2008). Vissing-Jorgensen (2021) argues that the committee structure leads to a 'communication arms race' among members seeking to influence market expectations.

potential benefits require members with a diversity of models who are sufficiently willing to entertain one another's views. One way to push members in such a direction is via structured deliberation, for example, as in FOMC meetings, where members discuss their interpretation of economic conditions and policy views before voting.

The results also highlight a role that the chair may play in deliberations. The framework emphasizes the importance of members not adhering too much to their own models, as well as the potential costs of conformity. The chair may play an important role in balancing the two to avoid underperforming models.

Lastly, our framework sheds light on policies affecting the cost of dissent, such as external transparency of deliberations. Increasing dissent costs can offset members' tendency to overly advocate for their own models but may increase the possibility of converging to a sub-optimal model and decrease the predictability of decisions.

6 Conclusion

We study group decision-making at the FOMC. We find that members of the FOMC have distinct models—each reflecting a different way of interpreting economic data—that lead to persistent differences in policy preferences. The committee tilts its decision towards the preferences of members whose models better fit the recent data. These findings provide a new perspective on group decision-making, moving beyond traditional theories of private information aggregation. Our theoretical analysis illustrates a novel trade-off associated with the FOMC's decision-making—the committee's tilting behavior may improve performance but also risks excess sensitivity to data.

A Proofs

Proof of Proposition 1

Proof. The game is a potential game with potential function

$$\Phi(\alpha_A, \alpha_B) = \log \mathcal{L}(D_t|m) - \lambda(KL(m_A \| \alpha_A m_A + (1-\alpha_A)m_B) + KL(m_B \| \alpha_B m_B + (1-\alpha_B)m_A)),$$

which is strictly concave given the concavity of $\log \mathcal{L}(D_t|m)$ and the strict convexity of $KL(m_i\|\cdot)$. $\Phi(\alpha_A, \alpha_B)$ has a unique maximizer given strict concavity, and since the strategy space is convex and compact, corresponding to the unique equilibrium. □

Proof of Proposition 2

Proof. Denote

$$\begin{aligned}\phi(w_A) \equiv& \mathbb{E}_\mu \log(w_A m_A + (1-w_A)m_B) \\ =& \int_\Omega \log(w_A(m_A(\omega) - m_B(\omega)) + m_B(\omega)) d\mu(\omega).\end{aligned}$$

By concavity of ϕ, an interior optimum exists if $\phi'(0) > 0$ and $\phi'(1) < 0$. Computing the derivatives, this yields two conditions:

$$\text{(i)} \int_\Omega \frac{m_A(\omega)}{m_B(\omega)} d\mu(\omega) > 1, \text{ and (ii)} \int_\Omega \frac{m_B(\omega)}{m_A(\omega)} d\mu(\omega) > 1.$$

These conditions are directly satisfied by the model diversification assumption. □

Proof of Proposition 3

Proof. Taking a second-order approximation of $\psi(w_A) = \mathbb{E}_\mu(\log \mathcal{L}(x|w_A m_A + (1-w_A)m_B))$ around w_A^*, observe that $\psi(w_A) \approx \psi(w_A^*) - \frac{c}{2}(w_A - w_A^*)^2$, where $c \equiv -\psi''(w_A^*) > 0$, and the linear term disappears because $\psi'(w_A^*) = 0$. Taking expectations, we have

$$\psi(w_A) \approx \mathbb{E}_\mu \psi(w_A^*) - \frac{c}{2}\mathbb{E}_\mu\left((w_A - w_A^*)^2\right). \tag{A.2}$$

Note that $\mathbb{E}_\mu\left((w_A - w_A^*)^2\right) = \underbrace{(\mathbb{E}_\mu(w_A - w_A^*))}_{\equiv\beta(\lambda)}^2 + \underbrace{\mathbb{V}_\mu(w_A)}_{\equiv\sigma^2(\lambda)}$, so,

$$\psi(w_A) \approx \psi(w_A^*) - \frac{c}{2}(\beta(\lambda)^2 + \sigma^2(\lambda)). \tag{A.3}$$

By assumption, we have that $\mathbb{E}_\mu(\log(w_A^* m_A + w_B^* m_B)) - \max\{\mathbb{E}_\mu(\log m_A), \mathbb{E}_\mu(\log m_B)\} > \frac{c}{2}(\beta(\lambda)^2 + \sigma^2(\lambda))$. □

Proof of Corollary 1

Proof. First, denote $w_{A,MLE} = \arg\max_{w_A \in [0,1]} \log \mathcal{L}(x|w_A m_A + (1-w_A)m_B)$, i.e., the weight on model A that maximizes the likelihood of the observed data. Then, given

the structure of the game, we can write

$$w_A(\lambda) = \alpha(\lambda) w_{A,MLE} + (1 - \alpha(\lambda)) \frac{1}{2}, \tag{A.4}$$

where $\alpha(\cdot)$ is a continuous and strictly decreasing function that satisfies $\alpha(0) = 1$ and $\lim_{\lambda \to \infty} \alpha(\lambda) = 0$.

Next, from Equation (A.2), we see that maximizing the expected likelihood is equivalent to minimizing $MSE(\lambda) = \mathbb{E}_\mu((w_A(\lambda) - w_A^*)^2)$. From Equation (A.3), we have that $MSE(\lambda) = \beta(\lambda)^2 + \sigma^2(\lambda)$. We can observe that $MSE(\infty) = (\frac{1}{2} - w_A^*)^2$ and $MSE(0) = (\mathbb{E}_\mu(w_{A,MLE} - w_A^*))^2 + \mathbb{V}_\mu(w_{A,MLE})$. Given the assumption that $\mathbb{V}_\mu > \left(\frac{1}{2} - w_A^*\right)^2$, and from Equation (A.4) we can see that $w_A(\lambda)$ is continuous in λ. It follows that $MSE(\cdot)$ is minimized for $\lambda^* > 0$. □

Proof of Proposition 4

Proof. For existence, the strategy space $S = [0, 1] \times [0, 1]$ is convex and compact. Each member's utility $U_i(\alpha_i)$ is continuous and strictly concave in α_i due to the concavity of the log-likelihood term and the strict convexity of KL divergences. By the Maximum Theorem, best responses are non-empty, closed-valued, and upper hemicontinuous. Applying Kakutani's fixed-point theorem guarantees existence.

For multiplicity, the equilibrium is unique when $\gamma = 0$ (Proposition 1). As $\gamma \to \infty$, all strategies $\alpha_A = 1 - \alpha_B$ are equilibria. Lastly, the correspondence that maps γ to the equilibrium set is upper hemicontinuous, due to members' best responses being single-valued and continuous. Upper hemicontinuity guarantees a threshold $\bar{\gamma}$. □

References

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. 2022. "Subjective models of the macroeconomy: Evidence from experts and representative samples." *Review of Economic Studies*, 89(6): 2958–2991.

Andre, Peter, Ingar Haaland, Christopher Roth, Mirko Wiederholt, and Johannes Wohlfart. 2025. "Narratives about the Macroeconomy." *Review of Economic Studies, Forthcoming*.

Andre, Peter, Philipp Schirmer, and Johannes Wohlfart. 2023. "Mental models of the stock market."

Aruoba, S Borağan, and Thomas Drechsel. 2024. "Identifying monetary policy shocks: A natural language approach."

Barahona, Ricardo, S Casella, Kristy AE Jansen, and Vincenzo Pezone. 2024. "Do Teams Alleviate or Exacerbate Overreaction in Beliefs?"

Barro, Robert J, and David B Gordon. 1983. "Rules, discretion and reputation in a model of monetary policy." *Journal of Monetary Economics*, 12(1): 101–121.

Bastianello, Francesca, Paul H Décaire, and Marius Guenzel. 2025. "Mental Models and Financial Forecasts."

Bauer, Michael D, and Eric T Swanson. 2023*a*. "An alternative explanation for the "fed information effect"." *American Economic Review*, 113(3): 664–700.

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

Bauer, Michael D, Carolin E Pflueger, and Adi Sunderam. 2024. "Perceptions about monetary policy." *Quarterly Journal of Economics*, 139(4): 2227–2278.

Belden, Susan. 1989. "Policy preferences of FOMC members as revealed by dissenting votes." *Journal of Money, Credit and Banking*, 21(4): 432–441.

Bernanke, Ben S, and Kenneth N Kuttner. 2005. "What explains the stock market's reaction to Federal Reserve policy?" *Journal of Finance*, 60(3): 1221–1257.

Binetti, Alberto, Francesco Nuzzi, and Stefanie Stantcheva. 2024. "People's Understanding of Inflation." National Bureau of Economic Research.

Blinder, Alan S. 1999. *Central Banking in Theory and Practice.* Mit press.

Blinder, Alan S. 2007. "Monetary policy by committee: Why and how?" *European Journal of Political Economy*, 23(1): 106–123.

Blinder, Alan S. 2008. *The Quiet Revolution: Central Banking Goes Modern.* Yale University Press.

Bobrov, Anton, Rupal Kamdar, and Mauricio Ulate. 2025. "Regional Dissent: Do Local Economic Conditions Influence FOMC Votes?" *American Economic Review: Insights*, 7(2): 268–284.

Bordo, Michael, and Klodiana Istrefi. 2023. "Perceived FOMC: The making of hawks, doves and swingers." *Journal of Monetary Economics*, 136: 125–143.

Bordo, Michael D, Klodiana Istrefi, and Humberto Martínez. 2024. "Rules vs. Discretion: Decoding FOMC Policy Deliberations." National Bureau of Economic Research.

Boukus, Ellyn, and Joshua V Rosenberg. 2006. "The information content of FOMC minutes." Working Paper.

Brainard, William C. 1967. "Uncertainty and the Effectiveness of Policy." *American Economic Review*, 57(2): 411–425.

Bybee, J Leland. 2025. "The ghost in the machine: Generating beliefs with large language models."

Chappell Jr, Henry W, Rob Roy McGregor, and Todd Vermilyea. 2004. *Committee Decisions on Monetary Policy: Evidence from Historical Records of the Federal Open Market Committee.*

Charness, Gary, and Matthias Sutter. 2012. "Groups make better self-interested decisions." *Journal of Economic Perspectives*, 26(3): 157–176.

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

Cieslak, Anna, and Michael McMahon. 2023. "Tough talk: The fed and the risk premium."

Cieslak, Anna, Stephen Hansen, Michael McMahon, and Song Xiao. 2023. "Policymakers' uncertainty." National Bureau of Economic Research.

Cogley, Timothy, and Thomas J Sargent. 2005. "The conquest of US inflation: Learning and robustness to model uncertainty." *Review of Economic Dynamics*, 8(2): 528–563.

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "Why are target interest rate changes so persistent?" *American Economic Journal: Macroeconomics*, 4(4): 126–162.

Condorcet, Marquis de. 1785. "Essay on the Application of Analysis to the Probability of Majority Decisions." *Paris: Imprimerie Royale*, 1785.

Dunn, Wendy, Ellen E Meade, Nitish Ranjan Sinha, and Raakin Kabir. 2024. "Using Generative AI Models to Understand FOMC Monetary Policy Discussions."

Enke, Benjamin, Thomas Graeber, and Ryan Oprea. 2023. "Confidence, self-selection, and bias in the aggregate." *American Economic Review*, 113(7): 1933–1966.

Farmer, Leland E, Emi Nakamura, and Jón Steinsson. 2024. "Learning about the long run." *Journal of Political Economy*, 132(10): 3334–3377.

Fos, Vyacheslav, and Nancy R Xu. 2025. "When Do FOMC Voting Rights Affect Monetary Policy?"

Fos, Vyacheslav, Tommaso Tamburelli, and Nancy R Xu. 2025. "Local Monetary Policy." National Bureau of Economic Research.

Gáti, Laura, and Amy Handlan. 2023. "Monetary communication rules."

Gerling, Kerstin, Hans Peter Grüner, Alexandra Kiel, and Elisabeth Schulte. 2005. "Information acquisition and decision making in committees: A survey." *European Journal of Political Economy*, 21(3): 563–597.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus. 2021. "Five facts about beliefs and portfolios." *American Economic Review*, 111(5): 1481–1522.

Gürkaynak, Refet S, Brian Sack, and Eric Swanson. 2005. "The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models." *American Economic Review*, 95(1): 425–436.

Hack, Lukas, Klodiana Istrefi, and Matthias Meier. 2023. "Identification of systematic monetary policy."

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

Handlan, Amy. 2022. "Text shocks and monetary surprises: Text analysis of fomc statements with machine learning."

Hansen, Lars Peter, and Thomas J Sargent. 2008. *Robustness.* Princeton university press.

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

Hansen, Stephen, Michael McMahon, and Carlos Velasco Rivera. 2014. "Preferences or private assessments on a monetary policy committee?" *Journal of Monetary Economics*, 67: 16–32.

Herbst, Edward, and Fabian Winkler. 2021. "The factor structure of disagreement."

Iaryczower, Matias, Gabriel López-Moctezuma, and Paola Moscariello. 2025. "Career Concerns in Collective Decision-Making: The Federal Open Market Committee."

Istrefi, Klodiana. 2019. "In Fed watchers' eyes: Hawks, doves and monetary policy."

Johannes, Michael, Lars A Lochstoer, and Yiqun Mou. 2016. "Learning about consumption dynamics." *Journal of Finance*, 71(2): 551–600.

Ke, Shikun. 2024. "Analysts' Belief Formation in Their Own Words."

Kydland, Finn E, and Edward C Prescott. 1977. "Rules rather than discretion: The inconsistency of optimal plans." *Journal of Political Economy*, 85(3): 473–491.

Li, Zigang, Stijn Van Nieuwerburgh, and Wang Renxuan. 2025. "Understanding Rationality

and Disagreement in House Price Expectations." *Review of Financial Studies, Forthcoming*.

López-Moctezuma, Gabriel. 2023. "Sequential deliberation in collective decision-making: The case of the FOMC."

Malmendier, Ulrike, Stefan Nagel, and Zhen Yan. 2021. "The making of hawks and doves." *Journal of Monetary Economics*, 117: 19–42.

Meade, Ellen E, and David Stasavage. 2008. "Publicity of debate and the incentive to dissent: Evidence from the US Federal Reserve." *Economic Journal*, 118(528): 695–717.

Nakamura, Emi, and Jón Steinsson. 2018. "High-frequency identification of monetary non-neutrality: the information effect." *Quarterly Journal of Economics*, 133(3): 1283–1330.

Orphanides, Athanasios. 2001. "Monetary policy rules based on real-time data." *American Economic Review*, 91(4): 964–985.

Primiceri, Giorgio E. 2006. "Why inflation rose and fell: policy-makers' beliefs and US post-war stabilization policy." *Quarterly Journal of Economics*, 121(3): 867–901.

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

Romer, David. 2010. "A new data set on monetary policy: the economic forecasts of individual members of the FOMC." *Journal of Money, Credit and Banking*, 42(5): 951–957.

Sargent, Thomas J. 1999. *The conquest of American inflation.* Princeton University Press.

Sarkar, Suproteem K, and Keyon Vafa. 2024. "Lookahead bias in pretrained language models."

Scanlan, Susannah. 2023. "Attention Allocation and the Factor Structure of Forecasts."

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

Schwartzstein, Joshua, and Adi Sunderam. 2021. "Using models to persuade." *American Economic Review*, 111(1): 276–323.

Schwartzstein, Joshua, and Adi Sunderam. 2024. "Sharing Models to Interpret Data." Mimeo.

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

Sibert, Anne. 2006. "Central banking by committee." *International Finance*, 9(2): 145–168.

Swanson, Eric, and Vishuddhi Jayawickrema. 2024. "Speeches by the fed chair are more important than fomc announcements: An improved high-frequency measure of us monetary policy shocks."

Taylor, John B. 1993. "Discretion versus policy rules in practice." Vol. 39, 195–214, Elsevier.

Thornton, Daniel L, David C Wheelock, et al. 2014. "Making sense of dissents: a history of FOMC dissents." *Federal Reserve Bank of St. Louis Review*, 96(3): 213–227.

Vissing-Jorgensen, Annette. 2021. "Central banking with many voices: the communications arms race."

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in neural information processing systems*, 35: 24824–24837.

Wu, Lingxuan. 2023. "Mental Macro-Finance Models: Evidence and Theory."

Internet Appendix for

Policy by Committee

Toomas Laarits, Ben Matthies, Kaushik Vasudevan, and Will Yang

The Internet Appendix consists of three sections. Section IA.A is a data appendix presenting details on the processing of the FOMC transcripts, the LLM prompts used to extract information from the transcripts, and a sentence-level validation exercise of our argument dataset. Section IA.B presents argument categories and categories of data sources cited by committee members in their arguments. Section IA.C presents additional analyses and robustness checks mentioned in the main text.

IA.A Data Appendix

IA.A.1 Processing FOMC Data

FOMC transcripts are provided in PDF form on the FOMC "Historical Materials by Year" page.[22] We construct machine-readable text files of the meeting transcripts using the Python library PyMuPDF. We further process the resulting text by removing special characters, page numbers, and diacritic characters.

Individual turns of speaking are identified in the transcripts by capitalized titles and names. For example, statements by Ben Bernanke are preceded by "MR. BERNANKE" or "CHAIRMAN BERNANKE", depending on his role in the particular meeting. By contrast, any references to Ben Bernanke by other speakers are not capitalized, making the task of assigning speakers to each part of the transcript easier.

In the first step of processing the transcript materials, we apply a regular expression pattern to read out the speaker's name and the associated title ("MR.", "MS.", "CHAIR", "VICE CHAIR", among others). At this stage, we inspect the data for obvious typos in the name/title (such as "CHARMAN") or issues with optical character recognition (OCR) processing of certain characters. We also inspect the data for missing spaces or special characters that would stop our regex pattern from recognizing a speaker's name. This procedure results in the baseline speaker-matched textual data.

For transcripts of meetings prior to 1997, prepared speeches and reports by staff members of the Board are not included in the transcript text but are uploaded sepa-

[22] https://www.federalreserve.gov/monetarypolicy/fomc_historical_year.htm

rately. We process these PDF files using PyMuPDF or ABBYY FineReader and associate each segment with the presenter. We then link the speaker-matched materials to the relevant place in the transcript using the speaker's name and meeting date, confirming that all references to the "Appendix" in the transcript text are matched to relevant content.

In the second step, we inspect the names produced by our regex pattern and correct any obvious typos (such as "GRENSPAN"). We also separate out the title of the speaker and confirm consistent use within a meeting; occasionally the transcripts mix up "MR." and "MS." or refer to the Chair or Vice Chair in an alternative way. The President of the New York Fed serves as the Vice Chair of the FOMC, and we check that this title is correctly assigned in each case. We also reassign the titles "CHAIRMAN" and "VICE CHAIRMAN" to "CHAIR" and "VICE CHAIR", respectively, reflecting current usage and making it easier to filter the data. On occasion, several speakers are noted for the same speech, for instance, when the text is only included in the appendix. In those instances, we create duplicate rows of the data for each of the indicated speakers. As a final check, we count the words attributed to each name form and focus on low-word-count names to look for potential typos.

In the third step, we construct a dataset of full names, affiliations, and the associated titles for all members of the Board of Governors as well as the presidents of the twelve regional Federal Reserve banks. We use data from the Federal Reserve History website[23] and construct a meeting-member-level dataset. The titles of members of the Board of Governors can be Chair, Vice Chair (of the Board of Governors, distinct from the Vice Chair of the FOMC, which is a position held by the President of the New York Fed), Vice Chair for Supervision (since 2017), or Governor. Regional bank presidents are identified as such. On rare occasions, a regional Fed is represented at the meeting by the First Vice President and the title is adjusted accordingly. Separately, we use the information in the header of the transcript file or in the vote roll call to identify the meeting-by-meeting list of voting members of the FOMC. We verify the total number of voters with data from Thornton, Wheelock et al. (2014), updated by the Federal Reserve Bank of St. Louis, and also merge the identities of dissenters at each meeting.

In the fourth step, we assign speaker identifiers that stay constant across meetings regardless of the role the individual may take in a given meeting. For instance, Donald Kohn is part of 194 distinct transcripts, first as economist, then as Governor, and finally as Vice Chair of the Board of Governors. In the rare cases where two participants at a given meeting share a last name, the transcript indicates one of the names with an initial. In those instances, we assign the speakers their respective full names as the stable identifier across meetings. We also confirm that the last names of speak-

[23] https://www.federalreservehistory.org/people/affiliation

ers across meetings do not incidentally coincide. Recall that for members of the Board of Governors and for regional Fed presidents, we separately collect their names and roles, which allows us to confirm we are not identifying distinct people who happen to share a name.

In the fifth step, we seek to identify the policy discussion section of the meeting. The policy discussion is usually led off by a presentation by a Board economist, typically the director of the Division of Monetary Affairs. We therefore look for a long speech by the economist, for instance, Donald Kohn in much of the Greenspan era, followed by a discussion of policy options. We check our work by downloading agendas for each of the meetings from the Federal Reserve website and passing the agendas and our dataset of meetings to Anthropic Claude in order to identify the speeches in the dataset that correspond to the policy discussion. We manually inspect instances where the two approaches yield different starting points.

In the sixth and final step, we use Anthropic Claude to construct a dataset of demographics for all participants who ever vote in any of the meetings in our sample. In particular, we collect their birthdates, genders, party affiliations, if any, undergraduate and graduate institutions, as well as their degrees. For members of the Board, we also record the names of the presidents who nominated them, as well as the names of the presidents who re-nominated them, if any.

IA.A.2 LLM Prompts

After processing the transcripts, we provide the LLM (Anthropic Claude) with a number of prompts to construct our dataset. We provide the prompts below. The prompts structure the output of interest in machine-readable XML form, which we then systematically process.

- Prompts to identify and score committee members' arguments for each variable are in Section IA.A.2.1.
- Given their arguments, prompts to identify the data cited in committee members' arguments are in Section IA.A.2.2.
- The prompts to assign each argument made by each member to a category are given in Section IA.A.2.3.
- The prompt to identify each policy decision made during the meeting is in Section IA.A.2.4.
- Given the policy decisions made, the prompts to identify each member's alignment with the adopted decisions and their influence on those decisions' adoption are in Section IA.A.2.5.
- Prompts to extract the tenor of the discussion of the Tealbook with regard to Inflation, Growth, Employment, the Stock Market, and Credit Markets, used as an alternative way of measuring recent data in our analysis, are presented in Section IA.A.2.6.

IA.A.2.1 Prompts to Identify and Score Committee Members' Arguments

For each speaker-meeting-variable, we first take all of the member's speeches and run a prompt to identify the relevant speeches in which the committee member discusses the variable of interest. Having done so, we then input the relevant speeches into the LLM and prompt the LLM to identify all relevant arguments that are made with respect to the variable of interest, with a particular focus on the implications for policy. The variables of interest are Inflation, Employment, Growth/Output, Credit Markets, and the Stock Market.

The first prompt to identify the relevant speeches is below.

```
"""
You are an expert financial analyst tasked with reviewing
 speeches from a Federal Open Market Committee (FOMC) meeting.
 Your goal is to identify which speeches discuss a specific
 variable of interest and provide a concise, machine-readable
 summary of the findings.
```

```
First, consider the following variable of interest:

<variable_of_interest>
{variable}
</variable_of_interest>

Now, carefully read through the following speeches from the FOMC
 meeting:

<speeches>
{speeches}
</speeches>

Your task is to analyze these speeches and determine which ones
 discuss the variable of interest. Follow these steps:

1. List all speeches that mention the variable of interest, even
 indirectly.
2. Summarize the main points of each relevant speech in 1-2
 sentences.
3. Group relevant speeches into ranges based on their IDs.
4. For each range of relevant speeches, assign a relevance score
 (0-10) and explain your reasoning.
5. Extract brief, relevant quotes that support your decision (
 limit to 1-2 quotes per range).

Before providing your final answer, wrap your comprehensive
 analysis inside <speech_analysis> tags. Be as concise as
 possible while maintaining accuracy and completeness:

<speech_analysis>
1. List all speeches mentioning the variable of interest by ID.
2. For each relevant speech:
- Provide a 1-2 sentence summary of main points related to the
 variable.
3. Group speeches into ranges based on relevance and proximity of
  IDs.
4. For each relevant range of speeches, provide:
- Speech ID range
- Relevance score (0-10)
- Explanation of the relevance score (1-2 sentences)
- 1-2 brief, supporting quotes
```

```
5. Briefly discuss any potential biases or limitations in your
 analysis.
</speech_analysis>

After your analysis, provide your final output using the
 following XML structure:

<output>
<relevant_speech_ranges>
    <range>
    <range_start>[Speech ID]</range_start>
    <range_end>[Speech ID]</range_end>
    <relevance_score>[0-10]</relevance_score>
    <brief_description>[5 word max]</brief_description>
    </range>
    <!-- Repeat <range> element for each relevant speech range
     -->
</relevant_speech_ranges>
</output>

If no speeches discuss the variable of interest, output an empty
 <relevant_speech_ranges> element.

Remember to keep your analysis and output as brief as possible
 while maintaining accuracy and completeness. Focus on providing
 clear, concise information that directly addresses the variable
 of interest."""
```

The second prompt wherein the LLM is prompted to identify the arguments made is given below.

```
"""
You are an expert economic analyst tasked with evaluating
 speeches from Federal Open Market Committee (FOMC) members. Your
  goal is to identify and assess arguments related to a specific
 variable of interest and determine their implications for
 monetary policy.

First, carefully read the following FOMC member speeches:

<fomc_speeches>
{speeches}
</fomc_speeches>
```

```
The variable of interest for this analysis is:

<variable_of_interest>
{variable}
</variable_of_interest>

Your task is to analyze the speeches and identify arguments
 related to the variable of interest. For each argument, you will
  provide a structured analysis including a description,
 supporting quotation, policy implication score, and explanation.

Before providing your final output, wrap your detailed analysis
 inside <analysis_breakdown> tags. In this section:

1. List all relevant quotes mentioning the variable of interest
 from the speeches. Number each quote.
2. For each numbered quote:
a. Analyze the argument for both dovish (supporting lower
 interest rates) and hawkish (supporting higher interest rates)
 interpretations of the variable or its impact on policy.
b. Determine if the argument is a forecast, preference, or a mix
 of both:
    - Forecast: A prediction about the future level of the
     variable of interest only.
    - Preference: How the current/future level of the variable
     compares to the desired level, including potential
     consequences of deviations.
    - Mixed: If the statement contains elements of both forecast
     and preference, explicitly state this.
c. Analyze any data cited by the speaker:
    - Quantitative data: Specific numbers or statistics
    - Qualitative data: General observations or trends
    - Anecdotal data: Personal experiences or specific examples
    - Source of the data (e.g., Bureau of Labor Statistics,
     Federal Reserve Economic Data, personal interactions)
d. If no data is cited, explicitly state this.
e. Assign a preliminary score based on this analysis.
3. Review your preliminary scores and adjust if necessary for
 consistency across all arguments.
4. Synthesize the arguments and identify any overarching themes
 or patterns.

Use the following scoring system to indicate whether an argument
```

```
 supports dovish policy (lower interest rates) or hawkish policy
 (higher interest rates):
-3: Provides strong support for dovish policy
-2: Provides moderate support for dovish policy
-1: Provides slight support for dovish policy
0: Neutral
+1: Provides slight support for hawkish policy
+2: Provides moderate support for hawkish policy
+3: Provides strong support for hawkish policy

Your final output should be in machine-readable XML format. Use
 the following structure:

<output>
<argument>
    <description>[Brief description of the argument]</description
     >
    <quotation>[Supporting quotation from the speeches]</
     quotation>
    <score>[Score from -3 to +3]</score>
    <explanation>[Brief explanation of the score]</explanation>
    <forecast_or_preference>
    <type>[Forecast, Preference, or Mixed]</type>
    <explanation>[Brief explanation in 1 sentence]</explanation>
    </forecast_or_preference>
    <cited_data>[Brief description of data cited, including
     source and type (quantitative/qualitative/anecdotal). If no
     data cited, state "No data cited." Max 1 sentence.]</
     cited_data>
</argument>
<!-- Repeat <argument> block for each identified argument -->
</output>

Here's an example of the desired output structure (with generic
 content):

<output>
<argument>
    <description>Generic argument description</description>
    <quotation>"Generic quote from the speech."</quotation>
    <score>0</score>
    <explanation>Generic explanation of the score.</explanation>
    <forecast_or_preference>
```

```
    <type>Forecast</type>
    <explanation>Generic explanation of why this is a forecast.</
     explanation>
    </forecast_or_preference>
    <cited_data>Generic data citation, including source and type
     .</cited_data>
</argument>
</output>

Please proceed with your analysis and provide the structured
 output as described."""
```

IA.A.2.2 Prompts to Identify and Categorize Data Cited

Having identified the arguments made, we next classify the data cited by each of these arguments. We first sample 1000 arguments, pass them into the LLM, and prompt the LLM to construct a set of 10-20 categories of types of data cited, using the prompt below.

```
"""
You are an expert economic analyst tasked with categorizing the
 types of data sources cited in Federal Open Market Committee (
 FOMC) meetings. Your goal is to create a comprehensive list of
 categories that capture the various types of data sources
 mentioned by FOMC speakers.

Here are the FOMC meeting quotations for your analysis:

<fomc_quotations>
{quotations}
</fomc_quotations>

Your task is to analyze these quotations and create a list of
 10-20 categories that capture the types of data sources cited by
  the speakers. Follow these steps:

1. Review the FOMC quotations thoroughly.
2. Identify explicit data sources mentioned in the quotations.
3. Infer implicit data sources from the context.
4. Group similar data sources together.
5. Create initial categories based on these groups.
6. Refine and consolidate categories to reach the 10-20 range.
```

```
7. Cross-check the final list against the original quotations to
 ensure no significant data sources are missed.
8. Confirm that the final number of categories falls within the
 10-20 range.

Before providing your final list of categories, conduct a
 detailed analysis of the data sources. Wrap your thought process
  in <categorization_process> tags. This analysis should include:

a. List of explicit data sources (number each item)
b. List of inferred implicit data sources (number each item)
c. Initial grouping of similar sources (number each group)
d. Preliminary categories (number each category)
e. Refined and consolidated categories (number each category)

Ensure that you number each item in these lists to demonstrate
 thoroughness.

After your analysis, present your final list of 10-20 categories
 in the following format:

<category_list>
<category>
<name>[Category Name]</name>
<description>[Brief description of the data source type,
 including examples if applicable]</description>
</category>
...
</category_list>

Ensure that there are no redundancies and that each category
 captures a distinct type of data source. Your final list must
 contain between 10 and 20 categories.

Here's an example of how your output should be structured (note
 that this is a generic example and should not influence your
 actual categorization):

<categorization_process>
[Your detailed analysis goes here, following the steps outlined
 above]
</categorization_process>
```

```
<category_list>
<category>
<name>Economic Indicator A</name>
<description>Description of Economic Indicator A and its
 relevance to FOMC discussions.</description>
</category>
<category>
<name>Financial Market Metric B</name>
<description>Explanation of Financial Market Metric B and how it'
 s used in FOMC context.</description>
</category>
[Additional categories would be listed here to reach the 10-20
 range]
</category_list>

Please proceed with your analysis and category creation based on
 the provided FOMC quotations."""
```

With these categories in hand, we next individually input each argument into the LLM, alongside the data category names and descriptions, and prompt Claude to categorize each type of data source cited by the speaker in the argument, as well as whether the data reflect private/specialized information or public information.

```
"""
You are an expert analyst specializing in Federal Open Market
 Committee (FOMC) discussions. Your task is to analyze a
 quotation from an FOMC meeting, categorize the data used in the
 argument, and classify the information as either private/
 specialized or public.

First, carefully read the following FOMC discussion quotation:

<fomc_quotation>
{quotation}
</fomc_quotation>

Next, familiarize yourself with the numbered categories for data
 classification:

<data_categories>
{categories}
</data_categories>

Instructions:
```

```
1. Analyze the quotation to identify key data points used in the
 argument.
2. Categorize each data point using the numbered categories
 provided above.
3. Determine if the information in the argument is private/
 specialized or public.

Your consideration of each of these instructions should be a
 maximum of two sentences.

Important considerations:
- Private/specialized information refers to data that other FOMC
 members could not reasonably be expected to know.
- Data available to all FOMC members is considered public.
- Information from regional Federal Reserve banks is considered
 private/specialized.
- Tealbook/Greenbook information is considered public.
- Unique interpretations of public data are considered public
 information.

Before providing your final output, use <thought_process> tags to
  break down your analysis. In your thought process:

After your thought process, provide your final output using the
 following XML structure:

<output>
<data_categories>[Comma-separated list of category numbers]</
 data_categories>
<information_classification>
    <type>[Private/Specialized Information OR Public Information
     ]</type>
</information_classification>
</output>

Here's a generic example of the desired output structure:

<output>
<data_categories>1, 3, 5</data_categories>
<information_classification>
    <type>Public Information</type>
</information_classification>
</output>
```

```
Remember to keep your analysis concise and focused on the key
 points. Aim for brevity while maintaining clarity and accuracy.

Please proceed with your analysis and provide your output in the
 specified format."""
```

IA.A.2.3 Prompts to Categorize Members' Arguments

For each variable, we randomly sample 500 arguments five times and run a prompt to identify categories among the randomly sampled arguments. We then run an additional prompt to consolidate the categories across the different runs. Finally, argument-by-argument, we run a prompt to assign each argument to a category.

The prompt to identify categories given a set of arguments is given below.

```
"""
You are an expert financial analyst tasked with categorizing
 quotations from Federal Open Market Committee (FOMC) meetings.
 Your goal is to create a comprehensive set of distinct
 categories that encompass the economic arguments and concepts
 discussed in these meetings, focusing on causal relationships
 involving a specific economic variable of interest.

First, carefully read through the following set of FOMC meeting
 quotations:

<fomc_quotations>
{quotations}
</fomc_quotations>

The variable of interest for this analysis is:

<variable_of_interest>
{variable}
</variable_of_interest>

Your task is to analyze these quotations and create categories
 that collectively cover all the themes present in the quotes.
 Each category should represent a distinct economic argument or
 concept directly relevant to the variable of interest, with an
 emphasis on causal relationships.
```

```
Please conduct your analysis and category development process
 within <category_development_process> tags. Follow these steps:

1. Identify key phrases, economic terms, and main economic
 concepts or arguments across all quotes. List these out,
 numbering each item. For each item, provide a relevant quote
 from the FOMC quotations.
2. Assess how the identified concepts relate to policy
 preferences.
3. Identify potential causal relationships, particularly in
 relation to the variable of interest. Number each relationship
 and explicitly state how it relates to the variable of interest.
4. Identify and tally common themes and topics across all quotes.
  List these with their frequency.
5. Brainstorm potential categories based on your analysis,
 explicitly considering arguments of differing interpretations
 within each category. Number each potential category and provide
  a brief rationale for its inclusion.
6. For each proposed category, explicitly state how it relates to
  the variable of interest. If it doesn't relate directly, revise
  or discard it.
7. Refine your categories, ensuring they meet the following
 criteria. For each category, explicitly state how it meets each
 criterion:
    - Categories are mutually exclusive and non-overlapping.
    - Each category supports within-category disagreement.
    - Categories are at the "lowest level" that still supports
     within-category disagreement.
    - The level of generality/specificity is consistent across
     all categories.
8. Review and consolidate categories that may be capturing the
 same ideas. Explain your reasoning for each consolidation.
9. Explicitly check for any potential overlap between categories
 by creating a simple matrix to compare each category pair.
 Adjust categories as necessary based on this comparison.
10. For each proposed category:
    - Consider potential counterarguments or alternative
     interpretations. List these out.
11. For each category, create a brief one-sentence description
 highlighting the causal relationship it represents.
12. Summarize how each final category relates to the variable of
 interest.
13. Rank the final categories in order of relevance to the
```

```
 variable of interest, providing a brief explanation for each
 ranking.

When developing your categories, consider the following types of
 disagreements that can occur in FOMC discussions:

1. Measurement/Assessment Disagreements: Members primarily
 disagree about current economic conditions or how to interpret
 recent data, even when using similar analytical frameworks.
2. Causal/Structural Disagreements: Members primarily disagree
 about economic relationships, transmission mechanisms, or how
 variables affect each other.
3. Risk/Forecast Disagreements: Members share similar views of
 current conditions and causal relationships but disagree on
 future probabilities, uncertainties, or the likelihood of
 different scenarios.
4. Priority/Preference Disagreements: Members disagree on the
 relative importance of different objectives or acceptable
 tradeoffs between goals.

After your analysis, provide your final categories using the
 following XML format:

<categories>
<category>
    <name>Category Name</name>
    <description>Brief one-sentence description of the category,
     highlighting the causal relationship it represents, and the
     nature of disagreement within the category.</description>
    <disagreement_type>Type of disagreement represented by this
     category (choose from the four types listed above)</
     disagreement_type>
</category>
<!-- Repeat for all categories -->
</categories>

An example category output:

<category>
    <name>Growth Forecast/Outlook</name>
    <description>Projections of future economic growth scenarios,
      encompassing disagreements on the likelihood of different
     growth outcomes and their potential impacts on the economy
```

```
     .</description>
    <disagreement_type>Risk/Forecast Disagreements</
     disagreement_type>
</category>


Remember:
- Categories must be comprehensive, mutually exclusive, and non-
 overlapping.
- Each category should represent a distinct economic argument or
 concept directly relevant to the variable of interest.
- Descriptions should be concise (one sentence) but informative,
 emphasizing the causal relationships involved and explaining
 potential disagreements within the category.
- Categories should accommodate within-category disagreement.
- Ensure that categories reflect intuitive and theoretical
 economic relationships.
- The number of categories should be determined by the content of
  the quotes, not by a predetermined limit.
- Each category should represent only one type of disagreement.

Please begin your analysis and category development process now.
 """
```

The prompt to consolidate categories is given below.

```
"""
You are an expert financial analyst tasked with consolidating and
  refining categories derived from Federal Open Market Committee
 (FOMC) meeting quotations. Your goal is to create a
 comprehensive, non-overlapping set of categories that encompass
 the economic arguments and concepts discussed in these meetings,
  with a focus on causal relationships involving a specific
 economic variable of interest.

First, carefully review the following sets of categories derived
 from previous analyses:

<categories>
{categories}
</categories>

The variable of interest for this analysis is:
```

```
<variable_of_interest>
{variable}
</variable_of_interest>

Your task is to analyze these existing categories, consolidate
 them, and create a refined set of categories that collectively
 cover all the themes present in the original FOMC quotes. Each
 category should represent a distinct economic argument or
 concept directly relevant to the variable of interest, with an
 emphasis on causal relationships.

Please conduct your analysis and category refinement process
 within <category_refinement_process> tags. Follow these steps:

1. List all existing categories.
2. Group similar categories together, explaining the key elements
  that make them similar.
3. For each group of similar categories, create a consolidated
 category that captures the essence of all included categories.
 Explain your reasoning.
4. Check each consolidated category's relevance to the variable
 of interest. If it's not directly related, explain how you'll
 revise or discard it.
5. Refine your categories, explicitly stating how each meets the
 following criteria:
    - Categories are mutually exclusive and non-overlapping.
    - Each category supports within-category disagreement.
    - Categories are at the "lowest level" that still supports
     within-category disagreement.
    - The level of generality/specificity is consistent across
     all categories.
6. Create a brief one-sentence description for each category,
 highlighting the causal relationship it represents.
7. Assign each category one of the following disagreement types:
    - Measurement/Assessment Disagreements: Members primarily
     disagree about current economic conditions or how to
     interpret recent data.
    - Causal/Structural Disagreements: Members primarily disagree
      about economic relationships, transmission mechanisms, or
     how variables affect each other.
    - Risk/Forecast Disagreements: Members share similar views of
      current conditions and causal relationships but disagree on
      future probabilities or uncertainties.
```

```
    - Priority/Preference Disagreements: Members disagree on the
     relative importance of different objectives or acceptable
     tradeoffs between goals.
8. Rank the final categories in order of relevance to the
 variable of interest, providing a brief explanation for each
 ranking.

It's okay for this section to be quite long, as it involves
 detailed analysis and explanation of your thought process.

After your analysis, provide your final categories using the
 following XML format:

<categories>
<category>
    <name>Category Name</name>
    <description>Brief one-sentence description of the category,
     highlighting the causal relationship it represents, and the
     nature of disagreement within the category.</description>
    <disagreement_type>Type of disagreement represented by this
     category (choose from the four types listed above)</
     disagreement_type>
</category>
<!-- Repeat for all categories -->
</categories>

Here's an example of how a category should be structured:

    <category>
        <name>Growth Forecast/Outlook</name>
        <description>Projections of future economic growth
         scenarios, encompassing disagreements on the likelihood
         of different growth outcomes and their potential impacts
          on the economy.</description>
        <disagreement_type>Risk/Forecast Disagreements</
         disagreement_type>
    </category>


Remember:
- Categories must be comprehensive, mutually exclusive, and non-
 overlapping.
- Each category should represent a distinct economic argument or
```

```
 concept directly relevant to the variable of interest.
- Descriptions should be concise (one sentence) but informative,
 emphasizing the causal relationships involved and explaining
 potential disagreements within the category.
- Categories should accommodate within-category disagreement.
- Ensure that categories reflect intuitive and theoretical
 economic relationships.
- The number of categories should be determined by the content of
  the original categories, not by a predetermined limit.
- Each category should represent only one type of disagreement.

Please begin your analysis and category refinement process now.
"""
```

Given a set of categories and an argument, the prompt to categorize the argument is below.

```
"""
You are an expert economic analyst tasked with classifying
 arguments from Federal Open Market Committee (FOMC) meetings.
 Your goal is to analyze a given quotation and associated data,
 then classify the argument along five dimensions. Review the
 following information:

<fomc_quotation>
{quotation}
</fomc_quotation>

<quote_description>
{description}
</quote_description>

<data_cited_description>
{data_cited}
</data_cited_description>

<date_of_quote>
{ymd}
</date_of_quote>

<potential_categories>
{categories}
</potential_categories>
```

```
Classify the argument along these five dimensions:

1. Qualitative versus Quantitative: Primarily qualitative or
 quantitative data?
2. Time-Horizon: Present, historical, forward-looking, or N/A?
3. Source: Federal Reserve, Government, Survey, Academic,
 Anecdotal, Other, or N/A?
4. Geographical Scope: National, Regional/District, or
 International?
5. Category: Choose ONE number from the <potential_categories>
 section that best fits.

Wrap your analysis inside <classification_analysis> tags. For
 each dimension:
1. List relevant evidence from the quotation and associated data.
2. Summarize your final decision based on the evidence.
Keep each dimension's analysis concise and to a maximum of one
 sentence.

Then, provide your final classification in the following XML
 format:

<classification>
<qualitative_quantitative>[Classification]</
 qualitative_quantitative>
<time_horizon>[Classification]</time_horizon>
<source>[Classification]</source>
<geographical_scope>[Classification]</geographical_scope>
<category>[Single best-fit category number]</category>
</classification>

Important:
- Use only the standardized terms provided for each dimension.
- Choose only ONE number for the Category classification.
- Base your classification solely on the provided information.
- Keep your analysis brief to fit within the 1000 token limit.

Begin your analysis now."""
```

IA.A.2.4 Prompt to Identify Policy Decisions Made

To identify the policy decisions made, we input the speeches from the policy discussion section of each FOMC meeting, alongside the policy alternatives presented in

the Bluebook/Tealbook B. The prompt to identify these decisions is below.

```
"""
You are an expert economic analyst tasked with identifying and
 evaluating policy decisions from a Federal Open Market Committee
  (FOMC) meeting. Your goal is to provide a comprehensive
 assessment of all successfully adopted policy decisions and
 their impact on monetary policy stance.

‘    First, review the following materials from the FOMC meeting:

1. Speeches from the policy discussion:
<fomc_speeches>
{speeches}
</fomc_speeches>

2. Policy alternatives discussed:
<policy_alternatives>
{alternatives}
</policy_alternatives>

Your task is to analyze these materials and provide a
 comprehensive assessment of all successfully adopted policy
 decisions. Follow these steps:

1. Identify all policy decisions mentioned in the materials.
 Focus only on policy decisions, excluding personnel decisions or
  other non-policy matters.

2. For each policy decision:
a. Determine whether it was adopted or not.
b. If adopted:
    - Categorize the decision type (communication, rate decision,
      or other).
    - Evaluate the level of accommodation provided by the policy
     decision using this scale:
        -3: Provides strong support for dovish policy
        -2: Provides moderate support for dovish policy
        -1: Provides slight support for dovish policy
        0: Neutral
        +1: Provides slight support for hawkish policy
        +2: Provides moderate support for hawkish policy
        +3: Provides strong support for hawkish policy
    - Provide a brief justification for the score assigned.
```

```
    - Indicate whether the decision is pertinent to the committee
     's overall policy stance (yes/no).

3. Count the number of adopted policies to ensure
 comprehensiveness.

4. Summarize the overall policy stance adopted in this meeting.

Before providing your final analysis, wrap your thought process
 in <policy_analysis> tags. In this section:
- List and number all policy decisions mentioned in both the
 speeches and policy alternatives.
- For each decision:
a. Quote relevant phrases that indicate the policy decision.
b. Determine whether it was adopted or not.
c. If adopted:
    - Categorize the decision type (communication, rate decision,
      or other).
    - Evaluate the level of accommodation using the -3 to +3
     scale.
    - Provide a brief (1-2 sentence) justification for the score
     assigned.
    - Indicate whether the decision is pertinent to the committee
     's overall policy stance (yes/no).
- Cross-check all adopted decisions to ensure they are distinct
 and unique. If any duplicates are found, combine them into a
 single entry.
- Explicitly count the number of adopted policies to ensure
 comprehensiveness.
- Conclude with a summary of the overall policy stance.

After your analysis, provide your final output in the following
 XML structure:

<adopted_policies>
<adopted_policy>
    <description>
    [Description of the adopted policy]
    </description>
    <type>
    [Type of decision: communication, rate decision, or other]
    </type>
    <score>
```

```
    [Assigned score based on the -3 to +3 scale]
    </score>
    <justification>
    [Brief justification for the assigned score]
    </justification>
    <pertinent>
    [Yes/No - Is this decision pertinent to the overall policy
     stance?]
    </pertinent>
</adopted_policy>
<!-- Add more <adopted_policy> elements as needed -->
</adopted_policies>

Example of the output structure (using generic content):

<adopted_policies>
<adopted_policy>
    <description>
    The committee decided to maintain the target range for the
     federal funds rate.
    </description>
    <type>
    rate decision
    </type>
    <score>
    0
    </score>
    <justification>
    This decision maintains the status quo, neither tightening
     nor loosening monetary policy.
    </justification>
    <pertinent>
    Yes
    </pertinent>
</adopted_policy>
<adopted_policy>
    <description>
    The FOMC agreed to update its forward guidance on asset
     purchases.
    </description>
    <type>
    communication
    </type>
```

```
    <score>
    -1
    </score>
    <justification>
    The updated guidance suggests a slightly more accommodative
     stance, hinting at potential future easing.
    </justification>
    <pertinent>
    Yes
    </pertinent>
</adopted_policy>
</adopted_policies>

Remember to:
- Provide a comprehensive assessment covering all successfully
 adopted policies.
- Focus only on policy decisions, not personnel decisions or
 other matters.
- Categorize each decision type (communication, rate decision, or
  other).
- Use the scoring system to indicate each policy's stance (-3 for
  strongly dovish to +3 for strongly hawkish).
- Include a brief justification for each score assigned.
- Indicate whether each decision is pertinent to the overall
 policy stance.
- Ensure your final output adheres to the specified XML structure
 .
- Verify that all adopted decisions in your output are distinct
 and unique.

Begin your analysis now."""
```

IA.A.2.5 Prompts to Identify Alignment and Influence

With the policy decisions identified, we next run prompts to identify each member's alignment with each of the policy decisions made, and the degree of influence they had in the adoption of those decisions.

The prompt to identify alignment is given below.

```
"""
You are an expert analyst tasked with evaluating the alignment
 between a specific speaker's policy preferences and the final
 policy decisions made in a Federal Open Market Committee (FOMC)
```

```
 meeting. Your analysis will be based on the following
 information:

1. Speeches made during the FOMC meeting:
<speeches>
{speeches}
</speeches>

2. Policy alternatives discussed in the meeting:
<alternatives>
{alternatives}
</alternatives>

3. The policy decisions that were successfully made in the
 meeting:
<policy_decisions>
{decisions}
</policy_decisions>

4. The speaker we are focusing on:
<focus_speaker>
{stablespeaker}
</focus_speaker>

Your objective is to thoroughly analyze this speaker's statements
  in the context of the policy decisions provided in the <
 policy_decisions> section. Then, you will rate how strongly
 aligned the speaker's policy preferences are with these final
 policy decisions using this scale:

3: Perfectly aligned - The speaker's preferences fully match the
 final decisions
2: Moderately aligned - The speaker's preferences largely align
 with the final decisions, with minor differences
1: Weakly aligned - The speaker's preferences somewhat align with
  the final decisions, but with significant differences
0: Neutral - The speaker's preferences neither align nor oppose
 the final decisions
-1: Weakly opposed - The speaker's preferences somewhat oppose
 the final decisions, but with some areas of agreement
-2: Modestly opposed - The speaker's preferences largely oppose
 the final decisions, with minor areas of agreement
-3: Strongly opposed - The speaker's preferences fully oppose the
```

```
  final decisions

Before providing your final rating and justification, conduct a
 thorough analysis using the following steps. Wrap your analysis
 in <analysis> tags, focusing on conciseness and avoiding
 redundancy:

1. Identify and quote key statements from the focus speaker that
 indicate their policy preferences. Number each quote for easy
 reference.
2. Extract and list all policy decisions from the <
 policy_decisions> section. Number each decision.
3. Summarize the focus speaker's overall stance based on their
 statements.
4. For each policy decision provided in the <policy_decisions>
 section:
    a. Quote the policy decision verbatim.
    b. Briefly describe the decision.
    c. Analyze how it aligns with the focus speaker's preferences
      (use a scale from -3 to 3).
    d. Quote relevant parts of the focus speaker's statements
     that support the alignment score.
5. Calculate an average alignment score for all policy decisions.
6. Analyze the focus speaker's overall alignment with the final
 policy decisions, considering only the specific policy decisions
  provided in the <policy_decisions> section.
7. Summarize the key points of agreement and disagreement between
  the speaker and the final decisions.
8. For each point on the rating scale (-3 to 3), list arguments
 for why this rating might be appropriate, considering only the
 specific policy decisions provided.

After completing your analysis, provide your final output in this
  format:

<rating>[Insert final rating here]</rating>

<proposed_decisions>
[For each policy decision:]
<decision>
<index>[Decision number]</index>
<description>[Brief description of the policy decision]</
 description>
```

```
<alignment>[Alignment score for this decision]</alignment>
<justification>[Brief justification for the alignment score,
 including relevant quotes from the focus speaker]</justification
 >
</decision>
[Repeat for each policy decision]
</proposed_decisions>

<speaker_alignment>[Briefly summarize the overall alignment
 between the speaker and the final policy decisions, considering
 only the analyzed policy decisions]</speaker_alignment>

Ensure that your output is structured and machine-readable, using
  the XML tags provided. Here's an example of the desired output
 structure (with generic content):

<rating>0</rating>

<proposed_decisions>
<decision>
<index>1</index>
<description>Generic policy decision description</description>
<alignment>0</alignment>
<justification>The speaker's statement "..." suggests a neutral
 stance on this policy decision. While they mentioned "...", they
  also noted "...", indicating a balanced view.</justification>
</decision>
<decision>
<index>2</index>
<description>Another generic policy decision description</
 description>
<alignment>1</alignment>
<justification>The speaker expressed mild support for this
 decision, stating "...". However, they also raised concerns
 about "...", leading to a weakly aligned score.</justification>
</decision>
</proposed_decisions>

<speaker_alignment>Overall, the speaker shows a neutral alignment
  with the final policy decisions. While there are areas of
 agreement in some specific policy decisions, there are also
 areas of disagreement, resulting in a balanced perspective. The
 speaker's statements on [topic] align closely with the final
```

```
 decisions, but their views on [another topic] diverge
 significantly.</speaker_alignment>

Please proceed with your analysis and provide the structured
 output as requested, focusing only on the policy decisions
 provided in the <policy_decisions> section.
"""
```

The prompt to identify influence is given below.

```
"""
You are an expert analyst tasked with evaluating the influence a
 specific speaker exerted over the final policy decisions made in
  a Federal Open Market Committee (FOMC) meeting. Your analysis
 will be based on the following information:

1. Speeches made during the FOMC meeting:
<speeches>
{speeches}
</speeches>

2. Policy alternatives discussed in the meeting:
<alternatives>
{alternatives}
</alternatives>

3. The policy decisions that were successfully made in the
 meeting:
<policy_decisions>
{decisions}
</policy_decisions>

4. The speaker we are focusing on:
<focus_speaker>
{stablespeaker}
</focus_speaker>

Your objective is to thoroughly analyze this speaker's statements
  in the context of the policy decisions provided in the <
 policy_decisions> section. Then, you will rate how much
 influence the speaker exerted over these final policy decisions
 using this scale:

0: No influence - The speaker exerted no influence on the final
```

```
 decision. For example, the committee may have made the decision
 over their objections. Or the speaker did not explicitly discuss
  the decision.
1: Limited influence - The speaker exerted limited influence on
 the decision. For example, they may have voiced support, but
 without providing strong advocacy or defense.
2: Moderate influence - The speaker exerted moderate influence on
  the decision. For example, they may have defended the policy
 and made some arguments in support of it.
3: Strong influence - The speaker exerted strong influence on the
  decision, and was instrumental to its adoption. For example,
 they may have proposed it, or made arguments that were pivotal
 in its adoption.

Before providing your final rating and justification, conduct a
 thorough analysis using the following steps. Conduct your
 analysis inside <detailed_analysis> tags, focusing on
 conciseness and avoiding redundancy:

1. Identify and quote key statements from the focus speaker that
 indicate their policy preferences. Number each quote for easy
 reference.
2. Extract and list all policy decisions from the <
 policy_decisions> section. Number each decision.
3. Summarize the focus speaker's overall stance based on their
 statements.
4. Create a table summarizing each speaker's statements, their
 relevance to policy decisions, and a preliminary influence score
 .
5. Compare the focus speaker's statements with other speakers to
 gauge relative influence.
6. Analyze any dissenting opinions and how they were addressed.
7. For each policy decision provided in the <policy_decisions>
 section:
a. Quote the policy decision verbatim.
b. Briefly describe the decision.
c. Analyze the speaker's influence on this decision (use a scale
 from 0 to 3).
d. Quote relevant parts of the focus speaker's statements that
 support the influence score.
8. Calculate an average influence score for all policy decisions.
9. Analyze the focus speaker's overall influence on the final
 policy decisions, considering only the specific policy decisions
```

```
  provided in the <policy_decisions> section.
10. Summarize the key areas where the speaker exerted influence
 and where they did not.
11. For each point on the rating scale (0 to 3), list arguments
 for why this rating might be appropriate, considering only the
 specific policy decisions provided.

After completing your analysis, provide your final output in this
  format:

<rating>[Insert final rating here]</rating>

<proposed_decisions>
[For each policy decision:]
<decision>
<index>[Decision number]</index>
<description>[Brief description of the policy decision]</
 description>
<influence>[Influence score for this decision]</influence>
<justification>[Brief justification for the influence score,
 including relevant quotes from the focus speaker]</justification
 >
</decision>
[Repeat for each policy decision]
</proposed_decisions>

<speaker_influence>[Briefly summarize the overall influence of
 the speaker on the final policy decisions, considering only the
 analyzed policy decisions]</speaker_influence>

Ensure that your output is structured and machine-readable, using
  the XML tags provided. Here's an example of the desired output
 structure (with generic content):

<rating>1</rating>

<proposed_decisions>
<decision>
<index>1</index>
<description>Generic policy decision description</description>
<influence>1</influence>
<justification>The speaker showed limited influence on this
 decision. While they stated "...", their arguments did not seem
```

```
 to significantly sway the committee. The final decision appears
 to have been made based on other factors.</justification>
</decision>
<decision>
<index>2</index>
<description>Another generic policy decision description</
 description>
<influence>2</influence>
<justification>The speaker demonstrated moderate influence on
 this decision. They proposed "..." and defended it with
 arguments such as "...". These points were reflected in the
 final decision, suggesting the speaker's input was considered.</
 justification>
</decision>
</proposed_decisions>

<speaker_influence>Overall, the speaker showed a moderate level
 of influence on the final policy decisions. While they had
 significant input on [specific area], their influence was
 limited in [another area]. The speaker's arguments on [topic]
 appeared to have some impact on the committee's decision-making
 process, but were not always the determining factor.</
 speaker_influence>

Please proceed with your analysis and provide the structured
 output as requested, focusing only on the policy decisions
 provided in the <policy_decisions> section."""
```

IA.A.2.6 Prompts to Score Incoming Data from the Tealbook

In Table IA.C.12, we present supplemental regression analyses that measure macroeconomic and financial data based on textual analysis of a meeting's Tealbook. Below, we present the prompts used to construct these measures.

```
"""
You are an expert economic analyst tasked with evaluating
 economic and financial conditions and their potential impact on
 monetary policy decisions. Your analysis will be based on a
 provided Tealbook text and focus on a specific variable of
 interest.

First, carefully read through the following Tealbook text:
```

```
<tealbook_text>
{tealbook_text}
</tealbook_text>

Now, consider the following variable of interest:

<variable_of_interest>
{variable}
</variable_of_interest>

Your task is to analyze how this variable is discussed in the
 Tealbook text and evaluate its potential impact on monetary
 policy decisions. Follow these steps:

1. Identify relevant statements in the Tealbook text related to
 the variable of interest.

2. For each relevant statement:
a. Analyze both dovish (supporting lower interest rates) and
 hawkish (supporting higher interest rates) interpretations.
b. Evaluate the strength of the factor using this scale:
    -3: Strong support for dovish policy
    -2: Moderate support for dovish policy
    -1: Slight support for dovish policy
    0: Neutral
    +1: Slight support for hawkish policy
    +2: Moderate support for hawkish policy
    +3: Strong support for hawkish policy
c. Provide explicit quotes that support your evaluation.
d. Explain your reasoning, specifically in relation to monetary
 policy implications.
e. Assess your confidence level in the evaluation (Low, Medium,
 High).

3. Wrap your detailed analysis in <detailed_variable_analysis>
 tags. In this analysis:
a. List all relevant quotes from the Tealbook.
b. For each quote, analyze its implications for the variable and
 its potential impact on monetary policy, considering both dovish
  and hawkish interpretations.
c. Consider arguments for different strength ratings (-3 to +3),
 explicitly listing pros and cons for each potential rating in
 terms of supporting an interest rate cut or hike.
```

```
d. Make a preliminary judgment on the variable's strength.
e. Summarize your overall findings, focusing on the collective
 implications for monetary policy.
f. Double-check that all supporting quotes directly match the
 given variable and clearly support your strength evaluations.
g. Make and justify your final decisions for the variable's
 strength rating, ensuring it reflects the potential impact on
 monetary policy decisions.

4. Provide your final output in the following XML format:

<economic_analysis>
<variable>
    <name>[Variable name]</name>
    <strength_rating>[Rating from -3 to +3]</strength_rating>
    <confidence_level>[Low/Medium/High]</confidence_level>
    <supporting_quotes>
    <quote>[Relevant quote from Tealbook]</quote>
    [Additional quotes as needed]
    </supporting_quotes>
    <analysis>[Your explanation of the rating and its
     implications for monetary policy]</analysis>
</variable>
</economic_analysis>

Important Notes:
- Base your analysis solely on explicit statements in the text
 with respect to the variable of interest. Avoid inferences or
 assumptions.
- Be thorough in your analysis and provide clear justifications
 for your conclusions, always relating them back to potential
 monetary policy decisions.
- Ensure that all reasons and supporting quotes are directly
 associated with the variable of interest before finalizing your
 output.
- The <detailed_variable_analysis> section should show your
 thought process and consideration of different interpretations
 before reaching your final conclusions, with a specific focus on
  how the variable might influence decisions to cut or hike
 interest rates."""
```

IA.A.3 Sentence-Level Validation Exercise

A primary concern with the argument dataset constructed via LLM prompting is whether our prompting methodology captures all arguments made in a meeting transcript—a long set of text. To address this concern, we pass subsets of the transcripts containing relevant text to the LLM, simplifying the extraction over the set of relevant inputs. To do so, we first prompt the LLM to identify the relevant parts of the text. We subsequently prompt the LLM to identify the arguments within the parts of the text identified as relevant.

To assess the comprehensiveness of our methodology in capturing all arguments, we implement a sentence-level prompting validation approach. This process validates whether the dataset successfully captures the full set of arguments expressed in the original meeting transcripts.

We construct a pipeline that independently classifies argument-bearing sentences directly from the transcript at the individual sentence level. Then, we check whether these sentences are already represented in our main dataset, either word by word, in paraphrase, or within a broader argument. To do this, we develop a three-tiered matching pipeline to reconcile arguments identified at the sentence level back to the main dataset, which

Tier 1 identifies direct quote matches based on string similarity;

Tier 2 links the nearby sentences within a context window if the LLM decides that they belong to the same continuous, coherent argument flow;

Tier 3 and uses a broader, more general semantic match to catch some edge cases.

Applying the argument identification and result reconciliation process to a random sample of 20 meetings, we find that 91.4% of all identified argument-bearing sentences are matched to arguments in our main dataset (Table IA.A.1); furthermore, the unmatched sentences tend to be either ambiguous or not policy-relevant. They do not appear to introduce systemic directional bias or invalidate our main dataset.

	Sentence Count	Match Rate	Cumulative Coverage
Total Sentence Count	46244		
Total Argument Sentence Count	5655		
Tier 1 Matched	3526	62.3%	62.3%
Tier 2 Matched	1319	23.3%	85.7%
Tier 3 Matched	326	5.7%	91.4%

TABLE IA.A.1: MATCH RATES ACROSS MATCHING TIERS

We describe the validation process and the error analysis in more detail in the following sections.

IA.A.3.1 Sentence-Level Argument Identification

We use Claude 3.7 to label whether a sentence mentions at least one of the target variables (Growth, Inflation, Employment, Credit Markets, Stock Markets) and expresses an argument that may have implications for monetary policy. The prompt is intentionally written to align with the criteria of the argument extraction prompt for the main dataset, except at a sentence level. This classification is done independently per sentence. The total number of sentences labeled as arguments constitutes the base set of the validation procedure.

IA.A.3.2 Three-Tiered Argument Matching Pipeline

For each sentence classified as an argument, we apply a three-step matching-and-reconciliation pipeline to check whether that sentence has been captured by one of the arguments in our main dataset.

Tier 1: Direct Quote Match

In the first tier of our matching pipeline, we attempt to directly reconcile each LLM-flagged argument sentence with the *"quote"* field in our main argument dataset. This step aims at capturing cases where the sentence appears exactly the same word-by-word or with minor variations in the extracted arguments.

LLMs may have truncated, paraphrased, or otherwise modified the original transcript when generating quotes at the speaker-meeting level; therefore, we allow for fuzzy matching rather than requiring exact string equality. We use similarity thresholds to identify potential matches and manually verify a subsample to validate quality. Specifically, the similarity thresholds are specified as follows:

1. First, we compute a full-text similarity score between the tested sentence and each candidate quote (restricted to have the same speaker and meeting date) using a normalized string similarity metric based on the Python library difflib.SequenceMatcher. The highest scoring quote is retained, and if the score exceeds 0.95, the sentence is considered a match (exact).

2. If no exact match is found in step 1, we proceed to identify fuzzy or partial matches. We first compute the longest common matching subsequence as an anchor, then align the full sentence to the most relevant subsegment of the quote and then compute a similarity score. If the score exceeds 0.80, this sentence is considered a match (fuzzy or partial). See the example below.

Example: Successful Partial Match

- **Date:** 1977-07-19
- **Speaker:** CHAIRMAN BURNS
- **Identified sentence:**
  "Now what I find most constructive is that, first, the private-enterprise economy has put out a remarkable performance here in the past two years or a little longer."
- **Argument quote in our dataset:**
  "...first, the private-enterprise economy has put out a remarkable performance here in the past two years or a little longer. Since March of 1975 we have added 6-1/2 million jobs in our economy."
- **Argument description:**
  Economy showing remarkable job creation performance

In this case, the identified sentence matches a partial excerpt in the dataset quote. Although the beginning of the sentence ("Now what I find most constructive is that") is not included in the extracted argument, and the quote contains multiple sentences, the core claim is indeed reproduced. This qualifies as a successful Tier 1 match.

Tier 2: Contextual Continuity Match

The sentence-level argument detection will typically fragment longer, multi-sentence arguments into independent pieces. In the actual transcript, speakers tend to pose longer arguments by first providing the data or context and then giving preference or conclusion. A problem with only considering the direct quote match is that, for a long argument in the transcript, our main dataset would typically extract only the most conclusive or policy-relevant portion, while the sentence-level detection might mark the surrounding sentences as arguments that discuss a certain economic variable. In this case, the dataset would lack a directly matched quote for those sentences.

To account for this, our tier 2 matching performs a contextual continuity match. Specifically, for each sentence that remains unmatched after Tier 1, we check whether there exists a neighbor sentence (within a ± 10 sentence window) that has already been matched to an argument in the dataset through direct quote. If a nearby match exists, we prompt the LLM with both sentences and the neighboring context to assess whether they form a single, coherent argument. If yes, we classify it as a valid match by continuity.

The purpose of this step is to capture cases where a sentence is a part of a longer argument, and the argument has only been partially quoted in the main dataset, making it unable to match the sentence directly. Because this process only uses local context within a window and relies on identifying structural continuity rather than broader

semantic similarity, it allows for a more reliable and accurate reconciliation than Tier 3, which involves more discretionary judgments.

In the following examples, the sentence wrapped by *** is the unmatched sentence, and the sentence wrapped by === is the nearby matched sentence.

Example 1: Positive Continuity Match (Matched) **Speaker:** MS. GEORGE
Date: 2012-10-24
Context:

> [1872] Thank you.
>
> [1873] President George.
>
> [1874] Thank you, Mr. Chairman.
>
> [1875] **=== Economic growth in my District has slowed since the previous meeting. ===**
>
> [1876] Notably, District activity in the private sector related to federal spending has dropped, especially in New Mexico.
>
> [1877] ***** Expectations for retail sales and tourism in coming months have edged lower, with contacts expecting weaker gains in holiday sales compared with last year. *****
>
> [1878] And activity in the energy sector has moderated as the drilling rig count for both natural gas and oil dropped in October and expectations weakened.
>
> [1879] In manufacturing, the Bank's October survey, which will be released later this week, showed a sharp decline in orders, and it was the lowest reading for its composite index since 2009.
>
> [1880] Many survey participants noted that customers have put orders on hold until uncertainties are resolved.

Explanation from the LLM response: These sentences are clearly part of the same economic argument about slowing growth in the speaker's District. Sentence A (1875) establishes the main claim that "Economic growth in my District has slowed since the previous meeting." Sentence B (1877) provides specific supporting evidence for this claim by detailing one sector (retail sales and tourism) where expectations have "edged lower." This follows a logical progression where the speaker first makes a general claim about slowing growth, then provides specific examples across different sectors to support this claim. The sentences between and after these also follow this same pattern - each providing sector-specific evidence (federal spending, retail/tourism, energy, manufacturing) to support the main claim about slowing district growth. This

represents a cohesive argumentative structure where the speaker is building a comprehensive case about economic conditions in their district through multiple related pieces of evidence.

Result: Matched

Example 2: Negative Continuity Match (Not Matched) **Speaker:** MR. COLDWELL
Date: 1977-07-19
Context:

> [761] And in that connection, the staff did moderate their projection, I think, with an offset on monetary stimulus or less restraint, however one wants to word it.
>
> [762] I think thats in the right direction.
>
> [763] I dont think their figures reflect this kind of prospective inventory development that perhaps you have been suggesting and [that] I think is possible.
>
> [764] ***** So, therefore, I just want to record, I would think the odds are–while they have a reasonable-looking projection–there is more chance of falling short of that than exceeding it in the near-term future, and I wouldnt be surprised to see a slower rate of growth in the latter part of this year. *****
>
> [765] Im not sure thats unhealthy, so long as it doesnt go too far.
>
> [766] But I find myself psychologically prepared, anyway, for a somewhat slower rate of growth than what the staff has projected.
>
> [767] **=== A more minor fact, following on Governor Jackson–Im not sure I see quite as much momentum in the residential construction area as they have continuing through the first quarter of 78. ===**

Explanation from the LLM response: These sentences represent distinct arguments, despite appearing in the same speaker's turn:

1. Sentence A (764) concludes a line of reasoning about overall economic growth projections, expressing skepticism about the staff's general economic forecast and predicting slower growth in the latter part of the year.

2. Sentence B (767) introduces a new, separate point with clear transition markers: "A more minor fact, following on Governor Jackson–" signals a topic shift to a specific sector (residential construction). The phrase "A more minor fact" explicitly indicates this is a separate, secondary point.

3. The speaker's language in B frames it as an additional observation rather than a continuation of the growth argument. It addresses a specific economic sector (residential construction) rather than the broader growth trajectory discussed in Sentence A.

4. The intervening sentences (765-766) complete the thought from Sentence A about overall growth, while sentence 767 pivots to a new, more specific topic.

This represents a clear shift from one complete argument to a new, distinct point about a specific economic sector.

Result: Not Matched

Tier 3: Semantic Retrieval Match

For the final stage of the matching pipeline, we attempt to match the remaining unmatched sentences through a broader semantic comparison. Occasionally, a speaker may illustrate the same argument at different points in a meeting. Additionally, since the criteria in Tier 1 are quite strict, the pipeline may fail to match sentences that are, in fact, word-for-word matches, particularly when multiple truncations are present in the quoted text. Finally, a sentence may involve two or more variables, and mismatches in primary variable classification can also lead to failure to match. To address these issues, we include a more general semantic match process as the final step in the pipeline.

For each remaining unmatched sentence, we filter through all arguments in the dataset with the same speaker and meeting date. Then, we prompt the LLM to identify whether the reasoning and conclusion in the sentence has already been expressed in any of the existing arguments in the dataset, even if not with the same exact wording.

Example 3: Semantic Retrieval Match (Matched) **Speaker:** MR. PARRY
Date: 1989-03-28
Unmatched Sentence:
"Ted, considering the strength of the dollar in the past year and your comment that you think the dollar will remain relatively firm in the near term, why wouldn't it be reasonable to expect that any significant improvement in net exports would be delayed to the second half of the year at the earliest and most likely to 1990?"
Argument Quote in Dataset:
"However, we have less strength in our forecast originating in the net export sector, largely as a result of the strength of the dollar that we've experienced during the past year."
Explanation from the LLM response: Both arguments focus on the impact of dol-

lar strength on net exports. The unmatched sentence is phrased as a question about delayed improvement in net exports, while the manual argument makes a direct statement about weaker contribution from net exports. Despite the difference in form, both contain the same core economic reasoning: that dollar strength is negatively affecting net export performance and that this effect is expected to continue. The unmatched sentence also includes a more precise timeline, whereas the manual argument is more general and forecast-oriented.

Result: Matched

Note that since this step involves some discretion of the LLM, it is not intended to reclassify large numbers of previously matched sentences or to drive the main validation conclusion. This serves as a last step to resolve edge cases and ensure that no clear, equivalent arguments are missed. Its inclusion results in a slight increase in coverage rather than significantly altering the overall match rate.

IA.A.3.3 Error Analysis

The key principle for this validation is that while the argument extraction is performed at the speaker-meeting-variable level to maximize the retention of contexts and enable the LLM to capture long argument chains, coverage validation is best done at the sentence level: when presented with only a single sentence, the LLM should be capable of determining whether it expresses an argument. This means we do not need to worry about LLM missing parts due to limited context. This hybrid design allows us to retain context when constructing the main dataset while using a more granular sentence-level analysis for validating coverage and detecting potential omissions.

To better understand the nature of the remaining unmatched sentences after the three tiers, we conduct some further analysis targeted to understand two questions: (i) whether any systematic directional position is underrepresented, and (ii) whether unmatched content could indicate substantial omission of arguments in the dataset.

We aggregate argument-bearing sentences by stance (hawkish, dovish, and neutral), and report how many unmatched sentences fall into each category in Table IA.A.2.

We observe a higher proportion of neutral sentences among the unmatched sentence set than in the full set of sentences; this is likely because neutral or descriptive atomic sentences are less likely to form substantive arguments captured in the dataset, which typically contains judgments or preferences. However, we can see that among all directional sentences, the proportion of positive is roughly the same across all sentences and unmatched sentences, which should not cause bias when constructing di-

rectional measures.

Stance	All Argument Sentences	Unmatched Sentences	Share (All)	Share (Unmatched)
Negative	2,409	186	42.6%	38.4%
Positive	1,871	158	33.1%	32.6%
Neutral	1,375	140	24.3%	28.9%
Total	5,655	484	100.0%	100.0%
Directional Total (Pos + Neg)	4,280	326	–	–
Positive / (Pos + Neg)	0.437	0.429	–	–

TABLE IA.A.2: STANCE DISTRIBUTION COMPARISON

IA.A.3.4 Examples of Unmatched Sentences

The observed coverage rates of arguments from our main sample versus those identified in the sentence-by-sentence pipeline are 85.7% (with the stricter first two tiers) and 91.4% (including the general semantic match), which illustrates the general comprehensiveness of our main prompting method. We also explore which arguments are identified in the sentence-level prompting exercise that are not covered in our main sample.

Table IA.A.3 presents several representative examples of sentences that are identified by LLM as containing economic variable arguments but were not matched. These cases help clarify what type of statements tend to fall outside of our main dataset and why we believe they do not materially affect the comprehensiveness of the dataset.

Sentence	Commentary
MR. BLACK (1980-02-05) *And a pretty good inflation record.*	This is an ambiguous and minimal claim that lacks a clear structure. When presented to LLM independently, they can be flagged due to clear mentioning and judgment of an economic variable; but with the actual lack of substantive economic reasoning within the context, it does not fall into the threshold of inclusion in our main dataset.
CHAIRMAN VOLCKER (1980-05-20) *If the economy gets stronger and credit demands increase and the money supply rises, interest rates would rise.*	This sentence represents overlapping topics with multiple variables; in the actual context, this comes from a brief discussion of economic mechanism rather than any specific economic variable, and therefore is not included in the main dataset.
MR. SCHULTZ. (1981-07-07) *Yes, but we're likely to have some offsetting weakness in the money market funds.*	This sentence is actually a conversational response to the previous speaker (the last sentence is *What about Chuck's point about the lifting of the Regulation Q ceilings on the longer-term CDs? We could get a surge in some of those.*) In this case, as long as the response does not add substantive content to the discussion, this would not count as a discrete argument for the speaker.

TABLE IA.A.3: UNMATCHED SENTENCE EXAMPLES

IA.B Categories of Arguments and Data Sources

Table IA.B.1 presents categories and associated descriptions used to categorize data sources cited by committee members in their arguments. Tables IA.B.2, IA.B.3, IA.B.4, IA.B.5, and IA.B.6 present the categories and associated descriptions used to categorizes arguments made by committee members with respect to growth, inflation, employment, credit markerts, and the stock market.

Name	Description
Official Government Statistics	Data produced by U.S. government agencies such as the Bureau of Labor Statistics (BLS), Bureau of Economic Analysis (BEA), and Census Bureau. Examples include unemployment rates, GDP figures, Consumer Price Index (CPI), Personal Consumption Expenditures (PCE) price index, housing starts, and industrial production data.
Federal Reserve Internal Analysis	Economic projections, forecasts, and analyses produced by the Federal Reserve staff, including the Greenbook/Tealbook, staff forecasts, Board staff research, and Federal Reserve economic models. These analyses often synthesize multiple data sources to provide comprehensive economic outlooks.
Regional Federal Reserve Data	Quantitative data and formal surveys collected and produced by regional Federal Reserve Banks, including manufacturing surveys, business outlook surveys, regional price indices, and district-specific economic indicators. Examples include the Philadelphia Fed Manufacturing Survey, the Empire State Manufacturing Survey, and the Chicago Fed's Financial Conditions Index.

External Business and Consumer Surveys	Survey data collected by organizations outside the Federal Reserve System, including business sentiment surveys and consumer confidence/expectations surveys. Examples include the University of Michigan Consumer Sentiment Survey, Conference Board Consumer Confidence Index, NFIB Small Business Optimism Index, and Manpower Employment Outlook Survey.
Financial Market Indicators	Data from financial markets that reflect investor sentiment, expectations, and risk assessments. Examples include stock market indices (S&P 500, NASDAQ), bond yields, credit spreads, LIBOR-OIS spreads, TIPS-based inflation expectations, currency exchange rates, and commodity prices.
Industry-Specific Data	Quantitative data focused on particular economic sectors or industries. Examples include auto sales and inventory figures, housing market data (home sales, prices, inventory), steel production and shipment data, energy sector statistics (rig counts, production levels), and retail sales figures for specific segments.
Anecdotal Business Intelligence	Qualitative information obtained directly from business contacts, including conversations with CEOs and other executives, corporate earnings reports and guidance, and feedback from specific companies. This category captures first-hand business insights that may not be reflected in official statistics.

Regional Qualitative Reports	Qualitative assessments about regional economic conditions, including observations from Federal Reserve Bank directors, advisory councils, and broader regional economic trends. This category captures insights about local economic conditions that may precede changes in national data.
FOMC Member Analysis	Individual calculations, projections, and analytical frameworks developed by FOMC members themselves, rather than by Federal Reserve staff. This includes personal economic forecasts, custom calculations based on existing data, and individual interpretations of economic relationships.
Media and External Reports	Information from news sources, think tanks, and other external organizations not captured in other categories. Examples include newspaper articles, financial press coverage, business news reporting, and economic analyses from nongovernmental organizations.
International Economic Data	Statistics, trends, and observations about economies outside the United States. This includes foreign GDP growth rates, unemployment figures, inflation rates, central bank policies, and cross-country economic comparisons that provide context for U.S. economic conditions.
Historical Economic Comparisons	References to past economic episodes, cycles, and trends used to provide context for current conditions. This includes comparisons to previous recessions, inflation episodes, policy tightening cycles, and long-term economic patterns that help interpret present circumstances.

Academic Research and Literature	References to scholarly economic research, academic papers, and established economic theories. Examples include citations of economic studies, references to economic literature on topics like the Phillips curve or monetary policy transmission, and applications of academic economic models like the Taylor Rule.

TABLE IA.B.1: CATEGORIES OF DATA SOURCES CITED BY MEMBERS

Note: The table lists different categories of data cited by committee members during FOMC meetings. These categories and descriptions are generated by the LLM by providing it with a sample of 1000 arguments made by committee members and prompting it to identify relevant categories. The categories and descriptions are then provided to the LLM in additional prompts, along with each of the arguments made, in order to categorize the data cited by each member in each argument.

Category	Disagreement Type	Description
Growth Outlook and Risks	Risk/ Forecast	Projections of future economic growth scenarios, encompassing disagreements on the likelihood of different growth outcomes, the reliability of forecasts, and the balance of upside versus downside risks to the growth outlook.
Growth Relative to Potential	Measurement/ Assessment	Assessments of whether the economy is growing above or below its sustainable rate, involving disagreements about the measurement of potential growth, estimates of output gaps, and the implications of growth-potential differentials for policy decisions.
Sectoral Growth Patterns and Capacity Constraints	Measurement/ Assessment	Analysis of growth variations across economic sectors and regions, and how capacity limitations affect overall growth, with disagreements about which sectors are driving aggregate growth and the extent to which capacity constraints are limiting economic expansion.
Monetary Policy and Growth Transmission	Causal/ Structural	Examination of how monetary policy actions affect economic growth through various channels, with disagreements about the effectiveness, timing, and mechanisms of monetary policy transmission to real economic activity.
Fiscal Policy and Growth	Causal/ Structural	Analysis of how government spending, taxation, and other fiscal measures influence economic growth, with disagreements about the magnitude, duration, and channels through which fiscal policy affects economic activity.
Consumer Spending, Wealth Effects, and Growth	Causal/ Structural	Examination of how consumer spending and changes in household wealth affect economic growth, with disagreements about consumption trends, the strength of wealth effects on spending, and their importance for overall growth dynamics.
Business Investment and Growth	Causal/ Structural	Analysis of how capital expenditures by businesses contribute to economic growth, with disagreements about investment trends, determinants of business spending decisions, and their significance for both short-term demand and long-term productive capacity.

Labor Markets and Growth	Causal/ Structural	Examination of the relationship between employment, unemployment, labor force participation, and economic growth, with disagreements about labor market conditions, the direction of causality between labor markets and growth, and the implications for wage pressures.
Inflation and Growth Dynamics	Causal/ Structural	Analysis of the relationship between economic growth and inflation, with disagreements about the growth-inflation tradeoff, how growth affects price pressures, and the implications for achieving both price stability and maximum sustainable growth.
External Factors and Growth	Causal/ Structural	Examination of how international trade, global economic conditions, energy prices, and other external factors affect domestic economic growth, with disagreements about the importance, transmission channels, and policy implications of these external influences.
Financial Conditions, Confidence, and Growth	Causal/ Structural	Analysis of how credit availability, asset prices, financial market conditions, business confidence, and consumer sentiment affect economic growth, with disagreements about the importance, direction of causality, and policy implications of these relationships.
Business Cycle Dynamics and Growth	Measurement/ Assessment	Examination of growth patterns during different phases of the business cycle, including recessions, recoveries, and inventory adjustments, with disagreements about cycle identification, the causes of cyclical fluctuations, and appropriate policy responses at different cycle stages.
Structural Factors and Long-term Growth	Causal/ Structural	Analysis of how productivity trends, technological change, demographic shifts, and other structural factors affect long-term economic growth potential, with disagreements about the nature, magnitude, and policy implications of these fundamental determinants of sustainable growth.

TABLE IA.B.2: CATEGORIES OF GROWTH ARGUMENTS

Note: The table displays the categories of arguments made with respect to growth and their descriptions.

Category	Disagreement Type	Description
Inflation Expectations and Anchoring	Causal/ Structural	The degree to which inflation expectations are anchored to the central bank's target influences actual inflation outcomes through wage and price setting behavior, with disagreements about how expectations form and their importance relative to other inflation determinants.
Phillips Curve and Economic Slack	Causal/ Structural	The relationship between resource utilization (particularly labor market slack) and inflation pressures determines how economic growth translates into price changes, with disagreements about the strength, stability, and non-linearity of this relationship.
Monetary Policy Transmission to Inflation	Causal/ Structural	Monetary policy actions affect inflation through various transmission channels with different strengths and lags, with disagreements about the effectiveness, timing, and relative importance of different transmission mechanisms.
Inflation Target Interpretation	Priority/ Preference	How the inflation target is interpreted (as a ceiling, a floor, or a symmetric target) affects policy decisions and communication, particularly when inflation deviates from target, with disagreements about the appropriate interpretation and implementation of the target.
Supply-Side and External Factors	Causal/ Structural	Non-demand factors such as productivity, commodity prices, exchange rates, and global conditions affect inflation through cost channels and import prices, with disagreements about the magnitude and persistence of these effects.
Inflation Measurement and Methodology	Measurement/ Assessment	How inflation is measured and which measures are emphasized affects policy assessments and decisions, with disagreements about the most appropriate measures and methodologies for different policy purposes.
Inflation Dynamics and Persistence	Causal/ Structural	The degree to which inflation changes are transitory or persistent affects appropriate policy responses and inflation forecasts, with disagreements about the underlying mechanisms of inflation persistence and mean reversion.

Inflation Forecasts and Projections	Risk/ Forecast	The reliability and implications of inflation forecasts influence forward-looking policy decisions, with disagreements about forecast accuracy, methodologies, and the appropriate weight to place on forecasts versus current data.
Inflation Risks and Uncertainty	Risk/ Forecast	The distribution of risks around the inflation outlook, including deflation risks, affects policy risk management, with disagreements about the balance of risks and appropriate policy responses to uncertainty.
Inflation and Policy Credibility	Causal/ Structural	Central bank credibility affects how policy actions influence inflation expectations and actual inflation outcomes, with disagreements about how credibility is built, maintained, and leveraged for policy effectiveness.
Price and Wage Setting Behavior	Causal/ Structural	How firms set prices and how wages are determined affects inflation dynamics and persistence, with disagreements about the microeconomic foundations of price changes and wage-price feedback mechanisms.
Inflation and Policy Tradeoffs	Priority/ Preference	How policymakers weigh inflation objectives against other goals affects policy decisions when objectives conflict, with disagreements about the appropriate balance between inflation control and other economic objectives.
Monetary and Fiscal Interactions	Causal/ Structural	How monetary and fiscal policies interact affects inflation outcomes and appropriate monetary policy responses, with disagreements about the nature and importance of these interactions in different economic environments.

TABLE IA.B.3: CATEGORIES OF INFLATION ARGUMENTS

Note: The table displays the categories of arguments made with respect to inflation and their descriptions.

Category	Disagreement Type	Description
Labor Market Slack, Participation, and NAIRU	Measurement/ Assessment	Assessments of labor market utilization relative to full employment, encompassing disagreements about the current level of slack, the natural rate of unemployment, and how labor force participation trends affect these measures.
Employment Growth and Business Hiring	Measurement/ Assessment	Analysis of job creation trends across regions and sectors, including disagreements about the factors influencing business hiring decisions and the pace, quality, and sustainability of employment growth.
Wage Growth and Labor Costs	Measurement/ Assessment	Examination of wage and compensation trends, with disagreements about their implications for inflation, business decisions, and as indicators of labor market tightness.
Productivity-Employment Relationship	Causal/ Structural	Analysis of how productivity growth affects employment needs and labor demand, with disagreements about the strength and nature of this relationship and its implications for job creation.
Labor Market Structure and Frictions	Causal/ Structural	Examination of long-term structural changes in the labor market and frictions that affect the matching of workers to jobs, with disagreements about the nature and significance of these factors for employment outcomes.
Quality and Distributional Effects	Measurement/ Assessment	Analysis of the types and quality of jobs being created and how employment conditions affect different demographic groups, with disagreements about the significance and policy implications of these considerations.
Employment, Consumption, and Inflation Dynamics	Causal/ Structural	Examination of how employment affects consumer spending, inflation, and overall macroeconomic dynamics, with disagreements about the strength and nature of these relationships, including the Phillips curve.
Monetary Policy Impact on Employment	Causal/ Structural	Analysis of how monetary policy affects employment, with disagreements about transmission mechanisms, effectiveness, timing, and potential side effects of policy actions on labor markets.

Non-Monetary Policy Impacts on Employment	Causal/ Structural	Examination of how fiscal policy, regulatory policy, and policy uncertainty affect employment decisions and outcomes, with disagreements about the significance and mechanisms of these effects.
Employment Indicators and Measurement	Measurement/ Assessment	Discussion of which indicators best measure employment conditions, with disagreements about data reliability, interpretation, and the relative importance of different measures for policy decisions.

TABLE IA.B.4: CATEGORIES OF EMPLOYMENT ARGUMENTS

Note: The table displays the categories of arguments made with respect to employment and their descriptions.

Category	Disagreement Type	Description
Credit Availability and Market Functioning	Causal/ Structural	How credit availability is influenced by market functioning, including securitization, liquidity, and structural features, with disagreements about the relative importance of supply versus demand factors or the resilience of market mechanisms.
Interest Rates and Credit Market Pricing	Causal/ Structural	How interest rates (both policy rates and market rates) affect credit pricing, including spreads and risk premiums, with disagreements about the magnitude, timing, or direction of these effects across different segments of credit markets.
Financial Institution Health and Credit Provision	Causal/ Structural	How the health of financial institutions affects their ability and willingness to provide credit, with disagreements about the role of capital, profitability, funding conditions, and competitive pressures in determining lending behavior.
Regulatory and Policy Interventions in Credit Markets	Priority/ Preference	How regulatory changes and policy interventions affect credit markets, with disagreements about the appropriate tools, timing, and scope of interventions, as well as potential unintended consequences or tradeoffs between different policy objectives.
Credit Conditions and Economic Activity	Causal/ Structural	How credit conditions affect economic activity through various channels including household consumption, business investment, and asset prices, with disagreements about the strength of these transmission mechanisms or their importance relative to other economic factors.
International and Expectational Factors in Credit Markets	Risk/ Forecast	How international developments and market expectations shape domestic credit market conditions and behavior, with disagreements about the likelihood of spillovers from foreign markets or how expectations about future credit conditions influence current lending and borrowing decisions.

TABLE IA.B.5: CATEGORIES OF CREDIT MARKET ARGUMENTS

Note: The table displays the categories of arguments made with respect to credit markets and their descriptions.

Category	Disagreement Type	Description
Stock Market Valuation and Fundamentals	Measurement/ Assessment	Stock prices may deviate from fundamental values due to various factors, creating potential risks to financial stability and economic activity when corrections occur, with disagreements centering on whether current valuations are justified by economic fundamentals.
Stock Market Volatility and Systemic Stability	Causal/ Structural	Stock market fluctuations can threaten financial system stability through various contagion channels, potentially requiring policy intervention to prevent broader economic damage, with disagreements about the severity of these transmission mechanisms and appropriate responses.
Monetary Policy and Stock Market Interactions	Causal/ Structural	Monetary policy affects stock prices through various channels, and stock market conditions influence monetary policy decisions, creating a two-way relationship that affects economic outcomes, with disagreements about the appropriate policy response to market developments.
Stock Market Effects on Household Behavior	Causal/ Structural	Stock market movements affect household wealth and confidence, which influence consumption decisions and overall economic activity, with disagreements about the magnitude, timing, and significance of these wealth and sentiment effects.
Stock Market Influence on Corporate Behavior	Causal/ Structural	Stock price movements affect corporate financial and investment decisions, including capital allocation, hiring, and innovation, with implications for economic growth and stability, with disagreements about how these incentives shape business activity.
Stock Market as Economic and Financial Indicator	Measurement/ Assessment	Stock prices reflect and predict economic and financial conditions, providing information that can guide policy decisions and private sector behavior, with disagreements about the reliability and interpretation of these market signals.

Stock Market Functioning and Liquidity	Causal/ Structural	Market operations, liquidity conditions, and trading dynamics in stock markets affect price discovery, volatility, and the transmission of shocks to the broader economy, with disagreements about the resilience of market structures and potential vulnerabilities.
Stock Market and Financial Risk-Taking	Causal/ Structural	Stock market conditions influence risk appetite in financial markets, affecting asset allocation, leverage, and potential financial vulnerabilities, with disagreements about whether current risk-taking behavior is excessive or appropriate.
Global Stock Market Linkages	Causal/ Structural	International stock market connections transmit economic and financial shocks across borders, affecting domestic conditions and complicating policy decisions, with disagreements about the strength and importance of these cross-border effects.
Fiscal Policy and Stock Market Interactions	Causal/ Structural	Fiscal policy changes affect stock valuations through various channels, and stock market reactions can influence fiscal decisions and their economic impact, with disagreements about the magnitude and persistence of these effects.
Stock Market, Monetary Aggregates, and Inflation	Causal/ Structural	Stock market activity affects monetary aggregates and may reflect or influence inflation expectations, complicating monetary policy implementation and interpretation, with disagreements about the stability and significance of these relationships.

TABLE IA.B.6: CATEGORIES OF STOCK MARKET ARGUMENTS

Note: The table displays the categories of arguments made with respect to the stock market and their descriptions.

IA.C Additional Empirical Analyses

IA.C.1 Policy Decision Scores and Monetary Policy Surprises

In this section, we link the monetary policy decisions we identify from the transcripts with monetary policy surprises measured from high-frequency asset price responses to monetary policy announcements. This link verifies our LLM-based method for measuring policy decisions, and highlights that the policy decisions with respect to which we measure alignment in our main analysis are highly relevant for understanding market perceptions of monetary policy.[24]

Following prior work (Gürkaynak, Sack and Swanson, 2005; Nakamura and Steinsson, 2018), we measure monetary policy surprises using the high-frequency responses of interest rate futures in a 30-minute window following the FOMC's announcement of its policy decision. Our main measure of monetary policy surprises is mps_t, constructed as the first principal component of changes in the one- to four-quarter ahead Eurodollar futures in the 30-minute window following FOMC announcements, from Swanson and Jayawickrema (2024) and Bauer and Swanson (2023*b*). The sample runs from 1988-2019. We also analyze the changes of n-quarter ahead Eurodollar futures (labeled EDn), as well as the near- and 3-month maturity Fed Funds futures in the 30-minute window following the FOMC announcement (labeled FF and $FF3m$).[25] The sample of Fed Funds futures runs from 1995-2019.

For each meeting, we assign a score from -3 (highly accommodative / dovish) to +3 (highly restrictive / hawkish) to each of the identified changes to the policy directive. We then compute two measures: $Avg(\{PolicyScores_t\})$ as the average score across all changes; and $Avg(\{ExRatePolicyScores_t\})$ as the average score across all changes excluding the target rate.[26] We estimate regressions of monetary policy surprises on the $Avg(\{PolicyScores\})$ variables. The $Avg(\{PolicyScores\})$ variables are standardized to have zero mean and unit standard deviation, and monetary policy surprise variables are in basis points.

The first two columns of Table IA.C.1 report results from univariate regressions of mps_t on $Avg(\{PolicyScores_t\})$ and $Avg(\{ExRatePolicyScores_t\})$. The coefficients

[24] This result also relates to Handlan (2022), who constructs text-based monetary policy shocks from analysis of the FOMC's statement, and shows that these have explanatory power for Fed Funds futures surprises; and to Gáti and Handlan (2023) who relate variation in the Fed's discussion of its expectations embedded in the post-meeting statement to monetary policy surprises.

[25] Following Bernanke and Kuttner (2005), we scale up the change in the near-maturity Fed Funds futures to account for the fact that the contract's settlement is based on the average federal funds rate over the entire month.

[26] There are fewer observations in $Avg(\{ExRatePolicyScores_t\})$, as for several meetings, the only relevant decision identified is with respect to the target rate decision.

from these regressions are 1.42 and 1.03, respectively, indicating that a one standard deviation change in the $Avg(\{PolicyScores\})$ measures captures a 1 to 1.4 basis point monetary policy surprise. As reported in the Table, the standard deviation of mps_t is about 5 basis points, indicating an economically substantial relationship between the independent variables and monetary policy surprises.

The mps_t variable captures the component of monetary surprises stemming from both the target rate decision and other elements of the committee's decision, most importantly forward guidance about the future path of interest rates. In columns 3-6 of Table IA.C.1, we explicitly separate the surprise of the target rate decision from forward guidance, in the spirit of Gürkaynak, Sack and Swanson (2005). Columns 3 and 4 report results from regressions of the high frequency change in four quarter-ahead Eurodollar futures on the $Avg(\{PolicyScores\})$ variables, controlling for the contemporaneous change in one quarter-ahead Eurodollar futures. The coefficients on $Avg(\{PolicyScores_t\})$ and $Avg(\{ExRatePolicyScores_t\})$ are 0.74 and 1.20 in these regressions, compared to the 5.7 bps standard deviation of $ED4$. Columns 5 and 6 report results from regressions of the high frequency change in 3-month ahead Fed Funds futures on the $Avg(\{PolicyScores\})$ variables, controlling for the contemporaneous change in the near maturity Fed Funds futures contract. The coefficients on $Avg(\{PolicyScores_t\})$ and $Avg(\{ExRatePolicyScores_t\})$ are 0.66 and 0.63 in these regressions, compared to the 3.7 bps standard deviation of $FF3m$. Taken together, these results indicate that the $Avg(\{PolicyScores\})$ variables are strongly related to surprise changes to the target rate, as well as to surprise changes in communication about the future path of interest rates.

The relationship between the $Avg(\{PolicyScores\})$ variables and monetary policy surprises is notable for two reasons. First, our evidence in the main text is based on comparing members' alignment with and influence on policy decisions to how well their models match recent economic data. The relationship with monetary policy surprises indicates that our prior evidence isn't just measuring internal deliberation but is capturing behavior reflected in financial market movements to the resulting policy.

Second, taken in conjunction with the evidence on the committee's tilt towards different members' models based on the data, the results also suggest that the committee structure may play an important role in generating monetary policy surprises. The committee's tilting behavior is consistent with it not using a fixed policy rule to make a decision, as argued by Hack, Istrefi and Meier (2025), with shifts in the policy stance arising from the shifting influence of committee members coming as a surprise to investors. The link with monetary policy surprises is also complementary to mounting evidence of monetary policy surprises arising from market participants

	$y = mps$		$y = ED4$		$y = FF3m$		$y = mps_{orth}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Avg(\{PolicyScores_t\})$	1.20		1.06		0.57		0.30	
	(2.98)		(3.59)		(2.56)		(0.87)	
$Avg(\{ExRatePolicyScores_t\})$		0.96		0.86		0.64		0.18
		(2.42)		(4.99)		(2.87)		(0.53)
$ED1$			0.91	1.02				
			(8.49)	(18.29)				
FF					0.71	0.73		
					(7.45)	(8.04)		
σ_y	4.95	4.95	5.70	5.70	3.74	3.74	4.63	4.63
N	262	254	262	254	190	187	262	254
R^2	0.06	0.04	0.53	0.81	0.53	0.53	0.00	0.00

TABLE IA.C.1: POLICY DECISION SCORES AND MONETARY POLICY SURPRISES

Note: The table reports results for regressions of high-frequency asset price-based measures of monetary policy surprises on the hawkish/dovish tone of FOMC decisions. The dependent variables are: mps, the first principal component of changes in one- to four-quarter ahead Eurodollar contracts within a 30-minute post-announcement window; $ED4$, the change in the four-quarter ahead Eurodollar contract; $FF3m$, the change in 3-month ahead Fed Futures yields; and mps_{orth}, computed as residuals from regressing the mps measure on a set of controls that capture incoming macroeconomic and financial news. The independent variables of interest—$Avg(\{PolicyScores_t\})$ and $Avg(\{ExRatePolicyScores_t\})$ (excluding the policy rate decision)—are the standardized meeting-level average scores of each of the decisions made during the meeting, with scores ranging from –3 for strongly dovish to +3 for strongly hawkish. The regressions include controls for the high-frequency change in one-quarter ahead Eurodollar futures ($ED1$) and near-maturity Fed Funds futures (FF). The mps and ED measures used in the regressions are from Bauer and Swanson (2023*b*). The table reports the standard deviation of the dependent variables as σ_y; t-statistics computed using robust standard errors are reported in parentheses.

having uncertainty about the Fed's reaction function in response to macroeconomic news (Bauer and Swanson, 2023*a*,*b*; Bauer, Pflueger and Sunderam, 2024). Consistent with this interpretation, columns 7 and 8 of Table IA.C.1 report results from regressions of mps_{orth}, measured as the residual of mps after controlling for macroeconomic and financial news arriving prior to the FOMC meeting, on the $Avg(\{PolicyScores\})$ variables. These results indicate a negligible relationship, suggesting that the component of the $Avg(\{PolicyScores\})$ variables that is related to monetary policy surprises stems from investors being surprised by the committee's response to macroeconomic and financial news.

IA.C.2 Argument Categories Time Series

This section presents the share of argument categories over time by variable. Each figure presents stacked area plots that visualize the share of argument categories within FOMC discussions from 1976 to 2019. Each figure represents one of five variables: Inflation, Growth, Employment, Credit Markets, and Stock Market. For each meeting t and category c within theme τ, we calculate:

$$p_{c,t}^{\tau} = \frac{n_{c,t}^{\tau}}{\sum_{c' \in C_{\tau}} n_{c',t}^{\tau}} \tag{IA.C.1}$$

where $n_{c,t}^{\tau}$ is the number of arguments in category c at meeting t for theme τ, and C_{τ} is the set of all categories within theme τ. Categories are ordered by their average proportion across all meetings (most common at bottom). The data are smoothed using a 4-meeting trailing average.

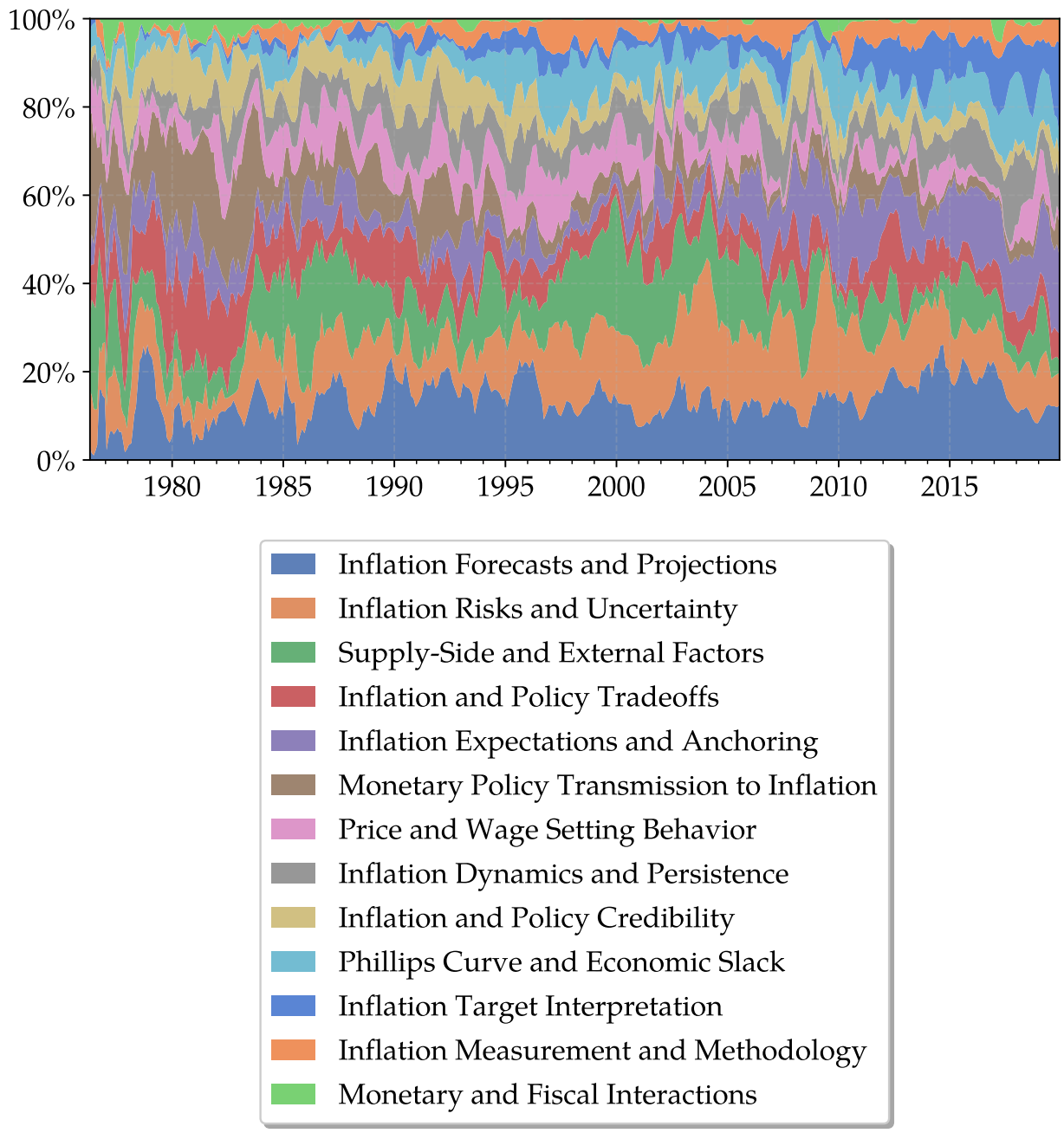


FIGURE IA.C.1: INFLATION ARGUMENTS BY CATEGORY

Note: The figure plots the 4-meeting trailing average of the proportions of arguments made regarding inflation in each of 13 categories during FOMC meetings.

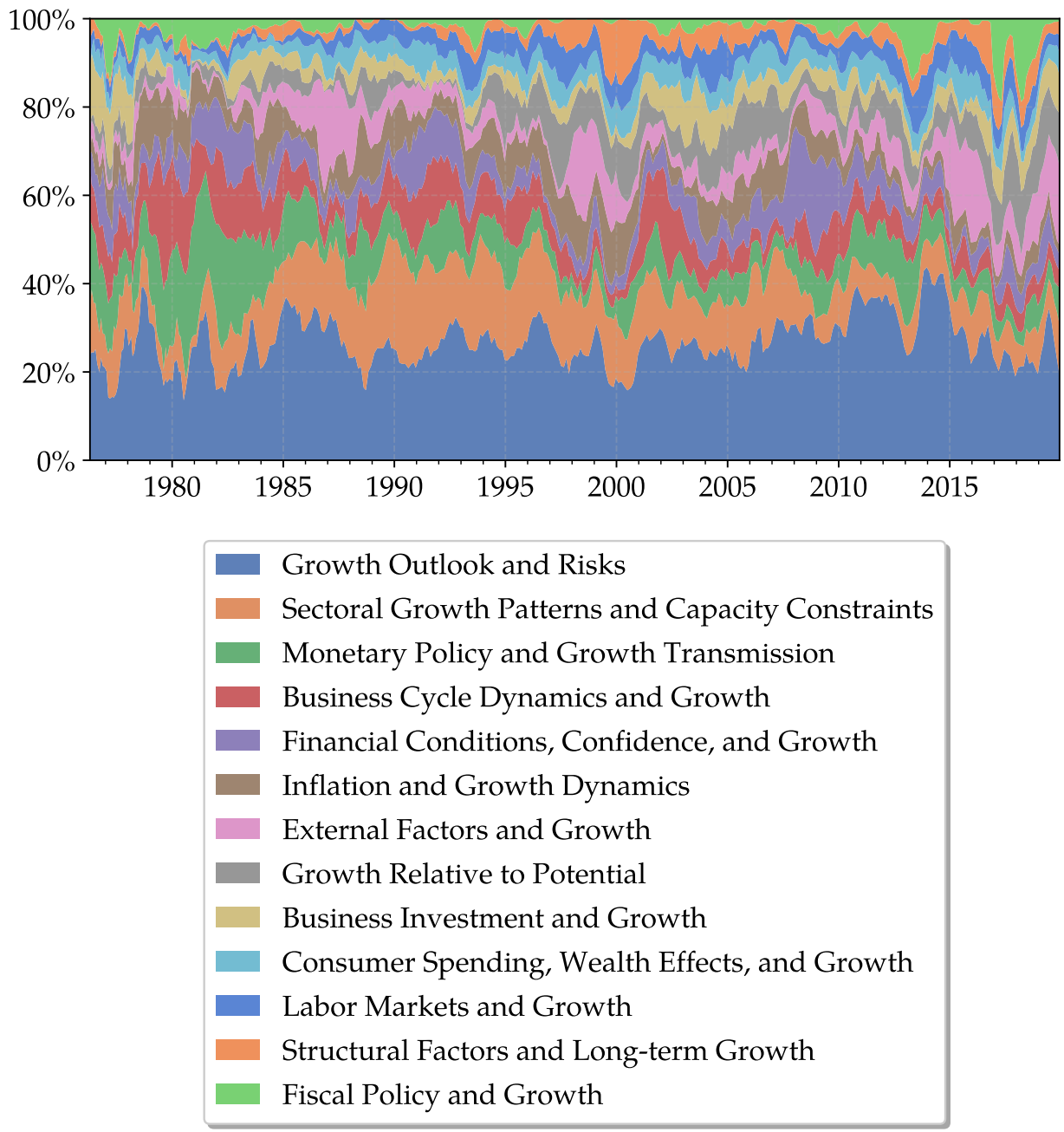


FIGURE IA.C.2: GROWTH ARGUMENTS BY CATEGORY

Note: The figure plots the 4-meeting trailing average of the proportions of arguments made regarding growth in each of 13 categories during FOMC meetings.

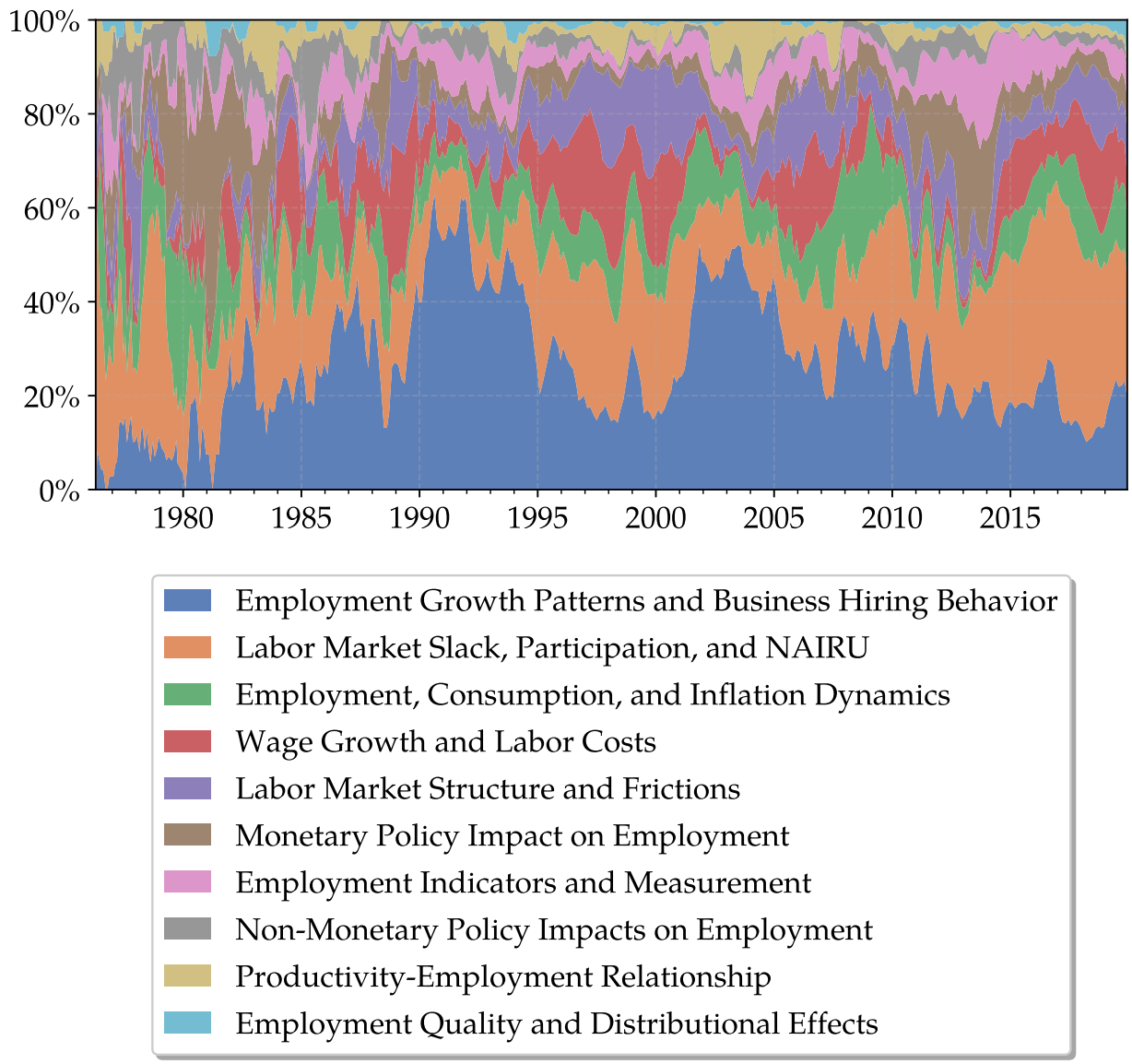


FIGURE IA.C.3: EMPLOYMENT ARGUMENTS BY CATEGORY

Note: The figure plots the 4-meeting trailing average of the proportions of arguments made regarding employment in each of 10 categories during FOMC meetings.

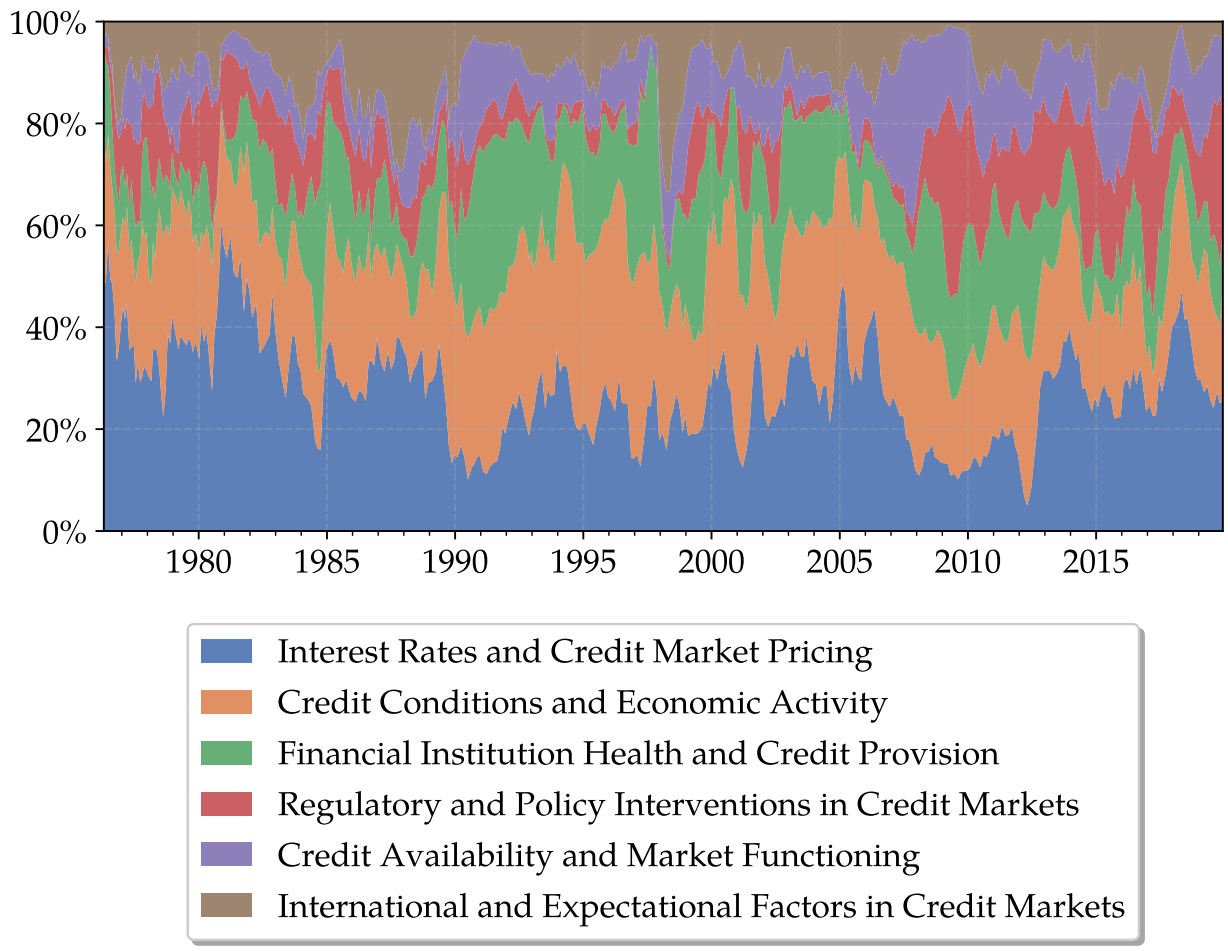


FIGURE IA.C.4: CREDIT MARKET ARGUMENTS BY CATEGORY

Note: The figure plots the 4-meeting trailing average of the proportions of arguments made regarding credit markets in each of 6 categories during FOMC meetings.

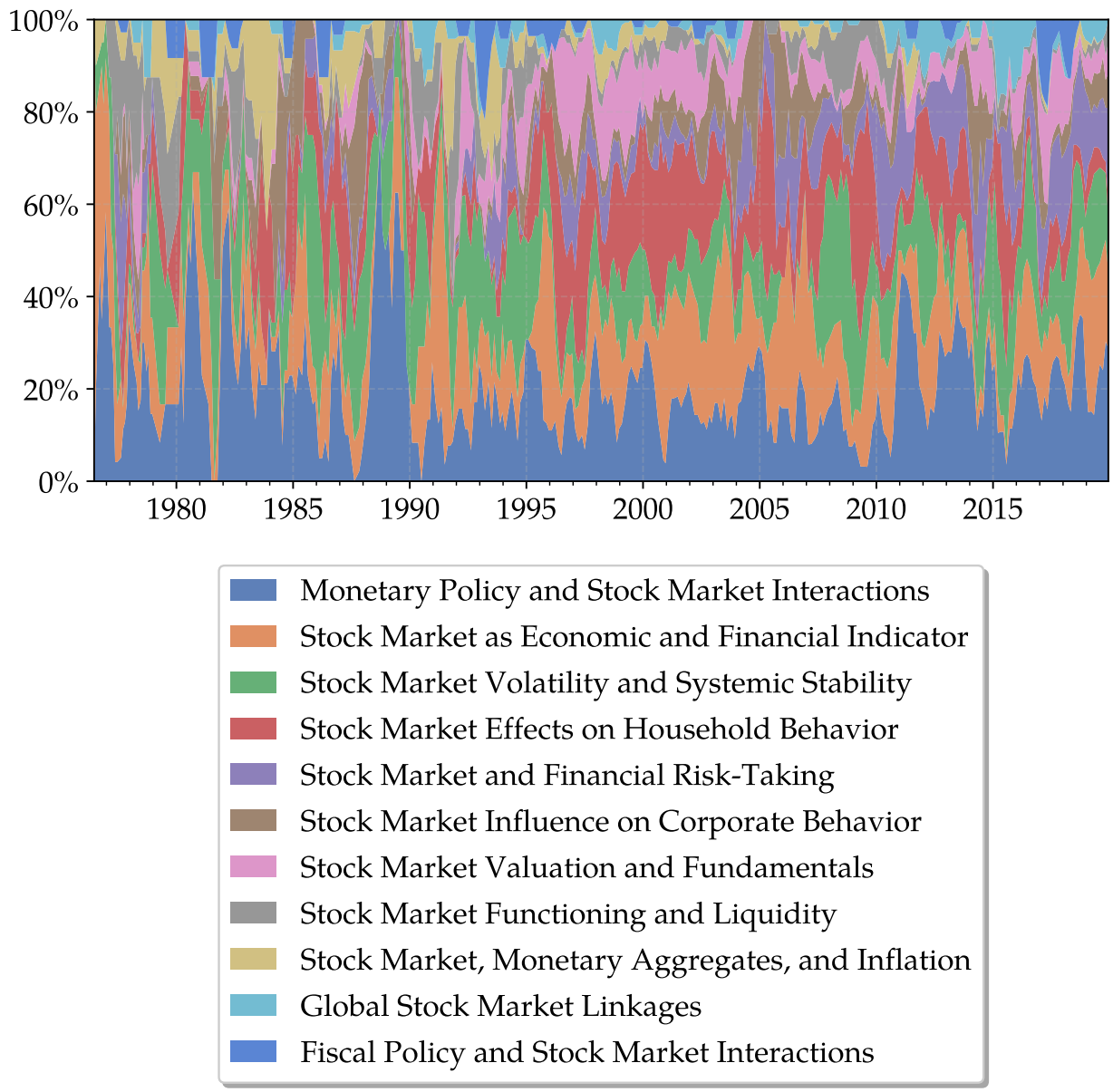


FIGURE IA.C.5: STOCK MARKET ARGUMENTS BY CATEGORY

Note: The figure plots the 4-meeting trailing average of the proportions of arguments made regarding the stock market in each of 11 categories during FOMC meetings.

IA.C.3 Additional Tables and Figures

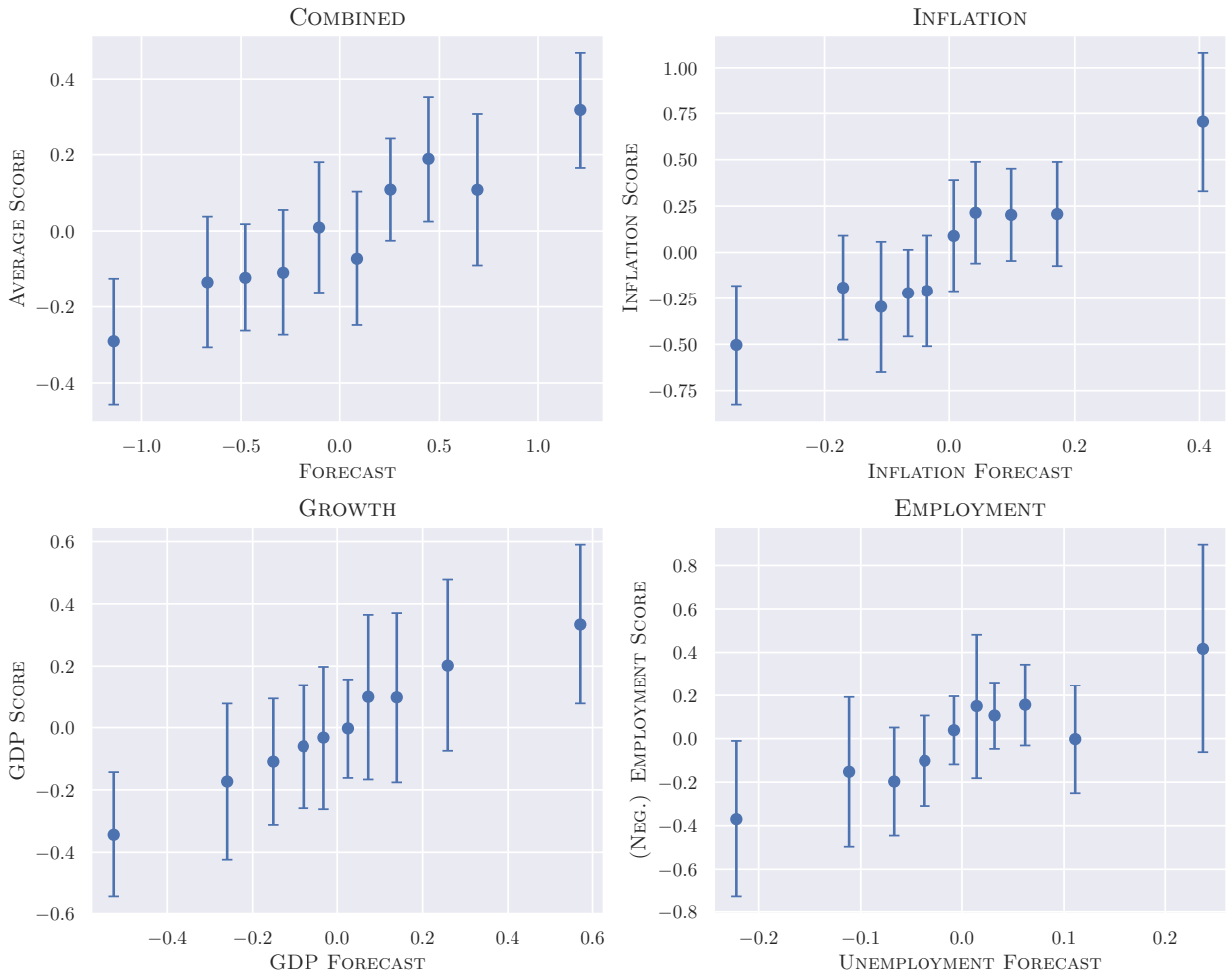


FIGURE IA.C.6: COMMITTEE MEMBERS' FORECASTS AND BELIEF ARGUMENT SCORES

Note: This figure replicates Figure 2, where the argument scores for each member are constructed using arguments that only reflect members' beliefs about the variable of interest.

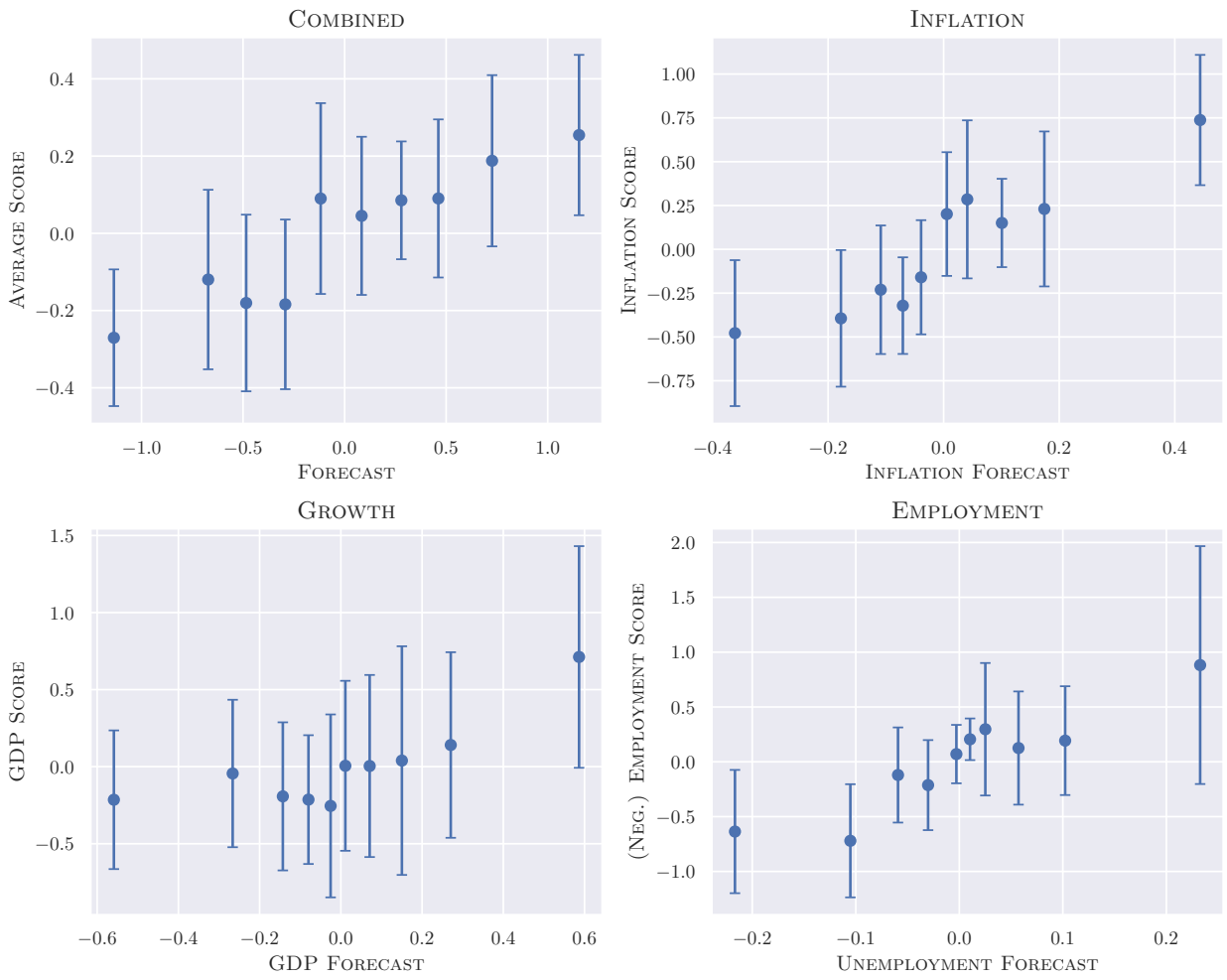


FIGURE IA.C.7: COMMITTEE MEMBERS' FORECASTS AND PREFERENCE ARGUMENT SCORES

Note: This figure replicates Figure 2, where the argument scores for each member are constructed using arguments that only reflect members' preferences about the variable of interest.

	Regression R^2					
	Inflation	Growth	Employment	Credit	Stocks	Average
Meeting FE	0.25	0.26	0.23	0.16	0.16	0.26
Member FE	0.28	0.20	0.20	0.16	0.07	0.28
Bootstrap null (99th pct.)	0.10	0.05	0.06	0.04	0.03	0.05
N	6402	6402	6402	6402	6402	6402

TABLE IA.C.2: MEMBERS' SCORES: FIXED EFFECTS WITH PUBLIC INFORMATION

Note: This table repeats the analysis in Table 5, only using arguments citing public information.

Comparison of Principal Components						
	Full Data			Member FE		
Principal Component Loadings	PC1	PC2	PC3	PC1	PC2	PC3
Growth	0.59	-0.28	0.08	0.55	-0.25	-0.18
Inflation	0.54	0.76	-0.34	0.59	-0.03	0.76
Employment	0.47	-0.07	0.74	0.43	0.81	-0.32
Credit Markets	0.34	-0.55	-0.57	0.38	-0.52	-0.50
Stock Market	0.14	-0.18	-0.10	0.12	0.08	-0.21
Proportion of Variance Explained	0.60	0.13	0.12	0.82	0.08	0.05

TABLE IA.C.3: PRINCIPAL COMPONENTS OF MEMBERS' SCORES WITH PUBLIC INFORMATION

Note: This table repeats the analysis in Table 6, only using arguments that cite public information.

	Main Effects (%)			Covariances (%)		
	Selection	Interpretation	Residual	Sel×Int	Sel×Res	Int×Res
All	10.3	94.6	0.8	-4.9	-0.1	-0.7
Inflation	16.1	88.2	2.4	-5.6	-0.4	-0.7
Growth	10.8	90.8	0.6	-1.6	-0.1	-0.5
Employment	18.7	90.8	5.7	-10.8	-1.8	-2.6
Credit Markets	16.0	89.1	1.6	-5.2	-0.3	-1.2
Stock Market	25.3	81.2	13.7	-13.4	-2.8	-4.0

TABLE IA.C.4: WITHIN-MEETING VARIANCE DECOMPOSITION, INCLUDING COVARIANCES

Note: The table repeats the variance decomposition in Table 8, but includes additional covariance terms omitted from the main table.

	Selection (%)	Interpretation (%)	Residual (%)
All	31.9	67.6	0.5
Inflation	34.7	63.7	1.6
Growth	32.9	66.6	0.5
Employment	37.5	59.4	3.0
Credit Markets	30.1	69.0	0.9
Stock Market	32.1	55.7	12.2

TABLE IA.C.5: VARIANCE DECOMPOSITION WITHOUT MEETING FIXED EFFECTS

Note: The table repeats the analysis in Table 8, without meeting fixed effects.

	$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$		$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model Fit: $\mathcal{L}(D_t \mid m_{i,t})$	0.26	0.12	0.16	0.05	-6.07	-3.43						
	(6.77)	(3.92)	(3.59)	(2.50)	(5.02)	(3.44)						
Model Fit: $\mathcal{L}(D_t \mid m_{i,FE})$							0.33	0.15	0.21	0.07	-7.07	-3.61
							(7.36)	(2.64)	(2.32)	(2.00)	(5.37)	(1.90)
N	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972
Meeting Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Member Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Role Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

TABLE IA.C.6: MODEL FIT AND POLICY DECISION ALIGNMENT, NO CONTROLS

Note: The table repeats the analysis in Table 9 excluding excluding role fixed effects.

	$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$		$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model Fit: $\mathcal{L}(D_t \mid m_{i,t})$	0.21	0.10	0.06	0.03	-5.19	-2.73						
	(5.80)	(3.31)	(3.16)	(1.66)	(5.22)	(3.26)						
Model Fit: $\mathcal{L}(D_t \mid m_{i,FE})$							0.30	0.16	0.11	0.09	-6.83	-3.74
							(7.72)	(2.98)	(4.26)	(3.46)	(5.31)	(2.00)
N	3903	3903	3903	3903	3903	3903	3972	3972	3972	3972	3972	3972
Meeting Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Member Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Role Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

TABLE IA.C.7: MODEL FIT AND POLICY DECISION ALIGNMENT, LAGGED MODEL

Note: The table repeat the analysis in Table 9 with members' models in the first four columns measured using their arguments in the meeting in $t-1$. The last four columns are identical to those reported in Table 9, and are reported for comparison.

	$Alignment_{i,t}$		$Influence_{i,t}$		$Alignment_{i,t}$		$Influence_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model Fit: $\mathcal{L}(D_t \mid m_{i,t})$	0.26	0.16	0.07	0.04				
	(8.02)	(6.25)	(4.20)	(2.95)				
Model Fit: $\mathcal{L}(D_t \mid m_{i,FE})$					0.29	0.13	0.11	0.08
					(8.79)	(2.90)	(4.88)	(2.91)
N	6397	6397	6397	6397	6397	6397	6397	6397
Meeting Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Member Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes
Role Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

TABLE IA.C.8: MODEL FIT AND POLICY DECISION ALIGNMENT, INCLUDING NON-VOTERS

Note: The table repeats the analysis in Table 9 including non-voters as well as voters. The $dissent_{i,t}$ results are not included, since only voting members can dissent.

	$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$		$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model Fit: $\mathcal{L}(D_t\|m_{i,t})$	0.20	0.10	0.05	0.03	-4.62	-2.48						
	(5.53)	(3.67)	(2.95)	(1.86)	(4.35)	(3.02)						
Model Fit: $\mathcal{L}(D_t\|m_{i,FE})$							0.30	0.19	0.11	0.10	-6.76	-3.91
							(9.01)	(3.39)	(4.48)	(3.31)	(5.72)	(1.96)
N	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972
Meeting Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Member Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Role Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

TABLE IA.C.9: MACRO MODEL FIT AND POLICY DECISION ALIGNMENT

Note: The table repeats the analysis in Table 9 using only macro variables (and not financial variables) in the construction of the model fit.

	$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$		$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model Fit: $\mathcal{L}(D_t\|m_{i,t})$	0.15	0.08	0.05	0.03	-4.21	-2.68						
	(4.22)	(2.55)	(2.50)	(2.11)	(4.08)	(2.99)						
Model Fit: $\mathcal{L}(D_t\|m_{i,FE})$							0.19	0.07	0.06	0.04	-4.79	-2.26
							(3.44)	(1.76)	(1.68)	(1.79)	(3.10)	(1.63)
N	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972
Meeting Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Member Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Role Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

TABLE IA.C.10: FINANCE MODEL FIT AND POLICY DECISION ALIGNMENT

Note: The table replicates the analysis in Table 9, using only financial variables (and not macro variables) in the construction of the model fit.

	$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$		$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model Fit: $\mathcal{L}(D_t \mid m_{i,t})$	0.24	0.13	0.04	0.01	-6.13	-3.70						
	(6.61)	(4.28)	(2.01)	(0.73)	(5.55)	(3.68)						
Model Fit: $\mathcal{L}(D_t \mid m_{i,FE})$							0.28	0.13	0.10	0.05	-6.24	-2.46
							(6.89)	(2.46)	(3.18)	(1.92)	(5.22)	(1.37)
N	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972
Meeting Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Member Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Role Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

TABLE IA.C.11: MODEL FIT AND POLICY DECISION ALIGNMENT, REALIZED TRAILING

Note: The table repeats the analysis in Table 9 using the last available quarter-of-quarter inflation and GDP growth and last available unemployment, rather than the Tealbook forecasts.

	$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$		$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model Fit: $\mathcal{L}(D_t \mid m_{i,t})$	0.19	0.12	0.08	0.06	-4.26	-2.44						
	(6.06)	(4.80)	(4.48)	(3.84)	(4.39)	(3.08)						
Model Fit: $\mathcal{L}(D_t \mid m_{i,FE})$							0.19	0.08	0.07	0.04	-3.97	-1.76
							(5.29)	(2.93)	(3.62)	(2.75)	(3.71)	(1.77)
N	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972	3972
Meeting Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Member Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Role Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

TABLE IA.C.12: MODEL FIT AND POLICY DECISION ALIGNMENT, TEALBOOK TEXT-BASED MEASURES

Note: The table repeats the analysis in Table 9, with the measures of incoming data constructed using textual analysis of the Tealbook. Each d_t variable is assigned a score from -3, indicating support for highly accommodative policy, to +3, indicating support for highly restrictive policy, based on the Tealbook discussion of the variable. The prompts to construct the measures are presented in Appendix IA.A.2.6.

	$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$		$Alignment_{i,t}$		$Influence_{i,t}$		$Dissent_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model Fit: $\mathcal{L}(D_t \mid m_{i,t})$	0.22	0.13	0.06	0.05	-5.70	-3.49						
	(5.80)	(4.07)	(3.25)	(3.06)	(4.60)	(3.59)						
Model Fit: $\mathcal{L}(D_t \mid m_{i,FE})$							0.30	0.20	0.10	0.11	-7.31	-5.26
							(6.56)	(3.15)	(3.49)	(3.81)	(4.93)	(2.29)
N	3731	3731	3731	3731	3731	3731	3731	3731	3731	3731	3731	3731
Meeting Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Member Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Role Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

TABLE IA.C.13: MODEL FIT AND POLICY DECISION ALIGNMENT, EX-GFC

Note: The table repeats the analysis in Table 9 excluding observations from the beginning of 2007 through the end of 2009.