Policy by Committee

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

January 8, 2026

Abstract

Committees shape numerous decisions, but empirical examination of group decision-making remains limited by measurement challenges. We address these challenges by analyzing Federal Open Market Committee members' macroeconomic forecasts alongside a novel dataset of their arguments from over forty years of meetings. Contrary to prevailing theories emphasizing information aggregation, we find that members present distinct models. Heterogeneous beliefs about key model parameters, such the Phillips curve slope and the natural unemployment rate, give rise to persistently different interpretations of public information. Committee decisions tilt toward members whose models better fit recent data. We develop a theoretical framework of model aggregation, showing that committees may improve decision quality but risk excess sensitivity to incoming data.

*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, 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. This paper was written while Vasudevan was visiting Yale University and the University of Chicago, whose hospitality he gratefully acknowledges.

1 Introduction

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

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

We find that, in contrast to prevailing theories emphasizing information aggregation, members present distinct and persistent models during committee meetings, defined as frameworks used to interpret public information and translate it into forecasts of key variables.[1] 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. We establish these findings through an empirical analysis of both FOMC member forecasts and meeting deliberations. Guided by our findings, we develop a theoretical framework that formalizes how committees aggregate heterogeneous models. The framework suggests that group decision-making may enhance decision quality when members are sufficiently willing to consider others' models, but it also carries the risk of excess sensitivity to incoming data.

Empirical Analysis. We now describe our empirical approach and findings. Using

[1] We broadly describe two members as having different "models" if we expect them to form different forecasts upon observing identical data. The underlying drivers may include: (i) different statistical/predictive models featuring different model parameters; (ii) different beliefs about causal relationships among different variables ("mental models"); (iii) heterogeneous priors; and (iv) systematic biases in interpretation, e.g., consistently adding an inflationary or disinflationary tilt to one's assessment.

both quantitative forecasts and a novel dataset of arguments extracted from FOMC transcripts, we document that members express persistently heterogeneous views despite regular information sharing, consistent with members employing different models to interpret the same data.

Our analysis proceeds in several steps, each designed to distinguish between differential information and heterogeneous models as sources of disagreement. First, we document strong and persistent heterogeneity in members' reported forecasts of inflation, growth, employment, and the Federal Funds rate. Member fixed effects explain substantial variation (R^2 values of 0.20–0.46 for one-year-ahead forecasts), with some members consistently more hawkish or dovish across time. These forecasts exhibit a strong factor structure: members forecasting high inflation tend to forecast high growth in the same meeting. Notably, we find especially strong member fixed effects in reported beliefs about long-term averages of the variables (R^2 of 0.34–0.64), indicating disagreements about steady-state parameters governing the economy. These disagreements are consistent with theoretical work showing that learning about model parameters can be extremely slow in macroeconomic environments (Johannes, Lochstoer and Mou, 2016; Farmer, Nakamura and Steinsson, 2024; Li, Van Nieuwerburgh and Renxuan, 2025).

To examine the sources of this persistent heterogeneity in forecasts, we analyze members' discussions during each meeting to understand whether these differences reflect different information or different models. We construct a novel dataset from FOMC transcripts using a large language model to identify policy arguments made by committee members from 1976 to 2019. Each argument expresses a view about inflation, growth, or employment, the same variables for which we observe forecast heterogeneity, supporting either accommodative or restrictive policy. We validate this dataset through comparisons with other LLMs and human coding.

Using these arguments, we provide direct evidence of model heterogeneity by extracting members' beliefs about two key model parameters that govern how labor market conditions translate into inflation: the slope of the Phillips curve and the non-accelerating inflation rate of unemployment (NAIRU). These parameters are central to monetary policy yet subject to substantial uncertainty and disagreement, making them natural candidates for examining model heterogeneity among policymakers. The Phillips curve slope determines how strongly labor market tightness generates inflation pressure; NAIRU determines the unemployment rate at which that pressure is zero. Members who disagree about either parameter will interpret identical labor market data differently when forecasting inflation.

We find that these parameter beliefs vary substantially across members and predict

their forecasts in the direction the underlying models imply. Members with steep Phillips curve beliefs respond more strongly to labor market conditions when forecasting inflation. Specifically, a one-standard-deviation increase in perceived labor tightness is associated with a 0.13 standard deviation higher inflation forecast for members who believe the Phillips curve is steep, compared to near-zero response for members who believe it is flat. Members who believe NAIRU is high forecast higher inflation because they perceive tighter labor markets at any given unemployment rate; this effect operates entirely through persistent cross-member differences in NAIRU beliefs rather than within-member variation. These patterns indicate that persistent forecast disagreements reflect genuine differences in members' models used to translate common information into forecasts.

Having documented evidence of model heterogeneity, we next examine the alternative hypothesis that disagreements reflect differential private information rather than heterogeneous models. First, we examine what information members cite: 82% of arguments reference publicly available information (government statistics, Fed staff analysis), while only 18% cite potentially private sources. When we construct scores using only public information arguments, they correlate at 0.95 with our full scores, indicating that private information has limited impact on the persistent heterogeneity we document. Second, we investigate whether disagreements arise from members focusing on different aspects of the economy or interpreting the same aspects differently. We classify arguments into topical categories designed to distinguish between different economic mechanisms (e.g., *Growth Forecasting and Outlook* versus *Financial Market Conditions*) while grouping together opposing interpretations of the same mechanism. We find that 78% of disagreement stems from heterogeneous interpretations of shared topics rather than topic selection. Members discuss the same economic mechanisms but reach different conclusions about their implications.

These findings demonstrate that persistent FOMC disagreements reflect members employing heterogeneous models rather than accessing different information. Members observe the same public data and discuss the same topics, yet consistently reach different conclusions. We trace these differences directly back to heterogeneous beliefs about key model parameters like the Phillips curve slope and NAIRU. Having established the presence of persistent model heterogeneity, we next examine how the committee aggregates these diverse models in its decision-making process.

We construct novel transcript-based measures that capture each member's alignment with and influence on committee decisions across all policy dimensions: target rates, forward guidance, balance sheet policies, and statement language. Our alignment mea-

sure ranges from complete opposition to perfect agreement, while our influence measure captures whether members lead policy formation or support established positions. These granular measures provide substantially more variation than traditional dissent data (which occur in only 6.9% of member-meeting observations) or preferred rate deviations. We validate these measures against monetary policy surprises from high-frequency asset price responses, dissenting votes, and members' stated rate preferences.

Our analysis reveals that the committee systematically tilts its decisions toward members whose models better fit recent economic data. We measure model fit using the accuracy of members' recent forecasts relative to realized outcomes. Members with one standard deviation better model fit show 0.05 to 0.19 standard deviations higher alignment with committee decisions and 0.04 to 0.15 standard deviations greater influence. The probability of dissent decreases by 2 to 7 percentage points per standard deviation improvement in fit which is economically meaningful given the 6.9% baseline dissent rate. These results, which control for meeting fixed effects and thus cannot be explained by all members updating toward incoming data, demonstrate that the committee aggregates heterogeneous models by weighting those that have proven more accurate, a pattern distinct from equal weighting or pure information aggregation.

Theoretical Framework. Our empirical findings suggest a new perspective on committees. Rather than aggregating private information, as in standard theories, committees aggregate different models for interpreting public information. To further explore its implications, we formalize this perspective in a simple theoretical framework of group decision-making under model uncertainty. Each member of a committee observes the same data but has a different, partially misspecified model, and the data provide a noisy signal about the true model.[2] The committee's objective is to agree on a model to make its policy decision. Each member advocates for a combination of all members' models, trading off a desire for a committee model that fits the observed data and a cost for supporting a model that deviates from their own. The committee's choice arises as a Nash equilibrium outcome. Given members' desire to match the observed data well, the committee tilts toward better-fitting models, matching our empirical evidence.

We evaluate the committee's performance based on how closely its chosen model matches the unknown true model in expectation. We find that the committee may outperform its best member's model under three basic conditions. First, the data must be sufficiently informative to distinguish among models. Second, members' models must

[2]There are numerous reasons that observed data provide only a noisy signal of the true model. For example, economic data may reflect measurement error and are often revised after the fact. Additionally, policy actions influence the dynamics of variables of interest and hamper identification of the true model (McLeay and Tenreyro, 2020; Cieslak, Hansen and Pang, 2025).

be "diverse," in that they capture different parts of the data-generating process. Third, members must be sufficiently willing to consider each other's perspectives. If the data are too noisy or members are too stubborn in advocating for their own perspectives, the committee may settle on a suboptimal model. Intriguingly, some stubbornness may improve the committee's performance by preventing overreaction to noisy data. This creates a fundamental trade-off in committee design—leveraging model diversity while tempering excess sensitivity to incoming information.

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

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

Our work relates to a long tradition in monetary policy and macroeconomics on decision-making in the face of model uncertainty. Prior work has presented numerous approaches, 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, and we evaluate the associated trade-offs.

Additionally, our paper relates to recent work on how the FOMC's policy decisions are affected by the composition of its voting members in terms of geographic region (Bobrov, Kamdar and Ulate, 2025; Fos, Tamburelli and Xu, 2025; Fos and Xu, 2025) and pol-

icy preferences (Malmendier, Nagel and Yan, 2021; Bordo and Istrefi, 2023; Hack, Istrefi and Meier, 2023). We take a different focus, examining how members' perspectives are weighted in the group decision-making process. We document that the committee tilts its decisions toward members with models that are better aligned with recent data. Given this behavior, our evidence is consistent with Hack, Istrefi and Meier (2025), who find that the FOMC appears to use a time-varying policy rule in making its policy decisions, and provides a micro-foundation for such behavior.

More broadly, our paper relates to empirical work studying central bankers' policy preferences and their determinants (Belden, 1989; Chappell Jr, McGregor and Vermilyea, 2004; Schonhardt-Bailey, 2013; Hansen, McMahon and Rivera, 2014; Istrefi, 2019; López-Moctezuma, 2023), and to work analyzing FOMC transcripts and minutes (Boukus and Rosenberg, 2006; Hansen, McMahon and Prat, 2018; Cieslak and Vissing-Jorgensen, 2021; Shapiro and Wilson, 2022; Cieslak and McMahon, 2023; Cieslak et al., 2023; Bordo, Istrefi and Martínez, 2024). Relative to this literature, we construct a new, granular dataset of the FOMC's decisions and individual members' beliefs, preferences, and economic arguments. In this regard, our paper is related to a small but growing literature that seeks to identify the mental models used by economic agents in experimental as well as macroeconomic and financial field settings (Andre et al., 2022; Andre, Schirmer and Wohlfart, 2023; Binetti, Nuzzi and Stantcheva, 2024; Andre et al., 2025; Bybee, 2025; Ke, 2025; Bastianello, Décaire and Guenzel, 2025; Barron and Fries, 2025; Charles and Kendall, 2025). Importantly, we characterize decision-makers' arguments and the resulting outcomes in a high-stakes setting.

2 Setting

Our sample consists of 365 regularly scheduled FOMC meetings from April 1976 through December 2019. The FOMC meets eight times per year to formulate monetary policy. The committee normally consists of 12 voting members who vote on the committee's decision: seven Governors of the Federal Reserve Board in Washington, DC, including a Chair; the President of the New York Fed; and four members selected on a rotating basis from the presidents of the 11 other regional Federal Reserve Banks.[3] 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

[3]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.

and actively participate in the meetings.

Since 1993, transcripts of FOMC meetings have been released to the public on a five-year lag and are available from the Federal Reserve website. These transcripts provide verbatim discussions from every FOMC meeting from 1976 onward. We describe the processing of the transcript data in Internet Appendix IA.A.1. Prior to each meeting, the Federal Reserve Board staff prepares Tealbook A (formerly the Greenbook), which provides discussions and forecasts of macroeconomic and financial conditions. We obtain the reports from the Federal Reserve Board website and the associated forecasts from the Philadelphia Federal Reserve website. The staff also prepares and distributes Tealbook B (formerly the Bluebook), which includes a set of monetary policy alternatives that 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.[4] The discussion of the economic and financial situation begins with a presentation by members of the Federal Reserve Board staff followed by questions, prepared statements by members of the committee, and a general discussion. The subsequent monetary policy discussion typically begins with the director of the Division of Monetary Affairs presenting monetary policy alternatives. The presentation is followed by a policy go-around where both voting and non-voting members discuss their policy preferences. Following this discussion, the committee votes on the proposed policy directive. In 1994, the committee began announcing the outcome of its meetings at the end of each meeting via a public statement; a substantial portion of the policy discussion segment thereafter concerns the wording of this statement.

Alongside the transcript data and outcomes from each meeting, we also study forecasts of real GDP growth, inflation, and the unemployment rate reported by members of the FOMC as part of the Monetary Policy Report to Congress (the "MPR") and the Summary of Economic Projections (the "SEP"). The MPR dataset, studied in Romer and Romer (2008); Romer (2010), provides forecasts submitted by members of the FOMC 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

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

"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 on a five-year lag.[5]

3 Members' Beliefs and Model Heterogeneity

We begin our analysis in this section by studying FOMC committee members' forecasts. We document strong and persistent disagreements in the forecasts of variables reported by FOMC committee members. We present multiple pieces of evidence that these persistent disagreements in expressed views are driven by heterogeneity in models—different ways of interpreting the same data. First, we show that members exhibit substantial disagreements about key model parameters, including the long-run average rates of unemployment, GDP growth, and inflation. We then turn to the FOMC transcripts, constructing a novel dataset of arguments made by members during meetings. From these arguments, we extract members' beliefs about the slope of the Phillips curve. We complement this with members' beliefs about the non-accelerating inflation rate of unemployment (NAIRU), measured from their long-run unemployment forecasts in the Summary of Economic Projections. We demonstrate that both sets of parameter beliefs predict members' forecasts in ways consistent with different underlying models. Finally, we present additional evidence consistent with model heterogeneity: members cite predominantly public information and disagree primarily through different interpretations of the same topics rather than by focusing on different aspects of the economy.

3.1 Overview and Framework

Before proceeding, we present a simple framework to guide our analysis in this section, and describe our results through this lens.

Consider an economy with an underlying state ω_t in period t. Monetary policy depends on certain aspects of this state, which we represent through key variables indexed by k (e.g., inflation, growth, employment). Committee members forecast these variables, and these forecasts shape their policy preferences.

All members observe the same public data D_t, which provide noisy signals about various components of ω_t. Member i interprets these data through their own probability

[5]The data were previously released on a ten-year lag. Because of this, the identities associated with forecasts made in 2015 have not yet been released and are expected to be released in 2026; see https://www.federalreserve.gov/monetarypolicy/files/FOMC20160111memo01.pdf.

model, m_i—their understanding of how the economy works—which induces a *persistent* mapping g_i from data to forecasts:

$$g_i(D_t) = (f^{pub}_{i,t,1}, \ldots, f^{pub}_{i,t,k})$$

Crucially, this mapping g_i represents member i's stable model of the economy, which persists across meetings.

In addition to public data, member i receives private information $P_{i,t}$ about the economy, which generates additional forecast components $(f^{priv}_{i,t,1}, \ldots, f^{priv}_{i,t,k})$.[6]

During FOMC meetings, members share their views and forecasts.[7] Member i's reported forecast for variable k combines three elements:

$$f_{i,t,k} = w_{i,1} f^{pub}_{i,t,k} + w_{i,2} f_{-i,t,k} + (1 - w_{i,1} - w_{i,2}) f^{priv}_{i,t,k} \tag{1}$$

where $w_{i,1}$ is member i's weight on their interpretation of public information, $w_{i,2}$ is their weight on other members' expressed views (where $f_{-i,t,k}$ represents some aggregate of others' reported forecasts), and $1 - w_{i,1} - w_{i,2}$ is their weight on their private information.

Our first result in this section is that member fixed effects explain substantial variation in forecasts of a variable $f_{i,t,k}$. Members who report higher forecasts of inflation at a given time period tend to do so persistently across time. Our second result is that members' forecasts are correlated across variables within a given time period: members forecasting high inflation also forecast high growth in a given meeting.

Equation (1) highlights two potential drivers of these results: (i) heterogeneous models, that is, members having persistent maps g_i that translate public information into forecasts in a consistent manner across variables; or (ii) members having differential private information that is correlated across dimensions.

For both results, given the structure of the FOMC where members publicly share their views each meeting, members' persistently different views are difficult to reconcile with simple models of differential information that are traditionally studied in macroeconomics, such as Woodford (2003) and Sims (2003), and are commonly applied to study macroeconomic forecast data, e.g., Coibion and Gorodnichenko (2015). In these models, members' private signals are noisy observations of the true macroeconomic state, and

[6] For simplicity, we abstract from whether member i processes private information through the mapping induced by their model, g_i, or observes signals directly.

[7] For the Summary of Economic Projections, committee members submit initial forecasts on the Friday prior to the FOMC meeting and may revise their forecasts until the morning of the second day of the FOMC meeting, after hearing other members express their views during the economic roundtable. See https://www.federalreserve.gov/monetarypolicy/fomc_projectionsfaqs.htm for additional details.

when sharing views and information, members' beliefs should converge as they learn from one another's expressed views.

In contrast, model heterogeneity is able to generate persistently different forecasts, even when members update their models in a Bayesian fashion, as learning about model parameters can be very slow in macroeconomic environments (Johannes, Lochstoer and Mou, 2016; Farmer, Nakamura and Steinsson, 2024; Li, Van Nieuwerburgh and Renxuan, 2025). Indeed, consistent with the mechanism emphasized by these papers, analyzing novel data from FOMC transcripts, we also report evidence of persistent heterogeneity in members' views about key model parameters for forecasting the variables of interest: their long-run average values, and the slope of the Phillips curve (which characterizes the strength of the tradeoff between maintaining low unemployment and low inflation). We further show that the public versus private nature of data cited by members and the topical composition of their arguments align with model-based rather than information-based differences. We accordingly interpret the results as evidence that members employ different models.[8]

Finally, we note that an additional interpretation of our results may involve strategic reporting of forecasts. Under this interpretation, our results indicate that members primarily communicate through different and persistent models when interpreting the same information in meetings. Taken in conjunction with the results we present in Section 4, this potential strategic element does not affect our broader message: the committee tilts its decisions in the direction of models that better match recent data. For expositional simplicity, throughout the rest of the paper, we discuss the beliefs reported by members in FOMC meetings as *their* beliefs, without emphasizing the potential strategic element.

3.2 Fixed Effects in Forecasts

With our conceptual framework in hand, we proceed with the analysis. Table 1 reports the variance of inflation, GDP growth, and unemployment forecasts from the Summary of Economic Projections and Monetary Policy Report to Congress explained by member fixed effects (controlling for meeting fixed effects). We use the Bonhomme et al. (2023) implementation of 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

[8]Our discussion of model heterogeneity focuses on differences in parameters that govern relationships across variables, e.g., the steepness of the Phillips curve or the steady state level of unemployment. This approach is also conceptually related to another representation of model heterogeneity—differences in beliefs about causal relationships between variables—as studied in other work (Eliaz and Spiegler, 2020; Charles and Kendall, 2025). For example, a member's perceived Phillips curve slope captures the extent to which they believe increased employment may cause price increases.

null of no true fixed effect heterogeneity.

The first set of rows reports results of forecasts for the remaining quarters of the present year, with R^2 values of 0.15, 0.09, and 0.11 for Inflation, Growth, and Employment.[9] The second set of rows reports R^2 values of forecasts for the subsequent year, which are 0.37, 0.20, and 0.24 for Inflation, Growth, and Employment. The results highlight the rigidity of reported beliefs at the member level. The last column of the table reports statistics for forecasts of the appropriate Fed Funds rate, reported in the Summary of Economic Projections starting in 2012; the R^2 values of member fixed effects are 0.34 and 0.46 for the current and subsequent year, indicating that the rigidity of forecasts for macroeconomic variables carries into members' forecasts of the appropriate Fed Funds rate.

To gauge whether the estimated R^2 values for member effects exceed what could arise by chance, we conduct 1,000 bootstrap simulations. Each simulation resamples forecasts with replacement inside every 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, indicating the strong presence of member fixed effects.

Next, we explore how members' forecasts are related across different variables. Table 2 reports the principal components of members' forecasts. The first three columns report principal components estimated from member-meeting observations. The first principal component loads positively on each of the variables and explains 47% of the variance. These positive loadings are consistent with the first principal component representing a hawk-versus-dove dimension, where committee members and meetings that forecast higher growth also forecast higher employment and inflation. The second and third principal components capture differential loadings on various variables, such as inflation versus growth and employment (the second principal component). The last three columns of the table report principal components from a matrix of member-level fixed effects for each variable, estimated after controlling for meeting fixed effects. The results are very similar, indicating the presence of a similar correlation structure for the persistent, time-invariant component of members' forecasts.[10]

The rigidity and correlation structure of members' forecasts suggest they employ heterogeneous models to interpret the same information. Notably, disagreements extend

[9]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.

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

	Inflation	Growth	Unemployment	Fed Funds*
Y0 Member FE	0.15	0.09	0.11	0.34
Y0 Bootstrap null (99th pct.)	0.05	0.05	0.06	0.06
$\sigma_{y,Y0}$	0.32	0.34	0.13	0.18
Y1 Member FE	0.37	0.20	0.24	0.46
Y1 Bootstrap null (99th pct.)	0.06	0.05	0.05	0.05
$\sigma_{y,Y1}$	0.27	0.35	0.23	0.47
LR Member FE	0.34	0.63	0.64	0.64
LR Bootstrap null (99th pct.)	0.06	0.04	0.04	0.05
$\sigma_{y,LR}$	0.10	0.16	0.28	0.30

TABLE 1: FORECASTS: R^2 OF MEMBER FIXED EFFECTS

Note: The 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 time 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 present year, rows labeled Y1 correspond to forecasts of the subsequent year. The regressions are estimated with the Leave-One-Out estimator of Kline, Saggio and Sølvsten (2020) implemented by Bonhomme et al. (2023). The rows labeled "Bootstrap null (99th pct.)" give 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. The rows labeled $\sigma_{y,Y}$ report the standard deviation of forecasts, after controlling for meeting fixed effects. The data for Fed Funds rates forecasts, reported in the last column, start in 2012 and the data on long-run averages for the variables start in 2009.

Comparison of Principal Components						
	Full Data			Member FE		
Principal Component Loadings	PC1	PC2	PC3	PC1	PC2	PC3
Inflation	0.41	0.91	-0.09	0.61	0.78	-0.13
Growth	0.65	-0.36	-0.67	0.64	-0.59	-0.49
Employment	0.64	-0.22	0.73	0.46	-0.22	0.86
Proportion of Variance Explained	0.47	0.31	0.22	0.48	0.31	0.21

TABLE 2: PRINCIPAL COMPONENTS OF COMMITTEE MEMBERS' FORECASTS

Note: The table reports principal components of the inflation, growth, and employment forecasts. The first three columns (Full Data) show the results where observations are member-meeting forecasts (standardized so that forecasts for each variable have zero mean and unit standard deviation over the full sample). The last three columns (Member FE) display results where observations are member fixed effects for each variable, estimated from regressions of the member-meeting forecasts for the variables on member and meeting fixed effects.

to fundamental parameters governing the economy: the last three rows of Table 1 report member fixed effects for long-run projections from the Summary of Economic Projections. The R^2 values of 0.34, 0.63, 0.64, and 0.64 for Inflation, Growth, Employment, and the Fed Funds rate indicate persistent disagreements about steady-state values—parameters that should be common knowledge if members shared a common model.[11] These disagreements about long-run parameters are consistent with theoretical work showing that learning about model parameters can be extremely slow in macroeconomic environments (Johannes, Lochstoer and Mou, 2016; Farmer, Nakamura and Steinsson, 2024; Li, Van Nieuwerburgh and Renxuan, 2025).

While these forecast patterns are consistent with heterogeneous models, they could also reflect differential private information that is correlated across variables and persistent over time. To distinguish between these explanations, we turn to a novel dataset of arguments made by committee members during FOMC meetings. This dataset allows us to directly examine what information members cite, what topics they discuss, and extract long time-series of the structural parameters underlying their models.

3.3 Members' Arguments During FOMC Meetings

To better understand the nature of forecast heterogeneity, we assemble a novel dataset of arguments made by committee members during FOMC meetings. We use these arguments to extract structural parameters that members express during meetings, evaluate whether disagreements stem from private information, and examine whether members focus on different aspects of the economy or interpret the same aspects differently.

In constructing our argument dataset, we focus on arguments made about growth, employment, and inflation, the same variables for which members provide quantitative forecasts, allowing us to directly connect members' verbal arguments with their numerical predictions. For each variable, we use a large language model, the 20250514 release of Claude Sonnet 4, to identify each policy-relevant argument made by FOMC committee members where they express forecasts or beliefs about the variable of interest.

We take several steps to ensure the reliability and reproducibility of the LLM's output. First, we use chain-of-thought prompting, wherein the LLM engages in stages of intermediate reasoning before answering a query.[12] Prior work has found chain-of-thought prompting to significantly improve performance in complex reasoning tasks (Wei et al., 2022). Second, we require the LLM to provide explicit transcript quotes to justify its rea-

[11] The lower explanatory power for long-run inflation reflects the committee's explicit 2% inflation target.

[12] For chain-of-thought prompts, we use the Anthropic prompt improver tool, discussed at https://www.anthropic.com/news/prompt-improver.

soning, ensuring that its analysis is grounded in the text. Third, we employ a multi-step procedure for each variable of interest: first identifying relevant passages, then analyzing the relevant subset of text. This approach ensures comprehensiveness in capturing all arguments and improves output stability (Bastianello, Décaire and Guenzel, 2025). We also set the temperature parameter to zero to ensure deterministic responses.

We score each argument where members express beliefs about the current or future level of a variable on a scale 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 rising unemployment or labor market weakening, while positive scores indicate expectations of declining unemployment or labor market tightening.[13]

We conduct extensive validation through multiple approaches. We hand-code a random sample of 293 arguments with multiple expert human coders, following best practices for LLM validation (Ludwig, Mullainathan and Rambachan, 2025). The argument-level scores show strong agreement (correlation = 0.80), and this rises to 0.93 at the member-meeting level—the unit at which we conduct our analyses—indicating that aggregation effectively reduces measurement noise. In a blind adjudication exercise where an independent judge evaluates disagreements between human coders and Claude without knowing the source, Claude's classifications prove competitive with expert human coding (44.6% vs 55.4% agreement). Cross-validation with Llama 3 70B yields a correlation of 0.93 with Claude at the member-meeting level, confirming that our approach captures genuine economic content rather than model-specific artifacts. The convergence of three independent measurement approaches at correlations exceeding 0.90 supports the robustness of these measures. We present comprehensive validation details in Appendix Section IA.B.1.

Table 4 presents examples of arguments and their scores from the August 9, 2011 FOMC meeting, which we use throughout the paper to illustrate our data. The meeting occurred during a period of zero interest rates and weak recovery from the Great Recession, with financial markets under pressure from U.S. and European fiscal concerns. The arguments displayed range from Narayana Kocherlakota and Charles Plosser pointing to falling unemployment and rising inflation measures (scores of $+2$), to Charles Evans and Daniel Tarullo warning about weak growth and recession risks (scores of -2 and -3).

[13]Specifically, a growth score of -3 indicates expectations of significant contraction or substantial downside risks, 0 indicates trend-level growth with balanced risks, and $+3$ indicates significantly above-trend growth or substantial upside potential. Similar scales apply for inflation (relative to target) and employment (relative to the natural rate of unemployment). Full scoring criteria and prompts are reported in Appendix IA.A.2.

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 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).

Table 5 reports summary statistics. In a given meeting, on average, 82% of members discuss inflation, 95% discuss growth, and 70% discuss employment. Unconditionally, a voter makes approximately 6 arguments about growth, 4 about inflation, and 3 about employment per meeting. Inflation arguments slightly lean toward supporting restrictive policy (mean score 0.2), whereas growth arguments lean accommodative (mean score -0.4). Non-voters show similar patterns, though their arguments support more restrictive policy on average, consistent with evidence that bank presidents tend to be more hawkish than governors (Malmendier, Nagel and Yan, 2021; Bordo and Istrefi, 2023). Because their arguments appear similar, we pool voters and non-voters when analyzing arguments.

Figure 2 plots the average of committee members' belief-based scores in each meeting for each variable. The scores demonstrate intuitive cyclical patterns, reaching troughs during NBER recessions. Concerns about high inflation are elevated in the 1970s and late 1990s, and were low in the years following the Global Financial Crisis.

To verify that argument scores capture relevant cross-sectional variation in members' beliefs, Figure 1 compares member-meeting scores with members' forecasts from the MPR and SEP. The figure sorts committee members into deciles based on their cross-sectionally demeaned forecasts and plots these against cross-sectionally demeaned argument scores. The combined panel pools all three variables after standardizing. The scores display a strong cross-sectional relationship with forecasts, confirming that our LLM-based measures capture the same belief heterogeneity observed in members' quantitative predictions.

Having established that our argument extraction captures meaningful variation in members' expressed beliefs, we next use this dataset to directly measure the structural parameters underlying members' models.

3.4 Belief Heterogeneity and Model Parameters

We provide direct evidence of model heterogeneity by extracting members' beliefs about two parameters central to macroeconomic forecasting: the slope of the Phillips curve and the non-accelerating inflation rate of unemployment (NAIRU). The Phillips curve slope governs how strongly labor market tightness translates into inflation and is a key input to policy design (Clarida, Gali and Gertler, 1999), yet its magnitude and stability remain subjects of considerable uncertainty and disagreement (Cogley and Sargent, 2005; Mavroeidis, Plagborg-Møller and Stock, 2014; McLeay and Tenreyro, 2020). NAIRU determines the unemployment rate consistent with stable inflation. Members holding dif-

Voting Members	Inflation	Growth	Employment
Members per meeting		——— 10.9 ———	
Avg prop. making argument per meeting	82%	95%	70%
Arguments per member-meeting	3.9	5.8	2.8
Avg prop. making belief argument per meeting	64%	88%	64%
Belief arguments per member-meeting	2.7	4.9	2.4
Avg argument score	0.2	-0.4	-0.3
Non-voting Members	Inflation	Growth	Employment
Members per meeting		——— 6.7 ———	
Avg prop. making argument per meeting	78%	89%	73%
Arguments per member-meeting	3.7	5.3	3.2
Avg prop. making belief argument per meeting	63%	84%	68%
Belief arguments per member-meeting	2.6	4.6	2.8
Avg argument score	0.3	-0.2	0.1

TABLE 5: ARGUMENT SUMMARY STATISTICS

Note: The table displays summary statistics for the arguments made related to each variable of interest. The top panel displays statistics for voting members of the committee and the bottom panel displays statistics for the non-voting members. Within each panel, the first row displays the average number of members per meeting. For a given variable, the second row displays the average number of members per meeting making an argument, and the third row displays the average number of arguments per member. The fourth and fifth rows report the average proportion of members per meeting making belief arguments, and the average number of belief arguments per member. The sixth row displays the average score across all arguments.

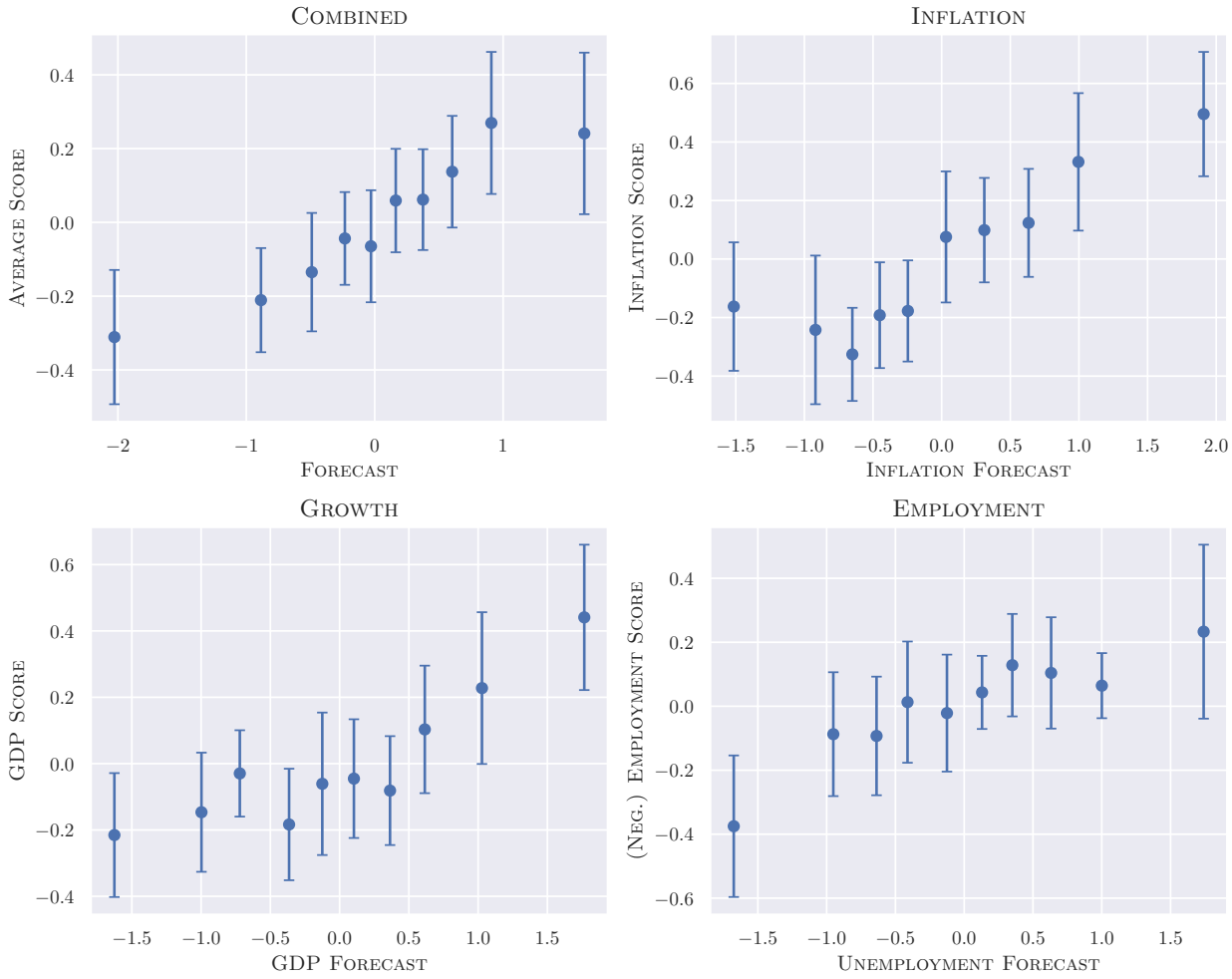


FIGURE 1: COMMITTEE MEMBERS' FORECASTS AND ARGUMENT SCORES

Note: The figure sorts committee members into deciles based on their cross-sectionally demeaned GDP, inflation, and unemployment forecasts, and plots the average of these forecasts versus the average cross-sectionally demeaned transcript-measured growth, inflation, and employment scores for the closest meeting. In the combined panel, observations are standardized to zero mean, unit variance within each variable, cross-sectionally demeaned by meeting, and then sorted into deciles; the panel plots the mean of each decile. Forecasts are from the Federal Reserve's Monetary Policy Reports to Congress, available semi-annually from 1992 to 2007 from Romer (2010); and from the Summary of Economic Projections, conducted prior to four meetings per year since 2008. The figure also plots plus and minus two standard error bars for each decile. Standard errors are computed by clustering by member and meeting within each decile.

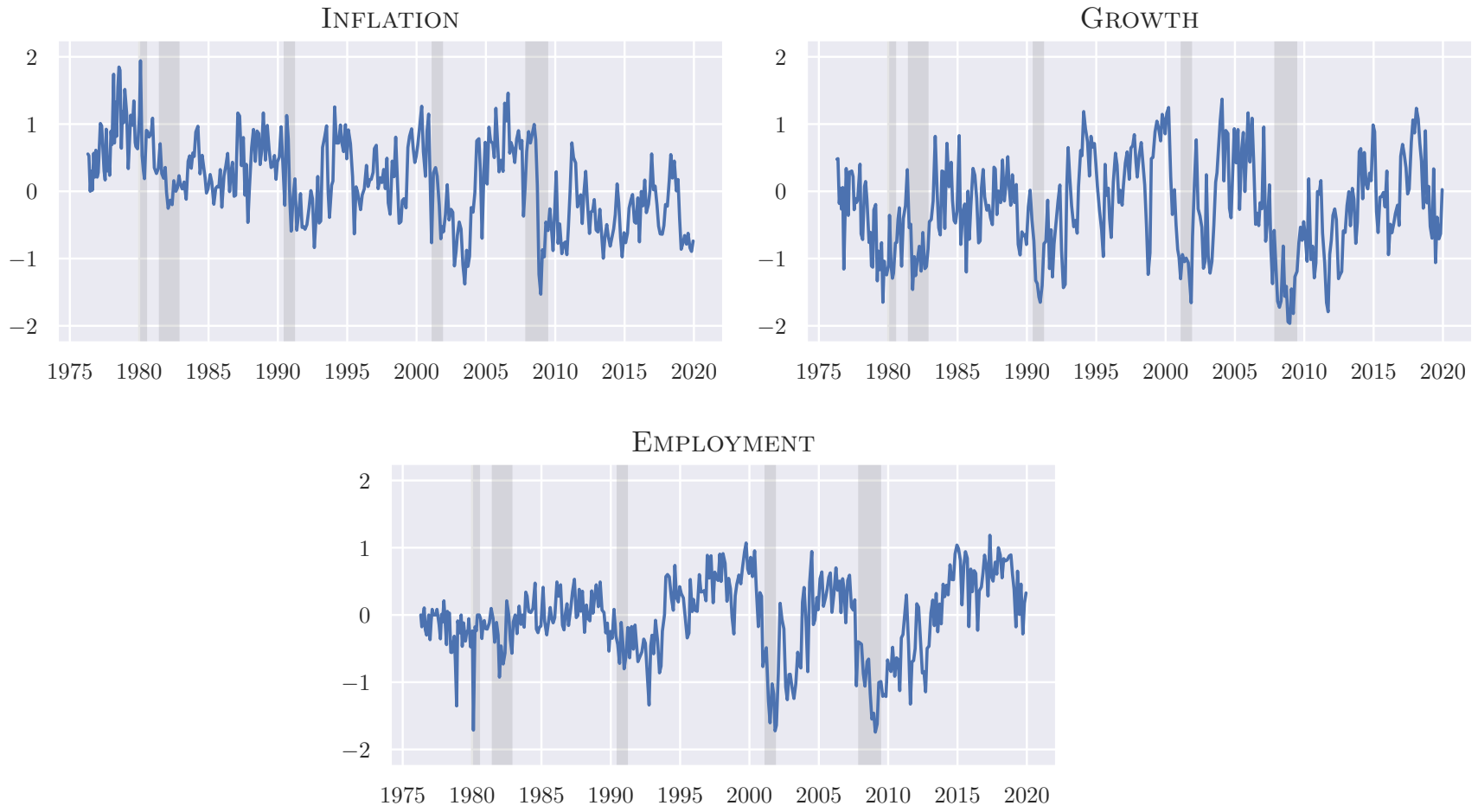


FIGURE 2: COMMITTEE MEMBERS' ARGUMENT SCORES OVER TIME

Note: The figure plots the average of 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.

ferent views about these parameters will interpret the same labor market data differently when forecasting inflation.

These parameter beliefs predict forecasts in the direction the underlying models imply. Members who believe the Phillips curve is steep respond more strongly to labor market conditions when forecasting inflation. Members who believe NAIRU is low see less inflation pressure at any given unemployment rate.

3.4.1 Extraction Methodology

We obtain NAIRU directly from FOMC members' long-run unemployment forecasts in the Summary of Economic Projections, beginning in April 2009. No analogous quantitative measure exists for Phillips curve beliefs, so we extract these from our arguments dataset using a large language model. The extraction follows conservative classification rules: a Phillips curve belief is recorded only when a speaker explicitly connects labor market conditions to inflation outcomes through causal language, not merely when both concepts appear in the same statement.

We classify Phillips curve slope beliefs on a three-point scale: $+1$ (steep), 0 (moderate), or -1 (flat). We aggregate argument-level classifications to the member-meeting level by averaging across all classified arguments for a given member at a given meeting. As with the forecasts documented above, members' beliefs about model parameters exhibit substantial persistence. Given this persistence, when a member does not express a Phillips curve belief in a given meeting, we carry forward their most recent prior expressed value.

We discuss the classification methodology in more detail in the Appendix.

Coverage Statistics

	Meetings with observation	Total members per meeting	Members per meeting with observation	Pct observations forward-filled	Median # of periods forward-filled
Raw	354	17.59	7.56	–	–
Fwd-filled	365	17.59	16.24	54.83	3.00

Coverage Statistics by Decade

	Pct members with observation, 1970s	Pct members with observation, 1980s	Pct members with observation, 1990s	Pct members with observation, 2000s	Pct members with observation, 2010s
Raw	17.66	22.59	49.72	47.58	62.77
Fwd-filled	75.70	90.75	96.26	94.51	97.98

Parameter Characteristics

	Average level	Median cross-sectional std
Raw	0.46	0.68
Fwd-filled	0.47	0.72

TABLE 6: PHILLIPS CURVE SLOPE BELIEFS: SUMMARY STATISTICS

Note: The table reports summary statistics for Phillips curve slope beliefs extracted from FOMC transcripts. Members per meeting with observation reports the average number of members per meeting for whom we extract a parameter belief. Pct of observations forward-filled and median # of periods forward-filled report the percentage and median number of periods that parameter beliefs are carried forward from previous meetings. The average level and median cross-sectional std rows report the average level and standard deviation of parameter beliefs across all member-meeting observations.

Speaker	Date	Value	Quote
Charles Plosser	Mar-2011	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..."
Eric Rosengren	Aug-2011	Steep	"Given the large degree of resource slack in the economy, it is quite possible that the inflation rate will soon be falling. Businesses are worried and labor markets are weak"
Sarah Raskin	Aug-2011	Steep	"There's going to continue to be considerable slack in labor markets over coming quarters, which will continue to exert significant restraint on wages and prices."
Ben Bernanke	Dec-2011	Moderate	"there are more factors than slack that affect inflation... a pretty strong global economy that has affected commodity prices and... has fed through into inflation numbers in various ways. These are some factors that are offsetting the slack."

TABLE 7: COMMITTEE MEMBERS' PHILLIPS CURVE SLOPE BELIEFS IDENTIFIED BY LLM: EXAMPLES

Note: The table presents examples of Phillips curve slope belief extractions. Phillips curve slope is classified as steep (inflation responds strongly to slack), moderate, or flat (inflation responds weakly to slack).

Table 6 reports summary statistics for the extracted beliefs among members of the committee. We extract an expressed belief about the slope of the Phillips curve in 354 of 365 meetings. On average, we extract beliefs for about 43% of members within a given meeting. After forward-filling, we obtain beliefs for 92% of member-meeting observations, with coverage consistently exceeding 90% from the 1980s onwards. The average value of the extracted Phillips curve slope is 0.47, indicating that expressed beliefs lean toward a steep Phillips curve on average, though the median cross-sectional standard deviation of 0.72 indicates substantial heterogeneity in these beliefs.

Table 7 presents illustrative examples from 2011. The excerpts demonstrate how members express parameter beliefs through distinct rhetorical approaches: President Charles Plosser directly casts doubt on relying on output gaps to forecast inflation, implying a flat Phillips curve. President Eric Rosengren and Governor Sarah Raskin link high labor market slack to low or falling inflation, implying a steep Phillips curve. Chairman Bernanke notes that factors other than slack may affect inflation and offset its effects, such as global commodity prices, which we classify as a moderate Phillips curve belief.

3.4.2 Parameter Beliefs and Forecast Responses

The central prediction of model heterogeneity is that members with different parameter beliefs should respond differently to the same economic conditions. A member who perceives a steep Phillips curve slope should forecast higher inflation when labor markets are tight, because they believe tight labor markets generate inflationary pressure. A member who believes the slope is flat should show little such response.

To assess this prediction, we first compute each member's perceived labor market tightness. For member i in meeting t, we define

$$\text{LaborTightness}_{i,t} = u^*_{i,t} - u_t, \tag{2}$$

where u_t is the most recent unemployment rate as of the meeting, and $u^*_{i,t}$ is member i's perceived NAIRU. From April 2009 onwards, members report long-run unemployment forecasts in the Summary of Economic Projections, and we use their most recently reported value as of meeting t. Prior to April 2009, we use the FOMC staff's estimate of NAIRU, which is common across members; in this period, identification of differential responses comes entirely from heterogeneity in Phillips curve slope beliefs.

With our measure of each member's perceived labor tightness, we then estimate re-

gressions of the form

$$f^1_{i,t,\text{inflation}} = \alpha_t + \beta_1 \text{pcslope}_{i,t} + \beta_2 \text{labortightness}_{i,t} + \beta_3 \left(\text{pcslope}_{i,t} \times \text{labortightness}_{i,t} \right) \tag{3}$$

where $f^1_{i,t,\text{inflation}}$ is member i's one-year-ahead inflation forecast at meeting t, α_t is a meeting fixed effect, and $\text{pcslope}_{i,t}$ is the most recently available perceived Phillips curve slope for member i expressed prior to meeting t. We standardize all variables to have zero mean and unit standard deviation, so coefficients can be interpreted as a one-standard-deviation change in inflation forecasts for a one-standard-deviation change in the independent variable. We estimate regressions using two measures of perceived Phillips curve slope: the last extracted value prior to t, and an expanding-window average over all extracted values prior to t. All regressions include meeting fixed effects, so we compare inflation forecasts across members within the same meeting based on their perceived Phillips curve slopes and labor market tightness.

The regression tests two predictions of model heterogeneity. First, the coefficient on labor market tightness (β_2) captures whether members with different NAIRU beliefs, and therefore different perceptions of labor market tightness, forecast inflation differently. Second, the interaction term (β_3) captures whether members with different Phillips curve slope beliefs respond differently to the same perceived labor market conditions. We expect both coefficients to be positive: members who perceive tighter labor markets should forecast higher inflation, and members who believe the Phillips curve is steep should be more sensitive to labor market conditions when forecasting inflation.

Table 8 reports the regression results. The first two columns measure perceived Phillips curve slope using the last extracted value prior to the meeting. The coefficient on labor market tightness is 0.61 with meeting fixed effects (t-statistic of 4.33), indicating that members who perceive tighter labor markets forecast higher inflation. With member fixed effects, this coefficient becomes statistically indistinguishable from zero, consistent with persistent NAIRU beliefs driving forecast differences: within-member variation in perceived tightness does not predict forecasts, but cross-member variation does. The coefficient on the interaction term is 0.13 with meeting fixed effects (t-statistic of 3.51) and 0.10 with member fixed effects (t-statistic of 2.94), indicating that members who perceive a steep Phillips curve forecast higher inflation when labor markets are tight. Unlike the NAIRU result, this effect survives member fixed effects, suggesting that both persistent differences in Phillips curve beliefs and within-member variation in expressed beliefs contribute to forecast heterogeneity. The last two columns measure Phillips curve slope using expanding-window averages and yield similar conclusions: the interaction term remains

	Last PC Slope		PC Slope (EW)	
	(1)	(2)	(3)	(4)
PC Slope × Labor Tightness	0.13	0.10	0.12	0.11
	(3.51)	(2.94)	(4.64)	(4.27)
Labor Tightness	0.61	-0.20	0.56	-0.02
	(4.33)	(0.87)	(4.05)	(0.12)
PC Slope	0.01	0.06	-0.00	0.10
	(0.64)	(2.37)	(0.14)	(2.08)
N	1014	1014	1053	1053
Meeting FE	Yes	Yes	Yes	Yes
Member FE	No	Yes	No	Yes

TABLE 8: MODEL PARAMETERS AND MACROECONOMIC FORECASTS

Note: The table reports coefficients from regressions of members' one-year-ahead inflation forecasts on their Phillips curve beliefs, labor market tightness (their perception of the non-accelerating inflation rate of unemployment, NAIRU, minus the most recent unemployment rate), and an interaction term. PC Slope (EW) denotes the expanding-window average of extracted Phillips curve slope beliefs. *t*-statistics in parentheses are based on standard errors clustered by member and meeting.

significant with both meeting fixed effects (0.12, *t*-statistic of 4.64) and member fixed effects (0.11, *t*-statistic of 4.27), again indicating that the persistent component of members' Phillips curve beliefs drives differences in their inflation forecasts.

Taken together, these results demonstrate that the model parameters we extract from meeting transcripts have predictive power for members' economic forecasts. Members with steep Phillips curve beliefs respond more strongly to labor market conditions when forecasting inflation. Members who believe NAIRU is high forecast higher inflation because they perceive tighter labor markets at any given unemployment rate. These patterns provide direct evidence that persistent forecast disagreements among FOMC members reflect genuine differences in their models of the economy. We next examine whether these disagreements could instead reflect differential private information, and find evidence inconsistent with this alternative.

3.5 Information Sources and Heterogeneous Interpretations

The preceding analysis provides direct evidence that FOMC members hold different beliefs about key model parameters, the slope of the Phillips curve and non-accelerating inflation rate of unemployment (NAIRU). An alternative interpretation, however, is that

persistent disagreements reflect differential private information rather than heterogeneous models. Under this view, members might have access to different data sources, for example, regional business contacts, specialized sectoral knowledge, or proprietary analyses, that lead them to different conclusions even if they share a common model. We present two pieces of evidence against this interpretation. First, members overwhelmingly cite publicly available information in their arguments. Second, members discuss the same economic topics but reach different conclusions, consistent with heterogeneous models for interpreting shared information rather than differential access to information.

3.5.1 Public versus Private Information

We examine what information members cite when making arguments to distinguish between differential information and heterogeneous models. If persistent disagreements primarily reflected private information, we would expect members to frequently cite unique, specialized data sources unavailable to other committee members.

We pass the quotations for each argument where members discuss beliefs about the current or future state of a variable to the LLM and prompt it to identify whether the cited data are qualitative or quantitative; whether they describe regional, national, or international quantities; and whether they reflect private or specialized information that other members would not reasonably know. Public information includes data readily available to all committee members, while private/specialized information includes data from sources unique to individual members or their districts. We construct categories of data by randomly sampling five sets of 500 arguments, prompting the LLM to identify the categories of arguments within each set, and then prompting it to consolidate the resulting five sets of categories. We then prompt the LLM to categorize the data cited in each argument. Internet Appendix Table IA.B.1 lists each of the categories and their descriptions.

Table 9 displays details on the information cited by committee members in their arguments. We identify that 82% of arguments reflect public information, while 18% reflect specialized or private information. About 78% of arguments reference data at the national level; 8% reference data that are specific to a region of the United States; 3% reference international data; and 21% reference data on the performance of a particular sector.[14]

Figure 3 plots the proportion of arguments in different data categories, broken down by the proportion that reflect public versus private information. The most common data category (51% of arguments) is Official government statistics (capturing GDP, CPI, etc.), of

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

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%

TABLE 9: INFORMATION CITED IN MEMBERS' ARGUMENTS

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

which 96% reflect public information. The next most commonly cited category is Federal Reserve Staff Analysis (28% of arguments, e.g., Greenbook/Tealbook forecasts), which also overwhelmingly reflects public information. The only category that primarily reflects private information is District Business Intelligence, e.g., reports produced by regional Federal Reserves or information from regional business contacts.[15]

To formally test whether private information drives disagreements, we compute member-meeting scores using only arguments that cite public information, excluding the 18% of arguments referencing private or specialized sources. If private information were important for explaining persistent heterogeneity, these public-only scores should differ substantially from our main scores. Instead, the correlations exceed 0.93 for each variable (average 0.95), demonstrating that members' persistent disagreements arise from how they interpret shared public information, not from differential private information.

[15]In Appendix Section IA.B.1, we perform a hand-coding exercise of the data categories for a subset of arguments for comparison with the LLM output. The results from this analysis reveal 90% agreement in data categorization, and also reveal that the proportion of private information cited is likely an upper bound or overestimate, relative to a similar classification by human coders.

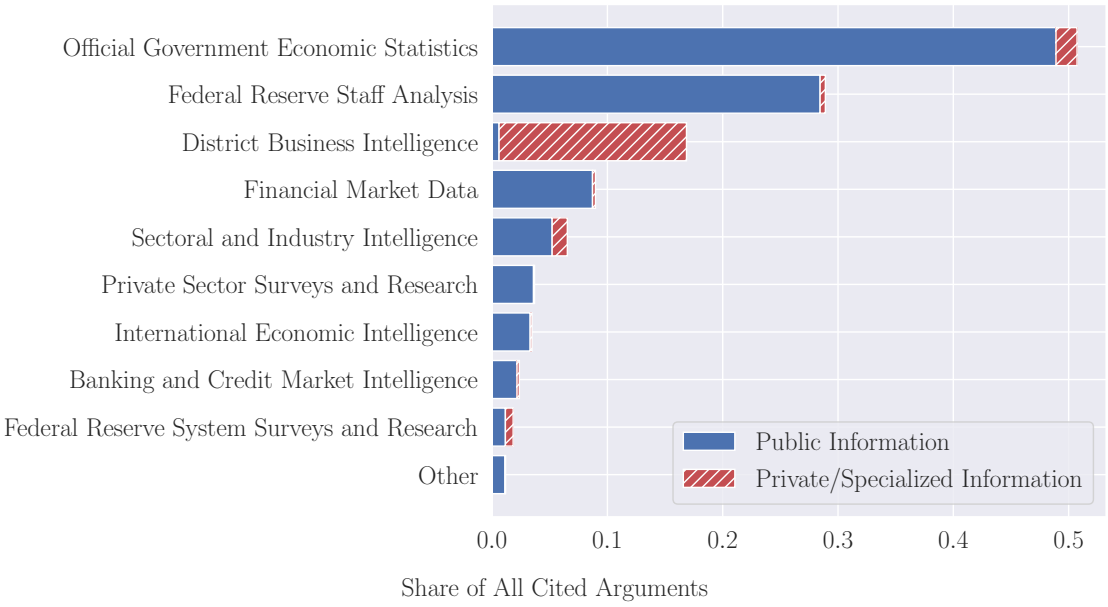


FIGURE 3: DATA CITED IN ARGUMENTS BY CATEGORY

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

3.5.2 Topic Selection versus Heterogeneous Interpretations

The predominance of public information raises the question of how persistent disagreements arise if members observe the same data. One possibility is that members focus on different aspects of the economy, for example, some emphasizing labor markets while others examine financial conditions. An alternative, consistent with model heterogeneity, is that members discuss the same economic mechanisms but interpret them differently. We find strong support for the latter: the vast majority of disagreement stems from differing interpretations of shared topics rather than from topic selection.

We classify each argument into topical categories designed to distinguish between different economic mechanisms while grouping together opposing interpretations of the same mechanism. For example, members interpreting the same strong GDP data as either reflecting structural productivity gains or temporary cyclical dynamics would both fall under *Growth Forecasting and Outlook*. We construct categories by randomly sampling five sets of 500 arguments, prompting the LLM to identify categories within each set, and consolidating the results. We identify 28 categories across the three variables; Appendix IA.B reports the categories and their descriptions.

Figure 4 displays examples of the categories, plotting the average number of arguments per meeting by category and their average scores. The first panel displays the most

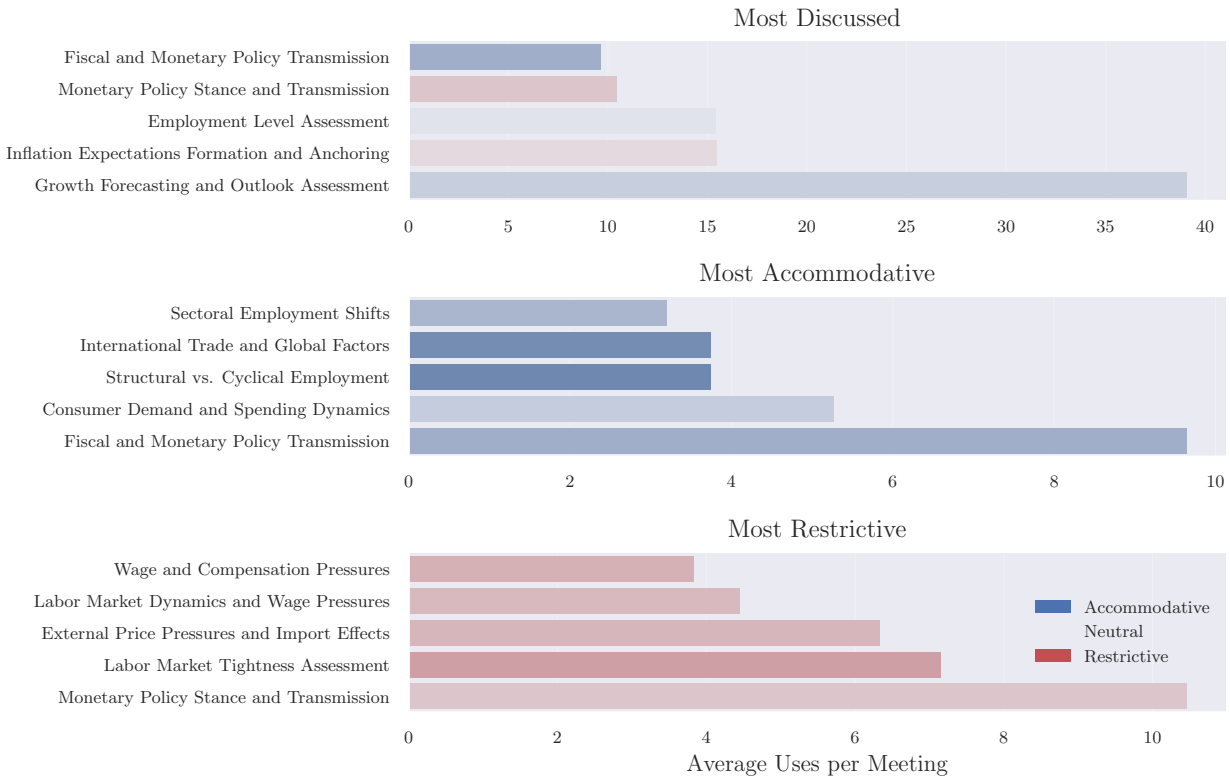


FIGURE 4: ARGUMENT TOPIC CATEGORIES

Note: The figure displays the average number of arguments made per meeting for selected categories. Bar color indicates average policy stance score, ranging from blue (accommodative) through white (neutral) to red (restrictive). Categories are grouped into most frequently discussed, most accommodative (average score < -0.25), and most restrictive (average score > 0.25).

frequently discussed categories, which primarily include reporting of forecasts of growth, inflation, and employment. The second panel plots the categories with the most negative average scores, indicating support for accommodative policy; these are primarily about growth and employment. The third panel plots arguments with the most positive average scores, indicating support for restrictive policy; these include arguments relating variables to inflation.[16]

To quantify the relative importance of topic selection versus interpretation, we decom-

[16]Section IA.C.3 in the Internet Appendix presents time series of argument category proportions by variable. The figures reveal that while some categories receive fairly constant discussion, such as *Growth Forecasting and Outlook Assessment*, others vary cyclically and exhibit secular trends, such as *Financial Market Conditions and Credit*.

pose each member's score into three components. For member i in meeting t, we write:

$$\gamma_{i,t,k} = \underbrace{\overline{\gamma}_{-i,t,k}}_{\text{Selection}} + \underbrace{(\widehat{\gamma}_{i,t,k} - \overline{\gamma}_{-i,t,k})}_{\text{Interpretation}} + \underbrace{(\gamma_{i,t,k} - \widehat{\gamma}_{i,t,k})}_{\text{Residual}}, \tag{4}$$

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.[17] 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. Nearly the entirety of Tarullo and Plosser's differences in argument scores stems from this interpretation gap.

Table 10 presents the variance decomposition across all member-meeting observations. The first row pools arguments across variables: interpretation effects account for 78.0% of total variance in member scores, while selection effects contribute 19.0%. Residual effects from exclusive categories explain only 3.0%, consistent with high topic overlap across members within meetings. The remaining rows demonstrate that this pattern is consistent across variables, with interpretation effects ranging from 80.3% (employment) to 87.6% (growth).[18]

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

[18] Internet Appendix Table IA.C.2 presents the variance decomposition without meeting fixed effects, where interpretation explains 57.2% of score variation and selection explains 40.6%. The increased selection variance is consistent with business cycle variation in topics explaining across-meeting score variation.

	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 10: WITHIN-MEETING VARIANCE DECOMPOSITION OF MEMBER SCORES

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

These results provide additional evidence for the model heterogeneity interpretation. Members attend to the same aspects of the economy—discussing the same topics and citing the same public information—yet reach persistently different conclusions. This pattern aligns with heterogeneous models, where members interpret shared information through different analytical frameworks, rather than with simple differential information models where sharing information should lead to convergence. Having characterized the nature of disagreements among committee members, we now examine how the committee aggregates these diverse views into collective policy decisions.

4 Monetary Policy by Committee

Our evidence indicates that committee members employ persistent heterogeneous models to interpret public data, with disagreements arising from different interpretations of shared information rather than differential private information. We now examine how the committee aggregates these models into collective decisions, focusing on the mechanisms that determine which members' views carry greater weight when persistent disagreements must be resolved.

We analyze this aggregation process using transcript-based measures that capture each member's position and influence across all committee decisions, providing granular variation in alignment and influence beyond existing measures such as voting outcomes. Our analysis reveals that the committee systematically tilts its decisions toward members whose models better align with recent economic data, suggesting that model performance

plays a key role in determining influence within the committee.

4.1 Overview and Framework

To examine how heterogeneous models translate into committee decisions, we extend our earlier stylized framework from Section 3. Each member i employs a persistent model m_i that maps public data D_t into forecasts. These forecasts, through member i's reaction function, yield a policy preference $\gamma_{i,t}$, their desired policy stance given their interpretation of economic conditions.

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.

Prior work typically assumes equal weighting across members (Malmendier, Nagel and Yan (2021)). In contrast, we test whether weights depend on how well each member's model fits recent data. Specifically, we define $\text{Fit}(D_t|m_i)$ as a measure of how well member i's model m_i aligns with recent economic realizations, and test whether better fit leads to greater influence.

We estimate regressions of the form:

$$w_{i,t} = \alpha_t + \eta_i + \beta \times \text{Fit}(D_t|m_i), \tag{6}$$

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

This specification allows us to distinguish between two theories of committee aggregation. 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 $\beta \approx 0$—no systematic relationship between model fit and influence once we control for fixed effects. 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. Crucially, because we include meeting fixed effects, this cannot be explained by all members updating toward the data; it reflects differential weighting across members within each meeting.

To interpret our findings, note that our alignment measure captures $y_{i,t} \propto \text{cov}(\gamma_{i,t}, \sum_j w_{j,t}\gamma_{j,t})$—high

alignment indicates member i's preferences are well-represented in the weighted committee decision. A positive β therefore indicates that better model fit (higher Fit$(D_t|m_i)$) leads to both greater influence ($w_{i,t}$) and higher alignment with final decisions.

4.2 Measuring Members' Influence in Committee Decisions

To test whether model fit determines influence, we require measures of the weight $w_{i,t}$ each member receives in committee decisions. We employ three complementary measures that capture different aspects of influence.

First, we use dissent votes ($Dissent_{i,t}$), obtained from Thornton, Wheelock et al. (2014) and continuously updated by the St. Louis Federal Reserve. A dissent indicates minimal weight in the committee's decision—the member's preference was overruled. Dissents occur in 6.9% of member-meeting observations.

Second and third, we construct transcript-based measures. $Alignment_{i,t}$ ranges from -3 to 3, where 3 indicates perfect alignment 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 (for example, by proposing the adopted policy) and 0 indicates no influence (for example, opposing the decision or not explicitly addressing it despite voting for it). These measures capture alignment and influence across all policy dimensions—target rates, forward guidance, balance sheet policies, and statement language.

To construct the transcript-based measures, we employ a two-step process. First, we identify all policy decisions in each meeting by providing an LLM with the policy discussion section and the pre-circulated policy alternatives. The LLM extracts decisions from both the alternatives and those proposed ad hoc during deliberation, and scores each decision from -3 (highly accommodative) to +3 (highly restrictive) to capture its policy stance.

Second, for each member-meeting pair, we assess that member's position on each decision. We provide the LLM with the policy discussion, alternatives, and identified decisions, then prompt it to score each member's alignment with each decision (-3 to 3) and their influence in its adoption (0 to 3). We construct $Alignment_{i,t}$ and $Influence_{i,t}$ as the average alignment and influence scores across all decisions made in that meeting.

Table 11 provides illustrative examples. Panel A presents the policy decisions and associated scores from the August 9, 2011 FOMC meeting. The LLM identifies five decisions made during the meeting from the policy alternatives (Tealbook B/Bluebook), with the assigned scores in parentheses: maintaining a target Fed Funds rate of 0–25 bps (-2); forward guidance about maintaining the low target rate through mid-2013 (-2); the addi-

tion of language that the committee has discussed and is prepared to employ the range of policy tools as appropriate (-1); and the decision to continue reinvesting principal payments from domestic securities holdings into Treasury securities (-1). The first decision is assigned a score of -2, reflecting significant accommodation at the zero lower bound. The remaining scores are also negative, as they introduce easier monetary policy conditions.

Panel B of Table 11 shows examples of alignment and influence scores for the August 2011 policy decisions for three voting members: Narayana Kocherlakota, Daniel Tarullo, and Governor 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 endorsing Alternative B. Kocherlakota is assigned alignment and influence scores of -3 and 0 for the mid-2013 guidance, which he explicitly argued against, seeing it as too accommodative and responsive to short-term fluctuations.

Daniel Tarullo is assigned alignment and influence scores of 1 and 2 for the mid-2013 forward guidance. The LLM identifies that he expressed concerns about the 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 we supported, but did not offer explicit arguments for. 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—demonstrates that our measures capture gradations of agreement and influence beyond binary voting outcomes. To establish that these measures meaningfully capture policy-relevant variation, we validate them in two ways.

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

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 Alt. 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 11: POLICY CHANGES AND POLICY ALIGNMENT: EXAMPLES

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

ing relevant dimensions of policy decisions. We report the results in Internet Appendix IA.C.2.

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

Table 12 sorts member-meeting observations into bins using $Alignment_{i,t}$ scores, and reports the average $Alignment_{i,t}$, $Influence_{i,t}$, and target rate deviation in each bin. The table also reports the percentage of dissents and number of voter-meeting observations for each bin. The table reveals a strong relationship between the alignment and influence measures, dissents, and target rate deviations. Of voters with $Alignment_{i,t}$ scores in the range [-3,-2], (-2, -1], and (-1, 0] indicating disagreement with the committee's decision, 91.7%, 79.6%, and 42.9% of them dissent, respectively, with dissent decreasing in likelihood by bin. The average target rate deviation similarly displays a generally decreasing relationship with $Alignment_{i,t}$, ranging from 27 basis points for alignment scores between -2 and -1, to 2 basis points for alignment scores between +2 and +3. That is, members who are more misaligned with the committee's policy decision prefer different target Fed Funds rates than those chosen.

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

4.3 Committee Decisions and Model Aggregation

Having established measures of member influence, we now test the hypothesis that members whose models better fit recent data receive greater weight in committee decisions.

We proceed with two sets of analyses. First, we proxy for a member's model using their forecasts of policy-relevant variables, and measure model fit based on how well

	$Alignment_{i,t}$ 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	0.22	0.52	0.83	1.19
Percentage dissenting votes	91.7%	79.6%	42.9%	16.7%	4.6%	0.7%
Target Rate Deviation* (bps)	22	27	19	7	4	2
Number of voter-meetings	12	49	184	516	1028	2195

TABLE 12: DECISION ALIGNMENT, INFLUENCE, AND DISSENT

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

their forecasts explain incoming data. We find that members with better fitting models tend to be better aligned with and more influential with respect to adopted decisions. Second, we specifically link members' alignment and influence over individual decisions with members' model parameters. We show that members' agreement and influence with specific decisions can be explained by their model parameters.

4.3.1 Model Fit and Weight in Committee Decisions

To test the relationship between a member's model fit to recent data and their role in committee decisions, we must first measure how well each member's model m_i aligns with incoming economic observations. We then examine whether members with better-fitting models exhibit greater alignment with, and exert greater influence over, the decision-making process. We construct three measures of model fit.

The first, $\text{Fit}^{\text{TB}}_{i,t}$, captures the proximity of a member's forecasts for inflation, output growth, and unemployment to the corresponding Federal Reserve Board staff forecasts in the Tealbook. Distributed shortly before each FOMC meeting, the Tealbook represents a comprehensive real-time synthesis of macroeconomic conditions and serves as a shared information set for committee discussions (Orphanides, 2001; Coibion and Gorod-

nichenko, 2012). Members whose forecasts align closely with this benchmark interpret incoming data similarly to the Board's dedicated forecasting staff, providing a contemporaneous measure of model fit before outcomes are realized.

Denoting $\mathcal{K} = \{\text{Inflation}, \text{Growth}, \text{Employment}\}$, for each $k \in \mathcal{K}$ and horizon $Y \in \{0, 1\}$, we define

$$\text{Fit}^{\text{TB}}_{i,t,k}(Y) = -\left| f^{\text{SEP}}_{i,t,k}(Y) - f^{\text{TB}}_{t,k}(Y) \right|. \tag{7}$$

Here $f^{\text{SEP}}_{i,t,k}(Y)$ is member i's SEP forecast made at meeting t for variable k at horizon Y, and $f^{\text{TB}}_{t,k}(Y)$ is the corresponding Tealbook forecast. We report results for $Y = 0$ (current-year outcomes) in the main text, and for $Y = 1$ (next-year outcomes) in the appendix.[19] We then rank members cross-sectionally, within each meeting t and variable k, by $\text{Fit}^{\text{TB}}_{i,t,k}(Y)$ (equivalently, by the absolute deviation), so that smaller deviations receive higher ranks and compute $\text{Fit}^{\text{TB}}_{i,t}(Y)$ as the average rank of $\text{Fit}^{\text{TB}}_{i,t,k}(Y)$ across variables.[20] While Fit^{TB} measures model fit by comparing forecasts to a staff benchmark, our second and third measures, $\text{Fit}^{\text{FE,Half}}_{i,t}$ and $\text{Fit}^{\text{FE,imputeQ}}_{i,t}$, capture *revealed* accuracy—they assign higher fit to members whose recent forecasts turned out to be correct once outcomes were realized.

Because FOMC participants primarily forecast annual outcomes, the most readily available measures of forecast errors would remain constant for a given member throughout the calendar year, excluding potentially important within-year variation. The two measures we construct address this challenge in different ways.

To construct $\text{Fit}^{\text{FE}}_{i,t}(H)$, we use half-year forecasts available in the SEP from 2008 onwards. In these releases, 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

$$\text{Fit}^{\text{FE}}_{i,t,k}(H) = -\left| v_k(H) - f^{\text{SEP}}_{i,t-1^{\text{SEP}},k}(H) \right|, \tag{8}$$

where H is the most recently realized half-year as of meeting t, $v_k(H)$ is the corresponding realized value, $t - 1^{\text{SEP}}$ denotes the most recent SEP vintage before meeting t, and $f^{\text{SEP}}_{i,t-1^{\text{SEP}},k}(H)$ is member i's forecast from that vintage. To aggregate to a member-meeting measure, we rank members cross-sectionally within each variable-meeting, with higher ranks corresponding to smaller forecast errors, and compute $\text{Fit}^{\text{FE,Half}}_{i,t}$ as the average rank across variables.

[19] Not every meeting has an SEP (or an associated forecast release). For meetings without a corresponding SEP, we carry forward the most recent available values of $f^{\text{SEP}}_{i,t,k}(Y) - f^{\text{TB}}_{t,k}(Y)$.

[20] The results do not rely on ranking; ranking provides a simple aggregation across variables. In the appendix, we also report specifications that use the raw values $\text{Fit}^{\text{TB}}_{i,t,k}(Y)$.

Our third measure, $\text{Fit}^{\text{FE,impute}}(Q)$, also captures forecast accuracy but addresses two limitations of $\text{Fit}^{\text{FE}}_{i,t}(H)$: it is available for the full sample (not just from 2008 onwards) and incorporates unemployment forecasts. The intuition is that we observe the Tealbook's forecast error for the most recent quarter and credit members who were positioned on the correct side of that error—those whose views, relative to the Tealbook, would have brought them closer to the realized outcome. For each $k \in \mathcal{K}$, we define

$$\text{Fit}^{\text{FE,impute}}_{i,t,k}(Q) = -\left| \underbrace{\left(f^{\text{TB}}_{t-1,k}(Q) - v_k(Q)\right)}_{\text{TB quarterly FE}} + \underbrace{\left(f^{\text{SEP}}_{i,t-1^{\text{SEP}},k}(Y) - f^{\text{TB}}_{t-1^{\text{SEP}},k}(Y)\right)}_{\text{Annual SEP vs. TB gap}} \right|, \tag{9}$$

where Q is the most recently realized quarter as of meeting t, $v_k(Q)$ is the realization, $f^{\text{TB}}_{t-1,k}(Q)$ is the last Tealbook forecast for that quarter, and $Y = 0$ is the current-year annual horizon. For inflation and GDP growth, both terms are annualized, consistent with how the Tealbook and SEP report forecasts. To aggregate to a member-meeting measure, we again rank members cross-sectionally within each variable-meeting and compute $\text{Fit}^{\text{FE,imputeQ}}_{i,t}$ as the average rank.

The annual SEP gap proxies for the member's near-term outlook because it pins down the average quarterly path over the remaining quarters of the calendar year (exactly so in $Q4$). Including the Tealbook's quarterly forecast error incorporates news arriving between t^{SEP} and t, under the assumption that members updated their near-term views in step with the Tealbook.

With these measures of fit in hand, we test whether model fit determines influence by estimating the regressions specified in Section 4.1. Table 13 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 ensure we identify influence from cross-sectional variation in model fit within each meeting. Variables except $Dissent_{i,t}$ are standardized to have zero mean and unit standard deviation. With $Alignment_{i,t}$ and $Influence_{i,t}$ as the dependent variables, coefficients represent standard deviation changes corresponding to a one standard deviation difference in model fit. With $Dissent_{i,t}$ as the dependent variable, coefficients correspond with the increase in dissent probability corresponding to a one standard deviation difference in model fit.

The results support our hypothesis that model fit to recent data predicts members' alignment with and impact on committee decisions. For dissent, a one standard deviation improvement in model fit reduces dissent probability by 2 to 6.9 percentage points

(depending on the fit measure), a substantial effect given the 6.9% baseline dissent rate. For alignment, coefficients range from 0.05 to 0.19 standard deviations, indicating that members with better-fitting models see their preferences substantially better reflected in committee decisions. For influence, coefficients range from 0.03 to 0.15 standard deviations, confirming that these members actively shape decisions rather than merely agreeing with them. In Appendix Table IA.C.4, we report regression results using fit measures constructed individually for each of inflation, growth, and unemployment. While there is some variation across the variables, the results are broadly consistent across variables.

4.4 Link to Monetary Policy Surprises and Shocks

Our results show that the committee's decisions tilt toward the preferences of members whose models better fit recent data. In Appendix IA.C.2, we link our policy decisions data to monetary policy surprises, measured as the high-frequency price response of interest rate futures in the 30-minute window following policy announcements (Gürkaynak, Sack and Swanson, 2005; Nakamura and Steinsson, 2018). This relationship underscores that the committee's tilting behavior has a material effect on policy, with the resulting decisions surprising financial markets. The result is also consistent with recent evidence of the FOMC using a time-varying policy rule (Hack, Istrefi and Meier, 2023), and suggests that monetary policy shocks may reflect "change in the political power of individuals on the FOMC," (Ramey, 2016), with such change driven by the fit of members' models to recent data.

5 Theoretical Framework

Our empirical analysis reveals that committee members employ persistently heterogeneous models to interpret largely public information, with the committee systematically favoring those whose models better fit recent data. To explore the implications of this finding, 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.

	Panel A: Tealbook-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{TB}}_{i,t,k}(Y)$	-3.13	0.087	0.092	0.120	0.088
	(2.85)	(2.97)	(3.49)	(3.55)	(3.63)
N	2179	2179	2179	3658	3658
Meeting FE	Yes	Yes	Yes	Yes	Yes
	Panel B: Half-year Forecast Error-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE}}_{i,t}(H)$	-6.89	0.185	0.146	0.170	0.110
	(2.44)	(2.97)	(2.49)	(2.91)	(2.79)
N	873	873	873	1456	1456
Meeting FE	Yes	Yes	Yes	Yes	Yes
	Panel C: Quarterly Imputed Forecast Error-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE,impute}}_{i,t}(Q)$	-2.23	0.048	0.043	0.052	0.031
	(2.77)	(2.24)	(2.03)	(2.08)	(1.70)
N	2123	2123	2123	3572	3572
Meeting FE	Yes	Yes	Yes	Yes	Yes

TABLE 13: MODEL FIT AND POLICY DECISION ALIGNMENT

Note: The table reports estimates from the regressions relating a member's model fit to recent data to their role in FOMC decisions. There are three dependent variables: $Dissent_{i,t}$ (an indicator for dissenting votes), $Alignment_{i,t}$ (a standardized measure of agreement with the adopted decision), and $Influence_{i,t}$ (a standardized measure of a member's role in shaping the decision). Variables except $Dissent_{i,t}$ are standardized, so coefficients in the $Alignment_{i,t}$ and $Influence_{i,t}$ columns represent standard-deviation changes associated with a one standard-deviation increase in model fit, while coefficients in the $Dissent_{i,t}$ columns represent percentage-point changes in dissent probability. Panel A uses Tealbook-based fit, Fit^{TB}, defined as the negative absolute deviation between member SEP forecasts and Tealbook forecasts for inflation, output growth, and unemployment (reported for current-year and aggregated via within-meeting ranks). Panel B uses half-year forecast-error fit, $\text{Fit}^{\text{FE}}(H)$, based on realized half-year outcomes for inflation and growth (available from 2008 onward). Panel C uses quarterly-imputed forecast-error fit, $\text{Fit}^{\text{FE,impute}}(Q)$, which combines the Tealbook's most recent quarterly forecast error with the member's SEP–Tealbook annual forecast gap to proxy near-term accuracy and is available for the full sample. "Voters" columns restrict the sample to voting members; "All Members" columns include all participants. N denotes the number of member-meeting observations. t-statistics are reported in parentheses.

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.[21]

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

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

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

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

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

$$v_i = \alpha_i m_i + (1 - \alpha_i) m_j, \text{ where } j \neq i \text{ and } \alpha_i \in [0, 1]. \tag{10}$$

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

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

of the committee, is given by averaging the members' advocated models:

$$m = \frac{1}{2}(v_A + v_B) = w_A m_A + (1 - w_A) m_B, \tag{11}$$

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

$$U_i(\alpha_i; \alpha_j) = \log \mathcal{L}(D_t|m) - \lambda KL(m_i \| v_i), \tag{12}$$

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

The first term, $\log \mathcal{L}(D_t|m)$, captures that members want a chosen committee model m that explains the data well, all else equal. This term may reflect that members are aware that their models are misspecified, and want to use data in the face of model uncertainty. Even if members are dogmatic that their own models are correct, such a term may still arise. For example, it may reflect difficulty advocating for one's own model when it is misaligned with the observed data, or it may reflect that the committee must justify its decision to constituencies that also observe D_t. This term leads the committee to tilt its decision toward models that better fit D_t, consistent with the empirical evidence.

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

5.2 Equilibrium

We study pure strategy Nash equilibria.

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

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

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

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

Proof. All proofs are provided in Appendix A. □

5.3 Committee Performance

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

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

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

Proposition 2 provides a necessary condition for committees to outperform their best member, namely that each member's model explains part of the state space better than the other's. This result is independent of the relative performance of the models. For example, a member B may have a model that is substantially more misspecified than member A's model, but as long as the models explain different parts of the state space, there can be benefits to combining them.[22]

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

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

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

on members' models are biased relative to the optimum (except when $w_A^* = \frac{1}{2}$). For intermediate values, as λ increases, the equilibrium model weights exhibit more bias and less variance relative to the optimum.

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

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

$$\Delta \equiv \mathbb{E}_\mu(\log(w_A^* m_A + (1 - w_A^*)m_B)) - \max\{\mathbb{E}_\mu(\log m_A), \mathbb{E}_\mu(\log m_B)\}, \tag{13}$$

$$\beta(\lambda) \equiv \mathbb{E}_\mu(w_A(\lambda) - w_A^*), \tag{14}$$

$$\sigma^2(\lambda) \equiv \mathbb{V}_\mu(w_A(\lambda)), \text{ and} \tag{15}$$

$$c \equiv -\mathbb{E}_\mu\left(\frac{\partial^2}{\partial w_A^2} \log \mathcal{L}(D_t | w_A m_A + (1 - w_A)m_B)|_{w_A = w_A^*}\right), \tag{16}$$

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

$$\Delta > \frac{c}{2}(\beta(\lambda)^2 + \sigma^2(\lambda)), \tag{17}$$

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

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

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

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

$$\mathbb{V}_\mu(w_A(0)) > \left(\frac{1}{2} - w_A^*\right)^2, \tag{18}$$

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^*,

$$\mathbb{E}_\mu\left[(w_A - w_A^*)^2\right] = \underbrace{(\mathbb{E}_\mu(w_A - w_A^*))^2}_{\text{Bias}} + \underbrace{\mathbb{V}_\mu(w_A)}_{\text{Variance}}. \tag{19}$$

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

Remark 1 (Model Variance and Communication) *Here, the cost of variation in the weights assigned to each member's model from randomness in the data is purely statistical. There may also be other costs, for example, related to time-inconsistency and the committee's credibility in the eyes of market participants, as emphasized in the rules versus discretion debate (Kydland and Prescott, 1977; Barro and Gordon, 1983). Variation in members' weights may also make it difficult for the committee to communicate its future stance, consistent with our empirics relating the committee's tilting behavior to monetary policy surprises, potentially stemming from market participants not understanding the FOMC's reaction function.*[23]

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

[23]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.

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 (12) to be

$$U_i(\alpha_i; \alpha_j) = \log \mathcal{L}(D_t|m) - \lambda KL(m_i\|v_i) - \gamma KL(m\|v_i). \tag{20}$$

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

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

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

5.5 Implications for Committee Design

Our framework illustrates novel tradeoffs and practical implications with committee decision-making of the form we document at the FOMC, where the committee's decisions tilt toward the preferences of members whose models better fit recent data. Committees may improve decision-making, but member selection is important, as potential benefits require members with a diversity of models who are sufficiently willing to entertain one another's views. One way to push members in such a direction is via structured deliberation, for example, as in FOMC meetings, where members discuss their interpretation of economic conditions and policy views before voting.

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

Lastly, our framework sheds light on policies affecting the cost of dissent, such as external transparency of deliberations. Increasing dissent costs can offset members' ten-

dency to overly advocate for their own models but may increase the possibility of converging to a sub-optimal model and decrease the predictability of decisions.

6 Conclusion

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

A Proofs

Proof of Proposition 1

Proof. The game is a potential game with potential function

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

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

Proof of Proposition 2

Proof. Denote

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

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

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

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

Proof of Proposition 3

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

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

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

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

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

Proof of Corollary 1

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

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

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

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

Proof of Proposition 4

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

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

References

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

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

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

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."

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

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

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

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

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

Binetti, Alberto, Francesco Nuzzi, and Stefanie Stantcheva. 2024. "People's understanding of inflation." *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.

Bonhomme, Stéphane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler. 2023. "How much should we trust estimates of firm effects and worker sorting?" *Journal of Labor Economics*, 41(2): 291–322.

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

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

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

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

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

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

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.

Clarida, Richard, Jordi Gali, and Mark Gertler. 1999. "The science of monetary policy: a new Keynesian perspective." *Journal of Economic Literature*, 37(4): 1661–1707.

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

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

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 Deci-

sions." *Paris: Imprimerie Royale*, 1785.

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.

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

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.

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

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

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

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

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

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

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

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

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

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

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

Ke, Shikun. 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.

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."

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

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

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.

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

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

Persico, Nicola. 2004. "Committee design with endogenous information." *The 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. 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.

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

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

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

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

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

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

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

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.

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.

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

Internet Appendix for

Policy by Committee

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

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

IA.A Data Appendix

IA.A.1 Processing FOMC Data

FOMC transcripts are provided in PDF form on the FOMC "Historical Materials by Year" page.[24] 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

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

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

In the third step, we construct a dataset of full names, affiliations, and the associated titles for all members of the Board of Governors as well as the presidents of the twelve regional Federal Reserve banks. We use data from the Federal Reserve History website[25] 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 demo-

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

graphics for all participants who ever vote in any of the meetings in our sample. In particular, we collect their birthdates, genders, party affiliations, if any, undergraduate and graduate institutions, as well as their degrees. For members of the Board, we also record the names of the presidents who nominated them, as well as the names of the presidents who re-nominated them, if any.

IA.A.2 LLM Prompts

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

- Prompts to identify and score committee members' arguments for each variable are in Section IA.A.2.1.
- Given their arguments, prompts to extract beliefs about the Phillips curve are presented in Sections IA.A.2.2.
- Given their arguments, prompts to identify the data cited in committee members' arguments are in Section IA.A.2.3.
- The prompts to assign each argument made by each member to a category are given in Section IA.A.2.4.
- The prompt to identify each policy decision made during the meeting is in Section IA.A.2.5.
- Given the policy decisions made, the prompts to identify each member's alignment with the adopted decisions and their influence on those decisions' adoption are in Section IA.A.2.6.

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

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

We turn our attention specifically to arguments identified 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.A.2.2 Prompts to Extract Phillips Curve Beliefs

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 when a speaker expresses a view about how labor market conditions affect inflation outcomes. We define three belief categories: *steep* (strong positive relationship between labor market tightness and inflation), *flat* (weak or no relationship), and *moderate* (qualified or partial relationship). The prompt includes automatic classification triggers for explicit Phillips curve mentions, detailed classification rules with strong indicator phrases for each category, extensive examples, and special case handling for forecasts, historical references, and policy discussions. The model returns a JSON 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"""
```

IA.A.2.3 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 for 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.A.2.4 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.A.2.5 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.A.2.6 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 policiesI 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.
"""
```

IA.B Categories of Arguments and Data Sources

Table IA.B.1 presents categories and associated descriptions used to categorize data sources cited by committee members in their arguments. Tables IA.B.2, IA.B.3, and IA.B.4 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.B.1: CATEGORIES OF DATA SOURCES CITED BY MEMBERS

Note: The table lists different categories of data cited by committee members during FOMC meetings. These categories and descriptions are generated by the LLM by providing it with a sample of 1000 arguments

made by committee members and prompting it to identify relevant categories. The categories and descriptions are then provided to the LLM in additional prompts, along with each of the arguments made, in order to categorize the data cited by each member in each argument.

Category	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.B.2: 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.B.3: 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.B.4: CATEGORIES OF EMPLOYMENT ARGUMENTS

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

IA.B.1 Validation of LLM Classifications

We validate our Claude-generated scores and classifications using two approaches: manual human coding and cross-validation with an alternative LLM. Ludwig, Mullainathan and Rambachan (2025) establish that using LLM outputs for estimation problems requires collecting benchmark data to empirically model potential measurement error. Without such validation data, plug-in estimation using LLM outputs recovers the target parameter if and only if there is no error in the LLM's outputs. We therefore implement a systematic validation exercise with human coding of a random sample of arguments.

Our validation methodology employs a two-stage random sampling procedure to ensure representativeness across speakers, meetings, and time periods. From our full dataset, we randomly sample 300 arguments per variable (growth, inflation, employment). The analyses in our paper are done at the member-meeting level (aggregated across all arguments made by the member in a given meeting), so we follow this structure when sampling. We first group arguments by member-meeting, randomly sample at the member-meeting level, and then take all arguments from selected meeting-member observations.[26] For the growth variable analyzed here, we obtain 293 arguments from 51 unique member-meeting combinations, with a mean of 5.7 arguments per member-meeting.

We develop a coding interface that presents arguments to human coders without revealing Claude's classifications (Figure IA.B.1). Two expert coders independently code all 293 growth arguments along four primary dimensions following the measures constructed by Claude. First, coders assign an outlook score from -3 to +3, capturing the argument's directional outlook on the economic variable, where -3 indicates strong negative outlook and +3 indicates strong positive outlook. Second, coders determine whether the speaker cited specific data or information, using the same criteria in the Claude prompt. When relevant, the coders then classify the data sources used in the argument using the same categories in the Claude prompt. Third, for arguments citing data, coders classify the information as either public (government statistics, market prices) or private/specialized (internal Fed analysis, business contacts). Fourth, coders assign each argument to one of the argument categories specific to the growth variable, such as "Growth Forecasting and Outlook Assessment" or "Consumer Demand and Spending Dynamics." The interface includes expandable reference panels with category definitions and coding guidelines to ensure consistency.

Inter-rater reliability between our two human coders is high. For the continuous outlook scores, we observe a correlation of 0.837 with a mean absolute difference of 0.567 points and exact agreement on 52.6% of arguments. For the binary classification of whether data are cited, agreement reaches 84.0%. These high agreement rates are consistent with coding protocols that are reliable.

To assess systematic bias in Claude's scoring, we estimate bias correction models following Ludwig, Mullainathan and Rambachan (2025). Specifically, we regress human

[26] We require that a member makes at least 3 arguments in a given meeting to be considered in our sample. When a speaker makes more than 10 arguments in a given meeting, we randomly sample 10 arguments made by the speaker in the meeting. Otherwise, we take all arguments.

Level of Analysis	N	α (SE)	β (SE)	R^2	Correlation
Member-Meeting Level					
Coder 1	51	0.091 (0.068)	0.885*** (0.060)	0.816	0.903
Coder 2	51	0.212*** (0.056)	0.809*** (0.050)	0.842	0.918
Average	51	0.151** (0.055)	0.847*** (0.053)	0.859	0.927
Argument Level					
Coder 1	293	0.037 (0.059)	0.688*** (0.034)	0.588	0.767
Coder 2	293	0.153*** (0.054)	0.642*** (0.031)	0.598	0.773
Average	293	0.095 (0.051)	0.665*** (0.029)	0.645	0.803

TABLE IA.B.5: BIAS CORRECTION MODEL ESTIMATES

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05. The null hypotheses are $\alpha = 0$ and $\beta = 1$.

scores on Claude scores:

$$V_H = \alpha + \beta \hat{V} + \varepsilon \tag{IA.B.1}$$

where V_H denotes the human score, $\hat{V}$ denotes Claude's score, and under the null hypothesis of no bias, we expect $\alpha = 0$ and $\beta = 1$. Table IA.B.5 presents the results at both the member-meeting level and argument level.

Since our main analysis operates at the member-meeting level, we focus on validating Claude's classifications after aggregating arguments to this level. The correlations at the member-meeting level range from 0.90–0.93, with R^2 values of 0.82–0.86, and the bias coefficient moves closer to one ($\beta = 0.847$ for averaged scores). This improvement from the argument-level results (correlations of 0.767 to 0.803) suggests that aggregation reduces measurement noise, with individual coding errors partially canceling out when averaged within speaker-meetings. Based on the averaged human scores at the member-meeting level, which show the highest correlation with Claude (0.927) and best model fit ($R^2 = 0.859$), we apply the correction formula to our full dataset:

$$\text{Score}_{\text{corrected}} = 0.151 + 0.847 \times \text{Score}_{\text{Claude}} \tag{IA.B.2}$$

Our validation reveals that Claude's classifications do contain measurement error, but also suggests that our estimates of effect magnitudes may be conservative. The convergence of three independent measurement systems—human expert coding, Claude, and Llama 70B—at correlations exceeding 0.90 at the member-meeting level provides strong evidence that our approach captures genuine variation in FOMC members' economic views rather than algorithmic artifacts.

For categorical classifications, we find substantial agreement between human coders and Claude across all dimensions (Table IA.B.6). Argument categories show 72–74% exact match rates with Cohen's kappa of 0.64–0.67, indicating substantial agreement beyond chance. Data category identification achieves 90% overall agreement across the eleven possible categories, though with systematic patterns of disagreement between the human coders and Claude: Claude over-identifies Federal Reserve Staff Analysis and Fi-

		Argument Category		Data Category	Information Type	
Coder	N	Match %	κ	Agree %	Match %	κ
Coder 1	293	74.4	0.669	89.8	86.7	0.507
Coder 2	293	72.0	0.639	90.4	84.8	0.432

TABLE IA.B.6: CATEGORY AGREEMENT SUMMARY

Note: N = number of arguments coded. Argument Category and Information Type show exact match percentages and Cohen's κ. Data Category shows overall agreement across all 11 categories (multiple categories can apply to each argument). For interpretation, $\kappa > 0.60$ indicates substantial agreement; $0.40 < \kappa < 0.60$ indicates moderate agreement. Cites Data agreement: 84.0% across all coders.

nancial Market Data, while under-identifying Official Government Statistics. For information type classification, agreement reaches 85–87%, though Claude systematically classifies more arguments as containing private/specialized information compared to human coders (19% vs. 11%).

Dimension	N	Human %	Claude %	p-value
Overall	570	55.4	44.6	0.011
By Classification Type:				
Cites Data	34	17.6	82.4	<0.001
Argument Category	74	47.3	52.7	0.610
Score	167	57.5	42.5	0.054
Data Category	262	58.4	41.6	0.006
Information Type	33	78.8	21.2	<0.001

TABLE IA.B.7: JUDGE ADJUDICATION RESULTS BY DIMENSION

Note: Shows percentage of cases where blind judge agreed with human coder versus Claude. P-values from two-sided binomial tests with a null hypothesis of a 50% agreement rate.

To provide additional validation and understand dimension-specific performance, we implement a novel blind adjudication exercise. When the human coder and Claude disagree on any classification dimension, we present both options to an independent judge who is blind to which option represents human versus Claude coding (Figure IA.B.2). The judge evaluates the original argument text and selects which classification is more appropriate, with option order randomized for each judgment to prevent systematic bias.

We identify 570 disagreements between human coding and Claude classifications across all dimensions. The judge interface presents these disagreements sequentially by argument, maintaining the same context as the original coder. For continuous scores, the judge

sees both values and selects the more appropriate one. For categorical variables, the judge sees full descriptions of both category options. The judge can select either option, indicate both are equally valid, indicate neither is correct, or skip difficult cases.

Table IA.B.7 presents the adjudication results. Overall, the judge agrees with the human coder 55.4% of the time versus Claude 44.6%. This indicates that Claude's performance is reasonably competitive with expert human coding. However, performance varies substantially across dimensions. Claude does relatively well at identifying data citations, with the judge agreeing with Claude 82.4% of the time ($p<0.001$). For argument categorization, the judge shows no significant preference (52.7% for Claude vs. 47.3% for human, $p = 0.610$), indicating comparable performance. Relative to human coding, Claude may overstate private information type assignments when classifying private versus public information, where the judge agrees with the human coder 78.8% of the time ($p<0.001$).[27]

IA.B.1.1 Inter-LLM Comparison as Robustness Check

As a robustness check, we assess whether our findings are specific to Claude or general to LLM-based classification by comparing Claude's classifications with two alternative large language models: Llama 3 7B and Llama 3 70B. Using the same 293-argument validation sample, we examine inter-LLM agreement on both quantitative scores and categorical classifications.

Model	N	α (SE)	β (SE)	R^2	Correlation
Member-Meeting Level					
Llama 7B	51	$-0.307 * *$ (0.092)	0.910 (0.083)	0.713	0.844
Llama 70B	51	0.053 (0.048)	0.757*** (0.043)	0.862	0.929
Argument Level					
Llama 7B	293	$-0.352 * **$ (0.081)	0.743*** (0.047)	0.472	0.687
Llama 70B	293	0.021 (0.048)	0.634*** (0.027)	0.656	0.810
For comparison:					
Human Average (MM)	51	0.151** (0.055)	0.847*** (0.053)	0.859	0.927

TABLE IA.B.8: INTER-LLM SCORE AGREEMENT: LLAMA VS. CLAUDE

Note: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$. The null hypotheses are $\alpha = 0$ and $\beta = 1$. MM = Member-Meeting level.

Table IA.B.8 presents alignment models regressing Llama scores on Claude scores at both the member-meeting and argument levels. Most notably, at the member-meeting level, the larger Llama 70B model achieves a correlation of 0.929 with Claude, exceeding the human-Claude correlation. This strong inter-LLM agreement on quantitative scores

[27] Analysis of these disagreements reveals that Claude systematically misclassifies certain public data sources as private/specialized information, particularly when Official Government Statistics are discussed in the context of Federal Reserve analysis. This pattern is consistent with our earlier finding that Claude over-identifies Federal Reserve internal sources while under-identifying government statistics.

suggests that our scoring approach captures meaningful economic content rather than model-specific artifacts.

The member-meeting level analysis reveals that aggregation dramatically improves the alignment between Llama models and Claude. Llama 70B's correlation increases from 0.810 at the argument level to 0.929 at the member-meeting level, exceeding the human-Claude correlation. Similarly, Llama 7B's correlation improves from 0.687 to 0.844. This improvement confirms that measurement error at the individual argument level is largely random and cancels out when aggregating to higher-level units. The persistent slope compression ($\beta = 0.757$ for Llama 70B at member-meeting level) indicates that bias correction remains necessary even at aggregated levels.

The models show divergent performance on categorical classifications. For information type classification (public vs. private/specialized), Llama 70B achieves 74.5% agreement with Claude (Cohen's $\kappa = 0.391$), indicating fair agreement, though it tends to over-identify private information (36.2% vs. Claude's 20.3%). Llama 7B achieves 79.7% agreement but with a critical flaw: it never classifies any argument as containing private/specialized information (100% classified as public, Cohen's $\kappa = 0$), suggesting no discriminatory ability. For argument categorization, Llama 70B shows weak agreement with Claude (37.2% exact match, $\kappa = 0.294$), while Llama 7B assigns the "Growth Forecasting" category to 98% of arguments, suggesting minimal discriminatory ability. For data citation identification, Llama 7B achieves 67.2% agreement with Claude while Llama 70B severely under-identified citations (41.6% agreement), with the latter identifying data in only 25.3% of arguments compared to Claude's 77.5%.

These inter-LLM comparisons provide two key insights. First, the high correlation between Llama 70B and Claude on quantitative scores at the member-meeting level (0.927), which matches the human-Claude correlation, supports the robustness of our scoring approach across different LLM architectures. Second, the dimension-specific performance variations—with Llama 70B excelling at scores but showing weak to fair performance at categorical coding (κ ranging from 0.170 for data categories to 0.391 for information type)—underscore that model size alone does not guarantee superior performance across all tasks, highlighting the importance of our human validation exercise for ensuring reliability across all classification dimensions.

The validation exercises collectively demonstrate that our LLM-based classifications are reliable for empirical analysis when properly corrected. The systematic nature of the biases we identify—compression of the score scale and specific patterns in categorical classifications—allows for straightforward correction. The high correlations between human and Claude scores, particularly at the member-meeting level (0.93), combined with the judge's assessment showing competitive overall performance and the robustness check showing similar patterns across different LLMs, support the validity of our approach. Most importantly, the substantial improvement in validation metrics when aggregating to the member-meeting level aligns well with our paper's primary unit of analysis, where measurement errors partially cancel out through averaging.

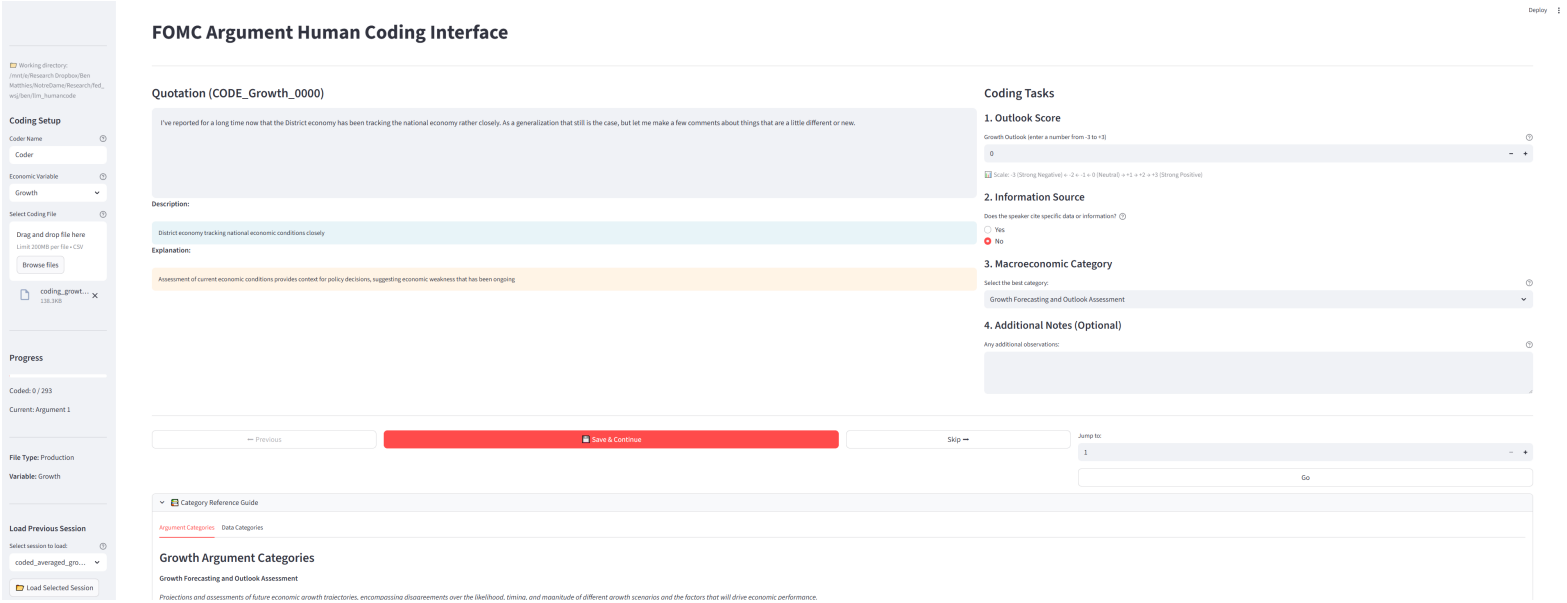


FIGURE IA.B.1: HUMAN CODING INTERFACE FOR FOMC ARGUMENTS

Note: The interface presents arguments individually with coding tasks for: (1) outlook scores ranging from -3 to +3, (2) identification of data citations with source classification, and (3) categorization of argument types. Reference panels with definitions are accessible via expandable sections.

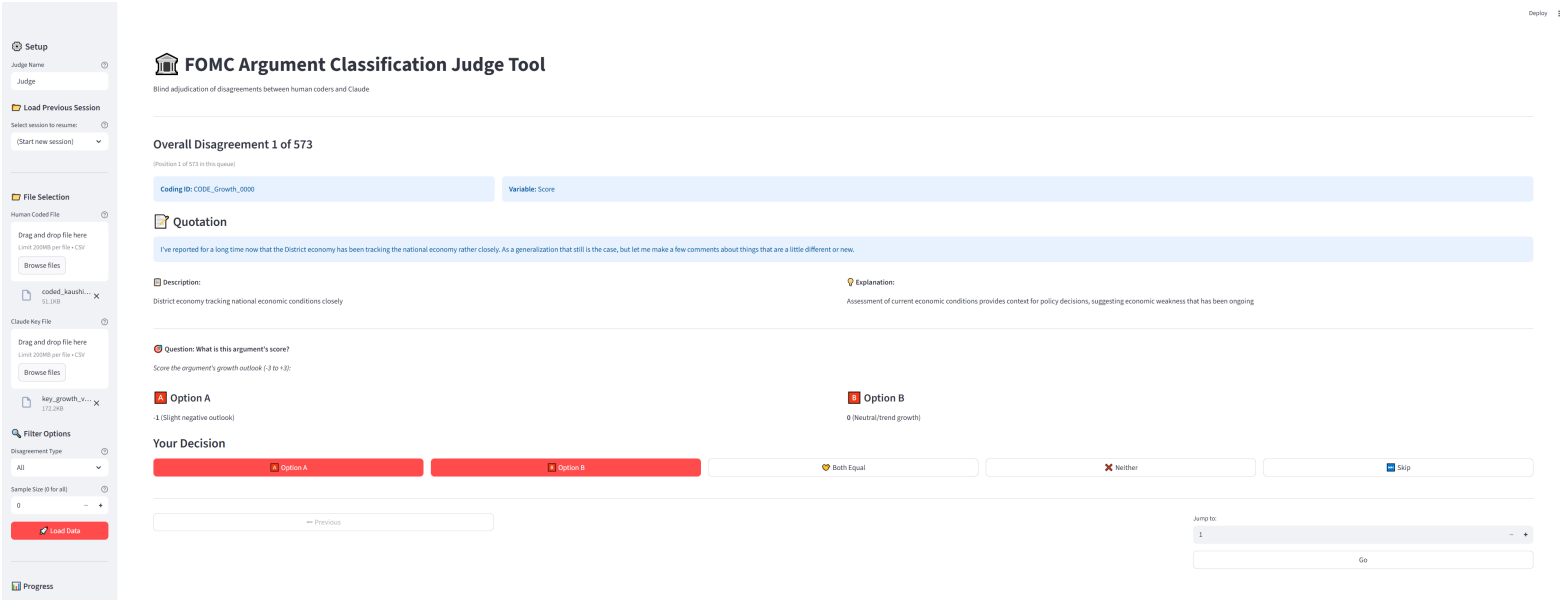


FIGURE IA.B.2: BLIND ADJUDICATION INTERFACE FOR CLASSIFICATION DISAGREEMENTS

Note: The interface presents arguments individually with coding tasks for: (1) outlook scores ranging from -3 to +3, (2) identification of data citations with source classification, and (3) categorization of argument types. Reference panels with definitions are accessible via expandable sections.

IA.C Additional Empirical Analyses

IA.C.1 Time-Series Properties of Phillips Curve Beliefs

Figure IA.C.1 displays the evolution of mean parameter beliefs over time. Phillips curve beliefs were predominantly steep, though highly volatile through the 1970s and 1980s, reflecting the prevailing view that unemployment-inflation tradeoffs were substantial. Beliefs declined in the 1990s and shifted sharply toward flat from 2017–2019, with the mean turning negative in the 2010s. This timing coincides with the "missing inflation" puzzle—the observation that inflation remained subdued despite historically low unemployment.

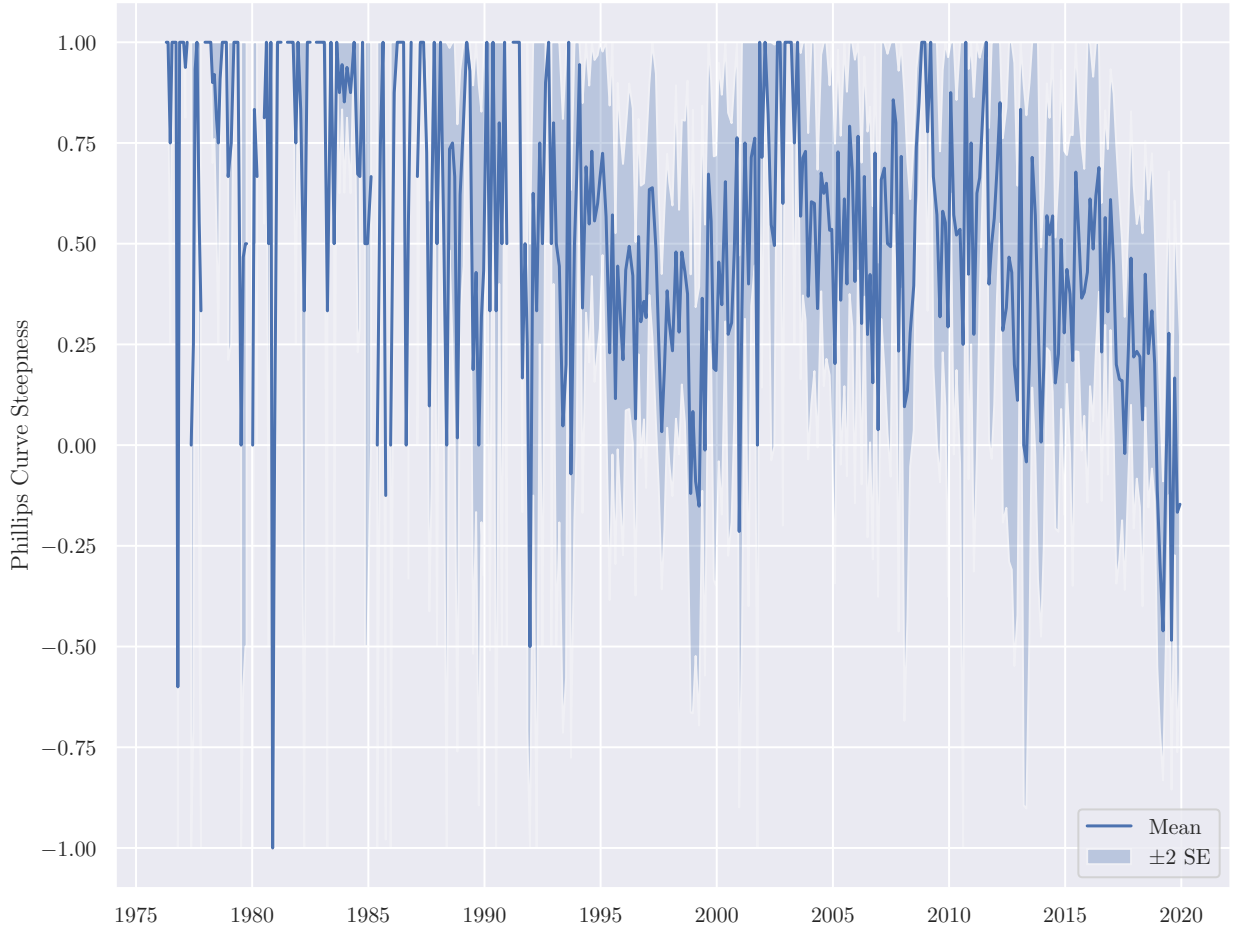


FIGURE IA.C.1: EVOLUTION OF FOMC MEMBERS' PERCEIVED PHILLIPS CURVE SLOPES OVER TIME.

Note: The figure plots the annual mean belief score (solid line) with ±2 standard deviation bands (shaded region). Phillips curve beliefs are coded from −1 (flat) to +1 (steep).

IA.C.2 Policy Decision Scores and Monetary Policy Surprises

In this section, we link the monetary policy decisions we identify from the transcripts with monetary policy surprises measured from high-frequency asset price responses to monetary policy announcements. This link verifies our LLM-based method for measuring policy decisions, and highlights that the policy decisions with respect to which we measure alignment in our main analysis are highly relevant for understanding market perceptions of monetary policy.[28]

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 of n-quarter ahead Eurodollar futures (labeled EDn), as well as the near- and 3-month maturity Fed Funds futures in the 30-minute window following the FOMC announcement (labeled FF and $FF3m$).[29] The sample of Fed Funds futures runs from 1995-2019.

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.[30] 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.C.1 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 0.7 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.

The mps_t variable captures the component of monetary surprises stemming from both the target rate decision and other elements of the committee's decision, most importantly forward guidance about the future path of interest rates. In columns 3-6 of Table IA.C.1, we explicitly separate the surprise of the target rate decision from forward guid-

[28] 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.

[29] 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.

[30] 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.

ance, 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 $ED4$. 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 isn't just measuring internal deliberation but is capturing behavior reflected in financial market movements to the resulting policy.

Second, taken in conjunction with the evidence on the committee's tilt towards different members' models based on the data, the results also suggest that the committee structure may play an important role in generating monetary policy surprises. The committee's tilting behavior is consistent with it not using a fixed policy rule to make a decision, as argued by Hack, Istrefi and Meier (2025), with shifts in the policy stance arising from the shifting influence of committee members coming as a surprise to investors. The link with monetary policy surprises is also complementary to mounting evidence of monetary policy surprises arising from market participants having uncertainty about the Fed's reaction function in response to macroeconomic news (Bauer and Swanson, 2023*a*,*b*; Bauer, Pflueger and Sunderam, 2024). Consistent with this interpretation, columns 7 and 8 of Table IA.C.1 report results from regressions of mps_{orth}, measured as the residual of mps after controlling for macroeconomic and financial news arriving prior to the FOMC meeting, on the $Avg(\{PolicyScores\})$ variables. These results indicate a negligible relationship, suggesting that the component of the $Avg(\{PolicyScores\})$ variables that is related to monetary policy surprises stems from investors being surprised by the committee's response to macroeconomic and financial news.

	$y = mps$		$y = ED4$		$y = FF3m$		$y = mps_{orth}$	
	(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)
$ED1$			0.92	1.02				
			(8.18)	(16.89)				
FF					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.C.1: POLICY DECISION SCORES AND MONETARY POLICY SURPRISES

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

IA.C.3 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 theme τ, we calculate:

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

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

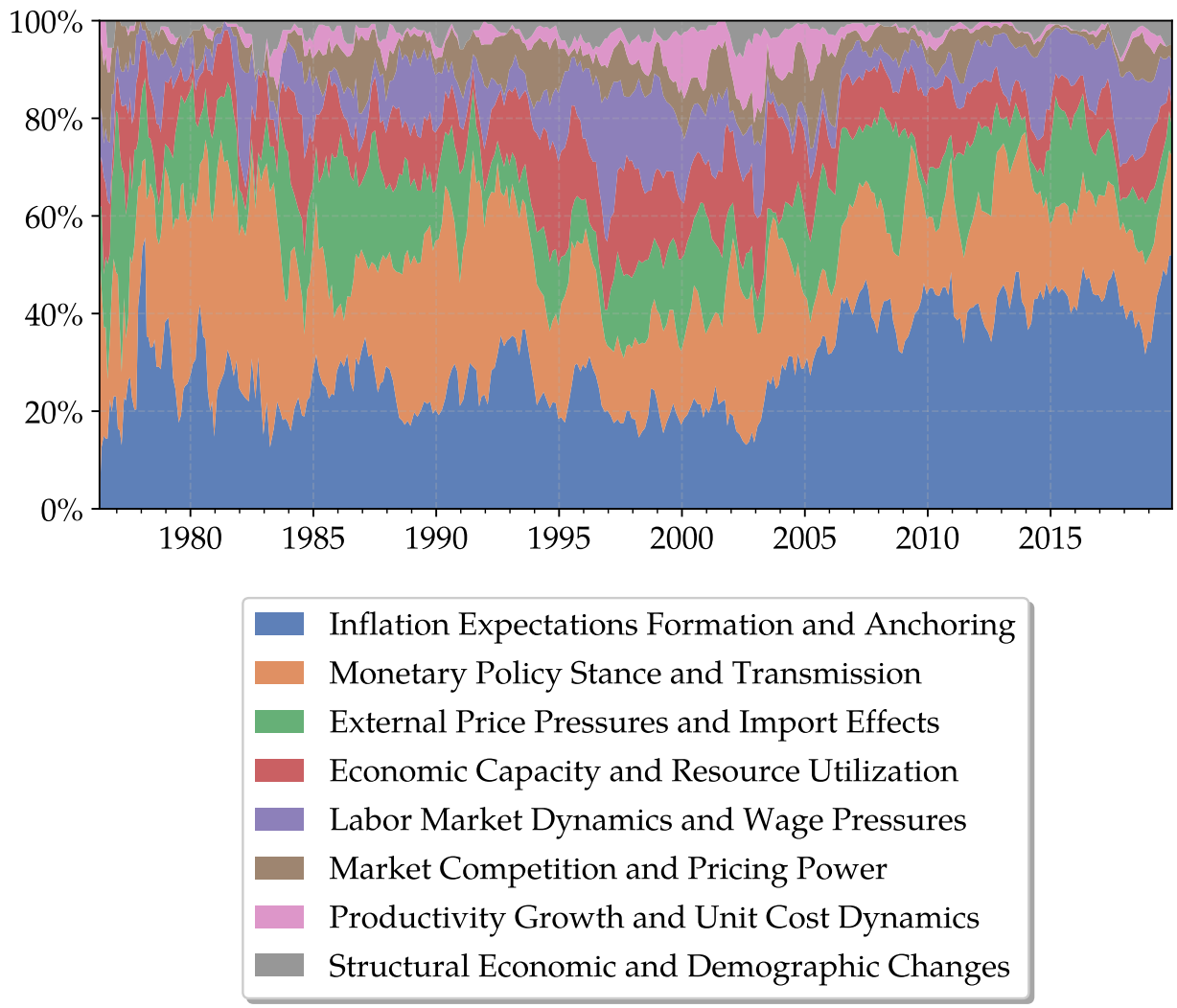


FIGURE IA.C.2: INFLATION ARGUMENTS BY CATEGORY

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.

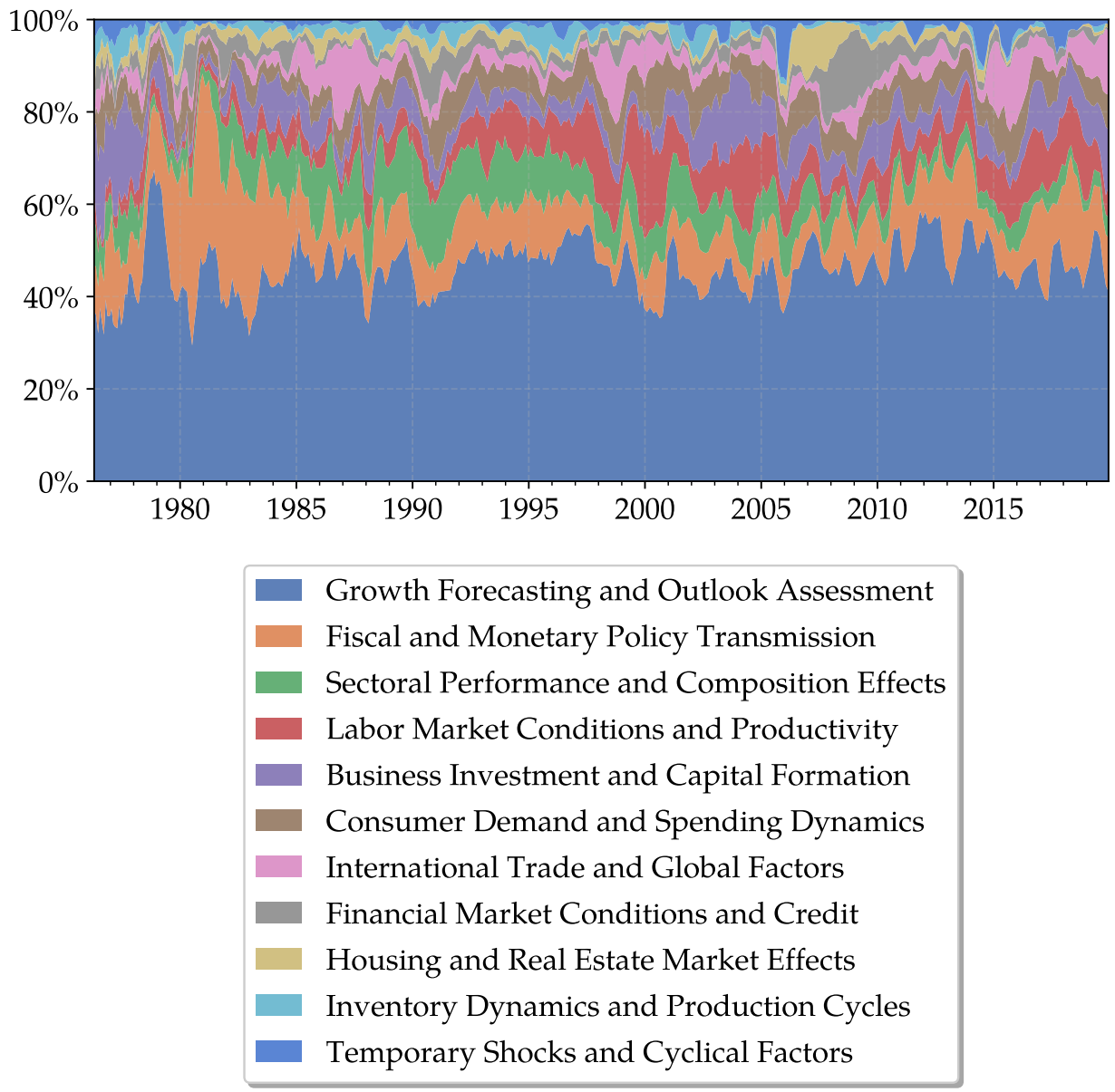


FIGURE IA.C.3: GROWTH ARGUMENTS BY CATEGORY

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.

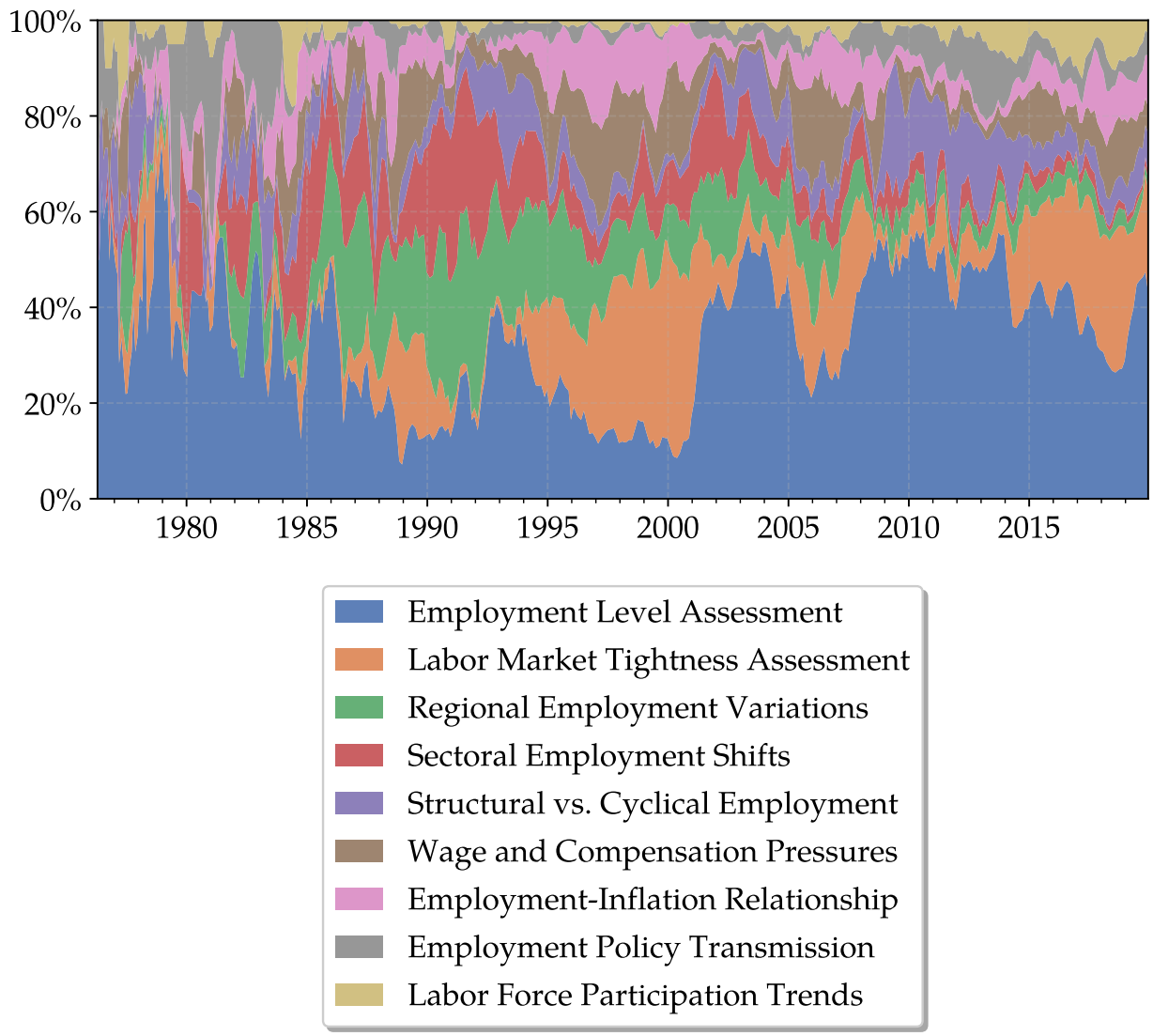


FIGURE IA.C.4: EMPLOYMENT ARGUMENTS BY CATEGORY

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.

	Panel A: Tealbook-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{TB}}_{i,t,k}(Y)$	-2.30	0.090	0.088	0.118	0.100
	(1.91)	(2.70)	(2.69)	(2.67)	(3.08)
N	1592	1592	1592	2677	2677
Meeting FE	Yes	Yes	Yes	Yes	Yes

TABLE IA.C.3: MODEL FIT AND POLICY DECISION ALIGNMENT

Note: The table repeats the analysis in Panel A of Table IA.C.3, using forecasts for the next-year (rather than the present year).

	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.C.2: VARIANCE DECOMPOSITION WITHOUT MEETING FIXED EFFECTS

Note: The table repeats the analysis in Table 10, without meeting fixed effects.

	Panel A: Tealbook-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{TB}}_{i,t,inflation}(Y)$	-2.53	0.085	0.065	0.109	0.065
	(2.18)	(2.37)	(2.21)	(3.46)	(2.95)
$\text{Fit}^{\text{TB}}_{i,t,growth}(Y)$	-1.97	0.047	0.069	0.055	0.054
	(2.36)	(2.02)	(2.76)	(2.22)	(2.59)
$\text{Fit}^{\text{TB}}_{i,t,employment}(Y)$	-1.84	0.044	0.052	0.069	0.052
	(2.39)	(1.95)	(2.78)	(2.40)	(2.71)
N	2179	2179	2179	3658	3658
Time FE	Yes	Yes	Yes	Yes	Yes
	Panel B: Half-year Forecast Error-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE}}_{i,t,inflation}(H)$	-3.93	0.117	0.089	0.122	0.071
	(1.83)	(1.84)	(2.14)	(2.09)	(2.16)
$\text{Fit}^{\text{FE}}_{i,t,growth}(H)$	-5.05	0.124	0.102	0.126	0.090
	(2.56)	(2.61)	(2.10)	(2.93)	(2.66)
N	873	873	873	1456	1456
Time FE	Yes	Yes	Yes	Yes	Yes
	Panel C: Quarterly Imputed Forecast Error-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE,impute}}_{i,t,inflation}(Q)$	-1.87	0.059	0.023	0.079	0.031
	(2.03)	(2.16)	(0.97)	(2.94)	(1.59)
$\text{Fit}^{\text{FE,impute}}_{i,t,growth}(Q)$	-0.18	-0.007	-0.011	-0.008	-0.012
	(0.25)	(0.31)	(0.63)	(0.43)	(1.02)
$\text{Fit}^{\text{FE,impute}}_{i,t,employment}(Q)$	-2.36	0.038	0.050	0.040	0.037
	(2.84)	(1.58)	(2.51)	(1.47)	(1.98)
N	2123	2123	2123	3572	3572
Time FE	Yes	Yes	Yes	Yes	Yes

TABLE IA.C.4: MODEL FIT AND POLICY DECISION ALIGNMENT, BY VARIABLE

Note: The table repeats the analysis presented in Table IA.C.3, running separate univariate regressions for the individual fit measures constructed for each of inflation, employment, and growth rather than the composite measure.

	Panel A: Tealbook-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{TB}}_{i,t,inflation}(Y)$	-5.09	0.157	0.101	0.177	0.109
	(2.81)	(2.95)	(2.39)	(4.11)	(3.69)
$\text{Fit}^{\text{TB}}_{i,t,growth}(Y)$	-3.38	0.072	0.096	0.068	0.070
	(2.55)	(2.14)	(2.97)	(2.23)	(2.88)
$\text{Fit}^{\text{TB}}_{i,t,employment}(Y)$	-2.47	0.076	0.070	0.103	0.069
	(2.32)	(2.42)	(2.67)	(2.58)	(2.80)
N	2179	2179	2179	3658	3658
Time FE	Yes	Yes	Yes	Yes	Yes
	Panel B: Half-year Forecast Error-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE}}_{i,t,inflation}(H)$	-7.62	0.271	0.148	0.260	0.142
	(2.12)	(3.20)	(2.64)	(2.48)	(2.08)
$\text{Fit}^{\text{FE}}_{i,t,growth}(H)$	-11.03	0.273	0.221	0.251	0.143
	(2.28)	(2.05)	(1.95)	(2.29)	(1.86)
N	873	873	873	1456	1456
Time FE	Yes	Yes	Yes	Yes	Yes
	Panel C: Quarterly Imputed Forecast Error-based Fit				
	Voters			All Members	
	$Dissent_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$	$Alignment_{i,t}$	$Influence_{i,t}$
$\text{Fit}^{\text{FE,impute}}_{i,t,inflation}(Q)$	-4.80	0.127	0.052	0.140	0.059
	(2.62)	(2.64)	(1.20)	(3.20)	(1.68)
$\text{Fit}^{\text{FE,impute}}_{i,t,growth}(Q)$	-1.33	0.043	0.005	0.006	-0.016
	(0.65)	(0.73)	(0.11)	(0.15)	(0.54)
$\text{Fit}^{\text{FE,impute}}_{i,t,employment}(Q)$	-3.01	0.060	0.071	0.069	0.047
	(2.47)	(1.80)	(2.37)	(1.61)	(1.79)
N	2123	2123	2123	3572	3572
Time FE	Yes	Yes	Yes	Yes	Yes

TABLE IA.C.5: MODEL FIT AND ALIGNMENT, BY VARIABLE (RAW)

Note: The table repeats the analysis presented in Table IA.C.4, 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.