Policy by Committee*

Toomas Laarits, Ben Matthies, Kaushik Vasudevan, Will Yang

February 5, 2026

Abstract

Consequential decisions are routinely delegated to groups, yet empirical examination of group decision-making remains limited due to measurement challenges. We address these by analyzing Federal Open Market Committee members' forecasts alongside a novel dataset of their arguments from over forty years of meetings. Departing from prevailing theories emphasizing information aggregation, we find that members hold distinct and persistent models: beliefs about causal relationships and steady-state values that generate divergent interpretations of the same macroeconomic data. We find that groups aggregate models by tilting toward members whose models better fit recent data. We show theoretically that fit-based model aggregation in committees may improve decision quality but risks excess sensitivity to incoming data.

JEL: D23, D71, D84, E52

Keywords: group decision-making, expectations formation, model heterogeneity, monetary policy committee

---

**First Draft Online: April 9, 2025.* 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 our discussants Anna Cieślak, Stephen Hansen, Matias Iaryczower, Adi Sunderam, and Nancy Xu; Nick Barberis, Chris Clayton, Zhi Da, Bill English, Joel Flynn, Paul Fontanier, Stefano Giglio, Peter Hansen, Sid Kankanala, Chad Kendall, Randy Kroszner, Aleksei Oskolkov, Stefano Pegoraro, Ferdinand Pieroth, Carolin Pflueger, Jesse Shapiro, John Shim, Alp Simsek, Jeremy Stein, and Johannes Wohlfart; seminar participants at Yale SOM, the University of Notre Dame, NYU Stern, Columbia Business School, Miami Herbert Business School, Chicago Booth, and Boston College; and conference participants at NBER (Behavioral Finance; Political Economy; Organizational Economics; Big Data, AI, and Financial Economics), NYU WAPFIN, and USC FIFI. We thank Myeongrok Doh, Giovanni Leopizzi, Seback Oh, and Lucas Ribeiro for outstanding research assistance. This paper was written while Vasudevan was visiting Yale University and the University of Chicago, whose hospitality he gratefully acknowledges.

1 Introduction

Organizations and governments routinely delegate consequential decisions to groups rather than to an individual decision-maker. How these groups aggregate the views of their members, and what this implies for decision quality, are central questions in organizational economics and political economy. While a long-standing literature following Condorcet (1785) argues that groups may improve decision quality by pooling diverse information, decisions are often made in groups even when all members observe similar information, settings where information-sharing gains may be limited.

In contrast to the rich theoretical literature on group decision-making, empirical evidence remains scarce, as systematic tests require data on a group's objectives, its decisions, its members' viewpoints and arguments, and sufficient variation to identify how different views are aggregated. We address these challenges by analyzing a setting that is uniquely well-suited for study: the Federal Open Market Committee (FOMC), which sets U.S. monetary policy. The FOMC offers two key advantages. First, it is a structured, high-stakes setting with a repeated cast of decision-makers operating under a clearly stated objective, the dual mandate to pursue maximum employment and stable prices, who make decisions based on a focused set of economic variables. Second, the Fed produces comprehensive data on the decision-making process. These include the full set of policy decisions, voting records, detailed meeting transcripts, and forecasts from committee members about key decision inputs: unemployment, inflation, and output. These detailed data, available from 1976–2019, allow us to analyze members' views, their underlying reasoning, and how these views are aggregated into decisions.

Using these data, we document a mechanism beyond the traditional information aggregation channel. We find that members hold distinct and persistent *models*, defined as beliefs about the data generating process for the macroeconomy. Heterogeneous models lead members to interpret the same incoming data differently, generating differences in macroeconomic outlooks and policy preferences. Examining how the committee aggregates its members' views, we find that its decisions give greater weight to the views of members whose models better fit recent economic data. Taken together, these findings suggest that a key mechanism of group decision-making is model aggregation, not just information aggregation. To evaluate the implications for decision-making quality, we present a theoretical framework that reveals that such model aggregation in groups may enhance decision quality when members are sufficiently willing to consider others' models, but also carries the risk of excess sensitivity to incoming data.

Heterogeneous Models. Our first main result is that FOMC members hold heterogeneous models that lead them to interpret the same data differently. Members' macroeconomic forecasts alone are suggestive of model heterogeneity. Member fixed effects explain between 20% and 44% of variation in one-year-ahead forecasts, after accounting for common time-series variation. This finding is difficult to reconcile with differential private information, as members share their views each meeting yet exhibit persistent disagreement. Instead, we show that members hold distinct beliefs about key *model parameters*: the causal relationships and steady-state values that underlie standard macroeconomic frameworks. We further show that differences in these model parameters explain differences in members' forecasts.

We measure beliefs about steady-state values, such as the neutral interest rate (r^*) and the non-accelerating inflation rate of unemployment (*NAIRU*), from survey data. We extract beliefs about causal relationships from meeting transcripts using a large language model. Specifically, we construct a dataset of all policy-relevant arguments, made by each member at each meeting, by identifying discrete claims about current conditions, causal relationships, or policy preferences. From these arguments, we measure beliefs about the sensitivity of inflation to labor market conditions (*Phillips curve slope*), the interest sensitivity of output (*IS curve sensitivity*), and the transmission of credit conditions to real activity (*financial–real linkages*).

Members hold persistently different beliefs about these parameters. Member fixed effects explain 32% to 64% of variation in long-run steady-state values, and differences in causal parameter beliefs persist for years. Moreover, these differences generate systematic forecast disagreements: members who perceive a higher steady-state unemployment rate forecast higher inflation at any given unemployment level. A distinctive prediction of model heterogeneity is that causal parameters should generate forecast differences in *state-dependent* ways. In tight labor markets, a member who believes the Phillips curve is steep should forecast higher inflation relative to other members in the meeting, but lower inflation when labor markets are slack. We document this pattern in our data. Beliefs about the IS curve and financial–real linkages similarly interact with policy stance and credit conditions to explain variation in growth forecasts. These state-dependent patterns are distinct from simple "hawk-versus-dove" differences, in which members persistently forecast higher or lower inflation regardless of conditions.

We further characterize the nature of disagreement by examining the information members cite in their arguments. We find that 82% of arguments reference publicly available sources such as government statistics and Fed staff analysis. We score each argument to capture its support for accommodative versus restrictive policy and also classify the arguments by economic topic. Scores constructed from arguments citing only public information correlate at 0.93 with scores from all arguments, indicating that arguments citing private sources do not push members toward systematically different positions. Moreover, 78% of disagreement reflects differing interpretations of shared topics rather than topic selection. While information sharing may play a role, these patterns indicate that expressed disagreements largely take the form of differing interpretations of common information, consistent with model heterogeneity.

We validate our transcript-based measures, including parameter beliefs, argument scores, and information sources, against human coders. Correlations between LLM and human classifications are comparable to, or exceed, correlations between pairs of human coders.[1] We also benchmark against two alternative large language models, Llama 70B and 8B. For our regressions relating parameter beliefs to forecasts, we further correct for potential bias in LLM classifications by estimating the conditional distribution of human classifications given LLM output, adjusting both for directional misclassification and for false positives, and find that our main inferences are unchanged.

Model Aggregation. Our second main result is that the committee systematically tilts its decisions toward the preferences of members whose models better fit recent economic data. To examine how the committee aggregates different models, we construct a new dataset of decisions adopted at each meeting. These include not only the target policy rate, but also the exact phrasing of the policy statement and unconventional policies, a broader range of outcomes than typically studied.

We measure each member's expressed alignment with these decisions and their influence over them. Alignment ranges from complete opposition to perfect agreement, capturing how well a member's preferences are represented in the final decision. This measure can also capture a member's impact on decisions through discussions prior to the meeting. Our measure of influence captures proactive contributions to shaping decisions within the meeting itself. Together, these measures provide substantially more granular variation than voting data alone.

We assess model fit based on forecast accuracy. Forecasts are the observable output of members' models; comparing them to outcomes realized over the subsequent intermeeting period reveals which models better fit recent data. This timing reflects the deliberative process members themselves describe, where views expressed in one meeting are evaluated as new data arrive before the next, and can thereby gain or lose traction in sub-

[1] We provide details on the validation methodologies for these measures and the decision-based measures described below, including human coding protocols, reliability metrics, and comparisons across large language models in Appendix IA.C.

sequent decisions (Schonhardt-Bailey, 2013). Concretely, consider a member who believes in a steep Phillips curve. Upon observing tight labor markets, this member forecasts high inflation. If realized inflation subsequently comes in high, the data validate this view. We test whether such validation translates into greater weight on that member's views in the next meeting's decision.

Across policy alignment, policy influence, and dissents, we find that members with better-fitting models have greater sway over committee decisions within a given meeting. Members with a one standard deviation better model fit show 0.10 to 0.18 standard deviations higher alignment with committee decisions and 0.05 to 0.14 standard deviations greater influence. The probability of dissent decreases by 3 to 6 percentage points per standard deviation improvement in fit, economically meaningful given the 6.9% baseline dissent rate.

Model Aggregation and Decision Quality. To explore the implications of fit-based model aggregation for committee decision quality, we provide a simple theoretical framework of group decision-making under model uncertainty. Committee members hold different, partially misspecified models but observe common data that provide a noisy signal about which model better approximates the true data-generating process. The committee must agree on a model to guide its policy decision. Each member advocates for a combination of all members' models, trading off a desire for a committee model that fits the observed data against a cost of supporting a model that deviates from their own. The committee's choice arises as a Nash equilibrium in which better-fitting models receive greater weight.

Within our framework, 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 competing models. Second, members' models must be "diverse," in that they capture different features 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 on a suboptimal model. Notably, some stubbornness can 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.

These results have practical implications for committee design. Committee performance depends on model diversity, suggesting value in selecting members with different perspectives. Structured deliberation, like the FOMC's systematic discussion of macroeconomic conditions and policy options, may help aggregate diverse models while limiting excess rigidity.

Literature Review. Our paper relates to the literature on collective decision-making in organizations and institutions, surveyed in Gibbons and Roberts (2013): in particular to the literature starting with Marschak and Radner (1972) that considers how tasks, information, and communication should be allocated across agents in order to optimize team performance, and to the broader literature emphasizing organizations as information-processing and communication structures (Arrow, 1974). Prior work takes an informational perspective, emphasizing that group decision-making may improve decisions by aggregating private information, with a focus on designing committees to foster information acquisition and aggregation (Li, Rosen and Suen, 2001; Persico, 2004; Gerling et al., 2005; Sibert, 2006; Gerardi and Yariv, 2008). In contrast, we document that committee members operate with different models to interpret the same information. We argue that a key function of committees is to bring together members with different models. We highlight statistical tradeoffs in the allocation of decision-making influence to members whose models better match recent data, providing a distinct but complementary perspective to work on the allocation of decision-making rights given incentive, control, and communication problems traditionally studied in the literature (Aghion and Tirole, 1997; Baker, Gibbons and Murphy, 1999; Van den Steen, 2010; Bolton and Dewatripont, 2013).

Our approach of focusing on different models for interpreting the same data is situated within a growing body of work on narratives and (mental) models. Recent theoretical and experimental work formalizes the idea of people sharing models to interpret the same data, with a particular focus on characterizing which models may persist and be adopted (Shiller, 2017; Eliaz and Spiegler, 2020; Schwartzstein and Sunderam, 2021, 2025; Montiel Olea et al., 2022; Montiel Olea and Prat, 2025; Barron and Fries, 2025; Charles and Kendall, 2025; Aina, 2025; Aina and Schneider, 2025). We provide evidence from a high-stakes field setting on persistent model heterogeneity, and document how groups make decisions in the face of members with competing models.[2] In characterizing the nature and heterogeneity of mental models in a novel field setting, we complement a small but growing literature that seeks to identify mental models using surveys (Andre et al., 2022; Andre, Schirmer and Wohlfart, 2023; Binetti, Nuzzi and Stantcheva, 2024; Andre et al., 2025; Mei and Wu, 2025) and textual analysis (Ke, 2025; Bastianello, Décaire and Guenzel, 2025). We extract key model parameters from the text of experts' discussions,

[2]Given that members' models may be misspecified, our work relates to a primarily experimental 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).

document evidence on the fit-based aggregation of members' models in a group setting, and consider the associated implications.

Our work also connects to a long tradition in monetary policy and macroeconomics on model uncertainty and its implications. Prior work has presented numerous approaches, such as 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. We document that the FOMC aggregates its members' models by tilting in the direction of models that best fit recent data, and present a framework to evaluate the trade-offs.

Our paper also relates to a large body of 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; Hack, Istrefi and Meier, 2023; López-Moctezuma, 2023), and to work analyzing FOMC transcripts and minutes (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). A substantial focus of this literature is on what makes members 'hawks' versus 'doves' (members who persistently favor tighter policy to curb inflation versus those who prefer easier policy to support economic activity), using educational background, early life experiences, and discussions extracted from transcripts (Malmendier, Nagel and Yan, 2021; Bordo and Istrefi, 2023; Howes et al., 2026).[3] Relative to these papers, we propose and provide evidence for a different source of heterogeneity: differences in members' models of the macroeconomy. This mechanism predicts that forecasts will vary systematically with economic conditions; a member may be relatively hawkish in some periods and relatively dovish in others, depending on her model and current conditions. We also document that the committee tilts its decisions toward the views of members with better fitting models, and theoretically characterize the resulting trade-offs.

[3]Two recent papers extract information from FOMC meeting minutes and transcripts. Kakhbod, Kermani and Maciel (2025) study Fed policy stances using FOMC meeting minutes to extract the FOMC's attribution of inflation to supply-, demand-, or expectations-driven factors. Howes et al. (2026) study heterogeneous policy preferences and the reasoning and justifications of hawks versus doves. Relative to these papers, we propose a different idea: that members hold heterogeneous models of the macroeconomy. We show how different models will generate differences in members' macroeconomic beliefs that are directly measurable using their forecasts, and characterize the committee's adoption of models based on model fit. Additionally, a recent body of work also documents the effect of FOMC's geographic composition on its decisions (Bobrov, Kamdar and Ulate, 2025; Fos, Tamburelli and Xu, 2025; Fos and Xu, 2025). Our focus differs in examining how members' perspectives are weighted in the group decision-making process.

2 Setting

Our sample consists of 365 regularly scheduled FOMC meetings from April 1976 through December 2019. The FOMC typically meets eight times per year to formulate monetary policy. The committee normally consists of 12 voting members who determine the committee's decision by vote: seven governors of the Federal Reserve Board in Washington, DC, including a chair; the president of the Federal Reserve Bank of New York; and four members selected on a rotating basis from the presidents of the 11 other regional Federal Reserve Banks.[4] 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 with 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 how we process the transcript data in Internet Appendix IA.C.10. Prior to each meeting, the Federal Reserve Board staff prepares Tealbook A (formerly the Greenbook), which provides discussion 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 serve 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.

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.[5] The former 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

[4] 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.

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

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.

Alongside the transcript data and outcomes from each meeting, we also study forecasts of real GDP growth, inflation, and the unemployment rate reported by FOMC members as part of the Monetary Policy Report to Congress (the "MPR") and the Summary of Economic Projections (the "SEP"). The MPR, studied in Romer and Romer (2008) and Romer (2010), provides forecasts submitted by committee members in conjunction with an FOMC meeting preceding the report's submission in February and July of each year. The dataset covers forecasts reported from 1992 through the end of 2007. The February forecasts are made for the current year and the July forecasts for the current year and the following year. Starting in 2008, the SEP reports projections quarterly (March/June/September/December), extends the horizon (including "long-run" values), and, since 2012, includes participants' assessments of the appropriate Federal Funds rate. The data are released in an anonymized manner, alongside the policy statement of the committee. The FOMC releases identities associated with each forecast with a five-year lag.[6]

3 Beliefs and Model Heterogeneity

In this section we provide evidence that FOMC members hold distinct and persistent models, defined as beliefs about causal relationships and steady-state values. After documenting persistent forecast disagreement, we show that members exhibit substantial disagreement about key model parameters, including the non-accelerating inflation rate of unemployment (NAIRU), r^*, g^*, and long-run inflation. Then, we use FOMC transcripts to construct a novel dataset of arguments made by members during meetings. From these arguments, we extract members' beliefs about key causal relationships: the sensitivity of inflation to labor market conditions (*Phillips curve slope*), the interest sensitivity of output (*IS curve sensitivity*), and the transmission of credit conditions to real activity (*financial–real linkages*). These parameter beliefs predict systematic differences in member forecasts conditional on observing the same incoming macroeconomic data. Lastly, we find that committee members make arguments that predominantly cite public, rather than private, information and that disagreements are voiced primarily through different

[6]Members are directed to make forecasts under their belief about "appropriate monetary policy," that is, under the future policy path each member deems most likely to achieve the committee's dual mandate. Relative to forecasts made under a commonly assumed policy, we expect this directive to attenuate forecast differences, e.g., a member who perceives a high degree of inflationary pressures may deem restrictive policy more appropriate, which would in turn temper their forecast of inflation. This may slightly attenuate our findings that rely on forecast heterogeneity, for example in Sections 3.2 and 4.4.

interpretations of the same economic topics rather than by focusing on different aspects of the economy.

3.1 Motivating Evidence: Fixed Effects in Forecasts

Committee members exhibit persistent disagreement in their macroeconomic forecasts. Table 1 reports the share of variation in inflation, GDP growth, and unemployment forecasts explained by member fixed effects, controlling for meeting fixed effects. We use the Kline, Saggio and Sølvsten (2020) leave-one-out estimator, which corrects for bias in the dispersions of estimated fixed effects and R^2 values even under the null of no true fixed effect heterogeneity.

Member fixed effects explain substantial cross-sectional variation in forecasts. For current-year forecasts, R^2 values are 0.18, 0.15, and 0.13 for inflation, growth, and unemployment; the cross-sectional standard deviations of these forecasts after controlling for meeting fixed effects are 0.31%, 0.32%, and 0.13%.[7] These within-meeting differences represent the relevant variation for committee decision-making and are economically large—a 0.31% standard deviation in inflation forecasts, for example, is substantial relative to the 2% target. Subsequent-year forecasts show similar patterns with more pronounced member fixed effects: R^2 values are 0.38, 0.20, and 0.24, and cross-sectional standard deviations of 0.27%, 0.34%, and 0.22%. This persistence extends to policy views. Starting in 2012, the Summary of Economic Projections includes members' assessments of the appropriate Fed Funds rate. Member fixed effects yield R^2 values of 0.30 and 0.44 for current-year and subsequent-year assessments, respectively.

To assess whether the estimated R^2 values for member effects exceed what could arise by chance, we conduct 1,000 bootstrap simulations. Each simulation resamples forecasts with replacement within each meeting, randomizing their assignment to members, and regresses the resampled forecasts on member fixed effects. The 99th-percentile R^2 values from these regressions, reported in the table as 'Bootstrap null (99th pct.),' provide a conservative benchmark under the null of no member fixed effects. The R^2 values far exceed this benchmark, consistent with significant member-level heterogeneity.

Given that FOMC members openly share their views at each meeting, their persistently different forecasts are difficult to reconcile with standard models of differential information, such as Woodford (2003) and Sims (2003), which are commonly applied to macroeconomic forecast data (Coibion and Gorodnichenko, 2015). In these models, members' private signals are noisy observations of the true macroeconomic state, and beliefs

[7]For flow variables (GDP growth and inflation), we net out quarters already realized at t^{SEP} and retain the implied annualized forecast for the remaining quarters of the year.

should converge as members learn from one another's expressed views.[8]

In contrast, model heterogeneity can generate persistently different forecasts even when members update in a Bayesian fashion, as learning about model parameters can be extremely slow in macroeconomic environments. For example, beliefs about steady-state values can generate substantial rigidity in forecasts, as emphasized in prior work (Johannes, Lochstoer and Mou, 2016; Farmer, Nakamura and Steinsson, 2024; Li, Van Nieuwerburgh and Renxuan, 2025; Crump et al., 2025). Consistent with this form of model heterogeneity, the last three rows of Table 1 display variation explained by member fixed effects for reported beliefs about long-run values from the Summary of Economic Projections. The R^2 values of 0.32, 0.63, 0.64, and 0.63 for inflation, growth, unemployment, and the Fed Funds rate indicate persistent disagreements about steady-state values, parameters that should be identical if members shared a common model.

3.2 Model Parameters and Forecasts

The long-run projections in Table 1 provide evidence of heterogeneity in beliefs about steady-state parameters such as r^*, NAIRU, and potential growth. Beyond steady-state values, members may also disagree about causal relationships: the Phillips curve slope, the interest sensitivity of output, and the transmission of credit conditions to real activity. There is substantial scope for disagreement about these model parameters.[9]

Unlike member beliefs about steady-state parameter values, no quantitative measure exists for beliefs about these causal parameters. We extract them from FOMC meeting transcripts, where members regularly discuss how labor market conditions, interest rates, and credit conditions affect inflation and output.

3.2.1 Constructing the Arguments Dataset

We construct a dataset of the arguments made by committee members during FOMC meetings. We define an argument as a discrete policy-relevant claim: a view about macroeconomic conditions, a belief about causal relationships, or a preference regarding policy outcomes. We extract arguments reflecting members' beliefs about growth, employment,

[8] An alternative interpretation is that members strategically misreport their forecasts. This does not affect our broader conclusions that: members argue for policy by sharing models, not just information; committee decisions are made by tilting toward the proposed models that better fit recent data. For expositional simplicity, we discuss the beliefs reported by members in FOMC meetings as their own beliefs.

[9] In large part due to empirical challenges in identifying their magnitudes, there is considerable scope for uncertainty and disagreement about the Phillips curve slope (Cogley and Sargent, 2005; Mavroeidis, Plagborg-Møller and Stock, 2014; McLeay and Tenreyro, 2020), the IS curve (Bernanke and Mihov, 1998; Romer and Romer, 2004; Uhlig, 2005; Nakamura and Steinsson, 2018), and strength of financial-to-real economic transmissions (Gilchrist and Zakrajšek, 2012; Caldara et al., 2016).

and inflation, the same variables for which members provide quantitative forecasts, as well as arguments regarding their beliefs about credit conditions. For each category, we use a large language model to identify policy-relevant passages and extract individual arguments.[10] To ensure reliability and reproducibility, we use chain-of-thought prompting (Wei et al., 2022), require the LLM to justify its reasoning and set the temperature parameter to zero.[11]

Table 2 reports summary statistics on the extracted arguments. Among voting members, 94% discuss growth in a given meeting, 75% discuss inflation, 68% discuss employment, and 59% discuss credit conditions. On average, a voter makes approximately 4.9 arguments about growth, 2.7 about inflation, 2.4 about employment, and 1.8 about credit per meeting. Non-voters show similar patterns. Because their arguments appear similar, we pool voters and non-voters when analyzing arguments.

3.2.2 Measuring Causal Parameter Beliefs

From the arguments dataset, we extract member beliefs about the slope of the Phillips curve, the sensitivity of output to interest rates, and the transmission of credit conditions to real activity. For each parameter, we instruct the LLM to identify statements in which the speaker connects both relevant components causally. Simply mentioning both concepts is insufficient; for the Phillips curve, for example, an argument referencing labor markets and inflation is classified only if the speaker expresses a view about how one affects the other. Appendix IA.C.11.2 provides the full extraction prompts.

Phillips Curve Slope. We identify statements linking labor market indicators to inflation outcomes and classify beliefs as steep, moderate, or flat based on the strength of the perceived relationship. A statement such as "tight labor markets are driving wage pressures that will feed into core inflation" implies a steep curve, while "despite unemployment below 4%, we've seen no acceleration in inflation" implies a flat curve.

Interest Rate Sensitivity. We identify statements linking monetary policy to output and classify beliefs as high, moderate, or low based on the perceived responsiveness. Statements describing policy as "gaining traction," "restraining demand," or expressing concern about overtightening indicate high sensitivity. Statements emphasizing resilience despite policy changes, such as continued growth "despite tightening" or describing transmission as "impaired," indicate low sensitivity. Statements describing qualified or partial

[10] We use the 20250514 release of Claude Sonnet 4. Appendix IA.C.11.1 provides the full extraction prompts and classification criteria. Table IA.4 provides examples.

[11] Another concern raised in other settings is look-ahead bias in LLMs, e.g., Sarkar and Vafa (2024). Look-ahead bias is unlikely to affect our results because our tasks primarily involve classifying textual content rather than predicting outcomes.

effects indicate moderate sensitivity.

Financial–Real Linkages. We identify statements linking credit conditions to real activity and classify beliefs as strong, moderate, or weak based on the strength of the perceived relationship. Statements describing significant feedback effects, for example, using language such as "self-reinforcing," "spiral," or "feedback loop," indicate strong linkages. Statements describing direct effects of credit conditions on activity without explicit feedback, or hedged amplification, indicate moderate linkages. Statements denying meaningful credit-activity connections or emphasizing resilience to financial stress indicate weak linkages.

Table 4 presents illustrative quotations and classifications. For instance, Philadelphia Fed president Charles Plosser directly casts doubt on using output gaps to forecast inflation, scored as reflecting a flat perceived Phillips curve. Boston Fed president Eric Rosengren links high labor market slack to low or falling inflation, scored as reflecting a steep perceived Phillips curve. Chairman Bernanke notes that other factors such as global commodity prices may affect inflation and offset the effects of slack, scored as indicating a moderate Phillips curve slope belief.

Constructing Parameter Measures We aggregate these classifications into member-meeting level parameter measures. First, we score each argument as 1 (steep/high/strong), 0 (moderate), -1 (flat/low/weak), or as not expressing a pertinent view. Then, for each member-meeting, we compute a parameter score by averaging the scores across relevant arguments.

For our analysis, we compute a member's parameter value in meeting t as the expanding window average of their meeting-level scores through meeting $t-1$. We focus on expanding averages because the extracted beliefs have both transitory and persistent components. To illustrate this persistence, Appendix Figure IA.1 sorts members into terciles based on quarterly and annual average parameter scores and plots average parameter values over the subsequent four years. For the Phillips curve and IS curve, top-tercile members maintain higher values than bottom-tercile members after four years; financial–real linkage estimates also persist but converge after four years.[12] We also report results using member fixed effects from expanding window regressions of parameter beliefs on member and meeting fixed effects, which identify parameter values based on differences across members within the same meeting.

[12] There is attrition over time on the committee. We observe at least one parameter value after four years for roughly half of the members; by the 16th subsequent quarter, we observe parameters for 30–44% of original members.

Table 3 reports summary statistics. On average, we extract beliefs for 42% of members within a given meeting regarding the Phillips curve slope, 63% regarding output sensitivity to policy, and 38% regarding financial–real linkages. The expanding window specifications extend coverage to over 90%. The raw parameter values display substantial cross-member heterogeneity, with median cross-sectional standard deviations of 0.68, 0.65, and 0.35 for the three parameters, respectively. The expanding averages retain much of this heterogeneity, with corresponding standard deviations of 0.34, 0.25, and 0.17.

Validation. We validate these classifications through human coding. Human coders, who are graduate students and researchers with backgrounds in economics, independently classified random samples of approximately 200 arguments per parameter without access to the LLM's output. We assess reliability using correlations between LLM and human classifications. Our parameter classifications achieve correlations of $r = 0.80$–0.83 with human coders, approaching or exceeding correlations achieved by two humans coding the same data. We also benchmark against Llama 70B and 8B, two other large language models, which perform below Claude on all parameter measures.[13] Finally, we assess the impact of potential bias on our main analyses via an exercise that imputes values by estimating the conditional distribution of human classifications given LLM output. Appendix IA.C reports comprehensive details.

3.2.3 Measuring Macroeconomic Conditions

A distinctive prediction of model heterogeneity is that causal parameter beliefs should interact with economic conditions to generate differences in forecasts. A steep Phillips curve belief generates different forecasts depending on whether labor markets are tight or slack; an IS curve belief matters differently when policy is restrictive versus accommodative. To test this prediction, we construct measures of relevant economic conditions perceived by members.

Perceived Labor Market Tightness (Phillips Curve). We measure perceived labor market tightness as the gap between the member's perceived NAIRU and the current unemployment rate:

$$\text{Perceived Labor Market Tightness}_{i,t} = u^*_{i,t} - u_t \tag{1}$$

[13] The validation procedures described here for the parameter measures are also applied to argument- and decision-related measures introduced in Sections 3.3 and 4. For decision measures, the complexity of synthesizing entire transcripts required an audit-based approach in which coders reviewed the LLM's extractions rather than coding independently.

where u_t is the unemployment rate as of meeting t and $u^*_{i,t}$ is member i's perceived NAIRU. A positive value indicates unemployment below NAIRU, implying a tight labor market. From April 2009 onwards, members report long-run unemployment forecasts in the Summary of Economic Projections (SEP), which we interpret as their perceived NAIRU. Prior to April 2009, we use the Federal Reserve staff's (Greenbook) most recent NAIRU estimate.

Perceived Policy Tightness (Interest Rate Sensitivity). We measure perceived monetary policy restrictiveness as the gap between the expected real interest rate and the member's perceived neutral real rate:

$$\text{Perceived Policy Tightness}_{i,t} = r_{i,t} - r^*_{i,t} \tag{2}$$

where $r_{i,t}$ is the one-year real interest rate perceived by the member and $r^*_{i,t}$ is the member's perceived neutral real rate. We estimate a member's perceived real rate using their one-year-ahead inflation forecast and Federal Funds rate forecast (from the Greenbook when member forecasts are unavailable). For $r^*_{i,t}$, we use member i's reported long-term Federal Funds rate and inflation forecasts where available (after 2009), the real-time Laubach and Williams (2003) r^* measure from 2005–2009, and an r^* value of 2% prior to 2005, as commonly assumed in prior work (Taylor, 1993).

Credit Stress (Financial–Real Linkages). We measure credit market stress using the spread between Moody's BAA corporate bond yield and the 10-year Treasury yield two days prior to the meeting, demeaned by its trailing three-year rolling average:

$$\text{Credit Stress}_t = (\text{BAA}_t - \text{Treasury10Y}_t) - \overline{(\text{BAA} - \text{Treasury10Y})}_t^{3\text{yr}} \tag{3}$$

Demeaning removes low-frequency trends in credit spreads, isolating cyclical variation. A positive value indicates tighter-than-normal credit conditions; we refer to such periods as exhibiting "high credit market stress." This measure is common across members.

3.2.4 Regression Specification

To test whether systematic forecast heterogeneity arises from the interaction of parameter beliefs and conditions, we estimate panel regressions of the form:

$$f^h_{i,t,v} = \alpha_t + \beta_1 \,\text{Parameter}_{i,t} + \beta_2 \,\text{Condition}_{i,t} + \beta_3 \,(\text{Parameter}_{i,t} \times \text{Condition}_{i,t}) + \varepsilon^h_{i,t,v} \tag{4}$$

where $f^h_{i,t,v}$ is member i's h-year-ahead forecast for variable v at meeting t, α_t is a meeting fixed effect, $\text{Parameter}_{i,t}$ is the member's parameter belief estimated from prior meetings, and $\text{Condition}_{i,t}$ is the relevant economic condition. The coefficient of interest is β_3. A significant estimate indicates that members with different parameter beliefs respond differently to the same conditions: steep Phillips curve believers forecast higher inflation than flat Phillips curve believers when labor markets are tight and forecast lower inflation when labor markets are slack, rather than forecasting uniformly higher inflation. This state-dependence identifies heterogeneity in beliefs about causal relationships. If forecast heterogeneity instead reflected persistent forecaster types—hawks who always forecast high inflation, doves who always forecast low—this would appear in β_1 alone. All variables are standardized to zero mean and unit variance, so coefficients represent standard deviation changes. Standard errors are clustered by member and meeting.

For Phillips curve regressions, the dependent variable is the one-year-ahead inflation forecast and we expect $\beta_3 > 0$. For interest rate sensitivity and financial–real linkage regressions, the dependent variable is the one-year-ahead GDP growth forecast; we expect $\beta_3 < 0$, as members who believe output responds strongly to policy or credit conditions should forecast lower growth when policy is restrictive or credit stress is elevated.

3.2.5 Parameter Beliefs and Forecast Heterogeneity

Table 5 reports estimates of the regression specification in Equation (4). All three interaction coefficients have the predicted signs, positive for the Phillips curve and negative for interest rate sensitivity and financial–real linkages. The magnitudes are economically meaningful. For the Phillips curve, the interaction coefficient of 0.20 implies that when perceived labor markets are one standard deviation tighter, a member with one standard deviation steeper slope belief forecasts 0.20 standard deviations higher inflation. The corresponding estimate in the financial–real linkages expanding average specification of -0.06 implies that a one standard deviation stronger belief in credit-activity linkages is associated with a 0.06 standard deviation lower growth forecast when credit stress is one standard deviation above average. For interest rate sensitivity, the estimate is -0.12. The main effect of perceived labor market tightness is also positive, reflecting another form of model heterogeneity: different beliefs about NAIRU lead members to perceive differing degrees of tightness from the same unemployment rate.

A unique implication of heterogeneity in perceived causal relationships is that the direction of the relationship between parameter beliefs and forecasts should flip depending on economic conditions. Within the same meeting, a member who believes the Phillips curve is steep should forecast higher inflation than a flat-curve believer when labor markets are tight, but lower inflation when labor markets are slack.

Figure 1 illustrates the state-dependent nature of these effects by splitting the sample based on perceived labor market tightness, perceived policy tightness, and credit stress, with all variables demeaned within meeting to isolate cross-sectional variation. The relationship between parameter beliefs and forecasts within meeting reverses sign across tight (blue) and slack (red) conditions, exactly as model heterogeneity predicts. Panel A pools across all three channels after z-scoring beliefs and forecasts within each regime.[14] The consistent pattern confirms that the state-dependence is systematic rather than idiosyncratic to any single parameter. This sign reversal is inconsistent with a simple hawk-dove interpretation, in which members persistently forecast higher or lower inflation regardless of conditions.

We report additional specifications of the regression analysis in the Appendix. Table IA.1 adds member fixed effects to the regressions and yields similar results, suggesting that time-series variation in forecasts can also be explained by interactions between model parameters and economic conditions. Table IA.2 replaces member-specific steady-state values with common estimates, for example using the Greenbook's r_t^* estimate in policy tightness measurements rather than member-specific $r_{i,t}^*$ values from the SEP. Interaction coefficients are similar in magnitude, indicating that heterogeneity in steady-state beliefs does not drive the relationship between causal parameter beliefs and forecasts.

Table IA.3 reports results using alternative parameter constructions: expanding averages that include zeros for meetings in which a member does not discuss a parameter, or forward-filling the most recent single-meeting value without averaging. Including zeros yields results similar to our main specification. Forward-filling yields directionally similar but weaker results, underscoring the importance of the persistent component of parameter beliefs.

Appendix Section IA.C.8 uses our human-coded sample to correct for potential measurement error. We estimate the conditional distribution of human classifications given LLM output, adjusting both for directional misclassification, where the LLM assigns a different score than humans would, and for false positives, where the LLM identifies a relevant belief but humans do not. We re-run the main regressions using bias-corrected parameter measures and find that interaction coefficient values attenuate slightly, but retain similar statistical strength.

[14]The sign of inflation forecasts is flipped so that all three channels have a common directional interpretation.

3.3 Information Content of Arguments

Model heterogeneity explains forecast differences, but members might also hold differential private information. In the FOMC setting, anecdotal accounts indicate that private information has limited impact. As committee vice chair Alan Blinder noted, "you can mostly ignore [differential information] in the monetary policy context because virtually all the data that matter are common knowledge" (Blinder, 2008). Nonetheless, members might have access to different data sources, such as regional business contacts or specialized sectoral knowledge, that lead to different conclusions. We use our arguments dataset to directly evaluate the presentation of private information during committee meetings.

From our arguments dataset, focusing on arguments in which members discuss their beliefs about unemployment, growth, or inflation, we classify each along two dimensions: whether the data it cites reflect public information, readily available to all committee members, or private information unique to individual members or their districts; and the stated source of the data, such as official government statistics, Federal Reserve staff analysis, or district business intelligence. Panel A of Table 6 reports the results. Of the arguments citing data, 82% reference public information while 18% reference private or specialized information. Appendix Figure IA.4 breaks down public and private information by data source category. The most common categories, official government statistics (51% of arguments) and Federal Reserve staff analysis (28%), are overwhelmingly public. Private information is concentrated in district business intelligence, such as reports from regional business contacts.[15]

To test whether private information drives disagreement, we score each belief-based argument from -3 to $+3$ based on its economic outlook. For inflation, positive scores indicate expectations of above-target inflation or upside risks; for growth, positive scores indicate above-trend expectations; for unemployment, positive scores indicate a tight labor market. Across all three, positive scores reflect views that would support tightening policy, while negative scores reflect views that would support easing.[16] We construct two argument score measures at the member-meeting level: an unrestricted measure averaging all arguments, and a restricted measure averaging only arguments that cite public

[15] In Appendix IA.C, we show that Claude systematically over-identifies private information relative to human coders—all disagreements involve Claude classifying as private when humans classify as public. That is, the true prevalence of private information is likely lower than our 18% estimate.

[16] Appendix Table IA.4 presents examples and Appendix IA.C.11.3 reports the full scoring criteria. Figure IA.3 displays that the scores capture cross-sectional variation in forecasts. The individual argument scores, before averaging arguments within member-meeting, correlate at $r = 0.80$ with human coding, approaching the human inter-coder ceiling of $r = 0.84$. Appendix IA.C reports details.

information. If private information provided distinct signals, these scores would diverge. Instead, correlations exceed 0.93 for each variable, indicating that arguments citing private sources do not push members toward systematically different positions.

Lastly, we characterize how disagreement manifests. Do members emphasize different aspects of the macroeconomy, or do they discuss the same aspects but reach different conclusions? To answer this, we classify each argument into topic categories designed to group different economic mechanisms while combining opposing interpretations of the same mechanism. For example, members interpreting strong GDP data as reflecting structural productivity gains and those viewing it as temporary cyclical dynamics would both fall under *Growth Forecasting and Outlook*. We identify 28 categories across the three variables; Appendix IA.A.3 reports the categories and descriptions.

We decompose each member's average argument scores for each variable into a selection effect, capturing how much reflects their choice of topics, and an interpretation effect, capturing how differently they interpret the same topics relative to other members. Panel B of Table 6 presents the results. Interpretation effects account for 78% of total variance in member scores, while selection effects contribute only 19%. This pattern holds across variables, with interpretation effects ranging from 80% (employment) to 88% (growth). Members attend to the same aspects of the economy but reach different conclusions.[17]

Overall, we find that members primarily cite public information in their arguments, that the policy stances are similar whether arguments cite public or private information, and that members attend to the same aspects of the economy yet reach different conclusions. These patterns do not rule out a role for private information, but they are consistent with model heterogeneity contributing to disagreement.

4 Monetary Policy by Committee

We now examine how the committee aggregates members' diverse views into decisions. We construct transcript-based measures of each member's alignment with, and influence over, committee decisions, which provide more granular variation than existing measures such as voting outcomes. Then, we construct measures of model fit using members' forecasts. We find that the committee systematically tilts its decisions toward members whose models better fit recent data.

[17]We describe this decomposition in more detail in Appendix IA.A.2. Appendix Table IA.5 presents the decomposition without meeting fixed effects, where interpretation explains 57% and selection explains 41%. The increased selection variance reflects business cycle variation in topics explaining across-meeting score variation.

4.1 Overview and Framework

We outline a simple regression framework to describe our analysis. Each member i employs a persistent model m_i that maps public data into forecasts. These forecasts yield a policy preference $\gamma_{i,t}$. The committee must aggregate these diverse preferences into a decision:

$$Decision_t = \sum_i w_{i,t}\gamma_{i,t}, \tag{5}$$

where $w_{i,t}$ represents the (implicit) weight given to member i's preference in period t.

We test whether weights depend on model fit. Defining $\text{Fit}_{i,t}$ as a measure of how well member i's model aligns with recent data, we estimate:

$$w_{i,t} = \alpha_t + \beta \times \text{Fit}_{i,t}, \tag{6}$$

where we proxy for $w_{i,t}$ using measures of alignment with and influence on committee decisions, α_t captures meeting fixed effects, and β is our coefficient of interest.

The inclusion of meeting fixed effects allows us to compare members within the same meeting, so a positive β cannot reflect all members updating uniformly toward incoming data; it must reflect differential weighting across members. Under canonical models with homogeneous priors but noisy private information (Sims, 2003; Woodford, 2003), members would converge to common beliefs after sharing information during meetings, predicting $\beta \approx 0$. In contrast, if the committee aggregates heterogeneous models based on their recent performance, we expect $\beta > 0$: members whose models better match recent data receive greater weight in decisions.

4.2 Measuring Members' Weight in Committee Decisions

We use three complementary measures that capture different aspects of influence. First, we use dissent votes ($Dissent_{i,t}$), which indicate minimal weight in the committee's decision: the member's preference was overruled.[18] Dissents occur in 6.9% of member-meeting observations. Second, we construct two transcript-based measures. $Alignment_{i,t}$ ranges from -3 to 3, where 3 indicates expressed agreement with the committee's decisions and -3 indicates complete opposition. $Influence_{i,t}$ ranges from 0 to 3, where 3 indicates the member played a pivotal role within the meeting (for example, by proposing the adopted policy) and 0 indicates no influence (for example, opposing the decision or

[18] We obtain dissent data from Thornton, Wheelock et al. (2014), continuously updated by the St. Louis Federal Reserve.

remaining silent on it). $Alignment_{i,t}$ captures the correspondence between a member's views and committee decisions, while $Influence_{i,t}$ credits proactive, within-meeting advocacy: proposing or actively supporting a decision rather than merely expressing agreement.

Constructing alignment and influence requires first identifying the policy decisions made in each meeting and then scoring each member's position on each decision. We describe each step in turn and then validate our measures against human coding and external benchmarks.

4.2.1 Identifying Policy Decisions

We identify all policy decisions in each meeting by providing an LLM with three inputs: the minutes of the meeting, the pre-circulated policy alternatives (Tealbook B), and the policy discussion section of the transcripts. The LLM extracts adopted decisions from the minutes and cross-references against the transcripts to capture decisions that emerged ad hoc during deliberation. We define monetary policy decisions to include interest rate targets, asset purchase programs, forward guidance, and statement language, which provides a richer set of decisions than the target rate changes traditionally studied in the literature. For auxiliary tests, we classify each decision based on whether it is an interest rate decision, or a non-rate decision, e.g., about communication or unconventional policies such as quantitative easing; and we score each decision based on its stance from -3 (highly accommodative) to +3 (highly restrictive). We report the full prompt in Appendix IA.C.11.6.

4.2.2 Scoring Alignment and Influence

For each member and meeting, we prompt the LLM to score alignment with each decision (-3 to 3) and influence over its adoption (0 to 3). Because we score alignment and influence at the decision level, these measures capture variation across policy dimensions that binary dissent votes cannot. A member may strongly support the target rate decision while opposing the forward guidance language adopted in the same meeting. That said, for our baseline specifications, we construct $Alignment_{i,t}$ and $Influence_{i,t}$ as averages across all decisions in that meeting. Full prompts appear in Appendix IA.C.11.7.

Alignment and influence capture distinct channels through which members shape decisions. A member may affect outcomes through arguments made in prior meetings or informal discussions outside the meeting (Schonhardt-Bailey, 2013), increasing alignment without actively advocating during the meeting itself. We interpret alignment as reflecting weight: if the committee's decision corresponds closely to a member's expressed views, we infer that member's preferences received substantial weight in the aggregation, whether through direct advocacy or through other members converging to similar positions. Influence, by contrast, captures active, within-meeting advocacy: proposing new actions, explicitly endorsing decisions, providing supporting arguments, or proposing language.

Table 7 provides illustrative examples from the August 9, 2011 FOMC meeting. Panel A presents four of the five policy decisions identified by the LLM, with the assigned scores in parentheses: maintaining a target Federal Funds rate of 0–25 bps (-2); forward guidance about maintaining the low target rate through mid-2013 (-2); the addition of language that the committee 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). All scores are negative, reflecting accommodative policy at the zero lower bound.

Panel B of Table 7 shows examples of alignment and influence scores for the August 2011 policy decisions for three voting members: president Narayana Kocherlakota, and governors Daniel Tarullo and Janet Yellen. Narayana Kocherlakota is assigned alignment and influence scores of 3 and 1 for the Fed Funds rate decision. All policy alternatives for the meeting included the same target Fed Funds rate of 0-25 bps, with no discussion or contention, and accordingly, little scope for exerting influence, though Kocherlakota explicitly endorsed the target by supporting Alternative B. Kocherlakota is assigned alignment and influence scores of -3 and 0 for the mid-2013 guidance, which he explicitly argued against, viewing 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. He expressed concerns about the 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. Janet Yellen is assigned alignment and influence scores of 3 and 1 for the Fed Funds rate target, which she supported without offering explicit arguments. Yellen is assigned alignment and influence scores of 3 and 2 for the language about the range of policy tools, explicitly helping craft the adopted language that the committee is "prepared to" employ these tools.

The variation in these scores, even among members who ultimately voted together, shows that our measures capture gradations of agreement and influence beyond binary voting outcomes.

4.2.3 Validation

We validate our LLM-based measures in three ways: human coding of the LLM outputs, comparison with existing measures of dissent and preferred policy rates, and asset price responses to policy announcements.

We use a human coded sub-sample of meetings to assess whether the LLM correctly identified the decisions at each meeting, as well as the decision scores, alignment scores, and influence scores. Correlations between human and LLM scores are 0.987 for decision stance scores, 0.97 for alignment, and 0.94 for influence.[19] We report details on the validation protocol, including sample sizes, inter-rater reliability, and tests for systematic bias, in Appendix IA.C.

We also compare our measures against dissent votes and members' preferred policy rates. Panels A and B of Table 8 sort member-meeting observations into bins by $Alignment_{i,t}$ scores, reporting the average $Alignment_{i,t}$, $Influence_{i,t}$, target rate deviation, percentage of dissents, and number of observations for each bin, separately for voting and non-voting members. Focusing on the voting members, the relationship between alignment and dissent is strong: among voters with $Alignment_{i,t}$ scores in the ranges [-3,-2], (-2, -1], and (-1, 0], dissent rates are 91.7%, 79.6%, and 43.6%, respectively. We also report comparisons with members' preferred target Fed Funds rates, hand-coded by Chappell Jr, McGregor and Vermilyea (2004) for two sub-samples: April 1976 through February 1978, and August 1987 through December 1996. Target rate deviations decrease with alignment, from 21 basis points in the lowest alignment bin to 1 basis point in the highest. The patterns are similar for non-voting members.

Panel C of Table 8 reports average $Alignment_{i,t}$, $Influence_{i,t}$, and dissenting behavior for the chair, governors, and regional presidents. The chair is most strongly aligned with and influential over decisions, with average alignment and influence scores of 2.52 and 1.66, consistent with their role in managing the meeting and consensus building (Blinder, 1999; Chappell Jr, McGregor and Vermilyea, 2004). Governors tend to be better aligned with and slightly more influential over decisions than presidents on average. Lastly, non-voting presidents tend to be the least aligned with decisions and least influential.

These results show that the LLM-based alignment and influence measures correspond

[19] Note, that unlike for parameter beliefs and arguments where human coders classified textual data independently without access to Claude's output, for decisions, coders reviewed Claude's extractions and assessed their accuracy; this audit-based approach was necessary given the complexity of synthesizing entire meeting transcripts for the decision classification exercises.

with existing data: low alignment captures dissenting behavior, and misalignment relates strongly to deviations of members' preferred rates from the chosen target. Given the rarity of dissents and the limited scope of the target rate deviation data, our measures provide more granular variation. We also validate our decision extraction: in Appendix Table IA.18, we show that the average stance scores assigned to policy decisions within a meeting relate strongly to monetary policy surprises measured from asset prices, confirming that the identified decisions capture economically meaningful policy variation.

4.3 Measuring Model Fit

A key feature of FOMC deliberations is that the impact of arguments is often realized across meetings: views expressed in one meeting shape how participants interpret data that arrive before the next meeting. Boston Fed president Cathy Minehan described this feature, noting

> "*I viewed the FOMC as a process with discussion evolving over a series of meetings... views expressed in one meeting could resonate and influence the discussion and decisions of subsequent meetings*" (Schonhardt-Bailey, 2013).[20]

In line with this inter-meeting perspective, we test whether members receive greater weight when their previously expressed views better fit the incoming data. We measure fit using *forecast errors*: the gap between a member's forecast stated in one meeting and the outcome realized before the next. Forecasts are the observable outputs of members' models (e.g., a member who believes in a steep Phillips curve and observes tight labor markets will forecast higher inflation); small subsequent errors indicate that the realized data were consistent with that model. These forecasts reflect underlying model parameter heterogeneity (documented in Section 3), with the advantage that they aggregate many parameters and interactions into a single observable output.

Specifically, we construct two forecast-error based measures of model fit. The first measure, $\text{Fit}^{\text{FE}}_{i,t}$, uses half-year forecasts available in the SEP from 2008 onwards. FOMC members forecast PCE inflation and GDP growth separately for the first and second halves of each year, providing more granular variation than annual forecasts. For $k \in \{\text{Inflation}, \text{Growth}\}$, we define

$$\text{Fit}^{\text{FE}}_{i,t,k} = -\left| v_{t,k} - f^{\text{SEP}}_{i,t-1,k} \right|, \tag{7}$$

[20] Schonhardt-Bailey (2013) also presents qualitative accounts from numerous former FOMC members echoing the cumulative, multi-meeting nature of FOMC deliberations.

where $v_{t,k}$ is the value of variable k for the half-year that has most recently been observed as of meeting t, and $f^{\text{SEP}}_{i,t-1,k}$ is member i's SEP forecast for that same half-year made in the most recent SEP vintage prior to meeting t. We rank members cross-sectionally within each variable-meeting, so that smaller deviations receive higher ranks, and compute $\text{Fit}^{\text{FE}}_{i,t}$ as the average rank across variables. Ranking allows for simple aggregation across variables but the results do not depend on ranking.[21]

The second measure, $\text{Fit}^{\text{FE,impute}}_{i,t}$, addresses two limitations of $\text{Fit}^{\text{FE}}_{i,t}$: limited sample availability (2008 onwards only) and exclusion of unemployment forecasts. For this measure, we impute half-year forecast errors using annual forecasts for inflation, growth, and unemployment and for the earlier sample period (1992–2007).

The intuition is that we observe the Tealbook's half-year forecast error once the half-year is realized, and we credit members who were positioned on the correct side of that error based on their annual forecast deviation from the Tealbook. Specifically, for each $k \in \{\text{Inflation}, \text{Growth}, \text{Unemployment}\}$, we define the imputed half-year forecast error as

$$\text{Fit}^{\text{FE,impute}}_{i,t,k} = -\left| \underbrace{\left(f^{\text{TB}}_{t^{\text{prior}},k} - v_{t,k}\right)}_{\text{TB half-year FE}} + \underbrace{\left(f_{i,t^{\text{prior}},k}(Y) - f^{\text{TB}}_{t^{\text{prior}},k}(Y)\right)}_{\text{Annual forecast gap at } t^{\text{prior}}} \right|, \tag{8}$$

where $v_{t,k}$ is the realized value for variable k for the half-year that has most recently been observed as of meeting t, $f^{\text{TB}}_{t^{\text{prior}},k}$ is the Tealbook's forecast for that same half-year from the most recent vintage prior to meeting t, and $Y = 0$ is the current-year annual horizon. For inflation and GDP growth, half-year forecasts are in annualized terms, so the scales of the two terms in Equation (8) are comparable. The annual forecast gap proxies for the member's half-year deviation from the Tealbook, assuming proportionality between half-year and annual deviations.

For example, suppose the Tealbook forecasted 2.5% inflation for the first half of the year and realized inflation was 3.0%, yielding a Tealbook error of -0.5 percentage points (too low). Member i forecasted 3.2% annual inflation versus the Tealbook's 2.8%, a gap of $+0.4$ percentage points. Under proportionality, member i likely also forecasted higher first-half inflation than the Tealbook—closer to what materialized. The imputed error is $|-0.5 + 0.4| = 0.1$, smaller than the Tealbook's error. A member who forecasted lower annual inflation than the Tealbook would have been even further from realized outcomes.

We rank members cross-sectionally within each variable-meeting and compute $\text{Fit}^{\text{FE,impute}}_{i,t}$ as the average rank across variables. For GDP and PCE from 2008 onwards, we use the

[21] In Section 4.4, we verify using raw values does not affect our main results (Appendix Table IA.13).

directly observed half-year forecast errors from the SEP rather than the imputed values. For unemployment and for the pre-2008 period, we use the imputed half-year forecast errors based on annual forecasts from the Monetary Policy Reports.

We report summary statistics for the fit measures across member characteristics in Appendix Table IA.10. Governors tend to have models that better fit recent data than regional Fed presidents, but there is substantial heterogeneity. Tenure on the committee and holding a PhD in economics bear little relationship to the model fit measures.

4.4 Committee Decisions and Model Aggregation

We evaluate the relationship between members' model fit and their weight in committee decisions by estimating the regressions specified in Section 4.1. Table 9 reports results using three dependent variables: $Dissent_{i,t}$ (an indicator for dissenting votes), $Alignment_{i,t}$ (degree of agreement with decisions), and $Influence_{i,t}$ (role in shaping decisions). All specifications include meeting fixed effects to identify effects from cross-sectional variation in model fit within each meeting. All variables except $Dissent_{i,t}$ are standardized to have zero mean and unit standard deviation, so coefficients represent standard deviation changes for a one standard deviation increase in model fit; coefficients on $Dissent_{i,t}$ represent percentage point changes in dissent probability.

The results show that model fit explains members' alignment with and influence over committee decisions. A one standard deviation improvement in model fit reduces dissent probability by 3.5 to 6.2 percentage points depending on the fit measure, a substantial effect given the 6.9% baseline dissent rate. For alignment, coefficients range from 0.11 to 0.18 standard deviations, indicating that members with better-fitting models are better represented in committee decisions. For influence, coefficients range from 0.06 to 0.15 standard deviations. Figure 2 plots the dynamic responses of a member's influence and alignment to an increase in their model fit in a quarter; a member's influence and alignment rise contemporaneously with improvements in that member's fit, with increased influence and alignment persisting into the following quarter.

The influence results are informative about the mechanism underlying model aggregation. By construction, $Influence_{i,t}$ captures active contributions during the meeting: proposing new decisions, advocating for specific language, or combining different aspects of the policy alternatives. Members whose models better fit recent data are not simply agreeing with policy, they actively shape committee decisions.[22]

[22]Our results are consistent with experimental evidence that documents model fit-based weighting of competing interpretations. For recent examples, see Barron and Fries (2025) Aina and Schneider (2025).

Figure 3 illustrates the link between model fit and decision alignment in 2019, which we discuss through the lens of the July 31, 2019 meeting. In that meeting, the most recent forecasts for 2019H1 GDP growth, from the June 19 SEP, showed considerable disagreement: presidents Kaplan and Mester predicted 2.2% while five members including chair Powell and governor Bowman predicted 2.5% and higher. The realized number in the July Greenbook was 2.8%. PCE forecasts similarly diverged, with the realized rate coming in at 1.54%. Averaging across 2019, Bowman emerges as the most accurate forecaster on both GDP and inflation, while Mester has the lowest model fit with the poorest GDP forecast and one of the poorest PCE forecasts.

These differences in model fit correspond to policy alignment. On July 31, the committee voted to cut the policy rate from 2.25% to 2%, citing forward-looking risks associated with trade policy, global growth, and low inflation. Bowman explicitly supported the decision, stating that "the balance of risks and available data support making a modest downward adjustment to our policy stance," while Mester disagreed with the cut. President Williams, who ranks high on both model fit and alignment, explicitly connected forecast errors to his preferred policy stance: "this pattern of consistently—and predictably, in a way—overpredicting inflation and being overly confident about the return of inflation to its goal, I think, should teach us... that the world has changed in important ways that should affect our decisions." His remarks articulate the evaluation criterion underlying model aggregation, fit to recent data determines which interpretations the committee treats as credible.

While Figure 3 shows a positive relationship between model fit and alignment, president George is a notable exception in the 2019 subsample. Despite high model fit, she dissented from the rate cut while similarly accurate presidents Williams and Daly supported it. George's dissent is a reflection of her interpretation of low inflation through a flat Phillips curve: "inflation has shown little interest in adhering to historical norms or Phillips curve theory." In contrast, Williams and Daly interpreted low inflation as a sign of inadvertently tight policy, noting "u^*... is clearly a lot lower than we thought. And r^* is lower than we thought." This exception illustrates that different underlying models may sometimes have similar explanatory power for recent past data, but generate different forecasts.

The baseline results shown in Table 9 extend to non-rate decisions, including forward guidance, statement language, and unconventional policy tools such as asset purchase programs. These decisions involve substantial discretion and generate considerable deliberation. Appendix Table IA.11 separates rate from non-rate decisions; model fit predicts alignment and influence across both. That members with better-fitting models shape even these discretionary policy choices reinforces the interpretation that model fit determines member weight in committee decision-making.

We report additional alternative specifications in the appendix. Appendix Tables IA.12 and IA.13 report results using fit measures constructed separately for inflation, growth, and unemployment; the results are broadly consistent across variables. Appendix Table IA.14 reports results from regressions including fixed effects indicating whether a member is the chair, a governor, or a regional Fed president; the majority of the effects persist after controlling for role, though role does explain some of the effect. Appendix Table IA.15 excludes the chair from the regressions and finds similar results. Appendix Table IA.28 reports results from using human-coded data to assess the impact of potential LLM-based bias by imputing alignment and influence scores using the conditional distribution of human coded values given LLM labels, and finds the main results persist.

Additional Analyses. Measures of forecasts are not systematically available prior to 1992. We construct alternative transcript-based measures of model fit that allow us to extend our analysis back to 1976. We use members' argument scores from Section 3.3 as proxies for expressed forecasts. Each belief-based argument is scored from -3 to $+3$ based on economic outlook: positive scores indicate expectations of above-target inflation, above-trend growth, or tight labor markets. We compute within-meeting averages by variable. Transcript-based fit compares a member's deviation from the cross-sectional average argument score at meeting $t-1$ to the subsequent change in the cross-sectional average from $t-1$ to t; members whose relative outlook at $t-1$ matches the direction of the committee's subsequent aggregate revision have better fit. Appendix Table IA.16 reports the results, which are qualitatively similar to those using forecast-error-based fit. This transcript-based approach offers a potential path for extending model fit analysis to settings where quantitative forecasts are unavailable.[23]

In Appendix Table IA.17, we show that members' forecasts correspond to the decisions they align with. Panel A uses our decision-level dataset to compare alignment across decisions within the same meeting: members with hawkish forecasts (high growth, high inflation, low unemployment) display greater alignment with and influence over restrictive decisions than accommodative decisions. Panel B shows that such members are more likely to dissent in favor of restrictive policy (using the Thornton, Wheelock et al. (2014) classification). These results provide further evidence that forecasts capture decision-relevant aspects of members' views.

Finally, we consider implications for understanding monetary policy surprises. In

[23] See also Bybee (2025) for using LLMs to proxy for macroeconomic forecasts using newspaper text.

Appendix Section IA.A.6, we show that our policy decision scores relate strongly to high-frequency asset price responses following FOMC announcements (Gürkaynak, Sack and Swanson, 2005; Nakamura and Steinsson, 2018). More restrictive decisions generate larger positive surprises, suggesting that the committee's tilting toward better-fitting models has material effects on policy. This result is consistent with evidence that the FOMC uses a time-varying policy rule (Hack, Istrefi and Meier, 2023), and suggests that monetary policy shocks may partly reflect "change in the political power of individuals on the FOMC" (Ramey, 2016)—with such changes driven by the fit of members' models to recent data.

5 Theoretical Framework

Our empirical analysis indicates that members hold persistent and heterogeneous models for interpreting public information, and that the committee favors those whose models better fit recent data. To explore the implications of this finding for decision-making quality, we develop a theory of group decision-making under model uncertainty. Unlike traditional theories that view committees as information aggregators, we formalize them as mechanisms for combining different models, where members advocate for their own approaches while seeking to match observed data, leading the committee to tilt in the direction of better fitting 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.[24]

[24] In line with our empirical evidence, we consider m_i to be 'structured models,' in the language of Hansen and Sargent (2022): beliefs about state probabilities arising from different parameters (such as the ones we study). Additionally, for expositional ease, we abstract from the model-updating process; the assumptions can be thought to apply 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, realized CPI and GDP growth, for example, 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,

$$v_i = \alpha_i m_i + (1-\alpha_i)m_j, \text{ where } j \neq i \text{ and } \alpha_i \in [0,1]. \tag{9}$$

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:

$$m = \frac{1}{2}(v_A + v_B) = w_A m_A + (1-w_A)m_B, \tag{10}$$

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

$$U_i(\alpha_i;\alpha_j) = \log \mathcal{L}(D_t|m) - \lambda KL(m_i \| v_i), \tag{11}$$

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 model.

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.[25]

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, or to missing hard-to-detect tail risks captured in members' models that did not manifest. 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 on 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.

[25]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 $(\frac{1}{3} + \epsilon, \frac{1}{3}, \frac{1}{3} - \epsilon)$.

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

$$\begin{aligned}
\Delta \equiv & \mathbb{E}_\mu(\log(w_A^* m_A + (1 - w_A^*) m_B)) - \max\{\mathbb{E}_\mu(\log m_A), \mathbb{E}_\mu(\log m_B)\}, && (12)\\
\beta(\lambda) \equiv & \mathbb{E}_\mu(w_A(\lambda) - w_A^*), && (13)\\
\sigma^2(\lambda) \equiv & \mathbb{V}_\mu(w_A(\lambda)), \text{ and} && (14)\\
c \equiv & -\mathbb{E}_\mu\left(\frac{\partial^2}{\partial w_A^2} \log \mathcal{L}(D_t | w_A m_A + (1 - w_A) m_B)|_{w_A = w_A^*}\right), && (15)
\end{aligned}$$

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 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

$$\Delta > \frac{c}{2}(\beta(\lambda)^2 + \sigma^2(\lambda)), \tag{16}$$

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.

Notably, 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*

$$\mathbb{V}_\mu(w_A(0)) > \left(\frac{1}{2} - w_A^*\right)^2, \tag{17}$$

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^*,

$$\mathbb{E}_\mu\left[(w_A - w_A^*)^2\right] = \underbrace{(\mathbb{E}_\mu(w_A - w_A^*))^2}_{\text{Bias}} + \underbrace{\mathbb{V}_\mu(w_A)}_{\text{Variance}}. \tag{18}$$

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 empirical findings relating the committee's tilting behavior to monetary policy surprises, potentially stemming from market participants not understanding the FOMC's reaction function.*[26]

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 (11) to be

$$U_i(\alpha_i; \alpha_j) = \log \mathcal{L}(D_t|m) - \lambda KL(m_i \| v_i) - \gamma KL(m \| v_i). \tag{19}$$

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.

[26]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.

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. However, member selection is important, as 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. At the same time, it is possible for the committee to tilt too far in the direction of models that better fit recent data, because limited data can only imperfectly reveal the true model. For example, a model may capture recent data well but do a poor job of capturing unobserved tails of the distribution. Tilting too far in the direction of such a model is likely to be suboptimal.[27]

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.

[27] For related discussion of committee members' concerns about risks, see Cieslak, Hansen and Pang (2025).

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

Aghion, Philippe, and Jean Tirole. 1997. "Formal and real authority in organizations." *Journal of Political Economy*, 105(1): 1–29.

Aina, Chiara. 2025. "Tailored stories."

Aina, Chiara, and Florian Schneider. 2025. "Weighting competing models."

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."

Arrow, Kenneth J. 1974. *The limits of organization.* WW Norton & Company.

Baker, George, Robert Gibbons, and Kevin J Murphy. 1999. "Informal authority in organizations." *Journal of Law, Economics, and Organization*, 15(1): 56–73.

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

Barron, Kai, and Tilman Fries. 2025. "Narrative Persuasion."

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."

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 Ilian Mihov. 1998. "Measuring monetary policy." *Quarterly Journal of Economics*, 113(3): 869–902.

Binetti, Alberto, Francesco Nuzzi, and Stefanie Stantcheva. 2024. "People's understanding of inflation." *Journal of Monetary Economics*, 148: 103652.

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.

Bolton, Patrick, and Mathias Dewatripont. 2013. "Authority in organizations." *Handbook of Organizational Economics*, 342–372.

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.

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."

Caldara, Dario, Cristina Fuentes-Albero, Simon Gilchrist, and Egon Zakrajšek. 2016. "The macroeconomic impact of financial and uncertainty shocks." *European Economic Review*, 88: 185–207.

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.*

Charles, Constantin, and Chad W Kendall. 2025. "Causal Narratives."

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, and Hao Pang. 2025. "Risk Management in Monetary Policy: A Review with Asset Pricing Implications."

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. 2015. "Information rigidity and the expectations formation process: A simple framework and new facts." *American Economic Review*, 105(8): 2644–2678.

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

Crump, Richard K, Stefano Eusepi, Emanuel Moench, and Bruce Preston. 2025. "How Do We Learn About the Long Run?"

Eliaz, Kfir, and Ran Spiegler. 2020. "A model of competing narratives." *American Economic Review*,

110(12): 3786–3816.
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.
Gerardi, Dino, and Leeat Yariv. 2008. "Information acquisition in committees." *Games and Economic Behavior*, 62(2): 436–459.
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.
Gibbons, Robert, and John Roberts. 2013. *The handbook of organizational economics.* Princeton University Press.
Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit spreads and business cycle fluctuations." *American Economic Review*, 102(4): 1692–1720.
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."
Hansen, Lars Peter, and Thomas J Sargent. 2008. *Robustness.* Princeton university press.
Hansen, Lars Peter, and Thomas J Sargent. 2022. "Structured ambiguity and model misspecification." *Journal of Economic Theory*, 199: 105165.
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.
Howes, Cooper, Marc Dordal i Carreras, Olivier Coibion, and Yuriy Gorodnichenko. 2026. "How Monetary Policy Is Made: Lessons from Historical FOMC Discussions." National Bureau of Economic Research.
Iaryczower, Matias, Gabriel López-Moctezuma, and Paola Moscariello. 2025. "Career Concerns in Collective Decision-Making: The Federal Open Market Committee."
Johannes, Michael, Lars A Lochstoer, and Yiqun Mou. 2016. "Learning about consumption dynamics." *Journal of Finance*, 71(2): 551–600.
Kakhbod, Ali, Amir Kermani, and Bernardo Maciel. 2025. "In the Fed's Mind."
Ke, Shikun. 2025. "Analysts' Belief Formation in Their Own Words."
Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-out estimation of variance components." *Econometrica*, 88(5): 1859–1898.
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.
Laubach, Thomas, and John C Williams. 2003. "Measuring the natural rate of interest." *Review of Economics and Statistics*, 85(4): 1063–1070.
Li, Hao, Sherwin Rosen, and Wing Suen. 2001. "Conflicts and common interests in committees." *American Economic Review*, 91(5): 1478–1497.
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.
Marschak, Jacob, and Roy Radner. 1972. *Economic Theory of Teams. Cowles Foundation for Research in Eco-*

nomics at Yale University. Monograph Series, New Haven and London:Yale University Press.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H Stock. 2014. "Empirical evidence on inflation expectations in the New Keynesian Phillips Curve." *Journal of Economic Literature*, 52(1): 124–188.

McLeay, Michael, and Silvana Tenreyro. 2020. "Optimal inflation and the identification of the Phillips curve." *NBER Macroeconomic Annual*, 34(1): 199–255.

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.

Mei, Pierfrancesco, and Lingxuan Wu. 2025. "Thinking about the economy, deep or shallow?"

Montiel Olea, José Luis, and Andrea Prat. 2025. "Competing Ideologies: Fit, Simplicity, and Fear." *Mimeo*.

Montiel Olea, José Luis, Pietro Ortoleva, Mallesh M Pai, and Andrea Prat. 2022. "Competing models." *Quarterly Journal of Economics*, 137(4): 2419–2457.

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

Persico, Nicola. 2004. "Committee design with endogenous information." *Review of Economic Studies*, 71(1): 165–191.

Primiceri, Giorgio E. 2006. "Why inflation rose and fell: policy-makers' beliefs and US postwar 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, Christina D, and David H Romer. 2004. "A new measure of monetary shocks: Derivation and implications." *American Economic Review*, 94(4): 1055–1084.

Romer, Christina D, and David H Romer. 2008. "The FOMC versus the staff: where can monetary policymakers add value?" *American Economic Review*, 98(2): 230–235.

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."

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. 2025. "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.

Shiller, Robert J. 2017. "Narrative Economics." *American Economic Review*, 107(4): 967–1004.

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

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of Monetary Economics*, 50(3): 665–690.

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.

Uhlig, Harald. 2005. "What are the effects of monetary policy on output? Results from an agnostic identification procedure." *Journal of Monetary Economics*, 52(2): 381–419.

Van den Steen, Eric. 2010. "Interpersonal Authority in a Theory of the Firm." *American Economic Review*, 100(1): 466–490.

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.

Woodford, Michael. 2003. "Imperfect common knowledge and the effects of monetary policy." *Knowledge, information, and expectations in modern macroeconomics: In honor of Edmund S. Phelps*, 25(1): 4.

Tables and Figures

	Inflation	Growth	Unemployment	Fed Funds
Y0 Member FE R^2	0.18	0.15	0.13	0.30
Y0 Bootstrap null (99th pct.)	0.05	0.05	0.06	0.05
$\sigma_{y,Y0}$ (percentage points)	0.31	0.32	0.13	0.19
Y1 Member FE R^2	0.38	0.20	0.24	0.44
Y1 Bootstrap null (99th pct.)	0.06	0.05	0.05	0.05
$\sigma_{y,Y1}$ (percentage points)	0.27	0.34	0.22	0.49
Long-Run Member FE R^2	0.32	0.63	0.64	0.63
Long-Run Bootstrap null (99th pct.)	0.06	0.04	0.04	0.05
$\sigma_{y,LR}$ (percentage points)	0.10	0.17	0.27	0.29

TABLE 1: HETEROGENEITY IN COMMITTEE MEMBERS' FORECASTS

Note: This table reports R^2 values from regressions of members' forecasts of inflation, growth, unemployment, and the Fed Funds rate on member fixed effects, controlling for meeting fixed effects. Forecasts are from the Monetary Policy Report to Congress (beginning in 1992) and the Summary of Economic Projections (beginning in 2008); rows labeled Y0 correspond to forecasts for the remainder of the present year, rows labeled Y1 correspond to forecasts of the subsequent year, and rows labeled Long run correspond to long-run forecasts. The regressions are estimated with the Leave-One-Out estimator of Kline, Saggio and Sølvsten (2020). Rows labeled "Bootstrap null (99th pct.)" report 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. Rows labeled $\sigma_{y,Y}$ report the standard deviation of forecasts, after controlling for meeting fixed effects, in percentage points. Data for Fed Funds rate forecasts start in 2012, and data on long-run averages for the variables start in 2009.

Voting Members	Inflation	Growth	Employment	Credit
Members per meeting		——— 10.9 ———		
Prop. making arg. per meeting	75%	94%	68%	59%
Args per member-meeting	2.7	4.9	2.4	1.8
Non-voting Members	Inflation	Growth	Employment	Credit
Members per meeting		——— 6.7 ———		
Prop. making arg. per meeting	72%	88%	71%	53%
Args per member-meeting	2.6	4.6	2.8	1.4

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 proportion of members per meeting making an argument, and the third row displays the average number of arguments per member per meeting.

	Phillips Curve			IS Curve			Fin–Real Linkages		
	Raw	Exp Avg	Exp FE	Raw	Exp Avg	Exp FE	Raw	Exp Avg	Exp FE
PANEL A: COVERAGE									
Number of Meetings	354	365	353	365	365	364	363	365	361
Speakers per Meeting	7.56	16.26	16.40	11.01	17.28	17.32	6.75	16.77	16.88
Coverage (%)	41.70	92.40	90.10	62.50	98.10	98.10	38.20	95.30	94.80
PANEL B: PARAMETER CHARACTERISTICS									
	Phillips Curve			IS Curve			Fin–Real Linkages		
	Raw	Exp Avg	Exp FE	Raw	Exp Avg	Exp FE	Raw	Exp Avg	Exp FE
Mean Level	0.45	0.53	1.26	0.58	0.58	-0.14	-0.07	-0.08	-0.10
Median Cross-Sectional Std. Dev.	0.68	0.34	0.34	0.65	0.25	0.26	0.35	0.17	0.17

TABLE 3: PARAMETER BELIEFS: SUMMARY STATISTICS

Note: The table reports summary statistics for model parameter beliefs extracted from FOMC transcripts. Panel A reports coverage statistics: the number of meetings with at least one parameter observation, the average number of speakers with an observation per meeting, and the percentage of all member-meeting observations with data. Panel B reports the mean level of parameter beliefs across all observations and the median within-meeting cross-sectional standard deviation. Columns show statistics for three specifications: Raw (original extracted values), Exp Avg (expanding window average of values by meeting), Exp FE (expanding window member fixed effect values controlling for meeting fixed effects).

Speaker	Date	Parameter	Value	Quote
Eric Rosengren	Aug-2011	Phillips Curve	Steep	"Given the large degree of resource slack in the economy, it is quite possible that the inflation rate will soon be falling. ...labor markets are weak"
Charles Plosser	Mar-2011	Phillips Curve	Flat	"The current experience in the United Kingdom... where inflation is rising in spite of what appear to be large output gaps, should make us at least somewhat uncomfortable about relying on these gaps too heavily..."
Ben Bernanke	Dec-2011	Phillips Curve	Moderate	"...there are more factors than slack that affect inflation.... we've seen a pretty strong global economy that has affected commodity prices"
Jack Guynn	Dec-2003	IS Curve	High	"Should we raise rates too high too quickly this time, we risk choking off demand for housing and durables..."
Jack Guynn	Mar-2001	IS Curve	Low	"a substantially easier monetary policy...would have a minimal impact on real output over the near term."
Frederic Mishkin	Sep-2007	Fin–Real Linkages	Strong	"...we have the potential for a vicious circle ... we have a financial disruption, which means that it's harder to allocate capital...and, as a result, you get a contraction in economic activity."
Donald Kohn	Aug-2007	Fin–Real Linkages	Weak	"Well-capitalized banks and opportunistic investors will come in and fill the gap...credit conditions will be tighter...and the effect on output will probably not be very large."

TABLE 4: PARAMETER BELIEF EXTRACTIONS: EXAMPLES

Note: The table presents examples of parameter belief extractions from FOMC transcripts. Phillips curve slope is classified as steep, moderate, or flat. Interest rate sensitivity is classified as high, moderate, or low. Financial–real linkages are classified as strong, moderate, or weak.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.:	Phillips Curve Inflation (Y+1)		Fin–Real Linkages GDP Growth (Y+1)		IS Curve GDP Growth (Y+1)	
Belief Measure:	Expanding Avg	Expanding FE	Expanding Avg	Expanding FE	Expanding Avg	Expanding FE
Belief × Condition	0.202	0.388	−0.059	−0.071	−0.120	−0.099
	(4.69)	(4.75)	(−3.21)	(−2.87)	(−2.19)	(−2.29)
Labor Market Tightness	0.494	0.251				
	(3.56)	(1.80)				
Perceived Policy Tightness					0.057	−0.074
					(0.38)	(−0.53)
Meeting FE	Yes	Yes	Yes	Yes	Yes	Yes
N	1,076	1,076	1,099	1,099	1,111	1,111

TABLE 5: PARAMETER BELIEFS AND FORECAST HETEROGENEITY

Note: The table reports coefficients from regressions of members' one-year-ahead forecasts on interactions between their parameter beliefs and relevant economic conditions. All variables are standardized to have zero mean and unit standard deviation. Belief × Condition denotes the relevant interaction term for each parameter: PC Slope Belief × Labor Market Tightness for Phillips Curve, Credit Channel Belief × Credit Stress for Financial–Real Linkages, and Interest Rate Sensitivity Belief × Perceived Policy Tightness for IS Curve. Expanding Avg computes parameter values as an expanding-window average of prior meeting parameter beliefs. Expanding FE computes parameter values as member fixed effects in expanding-window regressions of prior meeting parameter beliefs on member and meeting fixed effects. Labor Market Tightness equals member-specific NAIRU minus unemployment, using SEP long-run forecasts when available (April 2009 onwards) and Federal Reserve staff estimates otherwise. Perceived Policy Tightness equals the perceived real rate minus the perceived neutral real rate, constructed from members' SEP forecasts where available, and Greenbook forecasts otherwise. Credit Stress equals the BAA-10Y Treasury spread minus its trailing three-year average. t-statistics in parentheses are based on standard errors clustered by member and meeting.

Panel A: Information Cited in Members' Arguments

Summary	
Arguments citing data	60,857
Arguments *not* citing data	3,112
Total arguments	63,969
Information source	
Public information	82%
Private / Specialized info.	18%
Geographical scope	
National	78%
Sectoral	21%
Regional	8%
International	3%

Panel B: Within-Meeting Variance Decomposition of Member Scores

	Selection (%)	Interpretation (%)	Residual (%)
All	19.0	78.0	3.0
Inflation	13.6	84.6	1.8
Growth	10.8	87.6	1.6
Employment	16.9	80.3	2.8

Table 6: Information Content of Member Arguments

Note: Panel A reports summary statistics on the data cited in arguments made by committee members during FOMC meetings. Categories may sum to greater than 100% as a single argument may reference multiple pieces of data. Panel B displays variance decompositions of members' scores into (i) a selection effect capturing the average score of arguments made by other members with the same category composition; (ii) an interpretation effect capturing the average difference between a member's score and arguments made by other members in the same category; and (iii) a residual effect from arguments in categories not discussed by other members. The decomposition ignores covariances and normalizes the sum of variances to 100%.

Panel A: Example Policy Decisions			
Date	Change	Score	Justification
Aug-11	Target range for the Federal Funds rate at 0 to 1/4%.	-2	Significant Accommodation at ZLB.
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 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	1	Supported with Alternative B.
Aug-11	Mid-2013 Guidance	Kocherlakota	-3	0	Opposed calendar-date guidance.
Aug-11	Mid-2013 Guidance	Tarullo	1	2	Urged conditional ("based on projections").
Aug-11	Policy Tools Language	Tarullo	3	3	Proposed adopted language
Aug-11	Fed Funds Target	Yellen	3	1	Supported, not discussed.
Aug-11	Policy Tools Language	Yellen	3	2	Helped craft language.

Table 7: 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 minutes, the policy discussion section of the transcripts, and the pre-circulated policy alternatives (Tealbook B or Bluebook). The Justification columns report paraphrased output from the LLM justifying the scores.

Panel A: Voters						
	$Alignment_{i,t}$ range					
	[-3, -2]	(-2, -1]	(-1, 0]	(0, 1]	(1, 2]	(2, 3]
Average $Alignment_{i,t}$	-2.27	-1.29	-0.23	0.70	1.65	2.76
Average $Influence_{i,t}$	0.02	0.15	0.23	0.52	0.82	1.19
Percentage dissenting votes	91.7%	79.6%	43.6%	16.7%	4.6%	0.7%
Target Rate Deviation (bps)	21	27	19	7	4	1
Number of voter-meetings	12	49	181	514	1025	2193

Panel B: Nonvoters						
	$Alignment_{i,t}$ range					
	[-3, -2]	(-2, -1]	(-1, 0]	(0, 1]	(1, 2]	(2, 3]
Average $Alignment_{i,t}$	-2.34	-1.26	-0.10	0.67	1.64	2.75
Average $Influence_{i,t}$	0.01	0.07	0.17	0.41	0.73	1.02
Target Rate Deviation (bps)	32	30	18	9	6	2
Number of nonvoter-meetings	24	47	445	397	731	1103

Panel C: Cross-Member Statistics				
			Presidents	
	Chair	Governors	Voting	Non-voting
Average $Alignment_{i,t}$	2.52	2.04	1.88	1.72
Average $Influence_{i,t}$	1.66	0.90	0.86	0.75
Percentage dissenting votes	0.0%	5.5%	9.8%	–
Target Rate Deviation (bps)	0	4	5	7
N	365	1784	1825	2422

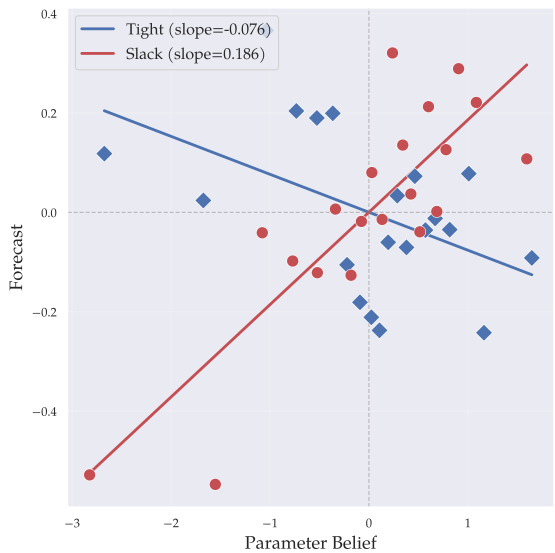
Table 8: Decision Alignment, Influence, and Dissent

Note: Panels A and B group member-meeting observations into bins based on $Alignment_{i,t}$ scores, where $Alignment_{i,t}$ ranges from -3 (strong opposition) to +3 (perfect agreement). For each bin, the table reports average $Alignment_{i,t}$; average $Influence_{i,t}$ (0 to 3); percentage of dissenting votes (Panel A only); and average Target Rate Deviation (the absolute difference between the chosen Fed Funds target and a member's preferred target rate, coded by Chappell Jr, McGregor and Vermilyea (2004), available only for April 1976–February 1978 and August 1987–December 1996). Panel C reports averages by member role: chair, governors, and regional Fed presidents (voting and non-voting).

	Panel A: Half-year Forecast Error-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}_{i,t}^{\text{FE}}$	-6.21	0.181	0.146	0.167	0.114
	(2.86)	(3.07)	(3.01)	(2.87)	(3.16)
N	893	893	893	1498	1498
Meeting FE	Yes	Yes	Yes	Yes	Yes
	Panel B: Half-year Imputed Forecast Error-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}_{i,t}^{\text{FE,impute}}$	-3.51	0.115	0.063	0.105	0.058
	(3.00)	(3.17)	(2.29)	(2.49)	(2.17)
N	2063	2063	2063	3484	3484
Meeting FE	Yes	Yes	Yes	Yes	Yes

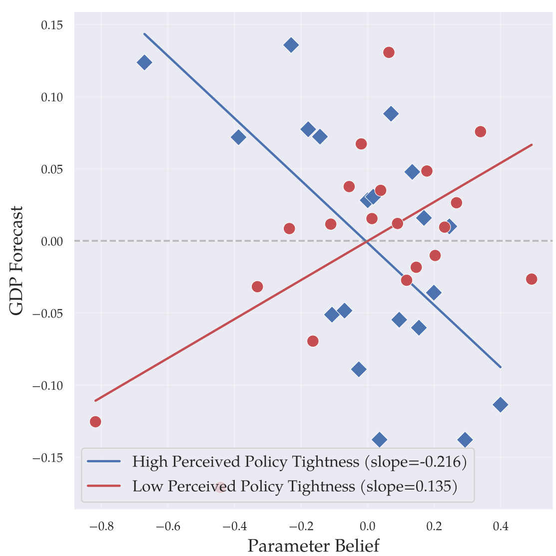
Table 9: Model Fit and Policy Decision Alignment

Note: The table reports estimates from regressions relating a member's model fit to their role in FOMC decisions. Dependent variables are $Dissent_{i,t}$ (an indicator for dissenting votes), $Alignment_{i,t}$ (agreement with the adopted decision), and $Influence_{i,t}$ (role in shaping the decision). $Alignment_{i,t}$ and $Influence_{i,t}$ are standardized, so coefficients represent standard-deviation changes for a one standard-deviation increase in model fit; $Dissent_{i,t}$ coefficients represent percentage-point changes in dissent probability. Panel A uses half-year forecast-error fit based on realized half-year outcomes for inflation and growth (available from 2008 onward). Panel B uses half-year-imputed forecast-error fit, which splices the data used in Panel A with imputed forecast errors computed using Tealbook's most recent half-year forecast error with the member's MPR/SEP–Tealbook annual forecast gap. "Voters" columns restrict to voting members; "All Members" includes all participants. Standard errors are clustered by member and meeting, and t-statistics in parentheses.

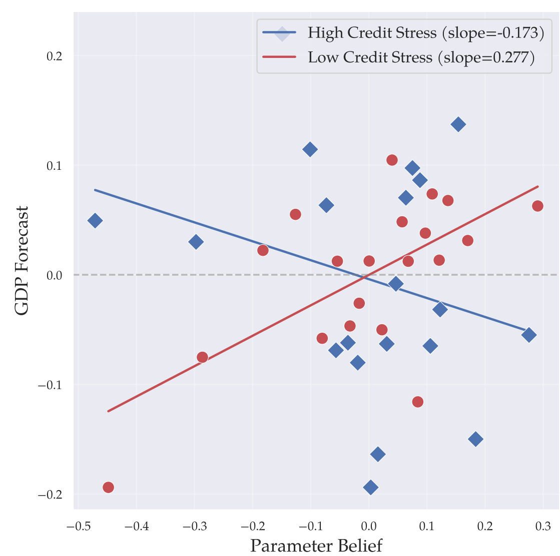


(A) POOLED

(B) PHILLIPS CURVE



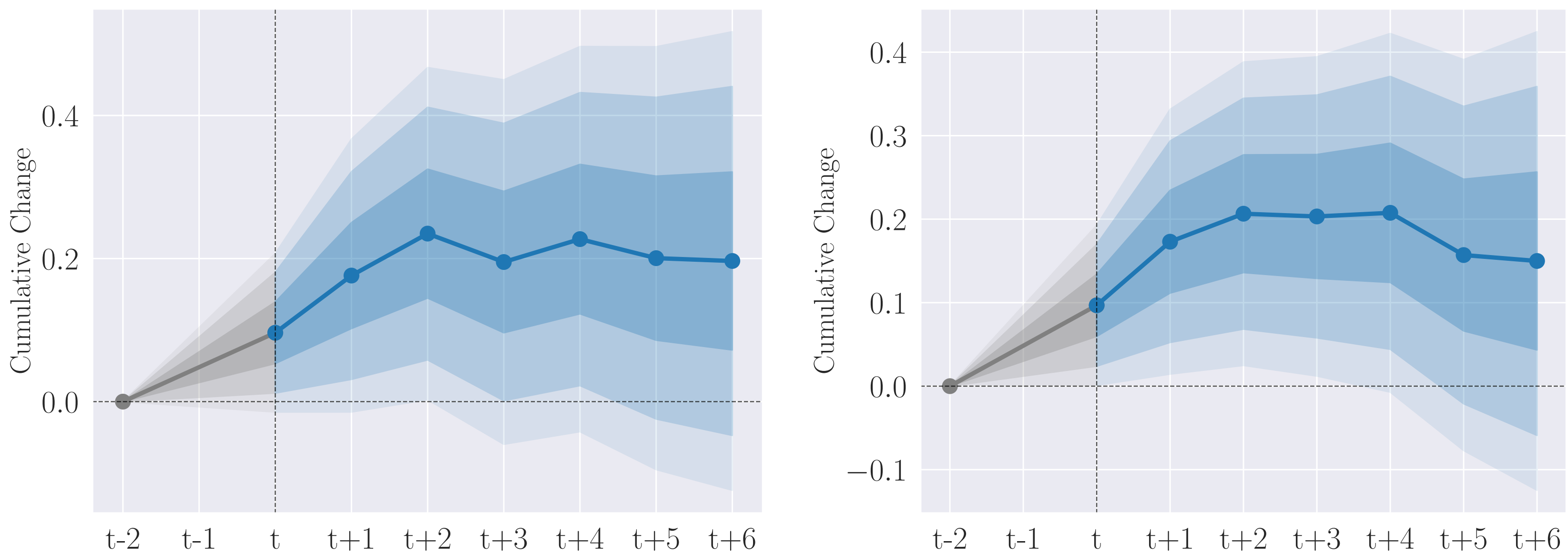
(C) INTEREST RATE SENSITIVITY



(D) CREDIT CHANNEL

FIGURE 1: PARAMETER BELIEFS AND STATE-DEPENDENT FORECASTS

Note: The figure shows how the relationship between members' parameter beliefs and their forecasts varies with economic conditions. Blue indicates tight conditions; red indicates slack conditions. Panel B: Phillips curve slope beliefs vs. inflation forecasts, split by labor market tightness. Panel C: Interest rate sensitivity beliefs vs. GDP forecasts, split by perceived policy tightness. Panel D: Credit channel beliefs vs. GDP forecasts, split by credit stress. Panel A pools across all three channels after z-scoring beliefs and forecasts within each regime; the sign of inflation forecasts is flipped so that all three channels have a common directional interpretation. All panels use 20 bins with meeting-demeaned variables. Slopes are estimated by OLS through binned means.



(A): Alignment (B): Influence

FIGURE 2: ALIGNMENT AND INFLUENCE IN RESPONSE TO CHANGES IN MODEL FIT

Note: The figure shows the cumulative response of a member's alignment with the committee (Panel A) and influence (Panel B) to a change in $\text{Fit}_{i,t}^{FE}$ in the previous quarter ($t-2$ to t). Cumulative changes are measured relative to the two-meeting sum of influence (or alignment) ending at $t-2$, normalized to zero in the figure. The grey region (from $t-2$ to t) captures the contemporaneous relationship between an increase in fit and changes in the dependent variable. The blue region (from t to $t+6$) shows cumulative alignment and influence in subsequent meetings. $\text{Fit}_{i,t}^{FE}$, $Alignment_{i,t}$ and $Influence_{i,t}$ are normalized to have zero mean and unit standard deviation. Regressions include meeting fixed effects and title fixed effects indicating whether a member is the chair, a governor, or a voting or non-voting regional Fed president. Standard errors are clustered by member and meeting date. Shaded bands indicate 68%, 95%, and 99% confidence intervals.

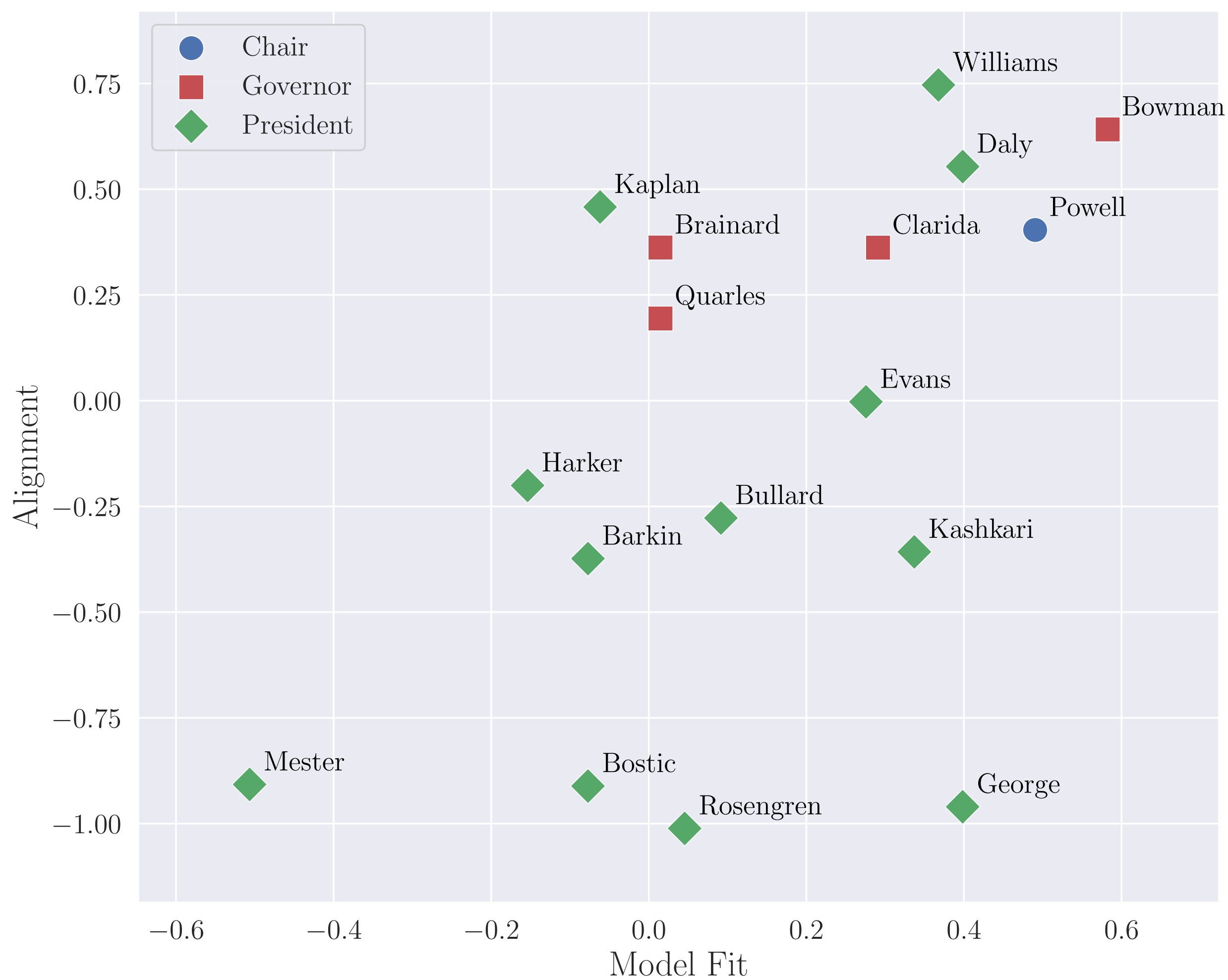


FIGURE 3: MODEL ALIGNMENT AND FIT ON THE COMMITTEE IN 2019

Note: The figure plots members' average model fit against their alignment with committee decisions in 2019. The horizontal axis shows model fit, measured as the rank of a member's half-year forecast errors relative to other members . The vertical axis shows $Alignment_{i,t}$, the average alignment of a member's expressed views with adopted policy decisions. Each point represents one committee member, with marker shapes indicating role: circles for the chair, squares for governors, and diamonds for regional bank presidents. Both variables are standardized to have zero mean and unit standard deviation in the sample.

Internet Appendix for

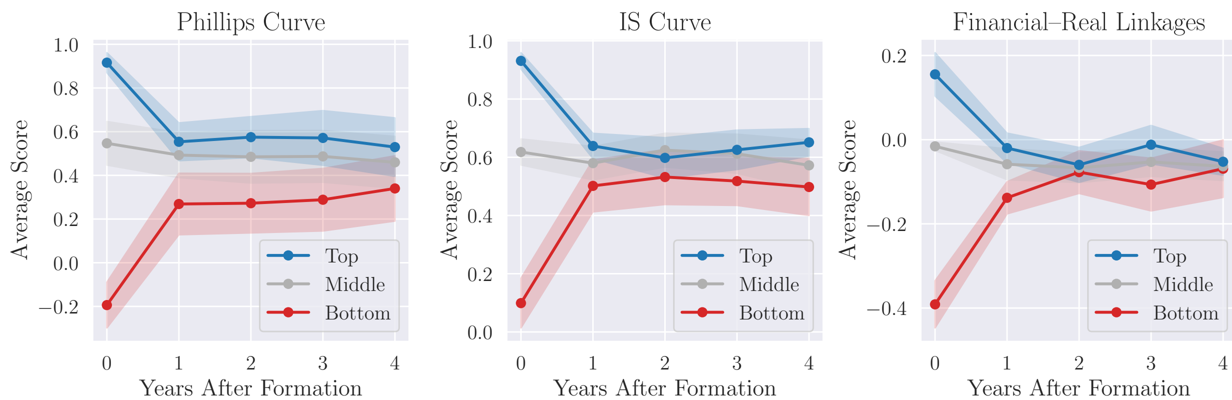
Policy by Committee

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

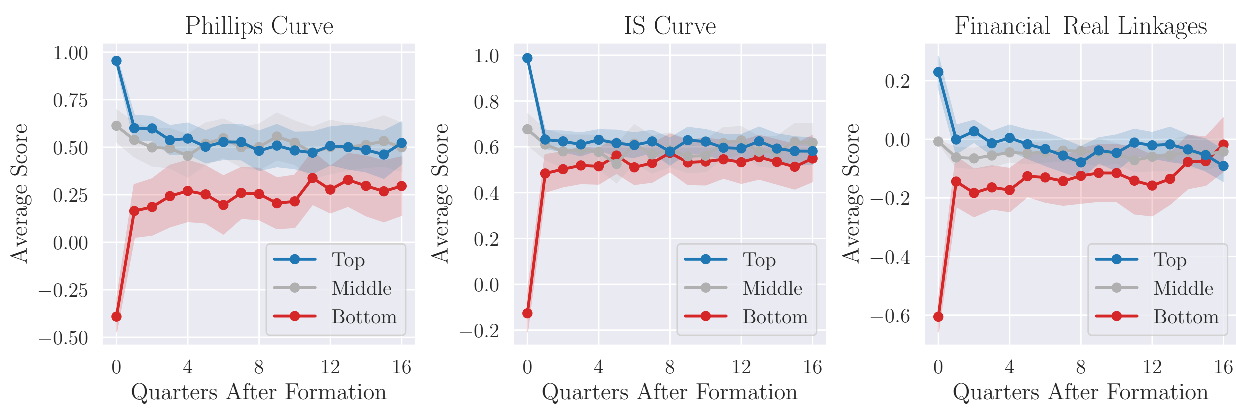
The Internet Appendix consists of two sections. Section IA.A presents additional analysis mentioned in the main text. Section IA.B is a data appendix presenting details on the processing of FOMC transcripts, the LLM prompts used to extract information from the transcripts, and a validation exercise of our argument dataset.

IA.A Additional Empirical Analyses

IA.A.1 Parameter Beliefs and Forecast Heterogeneity



PANEL (A): ANNUAL PARAMETER BELIEFS



PANEL (B): QUARTERLY PARAMETER BELIEFS

FIGURE IA.1: CAUSAL PARAMETER BELIEF PERSISTENCE

Note: This figure displays the evolution of parameter beliefs by initial tercile. Members are assigned to terciles based on their average scores for each parameter during a formation period ($t = 0$). The figure tracks each tercile's mean score over subsequent periods ($t = 1, 2, \ldots$) for the Phillips Curve, IS Curve, and Financial to Real Linkages, with 95% confidence intervals shaded. Panel (a) uses annual averaging; panel (b) uses quarterly. Standard errors are clustered by time and member.

	(1)	(2)	(3)	(4)	(5)	(6)
	Inflation (Y+1) Phillips Curve		GDP Growth (Y+1) Fin–Real Linkages		GDP Growth (Y+1) IS Curve	
Belief Measure:	Expanding Avg	Expanding FE	Expanding Avg	Expanding FE	Expanding Avg	Expanding FE
Belief × Condition	0.171	0.319	−0.066	−0.076	−0.105	−0.088
	(4.03)	(4.10)	(−3.55)	(−3.13)	(−1.83)	(−2.02)
Labor Market Tightness	−0.055	−0.253				
	(−0.33)	(−1.63)				
Perceived Policy Tightness					0.136	0.023
					(1.22)	(0.19)
Meeting FE	Yes	Yes	Yes	Yes	Yes	Yes
Member FE	Yes	Yes	Yes	Yes	Yes	Yes
N	1,076	1,076	1,099	1,099	1,111	1,111

TABLE IA.1: PARAMETER BELIEFS AND FORECAST HETEROGENEITY WITH MEMBER FIXED EFFECTS

Note: This table adds member fixed effects to the specifications in Table 5. All variables are standardized. See Table 5 for variable definitions. *t*-statistics in parentheses are based on standard errors clustered by member and meeting.

	(1)	(2)	(3)	(4)	(5)	(6)
	Inflation (Y+1) Phillips Curve		GDP Growth (Y+1) Fin–Real Linkages		GDP Growth (Y+1) IS Curve	
Belief Measure:	Expanding Avg	Expanding FE	Expanding Avg	Expanding FE	Expanding Avg	Expanding FE
Belief × Condition	0.220	0.414	−0.059	−0.071	−0.223	−0.094
	(4.10)	(3.94)	(−3.21)	(−2.87)	(−1.79)	(−1.95)
Meeting FE	Yes	Yes	Yes	Yes	Yes	Yes
N	1,076	1,076	1,099	1,099	1,111	1,111

TABLE IA.2: PARAMETER BELIEFS AND FORECAST HETEROGENEITY WITH COMMON STEADY-STATE VALUES

Note: This table repeats the analysis in Table 5 using common steady-state values (e.g., r_t^* and u_t^*) rather than member-specific estimates from the SEP. Labor Market Tightness and Perceived Policy Tightness are absorbed by meeting fixed effects in these specifications. All variables are standardized. t-statistics in parentheses are based on standard errors clustered by member and meeting.

	(1)	(2)	(3)	(4)	(5)	(6)
	Inflation (Y+1) Phillips Curve		GDP Growth (Y+1) Fin–Real Linkages		GDP Growth (Y+1) IS Curve	
Belief Measure:	Expanding Avg Zero	Last Realized	Expanding Avg Zero	Last Realized	Expanding Avg Zero	Last Realized
Belief × Condition	0.147	0.142	−0.045	−0.026	−0.069	−0.033
	(2.33)	(3.80)	(−2.82)	(−1.39)	(−2.00)	(−0.94)
Labor Market Tightness	0.618	0.695				
	(4.67)	(5.03)				
Perceived Policy Tightness					−0.010	−0.017
					(−0.07)	(−0.12)
Meeting FE	Yes	Yes	Yes	Yes	Yes	Yes
N	1,131	1,076	1,131	1,099	1,131	1,111

TABLE IA.3: PARAMETER BELIEFS AND FORECAST HETEROGENEITY WITH ALTERNATE MEASURES

Note: This table repeats the analysis in Table 5 using alternative methods to forward-fill members' parameter beliefs. Columns labeled Expanding Avg Zero use expanding-window averages of members' reported parameter beliefs, filling in zeros for meetings in which a member does not discuss a parameter. Columns labeled Last Realized use the most recent previously extracted parameter belief. All variables are standardized. t-statistics in parentheses are based on standard errors clustered by member and meeting.

IA.A.2 Argument Scores and Information Sources

This section provides additional details on our LLM-extracted argument outlook scores and our analysis of information sources.

IA.A.2.1 Examples of Arguments and Scoring

Table IA.4 presents examples of extracted arguments from the August 9, 2011 FOMC meeting. This meeting occurred during a period of zero interest rates and weak recovery from the Global Financial Crisis, with financial markets under pressure from U.S. and European fiscal concerns. The examples illustrate the range of views expressed by committee members. Kocherlakota and Plosser pointed to falling unemployment and rising inflation, receiving scores of $+2$. Evans and Tarullo warned about weak growth and recession risks, receiving scores of -2 and -3 respectively.

Figure IA.2 plots the average outlook score across committee members for each FOMC meeting. The scores display intuitive cyclical patterns, reaching troughs during NBER recessions. Concerns about high inflation were elevated in the 1970s and late 1990s and were subdued in the years following the Global Financial Crisis.

Figure IA.3 validates that outlook scores capture meaningful cross-sectional variation in members' beliefs. We compare argument scores with quantitative forecasts from the Monetary Policy Report to Congress and the Summary of Economic Projections. The figure sorts committee members into deciles based on their cross-sectionally demeaned forecasts and plots these against cross-sectionally demeaned argument scores from the corresponding meeting. The combined panel pools all three variables after standardizing. The strong positive relationship confirms that our LLM-based measures capture the same belief heterogeneity observed in members' numerical predictions.

Figure IA.4 presents the breakdown of information source types and public versus private information by information source type discussed in the text.

Speaker	Variable	Score	Quotes
Narayana Kocherlakota	Employment	$+2$	"The falling rate of unemployment and the rising rate of inflation suggest that ... resource slack is diminishing."
Charles Plosser	Inflation	$+2$	"Headline inflation and core inflation measures have been on an upward trend for most of the past year."
Janet Yellen	Growth	$+1$	"Supply chain disruptions hit autos last quarter, so it's reasonable to anticipate a pickup in production and sales this summer ..."
Janet Yellen	Growth	-2	"I've marked down my growth outlook substantially ... the economy has been running alarmingly close to stall speed ..."
Charles Evans	Growth	-2	"The weak incoming data indicate that it's highly unlikely the U.S. economy will achieve anything like a launch velocity."
Daniel Tarullo	Growth	-3	"Even the most sanguine view ... must acknowledge the economy is now sufficiently vulnerable that a modest shock could send us back into recession."

TABLE IA.4: COMMITTEE MEMBERS' ARGUMENTS IDENTIFIED BY LLM: EXAMPLES

Note: The table presents examples of arguments extracted from the August 9, 2011 FOMC meeting. Each argument is assigned a score from -3 to $+3$. For growth and inflation, negative scores indicate expectations below trend/target, while positive scores indicate expectations above trend/-target. For employment, negative scores indicate expectations of labor market weakening (rising unemployment), while positive scores indicate expectations of labor market tightening (falling unemployment).

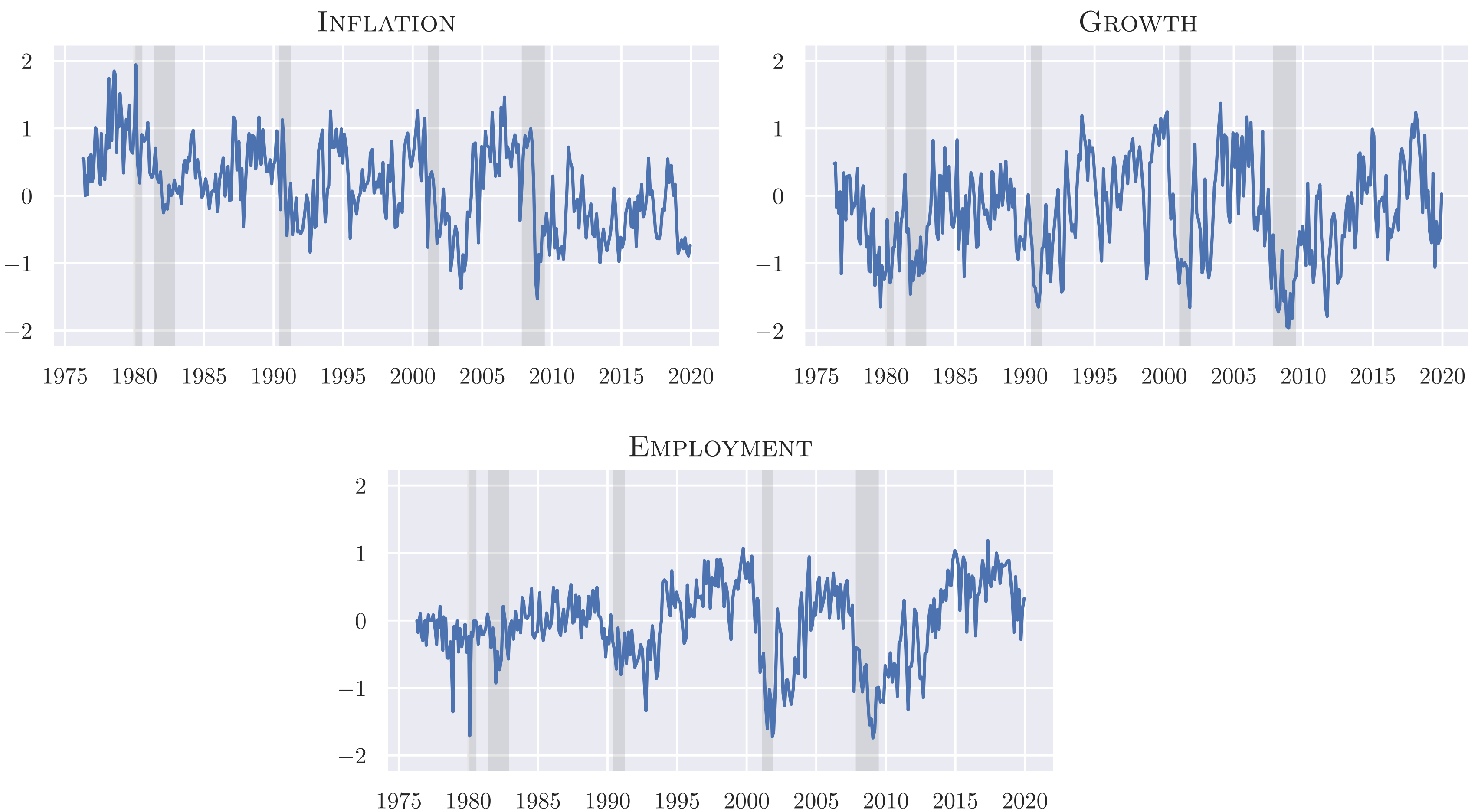


Figure IA.2: Committee Members' Argument Scores over Time

Note: The figure plots the average of outlook scores assigned to committee members for inflation, growth, and employment for each FOMC meeting. For each meeting, the forecast-based arguments made by members are assigned scores from -3 to $+3$. For growth and inflation, negative scores indicate expectations below trend/target, while positive scores indicate expectations above trend/target. For employment, negative scores indicate expectations of labor market weakening, while positive scores indicate expectations of labor market tightening. The dark gray bands indicate NBER recession periods.

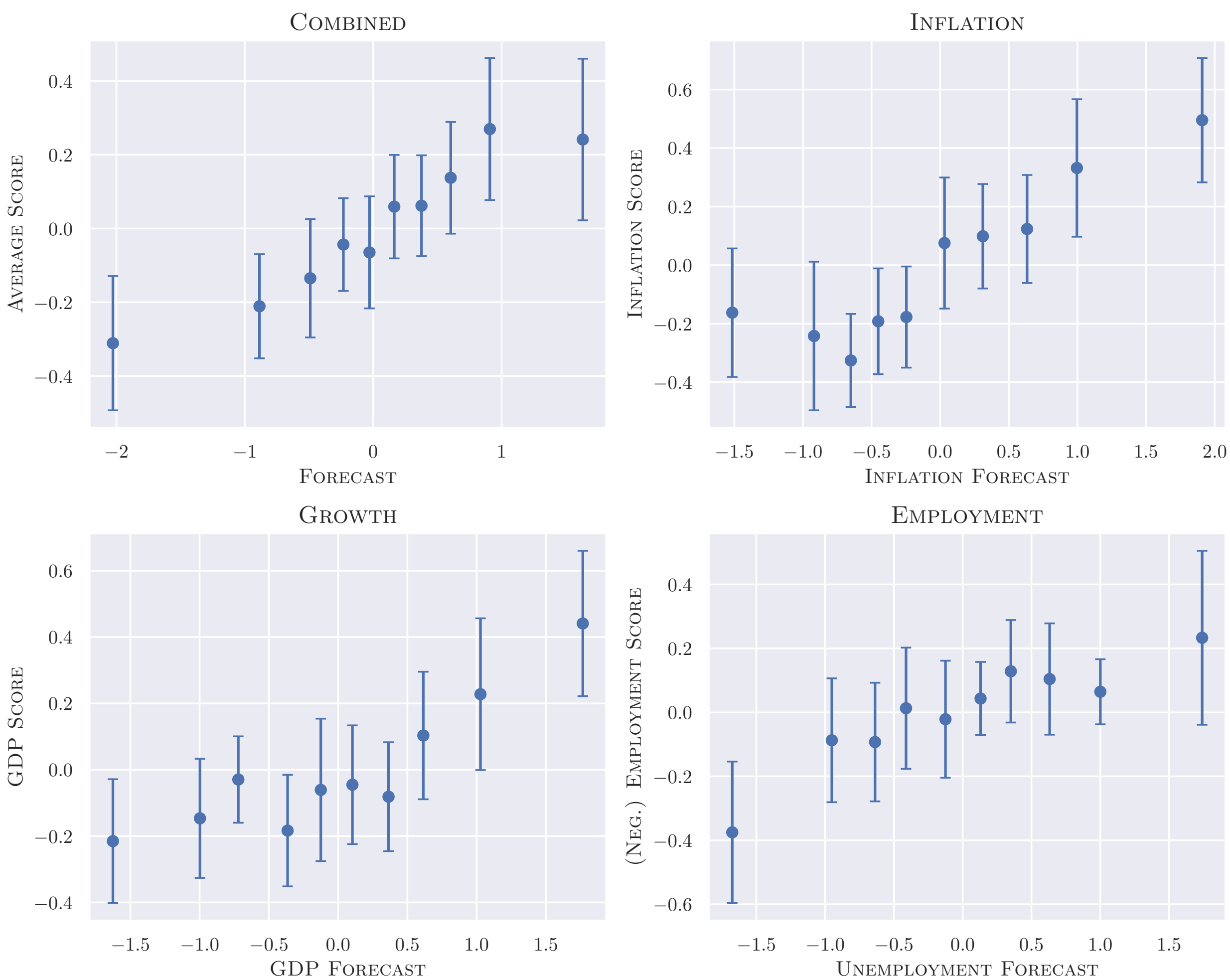


Figure IA.3: Committee Members' Forecasts and Argument Scores

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 outlook scores for the corresponding 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. Standard errors are computed by clustering by member and meeting within each decile.

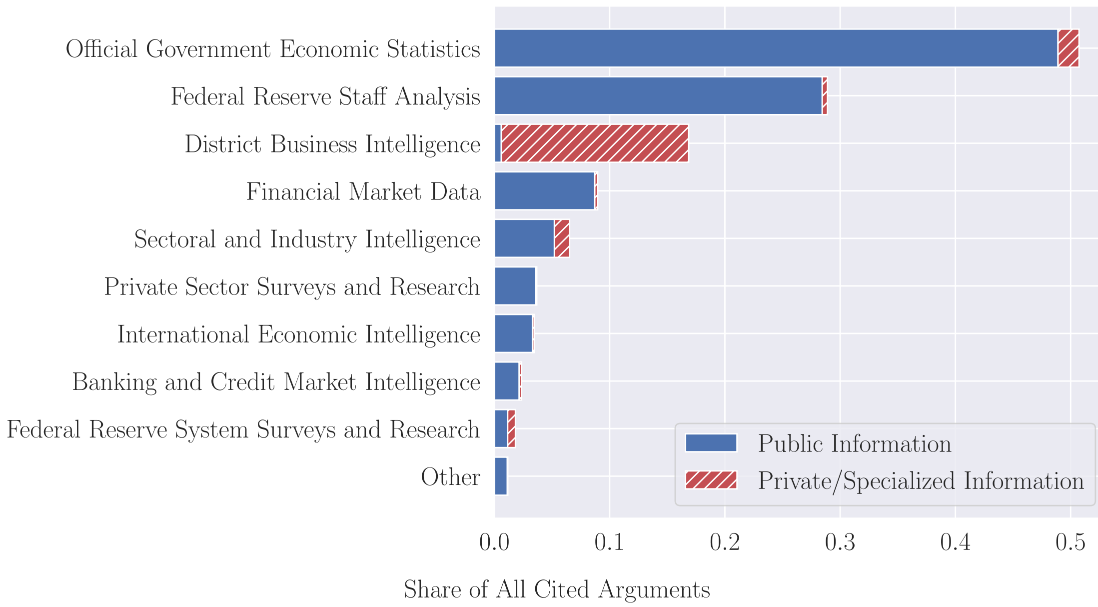


FIGURE IA.4: PUBLIC AND PRIVATE INFORMATION BY DATA SOURCE

Note: This figure shows the share of arguments by data source category, distinguishing public from private information. Detailed category descriptions are in Internet Appendix Table IA.6.

IA.A.2.2 Selection Versus Interpretation Effects

We classify each argument into topic categories to distinguish between topic selection and interpretation effects. The categories group together different economic mechanisms while combining opposing interpretations of the same mechanism. Figure IA.5 displays selected categories and their average policy stance scores. The first panel shows the most frequently discussed categories. The second panel shows categories with the most negative average scores, indicating support for accommodative policy. The third panel shows categories with the most positive average scores, indicating support for restrictive policy.

To quantify the relative importance of topic selection versus interpretation, we decompose each member's score into three components. For member i in meeting t, we write:

$$\gamma_{i,t,k} = \underbrace{\overline{\gamma}_{-i,t,k}}_{\text{Selection}} + \underbrace{(\widehat{\gamma}_{i,t,k} - \overline{\gamma}_{-i,t,k})}_{\text{Interpretation}} + \underbrace{(\gamma_{i,t,k} - \widehat{\gamma}_{i,t,k})}_{\text{Residual}}, \tag{IA.A.1}$$

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.[28] The selection effect captures how much of i's score is driven by their choice of topics. The interpretation effect captures the extent to which i's scores are driven by 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, we examine Charles Plosser's and Daniel Tarullo's arguments from the August 2011 meeting. With meeting fixed effects, Plosser's cross-sectionally demeaned score of $+0.33$ decomposes into a selection effect of -0.17 and an interpretation effect of $+0.49$. Tarullo's cross-sectionally demeaned score of -0.53 decomposes into a selection effect of -0.20 and an interpretation effect of -0.32. The similar selection effects reflect that both spent much of their discussion on the poor growth outlook. The interpretation gap is best illustrated by examining *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 0.0—a three-point interpretation gap on the same economic mechanism.

Table IA.5 presents the variance decomposition without meeting fixed effects. Selection effects account for 40.6% of total variance, compared to 19.4% in Panel B of Table 6. The difference reflects that topic selection is partly driven by current economic conditions common to all members within a meeting. Meeting fixed effects absorb this shared variation, isolating within-meeting differences in interpretation.

	Selection (%)	Interpretation (%)	Residual (%)
All	40.6	57.2	2.2
Inflation	37.9	60.8	1.3
Growth	38.2	60.7	1.1
Employment	39.4	58.6	2.0

TABLE IA.5: VARIANCE DECOMPOSITION WITHOUT MEETING FIXED EFFECTS

Note: This table repeats the variance decomposition in Panel B of Table 6 without meeting fixed effects.

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

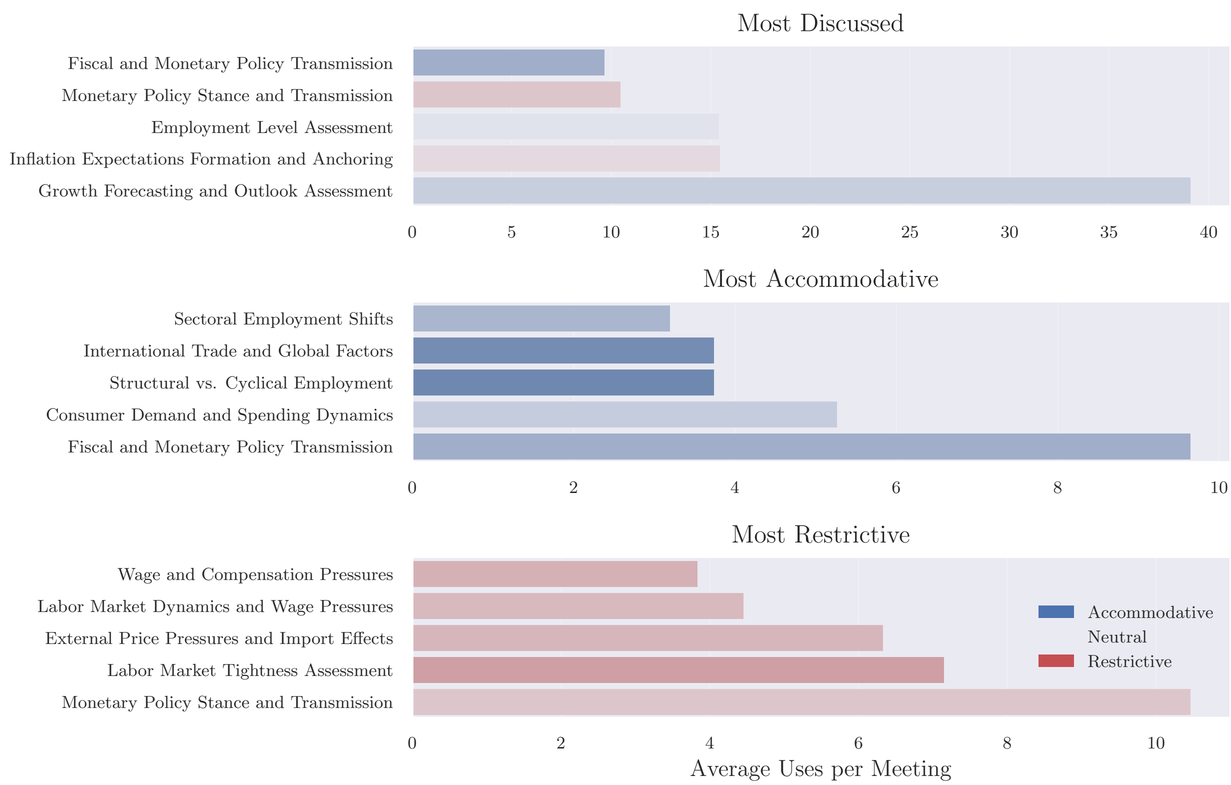


FIGURE IA.5: ARGUMENT TOPIC CATEGORIES

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 < -0.25), and most restrictive (average score > 0.25).

IA.A.3 Categories of Arguments and Data Sources

Table IA.6 presents categories and associated descriptions used to categorize data sources cited by committee members in their arguments. Tables IA.7, IA.8, and IA.9 present the categories and associated descriptions used to categorizes arguments made by committee members with respect to growth, inflation, and employment.

Name	Description
Official Government Economic Statistics	Formally published statistical data from government agencies, including employment figures (unemployment rates, payroll employment, labor force participation), national accounts (GDP, personal consumption, business investment), price indices (CPI, PCE, core inflation measures), industrial production, capacity utilization, retail sales, and other macroeconomic indicators. These sources are public data available to all committee members and the general public through agencies like the Bureau of Labor Statistics, Bureau of Economic Analysis, and Census Bureau.
Federal Reserve Staff Analysis	Internal Federal Reserve staff economic forecasts, projections, and analytical products including the Greenbook, Tealbook, Summary of Economic Projections (SEP), staff economic models, nowcasting models, and various analytical memos. These sources are specialized internal Fed analysis available only to FOMC participants and Fed staff, representing the institution's internal analytical capabilities.

District Business Intelligence	Information gathered through direct contact with businesses, executives, and economic actors within specific Federal Reserve districts, including anecdotal reports from business contacts, regional economic assessments, local labor market insights, hiring plans, pricing pressures, and on-the-ground intelligence about district economic conditions. These sources are specialized information unique to each Reserve Bank president's district network and business relationships.
Federal Reserve System Surveys and Research	Formal surveys, indices, and research conducted by Federal Reserve Banks and Board staff, including the Senior Loan Officer Opinion Survey, regional Fed business outlook surveys, Beige Book intelligence, reports from Federal Reserve Bank directors, coincident economic indices, diffusion indices, and flow-of-funds reports. These sources are specialized Fed institutional data, though some results may be publicly released.
Financial Market Data	Market-determined prices, rates, and indicators including stock indices, Treasury yields, corporate bond spreads, credit spreads, foreign exchange rates, swap rates, market volatility measures (like VIX), commodity prices (especially oil), energy market data, and other traded financial instruments. These sources are typically public data available in real-time through financial markets and data providers.

Banking and Credit Market Intelligence	Information about banking sector conditions, credit availability, lending standards, loan demand and growth, bank balance sheets, money supply growth, credit intermediation, and broader financial institution health. This includes both public regulatory data and specialized information gathered through the Fed's supervisory relationships with financial institutions. These sources are mixed between public regulatory data and specialized Federal Reserve supervisory intelligence.
Private Sector Surveys and Research	Economic data, forecasts, and sentiment measures from private organizations, academic institutions, and research firms, including ISM manufacturing and services indices, Conference Board consumer confidence, University of Michigan consumer sentiment, Blue Chip forecasts, National Federation of Independent Business (NFIB) surveys, ADP employment reports, and business outlook surveys from trade associations. These sources are typically public data originating from non-government entities.
Sectoral and Industry Intelligence	Industry-specific and sector-focused economic information covering housing markets (home sales, construction, prices, building permits), manufacturing activity, energy sector conditions, agricultural data, technology sector developments, retail conditions, corporate earnings reports, and intelligence from specific companies and trade associations. These sources are mixed between public industry data and specialized sectoral intelligence gathered through business contacts and industry relationships.

International Economic Intelligence	Information about foreign economic conditions, global growth trends, international trade data, overseas market developments, and economic conditions in other countries that affect the U.S. economy. This includes both public international statistics and specialized intelligence gathered through Federal Reserve international contacts, foreign central bank relationships, and international economic monitoring. These sources are mixed between public international data and specialized intelligence from Fed international relationships.

TABLE IA.6: 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	Description
Growth Forecasting and Outlook Assessment	Projections and assessments of future economic growth trajectories, encompassing disagreements over the likelihood, timing, and magnitude of different growth scenarios and the factors that will drive economic performance.
Consumer Demand and Spending Dynamics	Factors influencing household consumption decisions and their impact on economic growth, including disagreements over consumer confidence, income effects, wealth effects, and spending sustainability.
Business Investment and Capital Formation	Corporate investment decisions and their effects on productive capacity and economic growth, with disagreements over investment timing, magnitude, and the factors driving business capital allocation.
Labor Market Conditions and Productivity	Employment dynamics, wage growth, and productivity changes as drivers of economic growth, encompassing disagreements over labor market tightness, productivity trends, and their growth implications.
Financial Market Conditions and Credit	Financial sector conditions, credit availability, and market functioning as enablers or constraints on economic growth, with disagreements over the transmission mechanisms and magnitude of financial effects.
Fiscal and Monetary Policy Transmission	Government spending, taxation, and monetary policy effects on economic growth, including disagreements over policy effectiveness, timing of impacts, and appropriate policy responses to growth conditions.
Housing and Real Estate Market Effects	Residential investment, construction activity, and housing wealth effects on economic growth, with disagreements over the magnitude of housing market impacts and their persistence.
International Trade and Global Factors	External demand, trade flows, and global economic conditions as influences on domestic growth, encompassing disagreements over the degree of international economic integration and spillover effects.
Inventory Dynamics and Production Cycles	Inventory accumulation and depletion cycles as drivers of short-term growth fluctuations, with disagreements over the timing, magnitude, and sustainability of inventory-driven growth contributions.
Sectoral Performance and Composition Effects	Industry-specific performance variations and their aggregate impact on overall economic growth, including disagreements over sectoral strength, weakness, and the composition of economic activity.

Temporary Shocks and Cyclical Factors	Weather events, strikes, and other transitory influences on economic growth measurement and trends, with disagreements over the persistence and significance of temporary factors versus underlying growth fundamentals.

TABLE IA.7: CATEGORIES OF GROWTH ARGUMENTS

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

Category	Description
Labor Market Dynamics and Wage Pressures	The relationship between unemployment rates, labor market tightness, wage growth, and unit labor costs in driving inflationary pressures, with disagreements over the strength and timing of Phillips curve effects.
External Price Pressures and Import Effects	How commodity prices, energy costs, exchange rate movements, and import price changes transmit to domestic inflation, with varying views on the magnitude and persistence of these cost-push effects.
Inflation Expectations Formation and Anchoring	The role of inflation expectations in determining actual inflation outcomes, encompassing debates over how expectations are formed, measured, and whether they remain anchored to policy targets.
Economic Capacity and Resource Utilization	How output gaps, resource utilization rates, and economic slack influence inflationary pressures through demand-pull mechanisms, with disagreements over capacity estimates and their inflationary implications.
Monetary Policy Stance and Transmission	The effectiveness of monetary policy tools in controlling inflation, including debates over policy timing, magnitude of effects, and the transmission mechanisms through which policy affects price levels.
Productivity Growth and Unit Cost Dynamics	How productivity improvements and unit cost changes affect inflationary pressures, with varying assessments of productivity trends and their ability to offset wage and other cost pressures.
Market Competition and Pricing Power	The influence of competitive conditions and firms' pricing power on inflation, encompassing disagreements over whether businesses can successfully pass through cost increases to consumers.
Structural Economic and Demographic Changes	Long-term structural factors affecting inflation dynamics, including demographic shifts, globalization effects, and technological disruptions, with debates over their persistent impact on price-setting behavior.

TABLE IA.8: CATEGORIES OF INFLATION ARGUMENTS

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

Category	Description
Employment Level Assessment	Evaluations of current employment levels, job growth rates, and unemployment rates, encompassing disagreements about whether observed employment outcomes indicate economic strength, weakness, or appropriate progress toward full employment.
Labor Market Tightness Assessment	Assessments of labor market slack versus tightness conditions, encompassing disagreements about the degree of remaining capacity in labor markets and whether current conditions indicate full utilization of labor resources.
Employment-Inflation Relationship	Analysis of the causal relationship between employment levels and inflationary pressures, encompassing disagreements about whether tight labor markets will generate wage-price spirals or whether employment can expand without triggering inflation.
Wage and Compensation Pressures	Observations of wage growth and labor cost trends as outcomes of employment conditions, encompassing disagreements about whether current employment levels are generating meaningful wage pressures or remaining contained.
Structural vs. Cyclical Employment	Interpretations of whether employment changes reflect permanent structural shifts in the economy or temporary cyclical fluctuations, encompassing disagreements about the persistence and policy responsiveness of employment patterns.
Labor Force Participation Trends	Analysis of labor force participation rate changes and their causes, encompassing disagreements about whether participation declines are demographically driven, cyclically responsive, or reflect structural labor market changes.
Employment Policy Transmission	Assessments of how monetary policy actions affect employment outcomes, encompassing disagreements about the effectiveness, timing, and magnitude of policy impacts on job creation and labor market conditions.
Sectoral Employment Shifts	Analysis of employment changes across different industries and sectors, encompassing disagreements about whether sectoral shifts represent healthy economic reallocation or concerning structural disruptions to employment.

Regional Employment Variations	Observations of geographic differences in employment performance across Federal Reserve districts, encompassing disagreements about whether regional variations reflect local factors or broader national employment trends.

TABLE IA.9: CATEGORIES OF EMPLOYMENT ARGUMENTS

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

IA.A.4 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 three variables: Inflation, Growth, and Employment. For each meeting t and category c within variable v, we calculate:

$$p_{c,t}^{v} = \frac{n_{c,t}^{v}}{\sum_{c' \in C_v} n_{c',t}^{v}} \tag{IA.A.2}$$

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

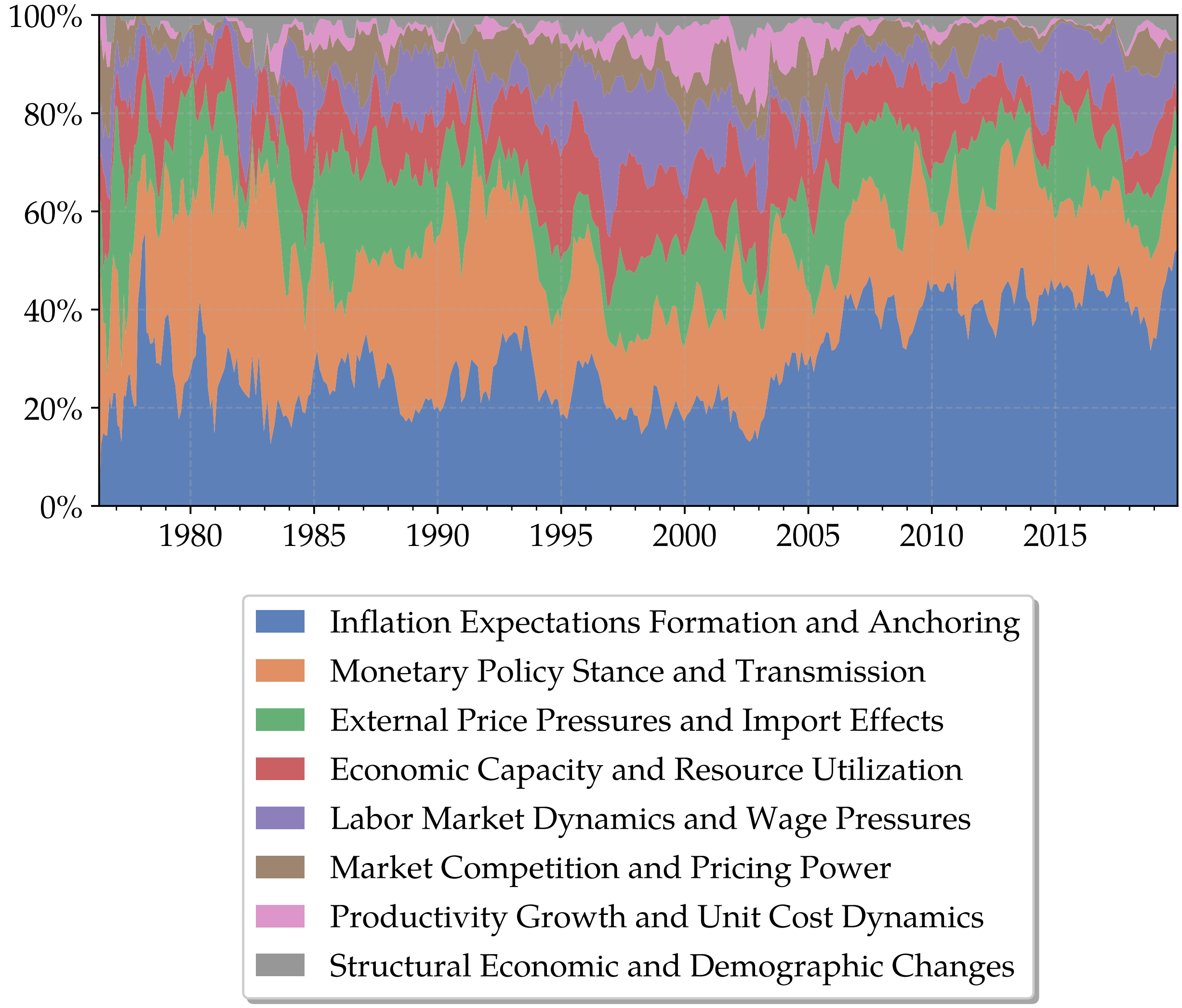


FIGURE IA.6: INFLATION ARGUMENTS BY CATEGORY

Note: The figure plots the 4-meeting trailing average of the proportions of arguments made regarding inflation in each of 8 categories during FOMC meetings.

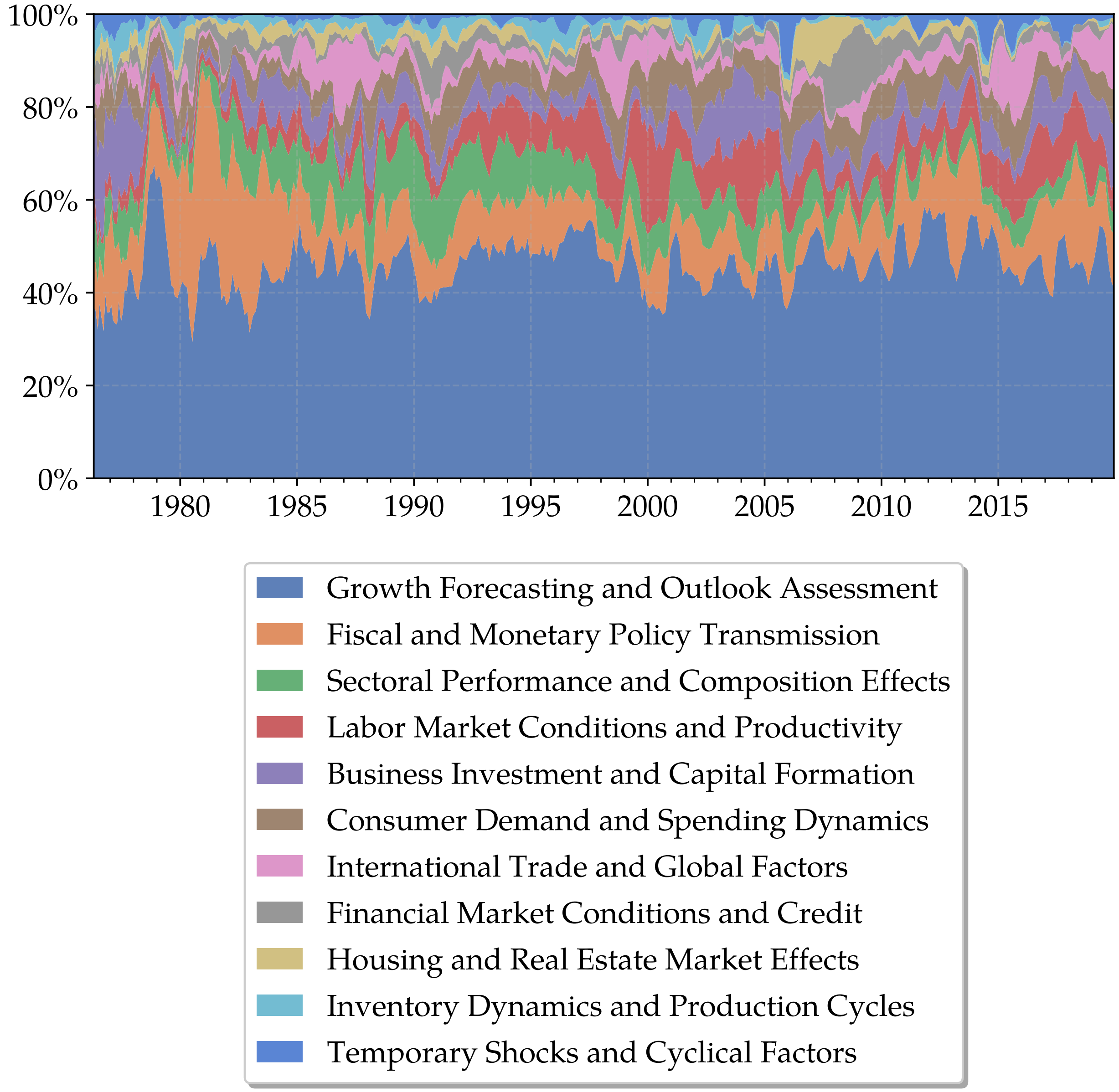


FIGURE IA.7: GROWTH ARGUMENTS BY CATEGORY

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

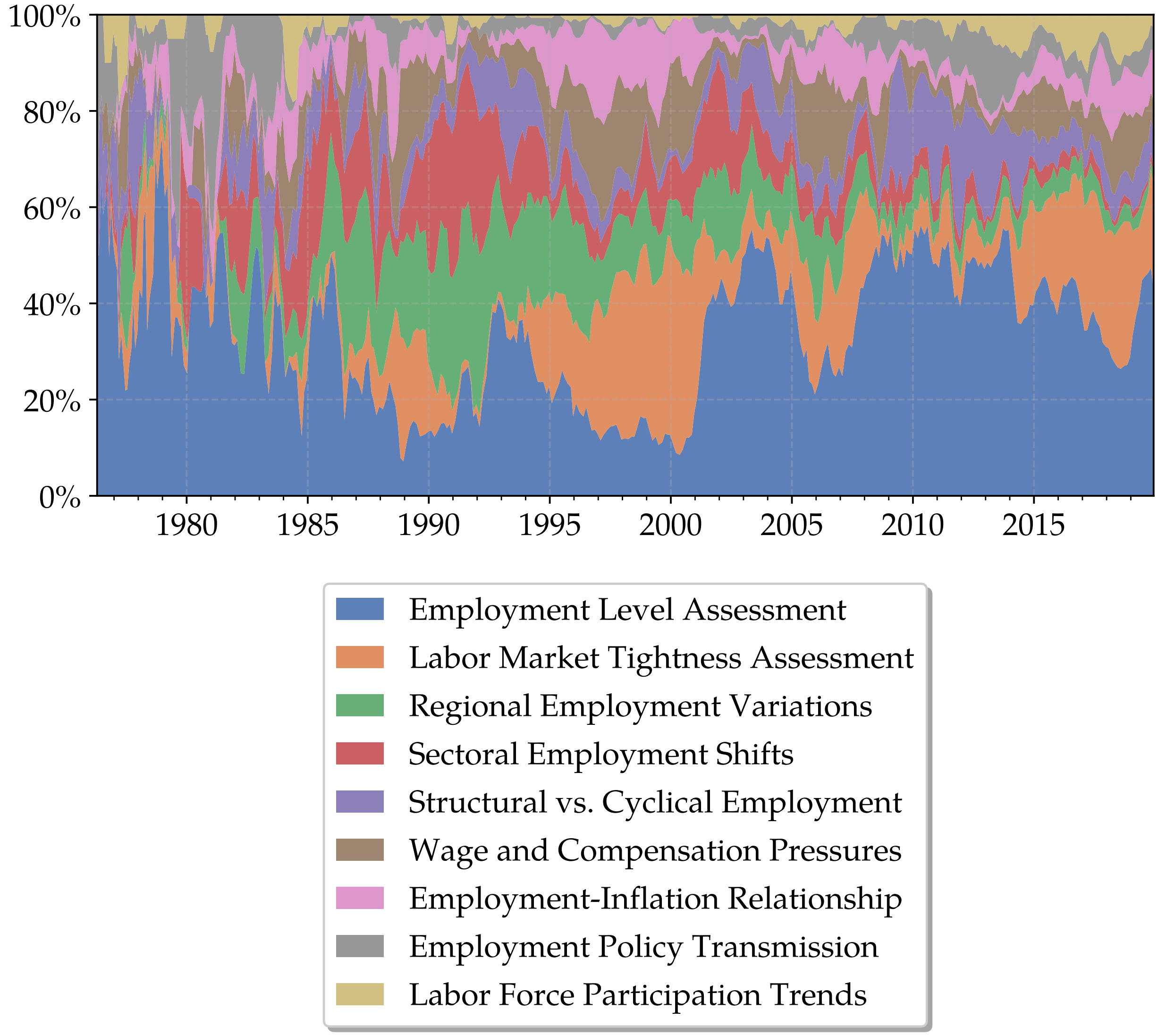


FIGURE IA.8: EMPLOYMENT ARGUMENTS BY CATEGORY

Note: The figure plots the 4-meeting trailing average of the proportions of arguments made regarding employment in each of 9 categories during FOMC meetings.

IA.A.5 Forecasts, Model Fit, and Policy Decision Alignment

	Role				Experience			Education	
	Chair	Governors	Voting Pres.	Non-voting Pres.	Low	Medium	High	Other	PhD Econ
FE-based Fit ($\text{Fit}_{i,t}^{\text{FE}}$)									
Mean rank	12.39	13.26	11.15	11.98	12.66	11.88	11.77	12.41	11.87
Std. dev.	3.70	3.80	4.01	4.16	3.64	4.18	4.25	4.08	4.06
FE-Imputed Fit ($\text{Fit}_{i,t}^{\text{FE,impute}}$)									
Mean rank	12.01	11.04	10.05	10.48	11.19	10.13	10.47	10.74	10.46
Std. dev.	3.07	3.39	3.19	3.24	3.19	3.33	3.25	3.42	3.22
Cross-sectional Heterogeneity									
Within-meeting std. (Fit^{FE})	–	2.02	3.00	3.38	2.68	3.23	3.41	2.93	3.35
Within-meeting std. ($\text{Fit}^{\text{FE,impute}}$)	–	2.41	2.74	2.78	2.49	2.77	2.76	2.55	2.81
Within-person std. (Fit^{FE})	3.45	3.10	3.06	3.48	3.22	3.43	3.45	3.18	3.36
Within-person std. ($\text{Fit}^{\text{FE,impute}}$)	2.63	2.45	2.42	2.64	2.31	2.57	2.40	2.67	2.48
Alignment and Influence									
Mean $Alignment_{i,t}$	2.52	2.03	1.87	1.72	1.93	1.93	1.84	1.92	1.89
Mean $Influence_{i,t}$	1.66	0.90	0.85	0.73	0.86	0.86	0.88	0.84	0.88
Dissent rate	0.0%	5.5%	9.8%	–	5.1%	8.1%	7.9%	5.7%	7.7%

TABLE IA.10: MODEL FIT AND ALIGNMENT SUMMARY STATISTICS SPLIT BY DEMOGRAPHICS

Note: The table reports summary statistics for model fit measures. $\text{Fit}_{i,t}^{\text{FE}}$ is based on half-year SEP forecast errors (2008–present). $\text{Fit}_{i,t}^{\text{FE,impute}}$ extends coverage by imputing pre-SEP errors using MPR and Greenbook forecast errors. Ranking ties are assigned the highest common score in Fit variable construction. Cross-sectional heterogeneity reports average within-meeting and within-person standard deviations. $Alignment_{i,t}$ and $Influence_{i,t}$ capture member i's alignment with and influence on policy decisions in meeting t; the dissent rate is the share of meetings in which voting members dissented. Categories split members by role (Chair, Governors, voting/non-voting Presidents), experience terciles (years in current position, within meeting), and education (PhD in Economics vs. other).

	Non-rate decisions				Rate decisions			
	Voters		All Members		Voters		All Members	
	$Alignment^{\text{ex-rate}}_{i,t}$	$Influence^{\text{ex-rate}}_{i,t}$	$Alignment^{\text{ex-rate}}_{i,t}$	$Influence^{\text{ex-rate}}_{i,t}$	$Alignment^{\text{rate}}_{i,t}$	$Influence^{\text{rate}}_{i,t}$	$Alignment^{\text{rate}}_{i,t}$	$Influence^{\text{rate}}_{i,t}$
Panel A: Half-year Forecast Error-based Fit								
$\text{Fit}^{\text{FE}}_{i,t}$	0.165	0.124	0.150	0.098	0.150	0.102	0.117	0.078
	(3.04)	(2.86)	(2.63)	(3.11)	(2.42)	(2.18)	(2.30)	(2.38)
N	893	893	1498	1498	853	853	1433	1433
Meeting FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Half-year Imputed Forecast Error-based Fit								
	Non-rate decisions				Rate decisions			
	Voters		All Members		Voters		All Members	
	$Alignment^{\text{ex-rate}}_{i,t}$	$Influence^{\text{ex-rate}}_{i,t}$	$Alignment^{\text{ex-rate}}_{i,t}$	$Influence^{\text{ex-rate}}_{i,t}$	$Alignment^{\text{rate}}_{i,t}$	$Influence^{\text{rate}}_{i,t}$	$Alignment^{\text{rate}}_{i,t}$	$Influence^{\text{rate}}_{i,t}$
$\text{Fit}^{\text{FE,impute}}_{i,t}$	0.114	0.055	0.105	0.048	0.084	0.051	0.071	0.044
	(3.41)	(2.01)	(2.62)	(1.86)	(2.46)	(2.22)	(2.23)	(2.25)
N	2063	2063	3484	3484	2003	2003	3386	3386
Meeting FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table IA.11: Model Fit and Alignment with Non-Rate Decisions

Note: The table repeats the analysis presented in Table 9, separating members' $Alignment_{i,t}$ and $Influence_{i,t}$ decisions with non-rate decisions and rate decisions.

Panel A: Half-year Forecast Error-based Fit					
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE}}_{i,t,\text{inflation}}$	-3.58	0.129	0.089	0.118	0.062
	(1.70)	(2.11)	(2.06)	(1.99)	(1.67)
$\text{Fit}^{\text{FE}}_{i,t,\text{growth}}$	-4.67	0.112	0.104	0.123	0.103
	(3.39)	(3.08)	(3.39)	(2.90)	(3.50)
N	893	893	893	1498	1498
Meeting FE	Yes	Yes	Yes	Yes	Yes

Panel B: Half-year Imputed Forecast Error-based Fit					
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE,impute}}_{i,t,\text{inflation}}$	-2.32	0.076	0.036	0.047	0.018
	(2.06)	(2.07)	(1.33)	(1.29)	(0.78)
$\text{Fit}^{\text{FE,impute}}_{i,t,\text{growth}}$	-2.73	0.081	0.048	0.087	0.058
	(3.41)	(3.23)	(2.47)	(3.09)	(3.00)
$\text{Fit}^{\text{FE,impute}}_{i,t,\text{unemployment}}$	-0.85	0.035	0.021	0.043	0.022
	(1.02)	(1.20)	(0.94)	(1.29)	(1.03)
N	2063	2063	2063	3484	3484
Meeting FE	Yes	Yes	Yes	Yes	Yes

Table IA.12: Model Fit and Policy Decision Alignment, by variable

Note: The table repeats the analysis presented in Table 9, running separate univariate regressions for the individual fit measures constructed for each of inflation, employment, and growth rather than the composite measure.

Panel A: Half-year Forecast Error-based Fit					
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE}}_{i,t,\text{inflation}}$	-7.65	0.282	0.178	0.266	0.146
	(2.16)	(3.27)	(2.91)	(2.45)	(2.12)
$\text{Fit}^{\text{FE}}_{i,t,\text{growth}}$	-11.12	0.262	0.239	0.240	0.171
	(2.72)	(2.35)	(2.72)	(2.32)	(2.25)
N	893	893	893	1498	1498
Meeting FE	Yes	Yes	Yes	Yes	Yes

Panel B: Half-year Imputed Forecast Error-based Fit					
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE,impute}}_{i,t,\text{inflation}}$	-6.90	0.204	0.099	0.099	0.018
	(2.67)	(2.58)	(1.60)	(1.47)	(0.37)
$\text{Fit}^{\text{FE,impute}}_{i,t,\text{growth}}$	-8.94	0.308	0.106	0.296	0.131
	(2.39)	(2.62)	(1.17)	(3.36)	(1.89)
$\text{Fit}^{\text{FE,impute}}_{i,t,\text{unemployment}}$	-1.77	0.109	0.057	0.102	0.061
	(1.16)	(1.95)	(1.11)	(1.35)	(1.16)
N	2063	2063	2063	3484	3484
Meeting FE	Yes	Yes	Yes	Yes	Yes

Table IA.13: Model Fit and Alignment, by variable (raw)

Note: The table repeats the analysis presented in Table IA.12, running separate univariate regressions for the raw individual fit measures constructed for each of inflation, employment, and growth, rather than a measure that cross-sectionally ranks at each point in time.

	Voters			All Members	
Panel A: Half-year Forecast Error-based Fit					
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}_{i,t}^{\text{FE}}$	-3.95	0.114	0.103	0.130	0.089
	(2.60)	(2.46)	(2.77)	(2.45)	(2.99)
N	893	893	893	1498	1498
Meeting FE	Yes	Yes	Yes	Yes	Yes
Panel B: Half-year Imputed Forecast Error-based Fit					
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}_{i,t}^{\text{FE,impute}}$	-2.67	0.082	0.036	0.084	0.041
	(2.73)	(2.52)	(1.56)	(2.16)	(1.76)
N	2063	2063	2063	3484	3484
Meeting FE	Yes	Yes	Yes	Yes	Yes

Table IA.14: Model Fit and Alignment, Including Title Fixed Effects

Note: The table repeats the analysis presented in Table 9, including title fixed effects in the regression, indicating whether a member is the chair, a governor, or a voting or non-voting regional Fed president.

	Panel A: Half-year Forecast Error-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}_{i,t}^{\text{FE}}$	-6.49	0.189	0.159	0.169	0.120
	(2.89)	(3.07)	(2.92)	(2.85)	(3.00)
N	801	801	801	1406	1406
Meeting FE	Yes	Yes	Yes	Yes	Yes
	Panel B: Half-year Imputed Forecast Error-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}_{i,t}^{\text{FE,impute}}$	-3.44	0.112	0.057	0.102	0.055
	(2.96)	(3.11)	(1.92)	(2.47)	(1.96)
N	1952	1952	1952	3373	3373
Meeting FE	Yes	Yes	Yes	Yes	Yes

Table IA.15: Model Fit and Alignment, Excluding Chair

Note: The table repeats the analysis presented in Table 9, excluding the chair from the analysis.

	Panel A: Transcript-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}_{i,t}^{\text{transcript}}$	-1.74	0.046	0.049	0.037	0.035
	(3.55)	(3.30)	(3.05)	(2.90)	(2.69)
N	3617	3617	3617	5701	5701
Meeting FE	Yes	Yes	Yes	Yes	Yes

	Panel B: Transcript-based Fit by Variable				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}_{i,t,\text{growth}}^{\text{transcript}}$	-1.04	0.046	0.049	0.037	0.037
	(2.68)	(3.61)	(3.01)	(3.03)	(2.99)
N	3617	3617	3617	5701	5701
$\text{Fit}_{i,t,\text{inflation}}^{\text{transcript}}$	-1.98	0.034	0.024	0.041	0.034
	(3.14)	(1.69)	(0.89)	(2.49)	(1.72)
N	2785	2785	2785	4400	4400
$\text{Fit}_{i,t,\text{unemployment}}^{\text{transcript}}$	-1.00	0.011	0.021	-0.005	0.003
	(2.00)	(0.66)	(1.05)	(0.27)	(0.16)
N	2481	2481	2481	4072	4072
Meeting FE	Yes	Yes	Yes	Yes	Yes

Table IA.16: Model Fit and Alignment, FOMC Transcript-based Measures

Note: The table reports estimates from regressions relating a member's transcript-based model fit to their role in FOMC decisions. Dependent variables are $Dissent_{i,t}$ (an indicator for dissenting votes), $Alignment_{i,t}$ (agreement with the adopted decision), and $Influence_{i,t}$ (role in shaping the decision). $Alignment_{i,t}$ and $Influence_{i,t}$ are standardized, so coefficients represent standard-deviation changes for a one standard-deviation increase in model fit; $Dissent_{i,t}$ coefficients represent percentage-point changes in dissent probability. Transcript-based fit compares a member's deviation from the cross-sectional average argument score at meeting $t-1$ to the change in the cross-sectional average from $t-1$ to t. Specifically, fit equals the absolute difference between the member's demeaned score at $t-1$ and the subsequent shift in the committee-wide average; members whose accommodative or restrictive lean at $t-1$ anticipated the direction of aggregate belief revisions have lower (better) fit values. Fit values are then ranked within each meeting, so higher rank indicates better fit. Panel A averages ranks across growth, inflation, and unemployment; Panel B reports each variable separately. "Voters" columns restrict to voting members; "All Members" includes all participants. t-statistics in parentheses are based on standard errors clustered by member and meeting.

PANEL A: DECISION-LEVEL

	Voters		All Members	
	$\text{Alignment}_{i,d,t}$	$\text{Influence}_{i,d,t}$	$\text{Alignment}_{i,d,t}$	$\text{Influence}_{i,d,t}$
$\text{Hawkish Decision}_{d,t}$ $\times$ $\text{Hawkish Forecast}_{i,t}$	0.069	0.023	0.071	0.020
	(3.10)	(2.34)	(2.78)	(2.54)
N	11,262	11,262	18,906	18,906
Member x Meeting FE	Yes	Yes	Yes	Yes

PANEL B: MEETING-LEVEL

	Voters			All Members	
	Dissent $\text{Direction}_{i,t}$	$\text{Alignment}_{i,d,t} \times$ $\text{Hawkish Decision}_{d,t}$	$\text{Influence}_{i,d,t} \times$ $\text{Hawkish Decision}_{d,t}$	$\text{Alignment}_{i,d,t} \times$ $\text{Hawkish Decision}_{d,t}$	$\text{Influence}_{i,d,t} \times$ $\text{Hawkish Decision}_{d,t}$
$\text{Hawkish Forecast}_{i,t}$	6.491	0.050	0.015	0.078	0.028
	(2.26)	(2.26)	(1.14)	(2.69)	(2.40)
N	2,201	2,201	2,201	3,693	3,693
Meeting FE	Yes	Yes	Yes	Yes	Yes

TABLE IA.17: FORECASTS AND DIRECTIONAL POLICY ALIGNMENT

Note: The table relates members' forecasts to the direction of decisions that they support. For each decision d in meeting t, $\text{Hawkish Decision}_{d,t}$ is a score from -3 to 3, where a score of -3 indicates the decision is strongly accommodative and $+3$ indicates the decision is strongly restrictive. The variable $\text{Hawkish Forecast}_{i,t}$ is constructed by averaging each member i's within-meeting t ranks based on their current year forecasts of unemployment, inflation, and growth, where higher ranks correspond to lower unemployment forecasts and higher growth and inflation forecasts. The variables $\text{Alignment}_{i,d,t}$ and $\text{Influence}_{i,d,t}$ capture member i's alignment and influence with decision d in meeting t. Dissent $\text{Direction}_{i,t}$ is constructed as $100\times$ the difference between a 0/1 indicator for whether voting member i dissented hawkishly and a 0/1 indicator for whether voting member i dissented dovishly in meeting t, where hawkish and dovish dissents are from Thornton, Wheelock et al. (2014). Panel A reports results from regressions of $\text{Alignment}_{i,d,t}$ and $\text{Influence}_{i,d,t}$ on $\text{Hawkish Forecast}_{i,t}$ interacted with $\text{Hawkish Decision}_{d,t}$ and includes member-by-meeting fixed effects, where observations are at the member-by-decision level. Panel B reports results from regressions where the dependent variables are dissent $\text{Direction}_{i,t}$ and meeting-by-member averages (over decisions) of the interactions of $\text{Alignment}_{i,d,t}$ (and $\text{Influence}_{i,d,t}$) with $\text{Hawkish Decision}_{d,t}$, and the independent variable is $\text{Hawkish Forecast}_{i,t}$. Standard errors are clustered by member and meeting, and t-statistics are reported in parentheses.

IA.A.6 Policy Decision Scores and Monetary Policy Surprises

In this section, we link the policy decisions we identify from the transcripts to high-frequency monetary policy surprises measured from asset price responses to FOMC announcements. This provides additional validation of our decisions dataset.[29]

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 to 2019. We also analyze the changes in 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 $FF3m$).[30] The sample of Fed Funds futures runs from 1995 to 2019.

For each meeting, we assign a score from -3 (highly accommodative) to +3 (highly restrictive) 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.[31] We estimate regressions of monetary policy surprises on the $Avg(\{PolicyScores_t\})$ variables. The $Avg(\{PolicyScores_t\})$ 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.18 report results from univariate regressions of mps_t on $Avg(\{PolicyScores_t\})$ and $Avg(\{ExRatePolicyScores_t\})$. The coefficients from these regressions are 1.11 and 0.81, respectively, indicating that a one standard deviation change in the $Avg(\{PolicyScores\})$ measures captures a 0.8 to 1.1 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.

[29]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.

[30]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.

[31]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.

The mps_t variable captures the component of monetary policy 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.18, 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 0.68 and 0.38 in these regressions, compared to the 5.7 bps standard deviation of $ED4_t$. 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 0.57 and 0.56 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 is not 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 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.18 report results from regressions of mps_t^{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 relation-

	$y = mps_t$		$y = ED4_t$		$y = FF3m_t$		$y = mps_t^{orth}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Avg(\{PolicyScores_t\})$	1.11		0.68		0.57		0.47	
	(3.61)		(2.68)		(3.51)		(1.76)	
$Avg(\{ExRatePolicyScores_t\})$		0.81		0.38		0.56		0.25
		(2.73)		(2.94)		(3.66)		(1.00)
$ED1_t$			0.92	1.02				
			(8.18)	(16.89)				
FF_t					0.70	0.71		
					(7.37)	(7.62)		
σ_y	4.95	4.95	5.70	5.70	3.74	3.74	4.63	4.63
N	262	262	262	262	190	190	262	262
R^2	0.05	0.03	0.51	0.79	0.53	0.52	0.00	0.00

TABLE IA.18: 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 restrictive or accommodative tone of FOMC decisions. The dependent variables are: mps_t, the first principal component of changes in one- to four-quarter ahead Eurodollar contracts within a 30-minute post-announcement window; $ED4_t$, the change in the four-quarter ahead Eurodollar contract; $FF3m_t$, the change in 3-month ahead Fed Funds futures yields; and mps_t^{orth}, computed as residuals from regressing the mps_t 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 accommodative to +3 for strongly restrictive. The regressions include controls for the high-frequency change in one-quarter ahead Eurodollar futures ($ED1_t$) and near-maturity Fed Funds futures (FF_t). The mps_t and ED_t 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.

ship, suggesting that the component of the $Avg(\{PolicyScores\})$ variables that is related to monetary policy surprises stems from investor surprise at the committee's response to macroeconomic and financial news.

IA.B Data Appendix

IA.C Validation of LLM-Extracted Measures

We validate our automated extraction of measures from FOMC transcripts by comparing against trained human coders and against two alternative large language models. This analysis helps to characterize the potential impact of measurement error arising from using LLMs to label our data. See, for example, Ludwig, Mullainathan and Rambachan (2025) and Carlson and Dell (2025), who discuss these problems and propose frameworks to address them.

As described in Section 3.2, we use Anthropic's Claude Sonnet 4 (version 20250514) to extract arguments and classify parameter beliefs from the transcripts. Here we compare Claude's classifications against those of trained human coders and against two open-source alternatives: Meta's Llama 70B and Llama 8B. We validate three types of measures that we constructed using Claude:

(i) *Parameter Beliefs:* beliefs on Phillips curve slope, IS curve sensitivity, Financial–Real Linkages strength.

(ii) *Arguments:* argument categories, outlook scores, and types of information cited.

(iii) *Decisions:* identifying decisions made at the meeting, and measuring alignment and influence of committee members regarding each decision.

In the validation exercise, we assess both detection, which measures whether the LLM correctly identifies when a belief, argument, or decision is present, and classification, which measures whether the LLM correctly categorizes the content conditional on detection.

We report Pearson correlations for ordinal and continuous variables (parameter beliefs, argument scores, influence, alignment), and Cohen's kappa for nominal categories (argument topics) and classifications without meaningful ordinal structure (information type, data source indicators).[32] We find agreement between LLM and human coding that is on the order of agreement between two trained human coders. Lastly, we use our human-coded sample to examine how inference for our main results on model heterogeneity driving forecasts and model aggregation may change from systematic differences between human coding and Claude-based coding, and find that inference remains intact.

[32]Cohen's kappa is defined as $\kappa = \frac{p_o - p_e}{1 - p_e}$ where p_o is observed agreement and p_e is expected agreement by chance.

IA.C.1 Human Coding Protocol

Human coders were graduate students and researchers with training in economics and familiarity with monetary policy. For parameter beliefs and arguments, human coders classified textual data independently without access to Claude's output, enabling true inter-rater reliability assessment. For decisions, coders reviewed Claude's extractions and assessed their accuracy; this audit-based approach was necessary given the complexity of synthesizing entire meeting transcripts to identify decisions.

For each parameter belief (Phillips curve, IS curve, Financial–Real Linkages), we construct stratified random samples of approximately 200 arguments with roughly 50 arguments per classification category. This stratified approach ensures sufficient observations in each category to estimate reliability, including for less common classifications. For argument validation, we randomly sampled 293 arguments across meetings and economic variables without stratification. For decision validation, we selected three FOMC meetings representing distinct policy contexts: August 1994 (tightening cycle), December 2008 (financial crisis emergency measures), and August 2011 (calendar-based forward guidance).

Coders received a detailed written manual defining each classification scheme with exact category definitions, illustrative examples, and guidance on edge cases. To ensure standardization, we developed custom web-based validation interfaces using Streamlit, an open-source platform. For parameter beliefs and arguments, the interface presented the quotation, description, and contextual explanation; Claude's classifications were stored in separate files never provided to coders. Coders selected classifications from drop-down menus, with the interface enforcing valid inputs and tracking progress. For the decision measures (alignment and influence), the interface presented Claude's extracted decisions alongside quoted evidence and justification, and coders assessed whether the evidence was accurately extracted, whether it supported Claude's interpretation, and provided their own scores.

Figure IA.9 displays the validation interface for parameter belief classifications. The interface presents the extracted quotation and a brief description of the argument's content. Coders select from the relevant classification categories using radio buttons; an expandable classification guide provides detailed category definitions.

IA.C.2 Machine Coding Protocol

To mirror the Claude-based coding setup, we used the same task prompts and output schemas and executed them with Llama on the NYU HPC cluster, keeping the Llama

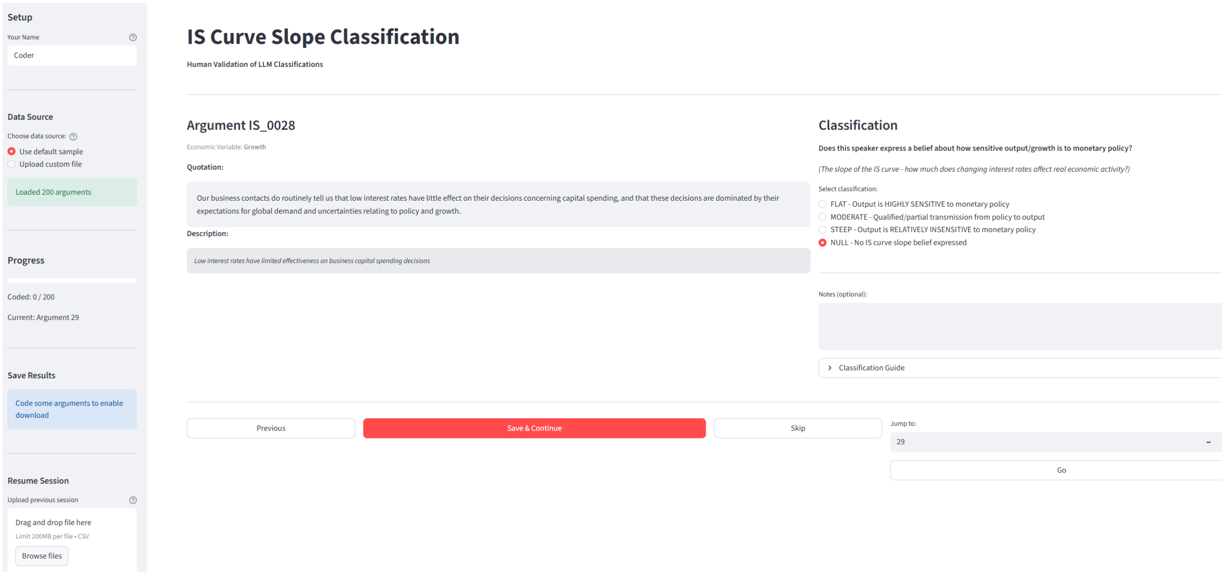


FIGURE IA.9: HUMAN CODING INTERFACE FOR PARAMETER BELIEF VALIDATION

Note: The figure displays the Streamlit-based interface used by human coders for parameter belief validation. The example shows an IS curve slope classification task. Coders view the quotation extracted from FOMC transcripts and a brief description, then select a classification (Flat, Moderate, Steep, or Null). The interface tracks progress and allows coders to save and resume sessions. Claude's classifications were stored in separate files not accessible through the interface.

runs blind to any Claude outputs. We ran two open-weight Llama variants via the self-hosted Ollama runtime: a smaller model (Llama 3.1 8B) and a larger, more powerful model (Llama 3.3 70B). The Llama 3.1 8B is a lightweight model targeted at basic text generation and simple coding tasks, while the 70B model is significantly more capable, handling complex reasoning and requiring substantially more computational power.

IA.C.3 Summary of Results

Table IA.19 presents consolidated validation results. Here, we report the concordance between the measures constructed by Claude and the alternatives from human coders.

Overall, Claude achieves high concordance with human classifiers. Claude achieves high correlations with human coders on parameter measures ($r = 0.80$–0.83). To put these numbers into context, we also report the correlation between the two human classifiers as "Human Ceiling". Claude exceeds this human inter-coder reliability for Financial–Real Linkages.

Moving to the middle panel, argument scores correlate at $r = 0.80$ with the human average, approaching the human ceiling of $r = 0.84$. Argument category classification matches human reliability ($\kappa = 0.65$ versus 0.64). The primary limitation is information type classification, where Claude achieves only moderate agreement ($\kappa = 0.49$) due to systematic over-identification of private information, where all information produced by regional Federal Reserve banks was classified as private information. Finally, decision-related measures show near-perfect agreement: influence scores correlate at $r = 0.94$ and alignment scores at $r = 0.97$, though as noted above, these validations used an audit-based approach without an independent human ceiling.

Measure Class	Measure	Metric	Claude	Human Ceiling
Parameter Beliefs	Phillips Curve Slope	r	0.811	0.856
	IS Curve Slope	r	0.795	0.800
	Financial–Real Linkages	r	0.825	0.784
Arguments	Argument Score	r	0.803	0.837
	Argument Category	κ	0.654	0.636
	Information Type	κ	0.485	0.741
Decisions	Decision Identification	% valid	90%	—
	Influence Scores	r	0.935	—
	Alignment Scores	r	0.972	—

TABLE IA.19: SUMMARY OF VALIDATION RESULTS

Note: Claude column reports correlation with human average. Human ceiling is inter-coder agreement among human raters. Decision measures used audit-based validation without independent human coding, so no human ceiling is available.

IA.C.4 Parameter Beliefs

We examine the parameter beliefs measures. We validate three measures capturing FOMC members' beliefs about economic relationships: the Phillips curve slope (inflation-unemployment tradeoff), the IS curve sensitivity (interest rate sensitivity of output), and Financial–Real Linkages (credit conditions transmission to the real economy). Each uses an ordinal scale with categories for strong, moderate, and weak relationships, plus a null category for arguments not expressing a pertinent belief.

Table IA.20 reports detection rates for parameter beliefs. By construction from the random sampling, Claude identifies parameter beliefs in 75% of arguments across all three parameters. Coder A detects beliefs in 50–70% of arguments, while Coder B detects beliefs in 37–45%. Notably, Claude rarely misses beliefs that humans detect. Across all comparisons, Claude fails to identify a belief when a human coder does in only 0–5 cases per parameter. This pattern indicates that Claude's higher detection rate reflects a lower threshold for identifying beliefs rather than a systematic error.

	Phillips Curve	IS Curve	Financial–Real Linkages
Non-null classifications (out of 200 arguments)			
Claude	150	150	150
Coder A	101	128	139
Coder B	76	74	90
Human non-null, Claude null			
Coder A	0	4	1
Coder B	3	5	1

TABLE IA.20: PARAMETER MEASURES: DETECTION RATES

Note: The top panel reports the number of arguments (out of 200) for which each rater identified a parameter belief. The bottom panel reports cases where a human coder identified a belief but Claude did not.

Table IA.21 presents classification performance. Claude achieves high correlations with human coders across all three measures ($r = 0.80$–0.83). For Financial–Real Linkages, Claude's correlation with the human average exceeds the human ceiling, the correlation between the two human coders. For the Phillips curve and IS curve, Claude maintains strong agreement. Llama 70B performs below Claude on all measures, while Llama 8B shows only moderate correlations, indicating that model scale matters for these nuanced classification tasks.

IA.C.5 Arguments Dataset

The arguments dataset captures individual points made by FOMC members during policy deliberations. We validate four dimensions: outlook scores, argument categories, information type, and data source categories. The validation sample consists of 293 randomly sampled arguments coded independently by two human coders.

Outlook Scores. For inflation, growth, and unemployment, outlook scores capture the stance on the economic outlook indicated by the argument on a scale ranging from -3 (strongly negative) to $+3$ (strongly positive). Table IA.22 presents the results. Claude achieves a correlation of $r = 0.80$ with the human average, compared to the human ceiling of $r = 0.84$. Individual coder correlations are $r = 0.773$ and $r = 0.767$. Llama 70B achieves a higher correlation ($r = 0.88$), exceeding the human ceiling.

	Phillips Curve	IS Curve	Financial–Real Linkages
Human Average	0.811	0.795	0.825
	($n = 111$)	($n = 129$)	($n = 141$)
Coder A	0.786	0.835	0.823
	($n = 101$)	($n = 124$)	($n = 138$)
Coder B	0.824	0.746	0.760
	($n = 73$)	($n = 69$)	($n = 89$)
Human Ceiling	0.856	0.800	0.784
Llama 70B	0.730	0.786	0.750
Llama 8B	0.556	0.615	0.580

TABLE IA.21: PARAMETER MEASURES: CLASSIFICATION PERFORMANCE

Note: The table reports correlations between Claude's parameter classifications and human coder classifications. "Human Average" averages available human coder scores (using a single coder when only one provided a non-null classification) before computing correlations. "Human Ceiling" is the correlation between the two human coders. Correlations computed on numeric encodings (e.g., steep=1, moderate=0, flat=-1 for Phillips curve). Llama models received identical prompts to Claude.

	Argument Scores
Human Average	0.803
	($n = 293$)
Coder A	0.773
	($n = 293$)
Coder B	0.767
	($n = 293$)
Human Ceiling	0.837
Llama 70B	0.876
Llama 8B	0.721

TABLE IA.22: OUTLOOK SCORES: CLASSIFICATION PERFORMANCE

Note: The table reports Pearson correlations between Claude's outlook score classifications (-3 to $+3$) and human coder classifications. "Human Average" is the correlation with the average of the two human coders' scores. "Human Ceiling" is the correlation between the two human coders. Llama models received identical prompts to Claude.

Argument Categories. Arguments are classified into 11 topic categories such as Growth Forecasting, Consumer Demand, and Labor Markets. Claude's pairwise agreement with individual humans ($\kappa = 0.65$) matches human-human agreement ($\kappa = 0.64$), demonstrating human-level reliability. The 28.7% human disagreement rate reflects inherent subjectivity in categorizing economic arguments. When compared only to cases where humans agree, Claude achieves $\kappa = 0.79$ with 84.2% raw agreement. Llama 8B performs near chance ($\kappa = 0.03$), assigning 97.6% of arguments to a single category.

Information Type. Information type classifies whether arguments rely on public information or private/specialized information. Claude's classification shows moderate agreement with human consensus ($\kappa = 0.49$, $N = 223$), below the human ceiling ($\kappa = 0.74$). The classifications follow a systematic pattern: all 30 disagreements involve Claude classifying as private when humans classify as public. Claude never misses private information (zero cases of Claude saying public when humans say private). This pattern suggests Claude applies overly broad criteria for private information, likely conflating category names containing terms like "District Intelligence" with non-public sources.

This systematic over-identification is conservative for our analysis. Our finding that private information plays a limited role in FOMC deliberations is robust; the true prevalence of private information is likely even lower than Claude's estimates suggest.

Data Source Categories. Table IA.23 presents validation results for the 11 binary data source indicators. Performance varies substantially across categories. Claude meets or exceeds the human ceiling for Fed System Surveys ($\kappa = 0.75$ versus 0.42), Official Government Statistics ($\kappa = 0.32$ versus 0.31), and Financial Market Data ($\kappa = 0.29$ versus 0.25). The relatively low kappa values even at the human ceiling for some categories indicate that data source identification is inherently difficult, likely because arguments often cite multiple sources implicitly.

IA.C.6 Decision Measures

We validate Claude's extraction of FOMC meeting decisions using a two-layer framework: decision identification, evaluating whether Claude correctly identifies policy decisions, and member-level scoring, evaluating how accurately Claude assigns influence and alignment scores. The validation sample is composed of speaker-decision pairs across 18 validated decisions from three meetings spanning different policy contexts. As noted above, decision validation uses an audit-based approach in which human coders reviewed Claude's extractions rather than coding independently, so no human ceiling is

Data Category	Claude vs. Consensus κ	Claude vs. Consensus Agreement	Human Ceiling κ	Human Ceiling Agreement
Fed System Surveys & Research	0.747	99.3%	0.417	97.3%
District Business Intelligence	0.740	93.5%	0.764	93.9%
International Economic Intelligence	0.541	96.2%	0.693	96.9%
Private Sector Surveys & Research	0.513	96.9%	0.697	98.3%
Federal Reserve Staff Analysis	0.452	78.5%	0.488	81.2%
Banking & Credit Market Intelligence	0.393	95.2%	0.518	94.9%
Sectoral & Industry Intelligence	0.336	89.4%	0.547	87.4%
Official Government Statistics	0.320	67.2%	0.305	66.2%
Financial Market Data	0.288	93.9%	0.254	91.5%

TABLE IA.23: DATA CATEGORY VALIDATION

Note: The table reports Cohen's kappa estimates for Claude's data source category classifications with human category classification. Categories sorted by Claude's kappa in descending order. "Other" and "No data cited" excluded due to near-zero variance.

available for comparison.

Decision Identification. Claude proposed 20 candidate decisions across the three meetings. A human coder validated 18 decisions (90%), with 17 exact matches and one requiring minor correction. The two rejected decisions from August 1994 were technical items (monetary aggregate ranges) that did not constitute substantive policy decisions. For validated decisions, the human coder agreed with Claude's type classification (rate decision, communication, other) in 100% of cases, and the correlation between Claude's and human hawkish/dovish scores is $r = 0.987$.

Llama models failed to provide a meaningful benchmark. Llama 70B identified only 6 of 18 validated decisions and completely missed the August 2011 meeting. Llama 8B failed to identify any decisions due to model limitations.

Influence and Alignment Scores. Table IA.24 presents validation results for member-level influence (0–3 scale) and alignment (-3 to $+3$ scale) scores. Claude achieves near-perfect correlations with the human coder for both influence ($r = 0.94$) and alignment ($r = 0.97$). Performance is consistent across all three meetings despite their different policy contexts.

The four high-disagreement cases in alignment scoring (defined as $|\text{diff}| \geq 2$) all occur in the December 2008 meeting, where unprecedented policy actions may have led to more nuanced member positions. Influence scoring shows no high-disagreement cases,

	Influence (0–3)	Alignment (-3 to $+3$)
1994-08-16	0.936	0.976
2008-12-16	0.925	0.962
2011-08-09	0.930	0.984
Overall	0.935	0.972

TABLE IA.24: DECISION SCORING VALIDATION

Note: The table reports Pearson correlations between Claude's scores and the human coder's scores for influence and alignment with decisions.

suggesting that influence criteria (first mover, reasoning provided, cited by others) are more objectively identifiable in transcripts than alignment.

IA.C.7 Comparison with Alternative LLMs

Table IA.25 summarizes Claude's performance relative to Llama models across all measures. Claude consistently outperforms Llama on parameter measures and complex classification tasks. The exception is argument scoring, where Llama 70B achieves a higher correlation, suggesting larger models may excel at quantitative assessment even when struggling with nuanced categorization. Llama 8B performs substantially worse across all measures, and both Llama models struggled with decision identification: Llama 70B found only 6 of 18 validated decisions while Llama 8B failed to identify any.

Measure	Claude	Llama 70B	Llama 8B
Phillips Curve (r)	0.811	0.730	0.556
IS Curve (r)	0.795	0.786	0.615
Financial–Real Linkages (r)	0.825	0.750	0.580
Argument Scores (r)	0.803	0.876	0.721
Argument Categories (κ)	0.654	—	0.033
Decisions Identified	18	6	—

TABLE IA.25: LLM COMPARISON

Note: The table reports correlations of LLM-assigned scores with human-coded scores for Claude, Llama 70B, and Llama 8b. Llama 8B assigned 97.6% of arguments to a single category. Decision identification reports count out of 18 validated decisions across three sample meetings; Llama 8B excluded due to model limitations.

IA.C.8 Assessing Potential Bias in Inference from LLM classifications

This section assesses whether systematic differences between LLM and human classifications could bias our estimates of how parameter beliefs predict forecast heterogeneity (Table 5) and how model fit relates to committee influence (Table 9). We estimate the conditional distribution of human classifications given LLM output, then apply bias corrections to our main estimates. We find little evidence that classification bias affects our inferences.

IA.C.8.1 Parameter Belief Heterogeneity

We first focus on members' parameter belief classifications. We seek to correct for two sources of potential systematic bias in LLM classifications that we observe relative to human classification: (i) *score misclassification*, where the LLM assigns a different ordinal score than humans would, conditional on the argument being relevant; and (ii) *relevance overclassification*, where the LLM marks an argument as relevant for a parameter when humans judge it irrelevant (a "false positive"). We note that these tests are likely conservative, given that along both dimensions, LLM–human disagreements were of the same order of magnitude as human–human disagreements.

For a given parameter score assigned by the LLM to argument j for parameter p, $s_{p,j} \in \{-1, 0, 1\}$, we use our human-coded validation sample to estimate two quantities: the probability that humans would classify the argument as irrelevant, $P(\text{null} \mid s_{p,j})$, and the expected human score conditional on relevance,

$$\mathbb{E}(h_{p,j} \mid s_{p,j}, h_{p,j} \neq \text{null}) = \frac{\sum_{h\in\{-1,0,1\}} h \times P(h_{p,j} = h \mid s_{p,j})}{1 - P(\text{null} \mid s_{p,j})}.$$

We estimate these probabilities using the empirical distribution of human scores for each LLM score bucket. For point estimates, we replace each LLM score $s_{p,j}$ with the corresponding $\mathbb{E}(h_{p,j} \mid s_{p,j}, h_{p,j} \neq \text{null})$. To construct confidence intervals to capture uncertainty in the conditional distribution of the human label for an argument given the LLM label, we use a bootstrap procedure that constructs 1,000 samples by resampling the validation data with replacement. For each bootstrap iteration, the procedure also probabilistically drops labeled observations with probability $P(h_{p,j} = \text{null} \mid s_{p,j})$ to reflect uncertainty from potential relevance overclassification.

Because our validation exercise employed two human coders, we consider two approaches for combining their assessments. The *average coder* method averages the two coders' ordinal scores, and when one coder assigns an ordinal score while the other marks the argument as irrelevant, uses the ordinal score. The *coder-pooled* method treats each (argument, coder) pair as an independent observation, effectively giving one half weight to arguments only coded by a single coder. We report results under both specifications because the two human coders exhibited substantial variation in their stringency at the extensive margin.

We then re-run the main specifications reported in Tables 5. Table IA.26 reports the results, and Table IA.27 reports associated bootstrap confidence intervals, which account for uncertainty in computing conditional probabilities of human labeling given LLM labeling. Panel A reports results from the average coder method. The odd columns report values in the regressions from replacing each argument score $s_{p,j}$ with a deterministic bias correction ($\mathbb{E}[h_{p,j}|s_{p,j}, h_{p,j} \neq \text{null}]$). The even columns report results from averaging over 1,000 bootstrap samples, which includes dropping observations with probability $P(h_{p,j} = \text{null} \mid s_{p,j})$. The coefficient values slightly drop compared with those reflected in the main text, for example: 0.20 versus 0.16, -0.059 versus -0.054, and -0.120 versus -0.101 for the Phillips curve, Financial–Real Linkages, and IS curve interaction terms. The *t*-statistics corresponding to those parameters, however, remain broadly similar: 4.69 versus 4.69 (deterministic) & 4.35 (average bootstrap); -3.21 versus -3.18 (deterministic) & -3.05 (average bootstrap); and -2.19 versus -2.16 (deterministic) and -1.91 (average bootstrap). Panel B reports results from the coder-pooled method: the results are broadly similar. Of note, the average bootstrap *t*-statistics are 4.13, -2.69, and -1.66 for the Phillips curve, Financial–Real Linkages, and IS curve interaction parameters. The drop relative to Panel A reflects the increased proportion of LLM identified arguments classified as null with this methodology, which introduces additional noise into estimated model parameters. The broad conclusion we draw is that directional bias in the LLM's classification has limited effect on the inference drawn.

IA.C.8.2 Model Aggregation Bias Correction

We perform a similar analysis for the model aggregation analysis reported in Section 4 of the main text. Here, the concern is about score misclassification instead of relevance overclassification. We find that inference is nearly identical.

Our analysis proceeds similarly: for member i's alignment (or influence) score with decision d in meeting t, $s_{i,d,t}$, we replace it with

$$\mathbb{E}(h_{i,d,t} \mid s_{i,d,t}) = \sum_h h \times P(h_{i,d,t} = h | s_{i,d,t}),$$

estimated using the human-coded data. We then re-run the regression results reported in

Panel A: Average Coder						
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.:	Phillips Curve Inflation (Y+1)		Fin–Real Linkages GDP Growth (Y+1)		IS Curve GDP Growth (Y+1)	
Belief Measure:	Expanding Avg (Point Est.)	Expanding Avg (Boot. Avg)	Expanding Avg (Point Est.)	Expanding Avg (Boot. Avg)	Expanding Avg (Point Est.)	Expanding Avg (Boot. Avg)
Belief × Condition	0.157	0.154	−0.054	−0.056	−0.101	−0.101
	(4.69)	(4.35)	(−3.18)	(−3.05)	(−2.16)	(−1.91)
Labor Market Tightness	0.563	0.580				
	(3.98)	(4.09)				
Perceived Policy Tightness					0.031	0.041
					(0.21)	(0.28)
Meeting FE	Yes	Yes	Yes	Yes	Yes	Yes
N	1,076	1,064	1,099	1,093	1,111	1,110

Panel B: Coder-Pooled						
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.:	Phillips Curve Inflation (Y+1)		Fin–Real Linkages GDP Growth (Y+1)		IS Curve GDP Growth (Y+1)	
Belief Measure:	Expanding Avg (Point Est.)	Expanding Avg (Boot. Avg)	Expanding Avg (Point Est.)	Expanding Avg (Boot. Avg)	Expanding Avg (Point Est.)	Expanding Avg (Boot. Avg)
Belief × Condition	0.151	0.140	−0.054	−0.051	−0.101	−0.087
	(4.69)	(4.13)	(−3.18)	(−2.69)	(−2.17)	(−1.66)
Labor Market Tightness	0.573	0.609				
	(4.03)	(4.25)				
Perceived Policy Tightness					0.029	0.027
					(0.20)	(0.20)
Meeting FE	Yes	Yes	Yes	Yes	Yes	Yes
N	1,076	1,054	1,099	1,085	1,111	1,105

Table IA.26: Parameter Heterogeneity: Adjusting for Potential Directional LLM Bias

Note: The table reports results from using a bias correction for LLM parameter scores using human-coded data. The Point Est. columns show coefficients using deterministic bias correction ($\mathbb{E}[h|s, h \neq \text{null}]$). Boot. Avg columns report mean coefficients and t-statistics across 1,000 bootstrap iterations with resampling of human-coded data, and randomly setting observations coded by the LLM to null the distribution of arguments scored as relevant by Claude but not scored as relevant by humans. All specifications include meeting fixed effects. Panel A reports results from the average coder approach, which averages the two coders' ordinal scores for each case, preserving partial disagreement (e.g., scores of $+1$ and 0 yield $+0.5$), and using the coded human's score if only one coder coded a given argument. Panel B reports results from the coder-pooled approach that treats each coder's classification as an independent draw, pooling both coders to construct $P(h|s)$.

PANEL A: AVERAGE CODER			
	Phillips Curve [1%, 99%]	Fin–Real Linkages [1%, 99%]	IS Curve [1%, 99%]
Belief × Condition (coef)	[0.120, 0.191]	[−0.079, −0.032]	[−0.158, −0.055]
Labor Market Tightness (coef)	[0.485, 0.680]		
Perceived Policy Tightness (coef)			[−0.003, 0.105]
PANEL B: CODER-POOLED			
	Phillips Curve [1%, 99%]	Fin–Real Linkages [1%, 99%]	IS Curve [1%, 99%]
Belief × Condition (coef)	[0.105, 0.182]	[−0.081, −0.017]	[−0.167, −0.019]
Labor Market Tightness (coef)	[0.492, 0.733]		
Perceived Policy Tightness (coef)			[−0.016, 0.103]

TABLE IA.27: PARAMETER HETEROGENEITY: BOOTSTRAP DISTRIBUTION FOR DIRECTIONAL BIAS CORRECTION

The table reports 1%–99% bootstrap confidence intervals for regression coefficients from the bias correction analysis. Each interval is computed from 1,000 bootstrap iterations, where each iteration resamples the human validation data and probabilistically drops observations based on $P(\text{null}|s)$, which is the probability that humans classify an LLM-identified argument as not relevant. Panel A reports intervals using the average coder approach, which averages the two coders' ordinal scores for each validation case, preserving partial disagreement (e.g., scores of +1 and 0 yield +0.5), and using the single coder's score when only one coder classified a given argument. Panel B reports intervals using the coder-pooled approach, which treats each coder's classification as an independent draw, pooling both coders to construct $P(h|s)$.

Panel A: Half-Year Forecast Error-Based Fit					
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE}}_{i,t}(H)$	−6.21	0.183	0.141	0.167	0.110
	(2.86)	(3.09)	(3.04)	(2.83)	(3.14)
N	893	893	893	1,498	1,498
Meeting FE	Yes	Yes	Yes	Yes	Yes

Panel B: Half-Year Imputed Forecast Error-based Fit					
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE,impute}}_{i,t}(H)$	−3.51	0.115	0.059	0.105	0.056
	(3.00)	(3.25)	(2.23)	(2.48)	(2.15)
N	2,063	2,063	2,063	3,484	3,484
Meeting FE	Yes	Yes	Yes	Yes	Yes

Table IA.28: Model Fit and Policy Decision Alignment, with Human-Coded Bias Correction

Note: The table reports regression results repeating the analysis in Table 9, using bias-corrected LLM-generated dependent variables (alignment and influence) based on human validation data. The bias correction replaces each Claude score s with its expected human score $\mathbb{E}[h|s]$.

Table 9. We also perform 1,000 bootstrap simulations, sampling human-coded scores with replacement in constructing $P(h_{i,d,t} = h|s_{i,d,t})$, to account for potential sampling variation in the human coding; we report 1st and 99th percentile values corresponding to these bootstraps.

Table IA.28 reports the main results and Table IA.29 reports the bootstrap confidence intervals. The coefficient values where alignment is the dependent variable are largely similar: 0.181 to 0.183 for forecast error-based fit, and constant at 0.115 for imputed forecast error-based fit. The coefficient magnitudes also similar for influence: 0.146 to 0.141 and 0.063 to 0.059 for forecast error-based fit and imputed forecast error-based fit, respectively. The t-statistics are also largely similar (3.07 to 3.09 and 3.01 to 3.04 for those two specifications with voters), indicating little change.

	Voters			All Members	
Panel A: Half-Year Forecast Error-Based Fit					
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}_{i,t}^{\text{FE}}(H)$	-6.21	[0.179, 0.185]	[0.137, 0.145]	[0.164, 0.171]	[0.106, 0.113]
	(2.86)	[3.06, 3.12]	[3.02, 3.06]	[2.79, 2.86]	[3.12, 3.15]
N	893	893	893	1,498	1,498
Meeting FE	Yes	Yes	Yes	Yes	Yes
Panel B: Half-Year Imputed Forecast Error-based Fit					
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}_{i,t}^{\text{FE,impute}}(H)$	-3.51	[0.113, 0.117]	[0.056, 0.062]	[0.104, 0.106]	[0.053, 0.057]
	(3.00)	[3.18, 3.32]	[2.17, 2.27]	[2.45, 2.52]	[2.10, 2.19]
N	2,063	2,063	2,063	3,484	3,484
Meeting FE	Yes	Yes	Yes	Yes	Yes

TABLE IA.29: MODEL FIT AND POLICY DECISION ALIGNMENT, WITH HUMAN-CODED BIAS CORRECTION CONFIDENCE INTERVALS

Note: The table reports bootstrap confidence intervals of regression results repeating the analysis in Table 9, using bias-corrected LLM-generated dependent variables (alignment and influence) based on human validation data. The bias correction replaces each Claude score s with its expected human score $\mathbb{E}[h|s]$. The bootstrap samples the human-coded sample with replacement to control for sampling variation in human coding.

IA.C.9 Discussion

Overall, Claude's automated extraction achieves reliability comparable to trained human coding across most measures. For parameter beliefs, Claude achieves high correlations ($r = 0.80$–0.83), meeting or exceeding human inter-coder reliability. For decision-related measures, Claude achieves near-perfect correlations ($r = 0.94$–0.97), though the audit-based methodology provides a less stringent test than blind independent coding. Argument scores and categories validate at human-level reliability. The primary difference in these classifications relative to humans, where Claude exhibits relative over-classification of private information, suggesting the true prevalence of private information in FOMC deliberations is likely lower than our estimates suggest. In simulation exercises, we use our human-coded sample to examine how inference for our main results on model heterogeneity driving forecasts and model aggregation may change from systematic differences between human coding and Claude-based coding, and find that inference remains the same.

IA.C.10 Processing FOMC Data

FOMC transcripts are provided in PDF form on the FOMC "Historical Materials by Year" page.[33] 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 separately. 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

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

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[34] 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 speakers 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 par-

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

ticular, 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.C.11 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 committee members' arguments for each variable are in Section IA.C.11.1.
- Given their arguments, prompts to extract beliefs about model parameters are presented in Section IA.C.11.2.
- Prompts to score members' belief-based arguments based on the outlooks they express are presented in Section IA.C.11.3.
- Given their arguments, prompts to identify the data cited in committee members' arguments are in Section IA.C.11.4.
- The prompts to assign each argument made by each member to a category are given in Section IA.C.11.5.
- The prompt to identify each policy decision made during the meeting is in Section IA.C.11.6.
- 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.C.11.7.

IA.C.11.1 Prompts to Identify 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. 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, and Growth.

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 where the LLM is prompted to identify the arguments made is given below.

```
"""You are an expert economic analyst specializing in Federal Reserve communications. Your
 task is to analyze FOMC member speeches to identify policy-relevant arguments about a
 specific economic variable and determine their implications for monetary policy.

Here are the FOMC member speeches you need to analyze:

<speeches>
{speeches}
</speeches>

Here is the specific economic variable you should focus your analysis on:

<variable_of_interest>
{variable}
</variable_of_interest>

## Task Overview

You need to identify policy-relevant arguments from these speeches that relate to the
 variable of interest. An argument qualifies as policy-relevant if it reflects the speaker's
  own perspective (not views they cite to disagree with) and fits into one or more of these
 five categories:

**1. Belief/Forecast - Cyclical**: A belief or forecast about recent past or current
 cyclical conditions or short-term cyclical movements in the variable of interest

**2. Belief/Forecast - Trend/Structural**: A belief or forecast about longer-term trends,
 potential levels, natural rates, or structural shifts in the variable of interest

**3. Preference/Consequences**: A preference for how the current/future level of the
 variable should compare to a desired level, including potential consequences of deviations
 from target levels

**4. Causal Statement**: A causal statement (potentially hypothetical) about what causes
 variation in the variable of interest

**5. Model Parameters**: A belief about policy-invariant constants that govern economic
 relationships and mechanisms related to the variable (distinct from describing current
 economic conditions or time-varying states)

## Critical Guidelines

- Include only the speaker's own views or views they clearly endorse
- Exclude arguments the speaker cites to refute or disagree with
- For structural parameters, focus on beliefs about stable economic mechanisms and
 relationships, NOT descriptions of current economic states or time-varying conditions
```

```
- For each policy-relevant argument, determine whether it supports accommodative monetary
 policy (lower interest rates) or restrictive monetary policy (higher interest rates)

## Analysis Process

Conduct your analysis using this four-step systematic process:

**Step 1: Comprehensive Quote Extraction**
- Read through each speech from beginning to end
- Extract every quote that mentions the variable of interest, even tangentially
- For each quote, provide: (a) the full quote with sufficient surrounding context (at least
 one full sentence before and after when available), and (b) a sequential number (Quote 1,
 Quote 2, etc.)
- Be exhaustively comprehensive - it's better to include too many quotes than to miss
 relevant ones
- It's OK for this section to be quite long

**Step 2: Quote-by-Quote Systematic Evaluation**
For each numbered quote from Step 1:
- State "Quote [number]:"
- First determine: Does this reflect the speaker's own view or an endorsed external view? Or
  is the speaker citing this to refute/disagree with it?
- If it reflects the speaker's perspective, systematically evaluate against each category by
  explicitly stating:
* Belief/Forecast - Cyclical: Y or N, with brief reasoning
* Belief/Forecast - Trend/Structural: Y or N, with brief reasoning
* Preference/Consequences: Y or N, with brief reasoning
* Causal Statement: Y or N, with brief reasoning
* Model Parameters: Y or N, with brief reasoning (Remember: parameters are not time-varying
 states or current conditions)
- If it fits any category, determine the monetary policy stance (accommodative vs
 restrictive) and explain your reasoning
- Conclude with: "Policy-relevant argument: Y or N"
- It's OK for this section to be quite long as you work through each quote systematically

**Step 3: Policy-Relevant Argument Identification**
- List all quotes from Step 2 that you determined to be policy-relevant arguments
- For each, write a brief description of the argument and its policy implications

**Step 4: Synthesis and Patterns**
- Identify overarching themes, patterns, areas of consensus, or disagreement among speakers
 regarding the variable of interest

## Output Format

After completing your systematic analysis, provide your final results in this exact XML
 structure. Each argument should be contained within an ‘<argument>‘ block:

‘‘‘xml
<output>
<argument>
    <description>[Brief description of the argument]</description>
    <quotation>[Supporting quotation from the speeches]</quotation>
    <belief_forecast_cyclical>[Y or N]</belief_forecast_cyclical>
    <belief_forecast_trend>[Y or N]</belief_forecast_trend>
```

~~~~
    <preference_consequences>[Y or N]</preference_consequences>
    <causal_statement>[Y or N]</causal_statement>
    <model_parameters>[Y or N]</model_parameters>
    <stance>[Accommodative or Restrictive]</stance>
    <explanation>[Brief explanation of the argument's implications for monetary policy]</
     explanation>
</argument>
<argument>
    <description>[Brief description of the second argument]</description>
    <quotation>[Supporting quotation from the speeches]</quotation>
    <belief_forecast_cyclical>[Y or N]</belief_forecast_cyclical>
    <belief_forecast_trend>[Y or N]</belief_forecast_trend>
    <preference_consequences>[Y or N]</preference_consequences>
    <causal_statement>[Y or N]</causal_statement>
    <model_parameters>[Y or N]</model_parameters>
    <stance>[Accommodative or Restrictive]</stance>
    <explanation>[Brief explanation of the argument's implications for monetary policy]</
     explanation>
</argument>
</output>
```

Work through the four-step systematic process in <systematic_analysis> tags, then provide
 your final structured output."""
~~~~

IA.C.11.2 Prompts to Extract Model Parameters.

Phillips curve: To extract FOMC members' beliefs about the Phillips curve relationship, we pass each argument through the LLM with a classification prompt. The prompt instructs the model to identify statements in which a speaker connects both a labor market indicator and an inflation outcome causally. Simply mentioning both concepts is insufficient; the speaker must express a view about how one affects the other. We classify beliefs as *steep* (labor market tightness significantly affects inflation), *flat* (little or no effect), or *moderate* (qualified or partial relationship). The prompt provides detailed classification rules with indicator phrases for each category and special case handling for forecasts, historical references, and policy discussions. The model returns an XML object containing the classification and a brief reasoning statement.

```
"""
You are tasked with extracting and classifying Phillips curve beliefs from Federal Open
 Market Committee (FOMC) speaker statements. The Phillips curve describes the economic
 relationship between labor market conditions and inflation outcomes.

# Your Task

Classify the speaker's Phillips curve belief based on their quote. You must determine
 whether the speaker believes that labor market conditions significantly affect inflation,
 have little effect, have a moderate effect, or if no belief is expressed.
```

```
# Classification Categories

**STEEP**: The speaker indicates that labor market conditions SIGNIFICANTLY affect inflation
 .

**FLAT**: The speaker indicates that labor market conditions have LITTLE or NO effect on
 inflation.

**MODERATE**: The speaker indicates a QUALIFIED or PARTIAL relationship between labor
 markets and inflation.

**NULL**: The speaker does NOT express a Phillips curve belief (this is the default).

# Classification Process

Follow this systematic process to classify the quote:

## Step 1: Check for Explicit Phillips Curve Mentions

If the quotation explicitly mentions "Phillips curve" and expresses a view about it,
 classify as follows:
- "Phillips curve is flat/dead/broken"     FLAT
- "Phillips curve remains valid/steep/strong"     STEEP
- "Phillips curve has weakened but exists"     MODERATE

If there is an explicit mention with a clear view, proceed directly to classification.
 Otherwise, continue to Step 2.

## Step 2: Identify Required Components

For a Phillips curve belief to be present, the quote must mention BOTH of these components:

**LABOR MARKET INDICATORS** (at least one of):
- Unemployment, employment, joblessness, jobless rate
- NAIRU, natural rate of unemployment
- Labor market tightness/slack/softness
- Full employment, maximum employment
- Capacity pressures, capacity utilization, capacity constraints
- Resource gaps, output gap
- Wage pressures, wage growth, wage increases, wage acceleration, wage conditions

**INFLATION OUTCOMES** (at least one of):
- Inflation expectations
- Price pressures, pricing power
- Actual or forecasted inflation changes/trends
- Core inflation, headline inflation, PCE
- Wage-price spiral, passthrough to prices
- Deflationary pressures, inflationary pressures

If BOTH components are NOT present, classify as NULL.

## Step 3: Identify Causal Connection

If both components are present, determine whether the speaker connects them causally. Simply
```

```
  mentioning both concepts without linking them is NOT sufficient for classification.

**Signs of NO causal connection** (classify as NULL):
- Only describes data without interpreting the relationship
- Discusses policy responses without explaining the economic mechanism
- Provides historical narrative without making causal claims
- Mentions both concepts in separate contexts without connecting them

**Signs of causal connection** (proceed to Step 4):
- Uses causal language linking the two concepts
- Expresses concern or confidence about transmission from one to the other
- Explains a mechanism by which one affects the other
- Shows surprise at presence or absence of expected relationship

## Step 4: Classify the Type of Relationship

If a causal connection exists, classify based on the language and strength of the
 relationship:

### STEEP Classification

Use when the speaker indicates labor markets SIGNIFICANTLY affect inflation.

**Strong indicators:**
- Causal language: "drives", "causes", "pushes", "leads to", "results in"
- Concern about transmission: "will feed into", "translate to", "spillover"
- Upside risk emphasis: "tight labor     higher inflation risk"
- Historical validation: "as we've seen before", "consistent with theory"

**Examples:**
- "Tight labor markets are driving wage pressures that will feed into core inflation"
- "Unemployment below NAIRU should push inflation higher"
- "Resource constraints will lead to price increases"
- "The strong labor market poses upside risks to inflation"

### FLAT Classification

Use when the speaker indicates labor markets have LITTLE or NO effect on inflation.

**Strong indicators:**
- Disconnection language: "despite", "even though", "hasn't translated", "failed to"
- Skepticism: "broken", "dead", "weakened", "no longer valid"
- Surprise or puzzle: "surprisingly", "puzzling that", "contrary to expectations"
- Other factors dominate: "driven by supply shocks not demand", "global factors matter more"

**Examples:**
- "Despite unemployment below 4%, we've seen no acceleration in inflation"
- "The Phillips curve appears quite flat in recent years"
- "Wage growth hasn't translated into price pressures"
- "Labor market tightness is not driving inflation"
- "Inflation dynamics are driven by supply factors, not labor markets"

### MODERATE Classification

Use when the speaker indicates a QUALIFIED or PARTIAL relationship.
```

```
**Strong indicators:**
- Hedging: "some", "modest", "limited", "to some extent"
- Conditionality: "depends on", "in certain circumstances", "may"
- Weakening: "less than before", "diminished", "not as strong as"
- Mixed signals: "on one hand... on the other", "both... and"

**Examples:**
- "Tight labor may generate some inflation pressure, but effects are modest"
- "The Phillips curve exists but has flattened considerably"
- "Labor markets affect inflation, but supply factors matter more"
- "We see limited passthrough from wages to prices"

## Step 5: Handle Special Cases

### Forecasts and Projections
Only classify if the reasoning reveals Phillips curve logic:
- "I forecast higher inflation"     NULL (no mechanism stated)
- "I forecast higher inflation because labor markets are tight"     STEEP (mechanism stated)

### Historical References
Only classify if the speaker endorses or rejects the relationship:
- "Inflation was high in the 1970s"     NULL (just description)
- "The 1970s showed how tight labor drives inflation"     STEEP (endorses relationship)

### Policy Discussions
Only classify if the speaker explains the policy through Phillips curve logic:
- "We should tighten policy"     NULL (no mechanism)
- "Tight labor will push inflation higher, so we should tighten"     STEEP (mechanism stated
 )

### Forward Guidance
Only classify if conditional on the labor-inflation link:
- "We'll keep rates low until 2024"     NULL (no condition)
- "We'll raise rates if tight labor starts pushing inflation up"     STEEP (conditional on
 link)

Provide your classification in well-formed XML format:

<classification>
<phillips_slope>[steep|flat|moderate|null]</phillips_slope>
<reasoning>[Brief explanation in 1-2 sentences referencing key phrases from the quote]</
 reasoning>
</classification>

**IMPORTANT REMINDERS:**
- Default to NULL unless the speaker clearly expresses a belief about how labor market
 conditions affect inflation
- Describing both concepts without connecting them causally is NOT sufficient for
 classification
- Your reasoning must cite specific phrases from the quotation
- Your XML output must be well-formed and machine-readable"""
```

IS curve: To extract FOMC members' beliefs about the IS curve relationship (the sensitivity of output to monetary policy), we pass each argument through the LLM with a classification prompt. The prompt instructs the model to identify statements in which a speaker connects both a monetary policy indicator and an output outcome causally. Simply mentioning both concepts is insufficient; the speaker must express a view about how one affects the other. We classify beliefs as *flat* (output highly sensitive to monetary policy), *steep* (output relatively insensitive), or *moderate* (qualified or partial transmission). The prompt provides detailed classification rules with indicator phrases for each category, disambiguation rules for constructions such as "resilient" and "despite tightening," and special case handling for forecasts, policy recommendations, and historical references. The model returns an XML object containing the classification and a brief reasoning statement.

```
"""
    You are tasked with extracting and classifying IS curve beliefs from Federal Open Market
      Committee (FOMC) speaker statements. The IS curve describes the relationship between
     monetary policy (interest rates) and real economic output/growth.

    ## Your Task

    Classify the speaker's IS curve belief based on their quote. You must determine whether
     the speaker believes that monetary policy significantly affects output, has little
     effect, has a moderate effect, or if no belief is expressed.

    ## Classification Categories

    **FLAT**: The speaker indicates that output is HIGHLY SENSITIVE to monetary policy
     changes.

    **STEEP**: The speaker indicates that output has LITTLE or NO response to monetary
     policy changes.

    **MODERATE**: The speaker indicates a QUALIFIED or PARTIAL transmission from policy to
     output.

    **NULL**: The speaker does NOT express an IS curve belief (this is the default).

    ---

    ## Classification Process

    Follow this systematic process to classify the quote:

    ### Step 1: Check for Explicit Transmission Mentions

    If the quotation explicitly discusses policy transmission effectiveness, classify as
     follows:

    - "policy is gaining traction / working / biting"     FLAT
```

```
- "pushing on a string / transmission impaired / channels clogged"     STEEP
- "some effect but limited / modest transmission"     MODERATE

If there is an explicit mention with a clear view, proceed directly to classification.
 Otherwise, continue to Step 2.

### Step 2: Identify Required Components

For an IS curve belief to be present, the quote must mention BOTH:

**MONETARY POLICY INDICATORS** (at least one):
- Interest rates, policy rate, fed funds rate, funds rate
- Rate increases/hikes/tightening, rate cuts/easing
- Policy stance: tight/restrictive/easy/accommodative
- Real rates, policy restraint/stimulus
- Monetary conditions, policy actions
- Money supply, money growth, monetary aggregates
- Balance sheet policies, asset purchases, QE

**OUTPUT/GROWTH OUTCOMES** (at least one):
- GDP, output, economic growth, real activity
- Aggregate demand, spending, consumption, investment
- Interest-sensitive sectors (housing, durables, capex)
- Economic slowdown, recession risk, expansion
- Business activity, economic momentum

**If BOTH components are NOT present, classify as NULL.**

### Step 3: Identify Causal Connection

If both components are present, determine whether the speaker connects them causally.
 Simply mentioning both concepts without linking them is NOT sufficient.

**Signs of NO causal connection (classify as NULL):**
- Only describes data without interpreting policy transmission
- Discusses policy decisions without explaining the mechanism
- Forecasts growth without attributing it to policy
- Mentions both concepts in separate contexts

**Signs of causal connection (proceed to Step 4):**
- Uses causal language linking policy to output
- Expresses concern about policy effects on growth
- Shows surprise at presence or absence of expected effects
- Explains mechanism by which policy affects activity

### Step 4: Classify the Type of Relationship

If a causal connection exists, classify based on language and signal words:

---

#### FLAT Classification

Use when the speaker indicates output is HIGHLY SENSITIVE to monetary policy.
```

```
**Strong indicators:**
- **Causal language:** "rate hikes are slowing," "policy is restraining," "tightening is
  reducing"
- **Traction language:** "gaining traction," "working through," "taking hold," "biting"
- **Concern about overtightening:** "risk of slowing too much," "could tip into
 recession"
- **Sector effects attributed to policy:** "housing is weakening due to rates," "
 investment responding to policy"
- **Effectiveness affirmation:** "policy is working," "already seeing impact," "easing
 providing support"

**Examples:**
- "Our rate increases are clearly slowing interest-sensitive sectors"
- "Policy tightening is gaining traction and restraining demand"
- "Further rate hikes risk tipping the economy into recession"
- "The 200 basis points of easing is providing significant support to growth"
- "Lower rates in mortgages are bolstering housing prices"
- "When the Federal Reserve decided to let up, conditions began to rebound"

---

#### STEEP Classification

Use when the speaker indicates output is RELATIVELY INSENSITIVE to monetary policy.

**Strong indicators:**
- **Resilience despite policy:** "despite tightening, growth continues," "resilient to
 rate increases," "even as we reduce accommodation"
- **Impairment language:** "transmission impaired," "channels clogged," "pushing on a
 string," "pipeline not full"
- **Skepticism:** "less effect than expected," "not seeing impact," "won't work," "
 cannot produce growth"
- **Liquidity trap references:** "liquidity trap," "zero bound ineffective"
- **Other factors dominate:** "fiscal policy driving growth," "structural factors," "
 global forces"

** CRITICAL PATTERN - "Despite/Even as" Construction:**
When speakers note that economic performance continues DESPITE or EVEN AS policy changes
 , this indicates STEEP:
- "even with several more rate increases, the economy should expand 3.5-4%"     STEEP
- "maintained solid momentum even as we reduce policy accommodation"     STEEP
- "growth seems solid and resilient and in less need of accommodation"     STEEP

The key insight: if growth continues regardless of policy direction, output is
 insensitive to policy.

**Examples:**
- "Despite 300 basis points of tightening, growth remains above trend"
- "The economy has proven surprisingly resilient to higher rates"
- "We simply do not have the capability to produce higher real growth through
 accommodative policy"
- "With credit channels impaired, monetary policy is pushing on a string"
- "Business fixed investment has remained consistently weaker despite balance sheet
 policies"
- "A further 25 basis point cut will do nothing to change the near-term outlook"
```

```
---

#### MODERATE Classification

Use when the speaker indicates a QUALIFIED or PARTIAL transmission effect.

**Strong indicators:**
- **Hedging:** "some," "modest," "incremental," "limited," "to some extent"
- **Mixed channels:** works in some sectors but not others
- **Conditionality:** "depends on," "in certain circumstances," "may"
- **Weakening transmission:** "less than before," "diminished effect"

**Examples:**
- "An incremental prod towards activity in the real economy"
- "Policies likely contributed to mortgage and auto borrowing, but business investment
 remained weak"
- "We see limited passthrough from policy to spending"
- "Some transmission is occurring, but effects are modest"

---

### Step 5: Handle Special Cases

#### Forecasts and Projections
Only classify if the reasoning reveals IS curve logic:
- "I forecast slower growth"     NULL (no mechanism stated)
- "I forecast slower growth because policy is restrictive"     FLAT

#### Policy Recommendations Without Mechanism
- "We should raise rates"     NULL (no transmission belief)
- "We should pause because hikes are already slowing the economy"     FLAT

#### Historical References
Only classify if the speaker endorses or rejects the transmission:
- "Growth slowed in 1982"     NULL (just description)
- "The 1982 episode showed how powerfully rate hikes affect output"     FLAT

#### Lag Discussions
- "Effects in 2-3 quarters"     indicates transmission is working (FLAT)
- "Won't see effects for 2+ years"     weak near-term transmission (STEEP)
- "Long and variable lags" without specifics     default to NULL

#### Inflation Channel vs Output Channel
Only classify if OUTPUT effects are discussed:
- "Rate hikes will reduce inflation"     NULL (inflation channel only)
- "Rate hikes will slow growth and thus reduce inflation"     FLAT (output channel
 explicit)
- "Tightening will cool demand"     FLAT (demand = output)

---

## Critical Disambiguation Rules

### 1. "Resilient" and "Solid Growth" Language
```

````
When speakers describe the economy as "resilient," "solid," or "maintaining momentum" in
  the context of policy changes, determine:

- **If despite policy:** "solid growth DESPITE tightening"     STEEP
- **If because of policy:** "solid growth BECAUSE of accommodation"     FLAT
- **If unconnected:** "solid growth" (no policy mention)     NULL

### 2. Distinguishing IS Slope from Policy Preference

A member can believe the IS curve is flat and support either restriction or
 accommodation.
- Accommodative + flat: "We should cut because lower rates will meaningfully boost
 growth"
- Restrictive + flat: "We should pause because our hikes are already slowing the economy
 "
- Accommodative + steep: "Cutting won't help much the problems aren't monetary"
- Restrictive + steep: "We can keep tightening the economy is resilient to rate
 increases"

**Focus on the TRANSMISSION BELIEF, not the policy preference.**

### 3. "Need for" vs "Effect of" Policy

- "We need more accommodation"     NULL (preference, not transmission belief)
- "We need more accommodation to boost growth"     FLAT (causal claim)
- "Economy needs less accommodation now"     NULL unless connected to policy already
 working

---

## Output Format

Provide your classification in well-formed XML format:
```xml
<classification>
<is_slope>[steep|flat|moderate|null]</is_slope>
<reasoning>[Brief explanation in 1-2 sentences referencing key phrases from the quote]</
 reasoning>
</classification>
```

**IMPORTANT REMINDERS:**
- Default to NULL unless the speaker clearly expresses a belief about how monetary
 policy affects output/growth
- Watch for the "despite/even as" construction - this typically indicates STEEP
- "Resilient" in the context of policy changes indicates STEEP
- Your reasoning must cite specific phrases from the quotation
- Your XML output must be well-formed and machine-readable"""
````

Financial–real linkage: To extract FOMC members' beliefs about financial–real linkages, we pass each argument through the LLM with a classification prompt. The prompt instructs the model to identify statements in which a speaker connects both credit market conditions and real economic activity causally. Simply mentioning both concepts is insufficient; the speaker must express a view about how one affects the other. We classify beliefs as *strong* (credit conditions significantly amplify economic shocks through feedback effects), *moderate* (direct effects of credit on activity without explicit feedback, or hedged amplification), or *weak* (credit conditions have little effect on economic activity). The prompt provides detailed classification rules with indicator phrases for each category and special case handling for forecasts, historical references, policy discussions, and wealth effects. The model returns a JSON object containing the classification and a brief reasoning statement.

```
"""
INSTRUCTION: Classify the speaker's belief about whether credit market conditions affect or
 amplify economic activity. Return null for unrelated content.

# Your Task

Classify the speaker's belief about the financial accelerator or credit channel based on
 their quote. You must determine whether the speaker believes that credit market conditions
 significantly amplify economic shocks, directly affect economic activity, have little
 effect, or if no belief is expressed.

# Classification Categories

**STRONG**: The speaker indicates that credit market conditions SIGNIFICANTLY AMPLIFY
 economic shocks through feedback effects between financial conditions and real activity.

**MODERATE**: The speaker indicates either (a) a QUALIFIED amplification mechanism, OR (b) a
  DIRECT EFFECT of credit conditions on real economic activity without explicit feedback.

**WEAK**: The speaker indicates that credit market conditions have LITTLE or NO effect on
 economic activity, or that financial-real economy linkages are minimal.

**NULL**: The speaker does NOT express a credit channel belief (this is the default).

# Classification Process

Follow this systematic process to classify the quote:

## Step 1: Check for Explicit Financial Accelerator/Credit Channel Mentions

If the quotation explicitly mentions "financial accelerator," "credit channel," "credit
 transmission mechanism," or "balance sheet channel" and expresses a view about it, classify
  as follows:
- "Financial accelerator is broken/minimal/insignificant"     WEAK
- "Financial accelerator remains strong/powerful/significant"     STRONG
- "Financial accelerator exists but is weaker than before"     MODERATE

If there is an explicit mention with a clear view, proceed directly to classification.
 Otherwise, continue to Step 2.
```

```
## Step 2: Identify Required Components

For a credit channel belief to be present, the quote must mention BOTH of these components:

**CREDIT MARKET CONDITIONS** (at least one of):
- Credit availability, credit access, lending standards, credit conditions
- Bank lending, loan growth, credit flows
- Credit spreads, risk premia, borrowing costs
- Financial conditions, financial stress, financial tightness/ease
- Balance sheet strength/weakness, net worth, collateral values
- Leverage, debt burdens, debt service
- Credit constraints, credit frictions, borrowing constraints

**REAL ECONOMIC ACTIVITY** (at least one of):
- Investment, business spending, capital expenditures
- Consumption, consumer spending, household expenditures
- GDP growth, output, economic activity
- Employment, hiring decisions
- Business formation, entrepreneurship
- Asset prices affecting wealth effects
- Real economy effects, economic momentum

If BOTH components are NOT present, classify as NULL.

## Step 3: Identify Causal Connection

If both components are present, determine whether the speaker connects them causally. Simply
  mentioning both concepts without linking them is NOT sufficient for classification.

**Signs of NO causal connection** (classify as NULL):
- Only describes data without interpreting the relationship
- Discusses policy responses without explaining the transmission mechanism
- Provides historical narrative without making causal claims
- Mentions both concepts in separate contexts without connecting them

**Signs of causal connection** (proceed to Step 4):
- Uses causal language linking credit conditions to real activity
- Expresses concern or confidence about transmission from one to the other
- Explains a mechanism by which one affects the other
- Shows surprise at presence or absence of expected relationship

## Step 4: Classify the Type of Credit-Activity Relationship

If a causal connection exists, determine what type of relationship the speaker describes:

### TYPE A     Amplification/Feedback

The speaker describes credit conditions AMPLIFYING shocks through feedback mechanisms.

**Indicators:**
- Feedback language: "self-reinforcing," "spiral," "vicious/virtuous cycle," "feedback loop"
- Amplification language: "magnify," "amplify," "exacerbate," "deepen," "compound," "worsen"
- Explicit causal chains that close the loop: "tighter credit     weaker activity
 tighter credit"
```

```
- Propagation language: "spillover," "contagion," "cascade"

**If Type A identified:** Proceed to Step 5 to classify strength.

### TYPE B     Direct Credit Effects

The speaker describes credit conditions CAUSING changes in real activity, without explicit
 feedback or amplification language.

**Indicators:**
- Credit conditions affecting spending, investment, or growth
- Balance sheet constraints limiting economic activity
- Lending standards impacting borrowing and spending decisions
- Financial stress weighing on the economy

**Examples:**
- "Tighter credit is slowing investment"
- "Weak balance sheets are restraining consumption"
- "Credit availability is affecting business spending"
- "Lending standards are holding back housing"
- "If net worth declines, consumers won't be able to maintain spending"
- "Balance sheet repair will take years, restraining growth"

**If Type B identified:** Classify as MODERATE.

### TYPE C     Denial of Credit Effects

The speaker indicates credit conditions are NOT meaningfully affecting economic activity.

**Indicators:**
- Disconnection language: "despite tight credit," "credit hasn't constrained," "minimal
 impact"
- Resilience emphasis: "strong balance sheets," "well-capitalized," "robust credit access"
- Other factors dominate: "driven by fundamentals not credit," "not a credit problem"
- Skepticism: "credit channel is weak," "limited transmission"

**Examples:**
- "Credit is not the constraint"
- "Despite tight credit, investment has remained robust"
- "Financial conditions haven't affected the real economy"
- "Spillovers have not materialized"
- "Firms have ample access to finance"

**If Type C identified:** Classify as WEAK.

### NOT Credit Channel     Classify as NULL

- Pure wealth effects without credit mechanism: "Lower asset prices reduce consumption
 through wealth effects"
- Pure interest rate transmission without credit frictions: "Lower rates stimulate demand"
- Monetary policy transmission without credit channel logic: "Easing will boost the economy"
- Description of conditions without causal interpretation

## Step 5: Classify Amplification Strength (Type A Only)
```

```
If the speaker describes amplification or feedback (Type A), classify the strength:

### STRONG Classification

Use when the speaker indicates SIGNIFICANT amplification through feedback mechanisms,
 without hedging about the mechanism's strength.

**Strong indicators:**
- Unhedged amplification language: "magnifies," "amplifies," "multiplies," "reinforces"
- Feedback/spiral language: "feedback loop," "self-reinforcing," "spiral," "cascade," "
 vicious cycle," "virtuous cycle"
- Explicit closed-loop chains: "tighter credit     weaker balance sheets     tighter credit"
- Historical validation: "as in 2008," "financial crisis showed"
- Concern about compounding effects: "will feed on itself," "mutually reinforcing"

**Examples:**
- "Credit conditions are amplifying the downturn"
- "We're seeing a self-reinforcing cycle of credit tightening and economic weakness"
- "Balance sheet effects are creating powerful feedback mechanisms"
- "The deterioration in financial conditions will magnify the impact on investment"
- "Falling asset prices tighten credit, which further depresses asset prices"

### MODERATE Classification

Use when the speaker describes amplification but with HEDGING about strength, scope, or
 certainty.

**Strong indicators:**
- Hedged amplification: "some amplification," "modest feedback," "may be reinforcing"
- Partial scope: "amplifying effects for some borrowers," "feedback in certain sectors"
- Weakening mechanism: "less amplification than past cycles," "muted feedback"
- Conditional: "could amplify if conditions worsen," "risk of feedback effects"

**Examples:**
- "Credit conditions are creating some amplification, but effects are modest"
- "The financial accelerator appears weaker than in past episodes"
- "There may be some self-reinforcing dynamics, but they appear contained"
- "Small businesses face amplifying credit constraints, though large firms have access"
- "We see limited feedback effects between financial and real sectors"

**Note on hedging language:** Classify based on the speaker's belief about the *mechanism's
 strength*, not their confidence about whether triggering conditions will materialize. "
 Credit conditions *could* amplify shocks significantly" is STRONG (believes in powerful
 mechanism, uncertain about trigger). "Credit conditions create *some* amplification" is
 MODERATE (believes mechanism is weak).

## Step 6: Handle Special Cases

### Forecasts and Projections
Only classify if the reasoning reveals credit channel logic:
- "I forecast slower growth"     NULL (no mechanism stated)
- "I forecast slower growth because credit tightening will slow investment"     MODERATE (
 direct effect)
- "I forecast slower growth because credit tightening will amplify the downturn"     STRONG
 (amplification)
```

```
### Historical References
Only classify if the speaker endorses or rejects the credit-activity relationship:
- "Credit was tight in 2008"     NULL (just description)
- "Credit tightening slowed the economy in 2008"     MODERATE (endorses direct effect)
- "The 2008 crisis showed how credit dynamics amplify shocks"     STRONG (endorses
 amplification)
- "Unlike 2008, credit conditions aren't affecting this downturn"     WEAK (rejects
 relationship)

### Policy Discussions
Only classify if the speaker explains the policy through credit channel logic:
- "We should ease policy"     NULL (no mechanism)
- "We should ease to support credit conditions and boost spending"     MODERATE (direct
 transmission)
- "We should ease to prevent credit tightening from amplifying the downturn"     STRONG (
 amplification)

### Wealth Effects
Wealth effects alone are NULL unless connected to credit mechanisms:
- "Falling home prices will reduce consumption through wealth effects"     NULL (wealth
 effect only)
- "Falling home prices will constrain home equity borrowing, reducing consumption"
 MODERATE (credit mechanism)
- "Falling home prices will tighten credit conditions, which will further depress home
 prices"     STRONG (feedback loop)

### Duration vs. Amplification
Prolonged direct effects are MODERATE; only feedback loops are STRONG:
- "Balance sheet repair will take years, restraining growth"     MODERATE (prolonged direct
 effect)
- "Balance sheet problems will create a self-reinforcing cycle of weak growth and tight
 credit"     STRONG (feedback)

# Output Format

Provide your classification in JSON format:

{
"credit_channel": "strong" | "weak" | "moderate" | null,
"reasoning": "Brief explanation in 1-2 sentences referencing key phrases from the quote"
}

# Classification Summary

| Relationship Type | Description | Classification |
|-------------------|-------------|----------------|
| Amplification (unhedged) | Explicit feedback/spiral/amplify language | STRONG |
| Amplification (hedged) | Feedback language with qualifications | MODERATE |
| Direct effect | Credit     activity, no feedback language | MODERATE |
| Denial | Credit doesn't matter / no effect | WEAK |
| No relationship | Missing components or no causal link | NULL |

**IMPORTANT REMINDERS:**
- Default to NULL unless the speaker expresses a clear belief about credit conditions
```

```
 affecting economic activity
- STRONG requires explicit amplification, feedback, or self-reinforcing language
- MODERATE captures both hedged amplification AND direct credit effects without feedback
 language
- WEAK captures denial or skepticism about credit-activity linkages
- Pure wealth effects and pure interest rate transmission (without credit frictions) are
 NULL
- Your reasoning must cite specific phrases from the quotation"""
```

IA.C.11.3 Prompts to Score Members' Arguments Based on Outlook

We score arguments for the analysis in Section 3.3 based on their economic outlook. We focus specifically on arguments the LLM identifies as forecasting the cyclical or trend component of our variables of interest. We then run an additional prompt to score each argument on the basis of the beliefs expressed about that variable's outlook. The prompt for inflation is given below.

```
"""You are an economic analyst tasked with classifying economic arguments about inflation
 using a standardized framework.

## Classification Framework

You must classify the quote across these four dimensions:

**1. Inflation Characterization** - How is current inflation described?
- **Low/Below Target**: Below the target inflation rate (includes deflationary pressures)
- **Moderate/Near Target**: At or near the target inflation rate
- **High/Above Target**: Above the target inflation rate
- **Mixed**: Different rates across measures, sectors, or components

**2. Inflation Outlook Indicator** - What inflation rate does the argument imply is expected
  for the future? Scale from -3 to +3

This dimension captures the speaker's forecast or outlook for future inflation relative to
 target, incorporating both their central expectation and their assessment of risks. Use the
  following scoring system:

-3: Expects significantly below-target inflation or deflation (or discusses substantial
 downside risks to inflation)
-2: Expects moderately below-target inflation (or discusses notable disinflationary risks)
-1: Expects slightly below-target inflation (or discusses some downside risks that could
 push inflation below target)
0: Expects inflation at target with balanced or minimal risks
+1: Expects slightly above-target inflation (or discusses some upside potential)
+2: Expects moderately above-target inflation (or discusses notable upside risks)
+3: Expects significantly above-target inflation (or discusses substantial upside risks/
 persistent high inflation)

Important considerations:
- Risk assessments should influence the score: discussion of upside risks should raise the
 score even if the central forecast is not explicitly above-target
```

```
- Arguments about current conditions may imply future outlook (e.g., "persistent inflation pressures suggest continued elevation")
- The speaker's emphasis on risks versus central tendencies should be reflected in the score
- References to "sticky" or "entrenched" inflation suggest higher scores; "transitory" or "temporary" suggest lower scores

**3. Risk Present** - Does the quote discuss inflation risks or vulnerabilities?
- **Yes**: Quote mentions risks, vulnerabilities, or potential scenarios regarding inflation
- **No**: Quote solely focuses on actual inflation conditions without risk language

**4. Scope** - What geographic or sectoral scope does the quote address? (Each dimension is independent Y/N)
- **National**: Does the quote discuss national/country-level aggregate inflation?
- **Regional**: Does the quote discuss sub-national geographic areas (states, cities, regions within a country)?
- **International**: Does the quote discuss cross-country comparisons or global inflation patterns?
- **Sectoral**: Does the quote discuss specific sectors, components, or categories (e.g., core vs. headline, goods vs. services, energy, food, housing, wages)?

Note: A quote may address multiple scopes simultaneously (e.g., national inflation with sectoral breakdowns). Mark "Y" for each scope that is substantively discussed in the quote.

## Instructions

First, conduct your analysis in ‘<analysis>‘ tags. Work through each classification dimension systematically, quoting specific phrases from the text and considering multiple options before deciding. Each section should be less than one sentence (partial sentences are acceptable).

- **Inflation Characterization**: Quote phrases describing current inflation levels/pressures and consider arguments for different classifications before deciding
- **Inflation Outlook Indicator**: Quote forward-looking phrases and explicit/implicit forecast language; assess momentum indicators, intensity modifiers (persistent, transitory, sticky, entrenched), and risk language (upside risks raise the score, downside risks lower it); map to the -3 to +3 scale considering both the expected inflation rate and the risk characterization
- **Risk Language Check**: Quote any phrases discussing vulnerabilities, potential scenarios, or uncertainties regarding inflation
- **Scope Analysis**: Quote phrases indicating national aggregate, regional, international, or sectoral focus and justify your assessment for each dimension

Then provide your classification using this exact XML format:
‘‘‘xml
<classification>
<inflation_characterization>[Low/Below Target|Moderate/Near Target|High/Above Target|Mixed]</inflation_characterization>
<inflation_outlook_indicator>[-3|-2|-1|0|1|2|3]</inflation_outlook_indicator>
<risk_present>[Yes|No]</risk_present>
<scope_national>[Y|N]</scope_national>
<scope_regional>[Y|N]</scope_regional>
<scope_international>[Y|N]</scope_international>
<scope_sectoral>[Y|N]</scope_sectoral>
</classification>
```

```
 Here is the economic argument you need to analyze:

<economic_argument>
{argument}
</economic_argument>"""
```

The prompt for growth is given below.

```
 """
You are an economic analyst tasked with classifying economic arguments about GDP growth
 using a standardized framework.

## Classification Framework

You must classify the quote across these four dimensions:

**1. Growth Characterization** - How is current growth described?
- **Low/Weak**: Below normal/trend growth rate (includes contraction/recession)
- **Medium/Stable**: Normal/trend growth rate
- **High/Strong**: Above normal/trend growth rate
- **Mixed**: Different rates across sectors or measures

**2. Growth Outlook Indicator** - What growth rate does the argument imply is expected for
 the future? Scale from -3 to +3

This dimension captures the speaker's forecast or outlook for future GDP growth relative to
 trend, incorporating both their central expectation and their assessment of risks. Use the
 following scoring system:

-3: Expects significantly below-trend growth or contraction (or discusses substantial
 downside risks to growth)
-2: Expects moderately below-trend growth (or discusses notable downside risks to growth)
-1: Expects slightly below-trend growth (or discusses some downside risks that could push
 growth below trend)
0: Expects trend-level growth with balanced or minimal risks
+1: Expects slightly above-trend growth (or discusses some upside potential)
+2: Expects moderately above-trend growth (or discusses notable upside potential)
+3: Expects significantly above-trend growth (or discusses substantial upside potential)

Important considerations:
- Risk assessments should influence the score: discussion of downside risks should lower the
  score even if the central forecast is not explicitly below-trend
- Arguments about current conditions may imply future outlook (e.g., "strong momentum
 suggests continued expansion")
- The speaker's emphasis on risks versus central tendencies should be reflected in the score

**3. Risk Present** - Does the quote discuss growth risks or vulnerabilities?
- **Yes**: Quote mentions risks, vulnerabilities, or potential scenarios regarding growth
- **No**: Quote solely focuses on actual growth conditions without risk language

**4. Scope** - What geographic or sectoral scope does the quote address? (Each dimension is
 independent Y/N)
- **National**: Does the quote discuss national/country-level aggregate GDP growth?
- **Regional**: Does the quote discuss sub-national geographic areas (states, cities,
 regions within a country)?
```

```
- **International**: Does the quote discuss cross-country comparisons or global growth
  patterns?
- **Sectoral**: Does the quote discuss specific sectors or industries (e.g., manufacturing,
  services, construction, consumer spending)?

Note: A quote may address multiple scopes simultaneously (e.g., national growth with
  sectoral breakdowns). Mark "Y" for each scope that is substantively discussed in the quote.

## Instructions

First, conduct your analysis in `<analysis>` tags. Work through each classification
  dimension systematically, quoting specific phrases from the text and considering multiple
  options before deciding. Each section should be less than one sentence (partial sentences
  are acceptable).

- **Growth Characterization**: Quote phrases describing current growth strength/weakness and
   consider arguments for different classifications before deciding
- **Growth Outlook Indicator**: Quote forward-looking phrases and explicit/implicit forecast
   language; assess momentum indicators, intensity modifiers, and risk language (downside
  risks lower the score, upside risks raise it); map to the -3 to +3 scale considering both
  the expected growth rate and the risk characterization
- **Risk Language Check**: Quote any phrases discussing vulnerabilities, potential scenarios
  , or uncertainties regarding growth
- **Scope Analysis**: Quote phrases indicating national aggregate, regional, international,
  or sectoral focus and justify your assessment for each dimension

Then provide your classification using this exact XML format:
```xml
<classification>
<growth_characterization>[Low/Weak|Medium/Stable|High/Strong|Mixed]</growth_characterization
  >
<growth_outlook_indicator>[-3|-2|-1|0|1|2|3]</growth_outlook_indicator>
<risk_present>[Yes|No]</risk_present>
<scope_national>[Y|N]</scope_national>
<scope_regional>[Y|N]</scope_regional>
<scope_international>[Y|N]</scope_international>
<scope_sectoral>[Y|N]</scope_sectoral>
</classification>```
Here is the economic argument you need to analyze:

<economic_argument>
{argument}
</economic_argument>"""
```

The prompt for unemployment is given below. Note: we multiply the scores by -1 after receiving the output from Claude, for consistency with inflation and growth in scoring arguments supporting restrictive policy having positive scores, and those supporting accommodative policy having negative scores.

```
"""
You are an economic analyst tasked with classifying economic arguments about unemployment
  using a standardized framework.
```
```

```
## Classification Framework

You must classify the quote across these four dimensions:

**1. Unemployment Characterization** - How is current unemployment described?
- **Low/Below Natural Rate**: Below the natural/target unemployment rate (includes references to tight labor markets)
- **Moderate/Near Natural Rate**: At or near the natural/target unemployment rate
- **High/Above Natural Rate**: Above the natural/target unemployment rate
- **Mixed**: Different rates across measures, demographics, or sectors

**2. Unemployment Outlook Indicator** - What unemployment trajectory does the argument imply is expected for the future? Scale from -3 to +3

This dimension captures the speaker's forecast or outlook for future unemployment relative to the natural rate, incorporating both their central expectation and their assessment of risks. Use the following scoring system:

-3: Expects significantly declining unemployment or substantial tightening of labor markets (or discusses substantial downside risks to unemployment)
-2: Expects moderately declining unemployment (or discusses notable risks of labor market tightening)
-1: Expects slightly declining unemployment (or discusses some downside risks)
0: Expects unemployment to remain at natural rate with balanced or minimal risks
+1: Expects slightly rising unemployment (or discusses some upside risks to unemployment)
+2: Expects moderately rising unemployment (or discusses notable risks of labor market weakening)
+3: Expects significantly rising unemployment or substantial labor market deterioration (or discusses substantial upside risks to unemployment)

Important considerations:
- Risk assessments should influence the score: discussion of upside risks (rising unemployment) should raise the score even if the central forecast is not explicitly above the natural rate
- Arguments about current conditions may imply future outlook (e.g., "persistent labor market weakness suggests continued elevation")
- The speaker's emphasis on risks versus central tendencies should be reflected in the score
- References to "structural" or "entrenched" unemployment suggest higher scores; "temporary" or "cyclical" suggest potential for lower scores

**3. Risk Present** - Does the quote discuss unemployment risks or vulnerabilities?
- **Yes**: Quote mentions risks, vulnerabilities, or potential scenarios regarding unemployment
- **No**: Quote solely focuses on actual unemployment conditions without risk language

**4. Scope** - What geographic or sectoral scope does the quote address? (Each dimension is independent Y/N)
- **National**: Does the quote discuss national/country-level aggregate unemployment?
- **Regional**: Does the quote discuss sub-national geographic areas (states, cities, regions within a country)?
- **International**: Does the quote discuss cross-country comparisons or global unemployment patterns?
- **Sectoral**: Does the quote discuss specific demographics, industries, or categories (e.g., by age, education, race, gender, industry sectors, full-time vs. part-time, U3 vs. U6 measures)?
```

````
Note: A quote may address multiple scopes simultaneously (e.g., national unemployment with
 demographic breakdowns). Mark "Y" for each scope that is substantively discussed in the
 quote.

## Instructions

First, conduct your analysis in `<analysis>` tags. Work through each classification
 dimension systematically, quoting specific phrases from the text and considering multiple
 options before deciding. Each section should be less than one sentence (partial sentences
 are acceptable).

- **Unemployment Characterization**: Quote phrases describing current unemployment levels/
 labor market conditions and consider arguments for different classifications before
 deciding
- **Unemployment Outlook Indicator**: Quote forward-looking phrases and explicit/implicit
 forecast language; assess momentum indicators, intensity modifiers (persistent, temporary,
 structural, cyclical), and risk language (upside risks to unemployment raise the score,
 downside risks lower it); map to the -3 to +3 scale considering both the expected
 unemployment trajectory and the risk characterization
- **Risk Language Check**: Quote any phrases discussing vulnerabilities, potential scenarios
 , or uncertainties regarding unemployment
- **Scope Analysis**: Quote phrases indicating national aggregate, regional, international,
 or sectoral/demographic focus and justify your assessment for each dimension

Then provide your classification using this exact XML format:
```xml
<classification>
<unemployment_characterization>[Low/Below Natural Rate|Moderate/Near Natural Rate|High/Above
  Natural Rate|Mixed]</unemployment_characterization>
<unemployment_outlook_indicator>[-3|-2|-1|0|1|2|3]</unemployment_outlook_indicator>
<risk_present>[Yes|No]</risk_present>
<scope_national>[Y|N]</scope_national>
<scope_regional>[Y|N]</scope_regional>
<scope_international>[Y|N]</scope_international>
<scope_sectoral>[Y|N]</scope_sectoral>
</classification>
```

Here is the economic argument you need to analyze:

<economic_argument>
{argument}
</economic_argument>"""
````

IA.C.11.4 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 five sets of 500 arguments, pass them into the LLM, and prompt it to construct categories based on the prompt below.

```
"""Here are the FOMC meeting quotations for your analysis:
```

```
<fomc_quotations>
    {quotations}
</fomc_quotations>

You are an expert economic analyst tasked with categorizing the types of data sources cited
 in Federal Open Market Committee (FOMC) meetings.

<data_source_definition>
A "data source" is the underlying source of information that was either:
(a) Explicitly cited by the speaker in their argument (e.g., "According to the latest GDP
 report..." or "Our district surveys show..."), OR
(b) Implicitly required for the speaker to make their claim (e.g., if a speaker says "
 inflation has accelerated," they must be referencing some measure of inflation, even if not
  explicitly named)

Data sources can be:
- **Public data**: Information available to all committee members and the general public (e.
 g., BLS employment reports, publicly released GDP figures, market indices)
- **Specialized data**: Information specific to a particular speaker, their Federal Reserve
 district, or their unique institutional access (e.g., "contacts in my district report...",
 internal Fed staff forecasts, proprietary surveys from a specific Reserve Bank)
</data_source_definition>

Your task is to analyze the FOMC quotations and create a list of 5-15 categories that
 capture the types of data sources cited by the speakers. Your categories should help us
 eventually classify whether each data source is "specialized" or "public."

Follow these steps:

1. Review the FOMC quotations thoroughly.
2. Identify explicit and implicit data sources mentioned in the quotations.
3. For each data source, note whether it appears to be public or specialized information.
4. Group similar data sources together.
5. Create initial categories based on these groups.
6. Refine and consolidate categories to reach the 5-15 range.
7. Ensure all categories are distinct and non-overlapping.
8. Cross-check the final list against the original quotations.

Before providing your final list of categories, conduct a detailed analysis of the data
 sources. Wrap your analysis in <data_source_analysis> tags to show your thought process. In
  your analysis:

a. Quote relevant passages from the FOMC quotations that mention or imply data sources.
b. List explicit data sources (with the actual information source clearly identified).
c. List inferred implicit data sources (what information the speaker must have been
 referencing).
d. For each source, note whether it appears to be public or specialized.
e. Group similar sources together.
f. Create preliminary categories and provide a brief argument for each category's inclusion.
g. Refine and consolidate categories.
h. Verify that categories are mutually exclusive and do not overlap.

It's OK for this section to be quite long and thorough.
```

```
Important: Ensure that each piece of data belongs to only one category, even if an argument
 might cite multiple pieces of data from different categories.

After your analysis, present your final list of 5-15 categories in the following format:

<category_list>
<category>
<name>[Category Name]</name>
<description>[Brief description of the data source type, including examples if applicable.
 Note whether sources in this category are typically public, specialized, or mixed.]</
 description>
</category>
...
</category_list>

Here's a generic example of how your output should be structured:

<category_list>
<category>
<name>Official Labor Market Statistics</name>
<description>Government-published employment and labor force data, such as unemployment
 rates, payroll numbers, and labor force participation rates from agencies like the Bureau
 of Labor Statistics. These sources are typically public data available to all committee
 members and the broader public.</description>
</category>
<category>
<name>District Business Intelligence</name>
<description>Information gathered from business contacts, surveys, and discussions within a
 specific Federal Reserve district. Examples include anecdotal reports from local firms,
 regional economic conditions, or district-specific surveys. These sources are typically
 specialized, as they represent information unique to a particular Reserve Bank president's
 district.</description>
</category>
</category_list>

Remember, your final list must contain between 5 and 15 distinct, non-overlapping categories
  that comprehensively capture the data sources mentioned in the FOMC quotations."""
```

We then prompt the LLM to consolidate the data category names and descriptions across the five runs into one combined list. We use the following prompt:

```
"""
You are an expert economic analyst tasked with consolidating multiple categorization schemes
  of data sources cited in Federal Open Market Committee (FOMC) meetings into a single,
 comprehensive taxonomy.

<data_source_definition>
A "data source" is the underlying source of information that was either:
(a) Explicitly cited by the speaker in their argument (e.g., "According to the latest GDP
 report..." or "Our district surveys show..."), OR
(b) Implicitly required for the speaker to make their claim (e.g., if a speaker says "
 inflation has accelerated," they must be referencing some measure of inflation, even if not
  explicitly named)

Data sources can be:
```

```
- **Public data**: Information available to all committee members and the general public (e.g., BLS employment reports, publicly released GDP figures, market indices)
- **Specialized data**: Information specific to a particular speaker, their Federal Reserve district, or their unique institutional access (e.g., "contacts in my district report...", internal Fed staff forecasts, proprietary surveys from a specific Reserve Bank)
</data_source_definition>

You have been provided with multiple category lists below, each produced from analyzing different sets of FOMC quotations. Your task is to consolidate these into a single, unified taxonomy.

<category_lists>
{category_lists}
</category_lists>

<consolidation_requirements>
1. **Eliminate redundancy**: Identify categories across different lists that describe the same type of data source and merge them into a single category.

2. **Favor breadth over specificity**: Create broad categories that encompass multiple types of variables or data points, rather than variable-specific categories. For example, prefer "National Economic Indicators" over separate categories for "GDP Data," "Inflation Data," and "Industrial Production Data."

3. **Ensure mutual exclusivity**: Each data source should belong to only one category. Categories must not overlap in their scope.

4. **Maintain comprehensiveness**: The final taxonomy should capture all types of data sources present across all provided category lists. No data source type should be left uncategorized.

5. **Preserve the public/specialized distinction**: Each category description should clearly indicate whether sources in that category are typically public, specialized, or mixed.

6. **Target range**: Aim for 8-15 categories in your final consolidated taxonomy. This range provides sufficient granularity while maintaining manageability.

7. **Maintain consistency**: Use clear, professional terminology consistent with Federal Reserve and economic analysis conventions.
</consolidation_requirements>

Follow these steps:

1. Review all provided category lists thoroughly.
2. Identify categories that are duplicates or near-duplicates across different lists.
3. Identify categories that are overly specific and could be combined into broader categories.
4. Group related categories together.
5. Create consolidated categories that subsume the grouped categories.
6. Verify that the consolidated categories are mutually exclusive.
7. Ensure all original data source types are represented in the final taxonomy.
8. Cross-check the final list for comprehensiveness and non-redundancy.

Before providing your final consolidated taxonomy, conduct a detailed consolidation analysis. Wrap your analysis in <consolidation_analysis> tags to show your thought process. In your
```

```
  analysis:

a. List all categories from the provided lists, noting which list each came from.
b. Identify clear duplicates or near-duplicates and explain why they should be merged.
c. Identify overly specific categories and propose broader categories that could encompass
 them.
d. For each proposed consolidated category, list which original categories it subsumes and
 provide justification.
e. Verify mutual exclusivity by checking potential boundary cases between categories.
f. Confirm that no data source types from the original lists are left uncategorized.

It is acceptable and expected for this section to be thorough and detailed.

After your analysis, present your final consolidated taxonomy in the following format:

<consolidated_taxonomy>
<category>
<name>[Category Name]</name>
<description>[Comprehensive description of the data source type, including the range of
 examples it encompasses. Clearly note whether sources in this category are typically public
 , specialized, or mixed. If this category consolidates multiple original categories, the
 description should reflect the full breadth of data sources now included.]</description>
</category>
...
</consolidated_taxonomy>

Example structure:

<consolidated_taxonomy>
<category>
<name>Official Government Economic Statistics</name>
<description>Formally published statistical data from government agencies, including
 employment figures, national accounts, price indices, trade statistics, and production
 metrics. Examples include Bureau of Labor Statistics reports, Bureau of Economic Analysis
 releases, Census Bureau data, and Federal Reserve statistical releases. These sources are
 public data available to all committee members and the general public.</description>
</category>
<category>
<name>Regional Business Intelligence</name>
<description>Information gathered through direct contact with businesses, surveys, and
 anecdotal reports within Federal Reserve districts. This includes district business
 contacts' assessments, regional economic conditions, local labor market insights, and
 district-specific surveys such as Beige Book contributions. These sources are specialized
 data, as they represent information unique to individual Reserve Bank districts and their
 business networks.</description>
</category>
</consolidated_taxonomy>

Your final consolidated taxonomy must contain between 8 and 15 distinct, mutually exclusive
 categories that comprehensively capture all data source types present in the provided
 category lists."""
```

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.

<data_citation_criteria>
A quotation "cites data" if the speaker's argument relies on or references empirical
 information, observations, or measurements. This includes:

**DOES cite data:**
- Explicit references to statistics, reports, or indicators (e.g., "GDP growth was 2.5%", "
 the unemployment rate fell")
- Implicit references to measured phenomena (e.g., "inflation has accelerated" implies
 reference to inflation data)
- Anecdotal evidence from specific sources (e.g., "contacts in my district report strong
 demand")
- References to trends, patterns, or conditions that require observational or measured
 evidence (e.g., "labor markets remain tight")
- References to forecasts or projections based on data (e.g., "staff projects growth will
 slow")
- Descriptions of current or past economic states that require empirical verification (e.g.,
  "credit conditions have tightened")

**DOES NOT cite data:**
- Pure policy preferences or recommendations without empirical support (e.g., "I believe we
 should raise rates")
- Theoretical or conceptual arguments not grounded in specific observations (e.g., "higher
 rates would reduce demand")
- Procedural or process-oriented statements (e.g., "I agree with the approach outlined")
- Hypothetical scenarios without reference to actual conditions (e.g., "if inflation were to
  rise, we should act")
- General principles or rules without empirical backing (e.g., "central banks must maintain
 credibility")
- Personal judgments about policy stance that don't reference observable conditions (e.g., "
 I support this decision")

If uncertain, ask: "What empirical information would the speaker need to know to make this
 statement?" If the answer is specific observable data, the quotation cites data.
</data_citation_criteria>

<data_source_definition>
A "data source" is the underlying source of information that was either:
(a) Explicitly cited by the speaker in their argument, OR
(b) Implicitly required for the speaker to make their claim

Important: Assign only ONE data category if the quotation cites only one piece of data (
 explicitly or implicitly). Assign MULTIPLE data categories only when the speaker cites
 multiple distinct pieces of data that fall under different categories.

If the quotation cites data but none of the provided categories adequately capture the data
 source, select "Other."
```

```
If the quotation does not cite any data according to the criteria above, select "No data cited."
</data_source_definition>

Instructions:
1. Determine whether the quotation cites data according to the criteria above.
2. If data is cited, identify the data source(s) used - both explicitly cited and implicitly referenced.
3. Categorize each distinct data source using the numbered categories, "Other," or "No data cited."
4. If data was cited, determine if the information is private/specialized or public based on the definitions below.

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

<classification_definitions>
**Public Information**: Data and information available to all FOMC members, regardless of whether it is available to the general public. This includes:
- Publicly released government statistics (BLS reports, GDP figures, etc.)
- Market data and financial indicators available to all
- Tealbook/Greenbook staff forecasts and projections
- Any data presented to the full committee
- Standard interpretations or analysis of publicly available data

**Private/Specialized Information**: Data and information specific to a particular speaker that other FOMC members could not reasonably be expected to know. This includes:
- District-specific business intelligence from a speaker's Federal Reserve district
- Anecdotal reports from contacts in a specific region
- District-level surveys or research unique to one Reserve Bank
- Any information that reflects a speaker's unique institutional access or regional perspective

Note: If "No data cited" is selected, information classification is not applicable.
</classification_definitions>

Before providing your final output, use <thought_process> tags to break down your analysis. In your thought process:
- First determine: Does this quotation cite data according to the criteria?
- If yes: Identify what data source(s) are being referenced (explicitly or implicitly)
- Determine which category/categories apply (or if "Other" is appropriate)
- If data was cited: Assess whether other FOMC members would have access to this information

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

<output>
<data_categories>[Single category number OR comma-separated list if multiple distinct data sources OR "Other" OR "No data cited"]</data_categories>
<information_classification>
    <type>[Private/Specialized Information OR Public Information OR Not Applicable]</type>
</information_classification>
</output>

Here are examples of the desired output structure:

Example 1 (Multiple categories with public data):
```

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

Example 2 (No data cited):
<output>
<data_categories>No data cited</data_categories>
<information_classification>
    <type>Not Applicable</type>
</information_classification>
</output>

Example 3 (Other category):
<output>
<data_categories>Other</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.

Perform your classification for the following FOMC discussion quotation:

<fomc_quotation>
{quotation}
</fomc_quotation>

The numbered categories for data classification are:

<data_categories>
{categories}
</data_categories>


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

IA.C.11.5 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, focusing on causal
  relationships involving a specific economic variable of interest.

First, read the following FOMC meeting quotations:
'''
<fomc_quotations>
{quotations}
</fomc_quotations>
'''

The variable of interest for this analysis is:
'''
<variable_of_interest>
{variable}
</variable_of_interest>
'''

## TASK
Analyze these quotations and create categories that collectively cover all themes present in
  the quotes. Each category should represent a distinct economic argument or concept
 directly relevant to the variable of interest, emphasizing causal relationships.

Conduct your analysis within '<category_development_process>' tags, following these steps:

1. Identify key phrases, economic terms, and main concepts across all quotes. Number each
 item and provide a supporting quote.

2. Identify causal relationships, particularly relating to the variable of interest. Number
 each relationship and explicitly state its connection to the variable.

3. Tally common themes and topics across all quotes. List these with their frequency.

4. Brainstorm potential categories based on your analysis. Number each category and provide
 a brief rationale for its inclusion, considering that categories should accommodate
 differing interpretations.

5. For each proposed category, explicitly state how it relates to the variable of interest.
 Revise or discard categories without direct relevance.

6. Refine categories to ensure they meet these criteria (explicitly state how each category
 meets each):
- Mutually exclusive and non-overlapping
- Support within-category disagreement
- At the "lowest level" that still supports within-category disagreement
- Consistent level of generality/specificity across all categories

7. Consolidate categories capturing the same ideas. Explain your reasoning for each
 consolidation.

8. Check for overlap between categories using a simple comparison matrix. Adjust as
 necessary.

9. For each category, create a one-sentence description highlighting the causal relationship
```

````
  it represents and the nature of disagreement within it.

10. Rank final categories by relevance to the variable of interest with brief explanations.

## FINAL OUTPUT FORMAT

Provide your final categories using this XML format:

```xml
<categories>
<category>
    <name>Category Name</name>
    <description>Brief one-sentence description highlighting the causal relationship and the
      nature of disagreement within the category.</description>
</category>
<!-- Repeat for all categories -->
</categories>
```

## EXAMPLE
```xml
<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>
</category>
```

## REQUIREMENTS
- Categories must be comprehensive, mutually exclusive, and non-overlapping
- Each category represents a distinct economic argument or concept directly relevant to the
 variable of interest
- Descriptions are concise (one sentence) but informative, emphasizing causal relationships
 and potential disagreements
- Categories accommodate within-category disagreement
- Categories reflect intuitive and theoretical economic relationships
- Number of categories determined by quote content, not predetermined limits

Begin your analysis 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 and the nature of disagreement within the category.

7. 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>
</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>
</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.

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 tasked with classifying a quotation into one of several predefined topic categories.

Here are the available categories to choose from:

<categories>
{categories}
</categories>

Here is the context for what this classification relates to:

<variable_description>
{variable}
</variable_description>

Here is the quotation you need to classify:

<quotation>
{quotation}
</quotation>

Instructions:
1. Carefully read the quotation and consider its main themes, topics, and content.
2. Review each of the available categories and consider how well each one matches the
 quotation.
3. Select the single category number that best fits the quotation's content.
4. First, provide your reasoning in <analysis> tags in two sentences or less, identifying
 the key theme and which category best matches.
5. Then output your final answer as the category number in XML format.

Your final answer should be formatted as follows:
```

```
‘‘‘xml
<category>X</category>"""
```

IA.C.11.6 Prompt to Identify Policy Decisions Made

To identify the policy decisions made, we input the minutes from the meeting, the speeches from the policy discussion section of each FOMC meeting, and the policy alternatives presented in the Bluebook/Tealbook B. The prompt to identify these decisions is below.

```
"""You are an expert economic analyst specializing in Federal Reserve policy. Your task is
 to analyze Federal Open Market Committee (FOMC) meeting materials, identify all
 successfully adopted monetary policy decisions, and assess each decision’s stance on the
 monetary accommodation-restriction spectrum.

You will work with three documents from a single FOMC meeting:

<adopted_policy_document>
{minutes}
</adopted_policy_document>

<policy_alternatives>
{alternatives}
</policy_alternatives>

<fomc_speeches>
{speeches}
</fomc_speeches>

## Critical Definitions

**Monetary policy decision**: A decision specifically related to monetary policy tools and
 communication, including:
- Interest rate targets or ranges
- Asset purchase programs (QE, balance sheet operations)
- Forward guidance about future policy path
- Reserve requirement changes
- Discount window or other lending facility terms
- Public statements or communications about monetary policy stance

**NOT monetary policy decisions**: Exclude all of the following:
- Personnel appointments or organizational matters
- Administrative procedures or committee operations
- Financial reporting or audit matters
- Any decision not directly related to monetary policy tools or communication

**Adopted decision**: A monetary policy decision that either:
1. Appears explicitly in the adopted policy document, OR
2. Is clearly described as having been adopted in the speeches

**Accommodation scoring scale**:
```

```
- -3: Strongly accommodative (provides strong monetary stimulus or support)
- -2: Moderately accommodative
- -1: Slightly accommodative
- 0: Neutral (neither accommodative nor restrictive)
- +1: Slightly restrictive
- +2: Moderately restrictive
- +3: Strongly restrictive (provides strong monetary tightening)

## Your Analysis Process

Conduct your complete analysis inside ‘<policy_analysis>‘ tags. Work through the following
 four steps systematically. It’s OK for this section to be quite long, as thorough analysis
 is essential.

**Step 1: Extract adopted decisions from the adopted policy document**

Review the adopted policy document carefully. For each potential decision you encounter:
1. Quote the relevant text verbatim
2. Ask: Is this a monetary policy decision? (Does it relate to interest rates, asset
 purchases, forward guidance, or other monetary policy tools/communication?)
3. If yes, add it to your numbered list with the quote
4. If no, explicitly note that you’re skipping it and explain why it’s not a monetary policy
  decision

Be systematic and thorough. This document is your primary source for adopted decisions. It’s
  OK for this step to be quite long.

**Step 2: Check speeches for additional adopted decisions**

Review the FOMC speeches to find any monetary policy decisions that are clearly described as
  having been adopted but may not appear in the adopted policy document.

For each potential decision:
1. Quote the relevant text from the speech
2. Ask: Is this a monetary policy decision?
3. If yes, ask: Is it clearly stated as adopted?
4. If both yes, add it to your numbered list with the quote
5. If no to either question, skip it and note why

**Step 3: Check for duplicate decisions**

Review your complete list of adopted monetary policy decisions. Create a systematic
 comparison by checking each decision against all others.

For each pair of decisions, ask: Do these describe the same underlying monetary policy
 action, even if worded differently?

Create a duplicate check table with these columns:
- Decision number and description
- Is this a duplicate? (Yes/No - if Yes, specify which other decision it duplicates)

After identifying duplicates, consolidate them into single entries with unified descriptions
 . List your final set of unique adopted monetary policy decisions with their consolidated
 text.
```

```
**Step 4: Analyze each unique adopted decision**

For each unique adopted monetary policy decision, complete the following analysis:

a) **Decision type**: Classify as one of:
- "communication" (forward guidance, statement language, communication about policy)
- "rate decision" (changes to interest rate targets or ranges)
- "other" (asset purchases, balance sheet operations, other monetary policy tools)

b) **Accommodation score analysis**:
- List all accommodative arguments: What aspects of this decision support monetary stimulus
 or easier policy? For each argument, quote the specific language from the decision that
 supports it. List each argument on a separate line.
- List all restrictive arguments: What aspects of this decision support monetary tightening
 or restrictive policy? For each argument, quote the specific language from the decision
 that supports it. List each argument on a separate line.
- Weigh these arguments against each other: Which side is stronger? By how much?
- Assign a score from -3 (strongly accommodative) to +3 (strongly restrictive) based on your
  weighing

c) **Justification**: Write 1-2 sentences explaining your score in terms of the decision's
 accommodation or restrictiveness

d) **Alternatives**: Review the policy alternatives document and note which alternative(s) (
 e.g., "Alternative A", "Alternative B", etc.) include this decision. Write "None" if the
 decision doesn't appear in any alternative document.

## Output Format

After completing your analysis in `<policy_analysis>` tags, provide your final results using
  the following XML structure:

<adopted_policies>
<adopted_policy>
    <description>
    [Clear, concise description of the adopted monetary policy decision]
    </description>
    <type>
    [Exactly one of: communication, rate decision, or other]
    </type>
    <score>
    [Single integer from -3 to +3]
    </score>
    <justification>
    [1-2 sentences explaining the score in terms of accommodation or restrictiveness]
    </justification>
    <alternatives>
    [List of alternatives containing this decision (e.g., "Alternative A, Alternative B") or
      "None"]
    </alternatives>
</adopted_policy>
<!-- Repeat the above structure for each unique adopted monetary policy decision -->
</adopted_policies>

Begin your analysis now. Remember to focus exclusively on monetary policy decisions and
```

```
 exclude any administrative, personnel, or non-monetary policy matters.
"""
```

IA.C.11.7 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 evaluating Federal Open Market Committee (FOMC) policy
 alignment. Your task is to assess how closely a specific speaker's stated policy
 preferences align with the final policy decisions adopted in an FOMC meeting.

You will analyze three pieces of information:

<speeches>
{speeches}
</speeches>

<policy_alternatives>
{alternatives}
</policy_alternatives>

<final_policy_decisions>
{decisions}
</final_policy_decisions>

# Task Overview

Analyze the focus speaker's statements and determine how strongly their policy preferences
 align with each final policy decision. Provide individual alignment scores for each
 decision and an overall alignment rating.

# Understanding Evidence Types

When assessing alignment, consider ALL THREE types of evidence:

**1. Direct statements**: Explicit comments about the specific decision itself

**2. Indirect evidence via alternatives**: If the speaker supports/opposes a policy
 alternative, this provides evidence about ALL decisions within that alternative, even if
 never explicitly discussed

**3. Stance alignment**: If the speaker expresses strong views about a policy stance or
 direction that directly relates to a decision, this provides evidence of alignment even
 without discussing the specific decision

# Alignment Rating Scale

Use this scale for both individual decisions and overall assessment:
```

```
- **3 (Full Agreement)**: Explicitly supports all key elements with no reservations
- **2 (Strong Agreement)**: Explicitly supports primary elements, minor reservations on
 secondary aspects
- **1 (Moderate Agreement)**: Supports the direction but has significant reservations or
 supports only some key elements
- **0 (Neutral)**: No clear statement, OR balanced support and opposition
- **-1 (Moderate Opposition)**: Opposes the direction but acknowledges some merit or
 suggests modifications
- **-2 (Strong Opposition)**: Explicitly opposes primary elements, though may concede minor
 points
- **-3 (Full Opposition)**: Explicitly opposes all key elements with no areas of agreement

# Scoring Rules

## When Speaker Explicitly Discusses the Decision
- Explicit support     **3**
- Explicit opposition     **-3**
- Partial support/opposition     Use full scale based on degree of agreement

## When Speaker Supports Alternative Containing the Decision
- Supports alternative + no concerns about decision     **3**
- Supports alternative + minor concerns about decision     **2**
- Supports alternative + significant concerns about decision     **1** or lower

## When Speaker Opposes Alternative Containing the Decision
- Opposes alternative + decision is central component     **-3**
- Opposes alternative + decision is peripheral component     **-2**
- Opposes alternative + expresses some support for decision     **-1** or higher

## When Speaker Expresses Stance Disagreement Without Discussing Decision Or Alternative
- Decision clearly embodies stance speaker opposes     **-3**
- Decision relates to but doesn't fully embody opposed stance     **-2**
- Relationship between opposed stance and decision is less direct     **-1**
- (Same logic applies in reverse for stance agreement)

## Combining Evidence
- Prioritize explicit statements about the specific decision
- Use indirect evidence (alternatives) or stance alignment as confirmatory evidence or when
 no direct statement exists
- If evidence conflicts, prioritize: Explicit > Indirect > Stance

# Analysis Instructions

Before providing your final output, conduct a thorough analysis in <analysis> tags. Work
 through these five steps systematically. It's OK for this section to be quite long,
 especially if there are many decisions to analyze.

**Step 1: Map speaker's positions on alternatives**
- Review all statements by the focus speaker
- List each alternative one by one (Alternative A, Alternative B, etc.)
- For each alternative, state whether the speaker supports it, opposes it, or has no clear
 position
- Quote specific passages revealing each position with sufficient context
```

```
**Step 2: Analyze each decision individually**

Work through each decision systematically, one at a time. For each decision:

a. Quote the decision verbatim from final_policy_decisions

b. Check for explicit statements: Quote any passages where the speaker directly discusses
 this decision. If none exist, state this clearly

c. Check for stance alignment: Identify broader statements about policy stances or
 directions that relate to this decision. Quote if they exist

d. Check for indirect evidence: List alternatives containing this decision and note whether
 the speaker supported/opposed them (reference Step 1)

e. Apply scoring rules: State which rule(s) you're applying and why based on the evidence

f. Assign alignment score (-3 to 3)

g. Write justification including: reasoning, supporting quotes, stance positions,
 alternative references, and which scoring rules you applied

**Step 3: Calculate numerical average**
- List each decision score on a separate line
- Show calculation explicitly (e.g., "(3 + 2 + 1 + (-1)) / 4 = 1.25")

**Step 4: Consider each possible overall rating**
- List out each rating value from -3 to 3
- For each rating value, assess whether it would be appropriate given:
- The numerical average
- Pattern of individual scores (clustered or varied?)
- Important areas of agreement/disagreement
- Whether certain decisions are more central/important
- Note which ratings are plausible candidates and which can be ruled out

**Step 5: Determine final overall rating**
- Choose the rating (-3 to 3) that best represents aggregate alignment
- Consider both numerical average and patterns in individual scores
- Overall rating should generally be close to numerical average but may adjust for
 meaningful patterns
- Explain why you chose this rating over nearby alternatives

# Output Format

After completing your analysis, provide output in this exact structure:

‘‘‘
<rating>[Single number from -3 to 3]</rating>

<proposed_decisions>
<decision>
<index>[Decision number]</index>
<description>[Brief description of the policy decision]</description>
<alignment>[Single number from -3 to 3]</alignment>
<justification>[Explanation including relevant quotes, alternative/stance references, and
```

```
 scoring rules applied]</justification>
</decision>
[Repeat for each policy decision]
</proposed_decisions>

<speaker_alignment>[2-3 sentences summarizing overall alignment, highlighting key areas of
 agreement and disagreement]</speaker_alignment>
‘‘‘

## Example Output Structure

‘‘‘
<rating>2</rating>

<proposed_decisions>
<decision>
<index>1</index>
<description>Decision to adjust the target rate by X basis points</description>
<alignment>3</alignment>
<justification>The speaker explicitly supported this rate adjustment, stating "[generic
 quote]". This direct statement with no reservations warrants a score of 3 under the
 explicit support rule.</justification>
</decision>
<decision>
<index>2</index>
<description>Decision regarding forward guidance language</description>
<alignment>1</alignment>
<justification>The speaker supported Alternative B which contained this decision, but
 expressed concerns about "[generic quote]". While supportive of the alternative overall,
 significant reservations about this element warrant a score of 1 under the indirect support
  with significant concerns rule.</justification>
</decision>
<decision>
<index>3</index>
<description>Decision on asset purchase program</description>
<alignment>-2</alignment>
<justification>The speaker did not discuss this specific decision but expressed strong
 opposition to the policy stance of "[generic stance]", which this decision embodies. This
 warrants a score of -2 under the stance disagreement rule.</justification>
</decision>
</proposed_decisions>

<speaker_alignment>The speaker showed strong overall alignment with the adopted policies,
 with an average score of 2.0. They particularly agreed with monetary policy rate decisions
 but had reservations about communication strategy. The main area of disagreement was with
 asset purchase decisions, where their stated preference conflicted with the adopted
 approach.</speaker_alignment>
‘‘‘

Here is the speaker you will focus your analysis on:

<focus_speaker>
{stablespeaker}
</focus_speaker>
```

```
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 specializing in Federal Open Market Committee (FOMC) deliberations
 . Your task is to assess how much influence a specific speaker exerted over the final
 policy decisions adopted in an FOMC meeting.

# Input Materials

Here are the policy decisions that were adopted in this meeting:

<policy_decisions>
{{decisions}}
</policy_decisions>

Here are the policy alternatives that were discussed during the meeting:

<alternatives>
{{alternatives}}
</alternatives>

Here are the speeches from the policy discussion section of the FOMC meeting, in
 chronological order:

<speeches>
{{speeches}}
</speeches>

Here is the specific speaker you will focus your analysis on:

<focus_speaker>
{{stablespeaker}}
</focus_speaker>

# Core Concept: Influence on Adopted Decisions

**Influence measures contribution to the adopted outcome.**

The key question is: Did this speaker help make this decision happen?

- A speaker who argued FOR a decision and provided compelling reasoning     HIGH influence
- A speaker who argued AGAINST a decision (even with excellent arguments)     ZERO influence
  on its adoption
- A speaker who didn't engage with a decision     ZERO influence

**Critical Rule**: Opposition does not count as influence, regardless of how substantive the
  arguments. If the decision was adopted despite the speaker's opposition, they did not
 influence its adoption     they failed to prevent it.

# Influence Scale (0-3)
```

```
| Score | Label | Core Question |
|-------|-------|---------------|
| **3** | Pivotal | Would this decision likely have been different or not happened without
 this speaker? |
| **2** | Contributory | Did the speaker add meaningful substance IN SUPPORT of a decision
 others were already advancing? |
| **1** | Supportive | Did the speaker explicitly endorse this decision without adding
 substance? |
| **0** | Absent/Opposed | Did the speaker not engage with OR oppose this decision? |

## Score 3 (Pivotal)

The speaker was essential to this decision being adopted. Evidence includes:
- **Originated the proposal**: First to propose this specific decision
- **Central persuader**: Provided the key argument that convinced others
- **Cited as source**: Other speakers explicitly referenced them as the authority or
 originator
- **Changed trajectory**: Shifted discussion toward this outcome

Score 3 requires counterfactual significance: the decision would likely have been different
 or not happened without this speaker.

## Score 2 (Contributory)

The speaker supported a decision that was already on the table AND added meaningful
 substance. Evidence includes:
- **Provided reasoning**: Gave substantive arguments, data, or analysis IN SUPPORT
- **Proposed language**: Suggested specific wording, numbers, or text that was incorporated
- **Defended against objections**: Responded to criticism raised by others
- **Offered refinements**: Suggested modifications that improved or shaped the final
 decision

Score 2 requires adding something beyond agreement     but the decision would likely have
 happened without them.

## Score 1 (Supportive)

The speaker explicitly endorsed this decision but added nothing substantive. Evidence
 includes:
- **Direct endorsement**: "I support this decision" without explanation
- **Alternative endorsement**: "I support Alternative B" (which contains this decision)
 without discussing it specifically

Score 1 indicates explicit support but no unique contribution.

## Score 0 (Absent/Opposed)

The speaker did not contribute to this decision's adoption. This includes:
- **Never mentioned**: Did not discuss this decision at all
- **Opposed**: Argued against this decision (regardless of argument quality)
- **Contradicted**: Took positions inconsistent with this decision

**Important**: A speaker who made excellent arguments AGAINST a decision that was adopted
 scores 0. Their opposition, however well-reasoned, did not influence the adoption     the
```

```
 decision happened despite them.

# Analysis Process

Work through your analysis in <analysis> tags before providing output.

## Step 1: List All Decisions

Count and list each decision in <policy_decisions>. You must score every one.

## Step 2: Determine Speaker's Stance on Each Decision

For EACH decision, determine whether the focus speaker:
- **Supported**: Explicitly endorsed (directly or via alternative endorsement)
- **Opposed**: Explicitly argued against
- **Absent**: Did not address

This is the critical first step. Quote the evidence for your stance determination.

**Rules for stance determination:**
- "I support Alternative B"     Supported all decisions contained in Alternative B (unless
 exception stated)
- "I oppose this language" / "I have concerns about X" / "I prefer we not do Y"     Opposed
- No mention of the decision or relevant alternative     Absent

## Step 3: Score Each Decision

For each decision, apply the scoring hierarchy IN ORDER. Stop at the first level that
 applies.

### If Opposed or Absent     Score 0
If you determined the speaker opposed or did not address this decision, assign Score 0. Do
 not assess contribution components.

### If Supported     Apply Hierarchy

**Test for Score 3 (Pivotal):**
Ask: Would this decision likely have been different or not happened without this speaker?
- Were they the first to propose it?
- Were they cited by others as the source or authority?
- Did they provide the central argument that persuaded the committee?

If YES to any     Score 3. If NO to all     Continue.

**Test for Score 2 (Contributory):**
Ask: Did the speaker add meaningful substance in support?
- Did they provide reasoning, data, or analysis (beyond mere preference)?
- Did they propose specific language that was adopted?
- Did they defend against objections or offer refinements?

If YES to any     Score 2. If NO to all     Continue.

**Test for Score 1 (Supportive):**
Ask: Did the speaker explicitly endorse without adding substance?
- Did they state support without explanation?
```

```
- Did they endorse an alternative containing this decision without discussing it?

If YES     Score 1.

### Justify Your Score

For each score, you must:
1. Quote the specific evidence from the focus speaker
2. Explain why this score level applies
3. **Explain why the next higher score does NOT apply** (unless scoring 3)

## Step 4: Assess Contribution Components (For Supported Decisions Only)

For decisions where the speaker scored 1, 2, or 3, assess these components:

| Component | Question | Only if Supporting |
|-----------|----------|-------------------|
| ‘first_mover‘ | Were they first to propose this decision? | Yes |
| ‘provided_reasoning‘ | Did they give arguments/data IN SUPPORT? | Yes |
| ‘cited_by_others‘ | Were they referenced by other speakers? | Yes |
| ‘proposed_language‘ | Did they suggest specific text that was used? | Yes |

**If the speaker opposed or was absent (Score 0), all components = no.**

Also assess:
- ‘discussion_extent‘: How much did they discuss this decision? (none/brief/moderate/
 extensive)
  - This applies regardless of stance     even opposition can be brief or extensive

## Step 5: Calculate Summary Measures

- ‘influence_components‘: Sum of the four component flags (0-4). Only meaningful when
 influence > 0.
- Note the distribution of scores across decisions.

# Output Format

‘‘‘xml
<influence_analysis>
  <meeting>{{ymd}}</meeting>
  <speaker>{{stablespeaker}}</speaker>
  <decisions>
    <decision>
      <id>[Decision number]</id>
      <description>[Brief description]</description>

      <!-- Stance determination (MUST come first) -->
      <stance>[supported, opposed, or absent]</stance>

      <!-- Influence score (0-3) -->
      <influence>[0, 1, 2, or 3]</influence>

      <!-- Component flags (all "no" if stance is opposed or absent) -->
      <first_mover>[yes or no]</first_mover>
      <provided_reasoning>[yes or no]</provided_reasoning>
```

````
      <cited_by_others>[yes or no]</cited_by_others>
      <proposed_language>[yes or no]</proposed_language>

      <!-- Additive component count (0-4) -->
      <influence_components>[0-4]</influence_components>

      <!-- Discussion quantity (applies regardless of stance) -->
      <discussion_extent>[none, brief, moderate, or extensive]</discussion_extent>

      <!-- Context -->
      <who_cited_them>[Speaker names if cited, else "N/A"]</who_cited_them>
      <speakers_before>[Speakers who discussed this before focus speaker, "None" if first, "
       N/A" if absent]</speakers_before>

      <!-- Evidence and reasoning -->
      <evidence>[Quote showing stance and contribution, or "Speaker did not address this
       decision." if absent, or "Speaker opposed this decision: [quote]" if opposed]</
       evidence>
      <justification>[Explain: (1) stance determination, (2) why this score applies, (3) why
        next higher score does NOT apply]</justification>
    </decision>
    [Repeat for EVERY decision]
  </decisions>
  <summary>[2-3 sentences: How many decisions scored 3/2/1/0? What was the speaker's overall
    role     pivotal leader, active contributor, passive supporter, or non-participant/
   dissenter?]</summary>
</influence_analysis>
```

# Examples

## Example: Pivotal Influence (Score 3)
```xml
<decision>
  <id>5</id>
  <description>Added "prepared to employ these tools" language</description>
  <stance>supported</stance>
  <influence>3</influence>
  <first_mover>yes</first_mover>
  <provided_reasoning>yes</provided_reasoning>
  <cited_by_others>yes</cited_by_others>
  <proposed_language>yes</proposed_language>
  <influence_components>4</influence_components>
  <discussion_extent>moderate</discussion_extent>
  <who_cited_them>Lockhart, Yellen</who_cited_them>
  <speakers_before>None</speakers_before>
  <evidence>"I wonder whether inserting an additional sentence in paragraph 4 might help. We
    might add: 'the Committee discussed the range of policy tools available' and 'is
   prepared to employ these tools as appropriate.'" Lockhart later stated: "I support
   Governor Tarullo's suggestion."</evidence>
  <justification>Stance: Supported     speaker proposed this language. Score 3 (Pivotal)
   because speaker originated this exact language (first_mover=yes), which was adopted
   verbatim, and was explicitly cited by Lockhart and Yellen. This decision would not have
   existed without this speaker. N/A for why not higher     this is the maximum score.</
   justification>
````

````
</decision>
```

## Example: Contributory Influence (Score 2)
```xml
<decision>
  <id>6</id>
  <description>MBS purchases at $40 billion per month</description>
  <stance>supported</stance>
  <influence>2</influence>
  <first_mover>no</first_mover>
  <provided_reasoning>yes</provided_reasoning>
  <cited_by_others>no</cited_by_others>
  <proposed_language>no</proposed_language>
  <influence_components>1</influence_components>
  <discussion_extent>moderate</discussion_extent>
  <who_cited_them>N/A</who_cited_them>
  <speakers_before>Williams, Yellen, Bernanke</speakers_before>
  <evidence>"On the agency MBS, I strongly favor $40 billion per month over $30 billion. The
    staff has concluded we have room to do $40 billion without impairing market function, so
    I don't see the reason for holding back."</evidence>
  <justification>Stance: Supported     explicitly favored $40B. Score 2 (Contributory)
   because speaker provided substantive reasoning (staff analysis, market function) beyond
   mere preference. Not Score 3 because speaker was not first to propose $40B (Williams and
   Yellen spoke earlier), was not cited by others, and didn't change the trajectory
   others were already advocating this amount.</justification>
</decision>
```

## Example: Supportive Only (Score 1)
```xml
<decision>
  <id>2</id>
  <description>Maintained target range 0-  %</description>
  <stance>supported</stance>
  <influence>1</influence>
  <first_mover>no</first_mover>
  <provided_reasoning>no</provided_reasoning>
  <cited_by_others>no</cited_by_others>
  <proposed_language>no</proposed_language>
  <influence_components>0</influence_components>
  <discussion_extent>brief</discussion_extent>
  <who_cited_them>N/A</who_cited_them>
  <speakers_before>English, Bernanke, multiple others</speakers_before>
  <evidence>"I support Alternative B"     Alternative B maintains the current target range
   .</evidence>
  <justification>Stance: Supported     endorsed Alternative B which contains this decision.
   Score 1 (Supportive) because speaker endorsed via alternative without discussing the rate
    decision specifically. Not Score 2 because no reasoning provided about the rate decision
   , no language proposed, no defense of the decision     just package endorsement.</
   justification>
</decision>
```

## Example: Opposition (Score 0)
````
````

````
```xml
<decision>
  <id>1</id>
  <description>Forward guidance through mid-2013</description>
  <stance>opposed</stance>
  <influence>0</influence>
  <first_mover>no</first_mover>
  <provided_reasoning>no</provided_reasoning>
  <cited_by_others>no</cited_by_others>
  <proposed_language>no</proposed_language>
  <influence_components>0</influence_components>
  <discussion_extent>extensive</discussion_extent>
  <who_cited_them>N/A</who_cited_them>
  <speakers_before>N/A</speakers_before>
  <evidence>Speaker opposed this decision: "I'm generally not in favor of time-dependent
    policies I prefer contingent approaches. The mid-2013 language concerns me because it
    may box us in."</evidence>
  <justification>Stance: Opposed     speaker explicitly argued against time-dependent
    forward guidance. Score 0 because influence measures contribution to the ADOPTED decision
    . Although the speaker provided extensive, substantive arguments, those arguments were
    against this decision. The decision was adopted despite the opposition, so the speaker
    had zero influence on its adoption. All component flags = no because opposition does not
    count as contribution.</justification>
</decision>
```

Example: Absent (Score 0)
```xml
<decision>
  <id>7</id>
  <description>Maintained reinvestment policy</description>
  <stance>absent</stance>
  <influence>0</influence>
  <first_mover>no</first_mover>
  <provided_reasoning>no</provided_reasoning>
  <cited_by_others>no</cited_by_others>
  <proposed_language>no</proposed_language>
  <influence_components>0</influence_components>
  <discussion_extent>none</discussion_extent>
  <who_cited_them>N/A</who_cited_them>
  <speakers_before>N/A</speakers_before>
  <evidence>Speaker did not address this decision.</evidence>
  <justification>Stance: Absent     speaker made no statement about reinvestment policy and
    did not endorse an alternative containing it. Score 0 because no contribution to this
    decision's adoption. All component flags = no.</justification>
</decision>
```

Critical Reminders

1. **Determine stance FIRST**   before assessing contribution level
2. **Opposition = 0**   even brilliant arguments against an adopted decision don't count
 as influence on its adoption
3. **Component flags only apply to supportive contributions**   if opposed/absent, all
 flags = no
````

```
4. **Score every decision**    verify your count matches <policy_decisions>
5. **Justify why not higher**    explicitly explain why the next score level doesn't apply
6. **Quote evidence**    every non-zero score needs a specific quote showing support
7. **discussion_extent is separate**    it measures quantity regardless of direction, doesn't affect score

Begin your analysis now.
"""
```

# References


**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.

**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.

**Carlson, Jacob, and Melissa Dell.** 2025. "A Unifying Framework for Robust and Efficient Inference with Unstructured Data." *arXiv preprint arXiv:2505.00282*.

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

**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.** 2025. "The Systematic Origins of Monetary Policy Shocks."

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

**Ludwig, Jens, Sendhil Mullainathan, and Ashesh Rambachan.** 2025. "Large language models: An applied econometric framework." National Bureau of Economic Research.

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

**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."

**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.