POLICY BY COMMITTEE*

Toomas Laarits
Ben Matthies
Kaushik Vasudevan
Will Yang

Abstract

Consequential decisions are routinely delegated to groups, yet empirical evidence on how groups aggregate their members' views is scarce. We study the Federal Open Market Committee (FOMC) using members' macroeconomic forecasts and a novel dataset of their arguments from nearly thirty years of meeting transcripts. Departing from theories emphasizing information aggregation, we find that members hold distinct and persistent models: beliefs about structural relationships and steady-state values that lead them to interpret macroeconomic data differently. The dominant dimension of this heterogeneity organizes members along a spectrum between the Keynesian and New Classical traditions, distinguished by beliefs about the effectiveness of aggregate demand stabilization. The committee aggregates different views by tilting its decisions towards members whose models better fit recent data. Such tilting can improve decision quality but risks overfitting to the prevailing macroeconomic regime.

JEL: D23, D71, D84, E52

Keywords: group decision-making, mental models, monetary policy

First Draft Online: April 9, 2025. This version: August 5, 2026. 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 Boston College, email: kaushik.vasudevan@bc.edu. Will Yang is with Yale University, email: will.j.yang@yale.edu. For helpful comments and conversations, we thank our discussants Michael Bauer, Anna Cieślak, Leyla Han, Stephen Hansen, Matias Iaryczower, Lubomir Litov, Michael McMahon, Adi Sunderam, and Nancy Xu; Nick Barberis, Chris Clayton, Zhi Da, Bill English, Peter Fisher, Joel Flynn, Paul Fontanier, Stefano Giglio, Peter Hansen, Cooper Howes, Sid Kankanala, Chad Kendall, Randy Kroszner, Aleksei Oskolkov, Stefano Pegoraro, Ferdinand Pieroth, Carolin Pflueger, Jonathan Rose, 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, Boston College, and the Chicago Federal Reserve; and conference participants at NBER (Behavioral Finance; Political Economy; Organizational Economics; Big Data, AI, and Financial Economics), NYU WAPFIN, USC FIFI, the UBC Winter Finance Workshop, the Adam Smith Workshop, the Rutgers Finance Conference, and the UW Summer Finance Conference. We thank Myeongrok Doh, Giovanni Leopizzi, Seback Oh, and Lucas Ribeiro for outstanding research assistance.

1 Introduction

Organizations and governments routinely delegate consequential decisions to groups rather than to a single decision-maker. Why they do so, and how the resulting decisions are reached, are central questions in organizational economics and political economy. A long-standing literature following Condorcet (1785) argues that groups may improve decisions by pooling diverse information. Yet such delegation is common even when all members observe similar information, limiting the scope for such gains.

Despite a rich theoretical literature on group decision-making, empirical evidence remains scarce, as systematic analysis requires data on a group's objectives and decisions, its members' viewpoints and arguments, and a sufficient history to identify how different views are aggregated. We address these challenges by studying the Federal Open Market Committee (FOMC) from 1992 through 2020. The FOMC, which sets U.S. monetary policy, offers two key advantages. First, it is a structured, high-stakes setting, with a clearly expressed dual mandate of maximum employment and stable prices, as well as a repeated cast of decision-makers who act on a focused set of economic variables. Second, the Fed produces extensive records of the decision-making process, including voting records, detailed meeting transcripts, policy alternatives circulated before each meeting, and committee members' forecasts of key decision inputs: unemployment, inflation, and output. From these records, we construct a comprehensive set of policy decisions adopted at each meeting, along with measures of members' arguments and their views, and examine how the committee aggregates these views into decisions.

Using this setting, we document a dimension of group decision-making beyond the traditional information aggregation channel. We find that members hold distinct and persistent *models*, defined as beliefs about the data-generating process for the macroeconomy. The primary axis of model heterogeneity maps onto a spectrum between the Keynesian and New Classical traditions. This spectrum reflects differences in beliefs about the effectiveness of aggregate demand stabilization: how strongly interest rates affect the output gap—the deviation of output from its potential value—and how strongly the output gap translates into inflation. Model heterogeneity leads members to interpret the same incoming data differently, generating disagreements over the macroeconomic outlook and appropriate policy. Examining how the committee resolves these disagreements, we find that its decisions tilt toward the policy positions of members whose models better fit recent economic data. This form of model aggregation may improve decision quality, but it may also overfit policy to the prevailing macroeconomic regime.

Heterogeneous Models. We estimate members' models by measuring their beliefs about key *model parameters* governing the joint dynamics of macroeconomic variables. We then test whether these models explain differences in how members translate observed economic conditions into forecasts. We focus on two relationships: the Phillips curve, which links inflation to labor market slack, and the Investment-Savings (IS) curve, which links output to interest rates. Together with a policy rule, these two relationships underpin much of modern monetary policy analysis and capture the tradeoffs between inflation, unemployment, and growth that are central to FOMC deliberations.

We study two types of parameters: steady-state values and slopes. Steady-state values describe members' long-run projections for the economy: the non-accelerating inflation rate of unemployment (*NAIRU*, or u^*), the neutral real interest rate (r^*), and potential output growth (g^*). Slopes govern the strength of the relationships between slack and inflation and between interest rates and output. We measure beliefs about the steady-state values using members' long-run forecasts reported in the Summary of Economic Projections (SEP). We estimate beliefs about the slope of the Phillips curve and the interest-rate sensitivity of output from members' statements in FOMC transcripts.

To construct these transcript-based measures, we use a large language model (LLM) to segment each member's statements at each meeting into distinct quotations that make a claim about unemployment, inflation, or growth. We then prompt the LLM to identify the arguments that express a view about the Phillips curve slope or rate-sensitivity of output, and to score the strength implied by the argument. We validate the transcript-based measures against human coders: the large language model's classifications agree with those of individual coders about as closely as the coders agree with one another.[1] To translate scored arguments into parameter beliefs, we estimate a state-space model, which treats each argument as a noisy signal of the member's latent belief and pools information across members and meetings.

The resulting estimates show persistent differences in members' models. These differences are organized along a primary axis that maps broadly onto a spectrum between the Keynesian and New Classical traditions. Members closer to the Keynesian tradition perceive a steeper Phillips curve and greater interest-rate sensitivity, implying a larger role for output gaps in driving inflation and greater scope for monetary policy to stabilize activity. In the post-GFC period, these views are also associated with lower u^* and r^*. Moreover, differences in models are highly persistent. For all five parameters, members in the top and bottom belief terciles remain separated four years later, and quarterly persistence estimates range from 0.85 to 0.99. Member fixed effects explain 58% to 75% of

[1] Appendix IA.E reports coding protocols and reliability metrics for large language model classifications.

the variation in the first principal component of parameter beliefs—the Keynesian versus New Classical dimension—after accounting for meeting effects. Examining the sources of model heterogeneity, we find that it is partially explained by macroeconomic experience and education, as studied in prior work, as well as institution and career background.

To test whether members' heterogeneous models lead them to draw different conclusions from the same data, we compare forecasts across members within the same meeting. A distinctive prediction of model heterogeneity is that members' models should systematically govern how their forecasts respond to economic conditions. We find that they do: members' Phillips curve beliefs and IS curve beliefs determine the sensitivity of their inflation and growth forecasts to labor market and policy conditions, respectively. As labor market slack rises, members who perceive a steeper Phillips curve lower their inflation forecasts by more than those who perceive a flatter curve. As policy tightness rises, members who perceive greater interest-rate sensitivity likewise lower their growth forecasts by more than those who perceive lower sensitivity. The resulting state-dependent forecast differences are substantial: a one-standard-deviation difference across members in the relevant interaction corresponds to 23–25% of average within-meeting dispersion in inflation forecasts and 28% in forecasts of GDP growth relative to potential. Beliefs about steady-state values also shape forecast differences: at a given unemployment rate and real policy rate, lower u^* implies more perceived labor market slack and lower r^* implies tighter perceived policy, leading members to forecast lower inflation and growth.

We consider three alternative explanations for forecast heterogeneity and find that while they may complement the role of model heterogeneity, they do not explain it: "hawk" versus "dove" types, differences in perceived shocks, and differential private information. Under a fixed hawk–dove typology, the ranking of members' forecasts should be stable across economic conditions, whereas the forecast differences we document vary systematically with labor market slack and policy tightness. Using additional measures constructed from meeting transcripts, we also show that members' perceptions of supply versus demand conditions explain substantial variation in inflation and growth forecasts but leave the relationship between parameter beliefs and forecasts unchanged. Finally, classifying the information sources cited in members' arguments, we find that approximately 82% of information sources drawn on are commonly observed by members, including official statistics and Federal Reserve staff analysis, and that excluding arguments based on private information has little effect on our measures of parameter beliefs.

Model Aggregation. Members hold persistently different models, yet the committee must aggregate the resulting policy positions into decisions at each meeting. We study

this aggregation process and find that adopted decisions tilt toward the policy positions of members whose models better fit recent economic data.

We illustrate this process with the committee's December 2012 adoption of the Evans rule, named for Chicago Fed President Charles Evans. The rule tied the duration of the near-zero policy rate to explicit thresholds for unemployment and inflation, marking a major shift from the FOMC's prior, calendar-based communication strategy. Members' positions on the rule were closely related to their models. Early proponents held models closer to the Keynesian tradition, characterized by a steeper Phillips curve, greater interest-rate sensitivity, and lower u^* and r^*. For these members, weak growth and low inflation indicated an aggregate-demand shortfall that policy could address. Support was initially limited but broadened during 2012 as incoming data on inflation and growth became more consistent with this view. The committee ultimately adopted the rule, yet late supporters' models never converged to those of the early advocates. These members appear to have acquiesced in the face of the data, accepting the policy while retaining their own models of the economy.

This episode motivates a systematic analysis of the committee's decision-making. We construct a novel and comprehensive dataset of decisions adopted at each meeting, extending beyond target rate choices to statement language, forward guidance, asset purchases, and other policy actions. Our main analysis asks whose views these decisions reflect. We relate each member's alignment with a decision and contribution to its adoption to how well the member's model fits recent data.

We construct the alignment and adoption contribution measures from the meeting transcripts. Alignment ranges from clear opposition to clear support, capturing the degree to which a decision accords with a member's stated views. Because important deliberation may occur outside the formal meeting, alignment may also reflect influence that is not directly observable in the transcripts. Adoption contribution, by contrast, measures the member's observable role in advancing the decision during the meeting. Although these transcript-based measures capture substantially more variation than voting records alone, we also use recorded dissents among voting members.

We assess model fit on the basis of forecast accuracy. Forecasts provide observable implications of members' models; comparing them with subsequently realized outcomes indicates which models have fit recent data more closely. Our focus on recent forecast performance mirrors members' own description of the deliberative process, in which views expressed at one meeting gain or lose traction as incoming data support or undercut them (Schonhardt-Bailey, 2013). Our first measure, *realized forecast fit*, is based on forecast errors and summarizes how closely each member's recent forecasts of inflation and growth

match realized outcomes. Our second measure, *parameter-implied forecast fit*, connects forecast errors to member model parameters and isolates the component of relative forecast performance predicted by these parameters.

Our main finding is that adopted decisions tilt toward the policy positions of members with better-fitting models. A one-standard-deviation improvement in realized forecast fit is associated with roughly 0.1 standard deviations greater alignment and adoption contribution among voters, and a 3.3-percentage-point lower probability of dissent, a substantial effect given the 5.1% unconditional dissent probability in our sample. The same pattern holds when we restrict attention to the component of fit predicted by parameter beliefs, indicating that a substantial part of the relationship between fit and decisions runs through the Phillips curve and IS curve beliefs we measure.

Our baseline measures are constructed from statements at the meeting a decision is adopted, yet major decisions often take shape over many meetings. We therefore examine 16 major policy innovations for which pre-adoption deliberation can be traced in the transcripts. Members who express support earlier in the deliberative process have higher alignment and adoption contribution at the decision date, showing that the decision-meeting measures track positions developed over time. Viewed alongside the persistence of parameter beliefs, this pattern is consistent with incoming data changing the relative strength of policy arguments arising from members' underlying models of the economy.

Lastly, we examine the potential costs and benefits of tilting towards models that better fit recent data. The adoption of the Evans rule illustrates the benefit: after repeatedly postponing the anticipated date of liftoff as weak demand conditions persisted, the committee shifted from calendar-based guidance to state-contingent guidance tied to explicit economic thresholds. The committee's 2020 adoption of Flexible Average Inflation Targeting illustrates the risks. The framework was motivated by the preceding decade's low-r^*, below-target inflation regime, but the conditions underlying it changed sharply when inflation rose far above target in 2021 and 2022. In the appendix, we present a simple model of committee aggregation that characterizes when weighting members by recent fit may improve decisions and when it may lead to overfitting to the prevailing regime.

Literature Review. Our paper relates to work on collective decision-making in organizations and institutions, surveyed in Gibbons and Roberts (2013), in particular to the literature studying organizations as information-processing and communication structures (Marschak and Radner, 1972; Arrow, 1974). Prior work takes an informational perspective, emphasizing that group decision-making may improve outcomes by aggregating private information, with a focus on designing committees to foster information acquisi-

tion and aggregation (Li, Rosen and Suen, 2001; Persico, 2004; Gerling et al., 2005; Sibert, 2006; Gerardi and Yariv, 2008). In contrast, we argue that a key function of committees is to bring together members with different models to interpret the same information.

A growing body of theoretical and experimental work formalizes the study of narratives and (mental) models, with a particular focus on characterizing their persistence and adoption (Shiller, 2017; Eliaz and Spiegler, 2020; Schwartzstein and Sunderam, 2021, 2025; Montiel Olea et al., 2022; Montiel Olea and Prat, 2025; Barron and Fries, 2025; Charles and Kendall, 2025; Aina, 2025; Aina and Schneider, 2025). We document persistent model heterogeneity in a high-stakes field setting.[2] Our characterization of FOMC members' models complements a nascent literature that identifies mental models using surveys (Andre et al., 2022; Andre, Schirmer and Wohlfart, 2023; Binetti, Nuzzi and Stantcheva, 2024; Andre et al., 2025; Mei and Wu, 2025) and textual analysis (Ke, 2025; Bastianello, Décaire and Guenzel, 2025). We estimate expert policymakers' models and show that they lie along a spectrum between the Keynesian and New Classical traditions.

Our work also connects to a tradition in monetary policy that studies different approaches to decision-making in the face of model uncertainty, 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). In practice, we find that policymakers address model uncertainty through committees that tilt toward models that best fit recent data, and we discuss the risks and benefits of this approach.

A large body of work studies central bankers' policy preferences and their determinants (Belden, 1989; Chappell Jr, McGregor and Vermilyea, 2004; Schonhardt-Bailey, 2013; Hansen, McMahon and Rivera, 2014; Hack, Istrefi and Meier, 2023; López-Moctezuma, 2023), often by analyzing FOMC transcripts and minutes (Hansen, McMahon and Prat, 2018; Cieslak and Vissing-Jorgensen, 2021; Shapiro and Wilson, 2022; Cieslak and McMahon, 2023; Cieslak et al., 2023; Bordo, Istrefi and Martínez, 2024). Much of this literature asks what makes policymakers persistently hawkish or dovish, tracing these orientations to education and macroeconomic experience (Malmendier, Nagel and Yan, 2021; Bordo and Istrefi, 2023). We instead show that differences in members' models generate state-dependent variation in forecasts: the same member may appear hawkish when labor markets are tight but not when they are slack. These model differences also reflect members' institutions and career backgrounds, beyond the factors previously explored.

Two recent papers examine FOMC members' macroeconomic reasoning and mod-

[2]Given that members' models may be misspecified, our work relates to a primarily experimental literature considering how group decision-making affects decision quality in light of behavioral biases (Charness and Sutter, 2012; Enke, Graeber and Oprea, 2023; Barahona et al., 2024).

els using meeting minutes and transcripts: Kakhbod, Kermani and Maciel (2025) study the FOMC's attribution of inflation to supply-, demand-, or expectations-driven factors, while Howes et al. (2026) show that members differ in whether they perceive policy as transmitting primarily into prices versus output. Relative to this work, we propose and provide evidence that members hold different models distinguished primarily by beliefs about the effectiveness of aggregate demand stabilization. We also document that the committee tilts its decisions toward the views of members with better-fitting models. Finally, recent work finds that the geographic composition of FOMC voters affects the committee's decisions (Bobrov, Kamdar and Ulate, 2025; Fos, Tamburelli and Xu, 2025; Fos and Xu, 2025). Our paper differs in its focus on how members' perspectives are reflected in group decision-making.

2 Heterogeneous Models of the Macroeconomy

We study FOMC members' models of the macroeconomy, which we define as their beliefs about the parameters governing the joint dynamics of macroeconomic variables. After presenting a framework to motivate our measurement of these models, we show that members' models persistently differ along a primary axis distinguished by beliefs about the effectiveness of aggregate demand stabilization. Observable characteristics such as institution, career background, macroeconomic experiences, and education explain some, but not all, of these differences.

2.1 Framework

We focus our study on beliefs about the parameters governing the Phillips curve, which defines how inflation evolves in response to output and the labor market, and the Investment-Savings (IS) curve, which defines how output responds to monetary policy. Together with a policy rule, these two relationships underlie much of modern monetary policy analysis (Woodford, 2003; Galí, 2008) and capture the central tradeoffs between output growth, inflation, and unemployment that dominate FOMC deliberations.

Below, we present a standard New Keynesian formulation of the Phillips and IS curves, based on Clarida, Gali and Gertler (1999), to make explicit the parameters governing these relationships and motivate our choice of parameters to study. For the Phillips curve, we focus on two parameters: the slope of the Phillips curve and the unemployment rate consistent with stable inflation (NAIRU or u^*). For the IS curve, we study three parameters: the interest-rate sensitivity of output, the natural real interest rate (r^*), and potential

growth rate of output (g^*).

We denote output and potential output as y_t and y_t^*, respectively, and denote unemployment and the non-accelerating inflation rate of unemployment (NAIRU) as u_t and u_t^*. We define the growth rates g_t and g_t^* as $g_t \equiv y_t - y_{t-1}$ and $g_t^* \equiv y_t^* - y_{t-1}^*$, and the output gap as $x_t \equiv y_t - y_t^*$.

Under the assumption that output gaps translate to inflation with a one-period lag, the Phillips curve equation is

$$\pi_t = \kappa x_{t-1} + \phi \pi_{t-1} + (1-\phi)\beta \mathbb{E}_t(\pi_{t+1}) + s_t \tag{1}$$

where π_t is the time t inflation rate and s_t is a supply or cost-push shock. The parameter κ captures the slope of the Phillips curve. Expressing the output gap as a function of the deviation of the unemployment rate u_t from u^* results in

$$\pi_t = -\kappa\chi\left(u_{t-1} - u_{t-1}^*\right) + \phi\pi_{t-1} + (1-\phi)\beta\mathbb{E}_t\left(\pi_{t+1}\right) + s_t, \tag{2}$$

where the constant χ maps unemployment to output. The key parameters governing the Phillips curve in Equation (2) are the slope κ and the level of u_t^*.

Under the assumption that real interest rates transmit to output with a one-period lag, the IS curve is given by

$$x_t = -\varphi\left(i_{t-1} - \mathbb{E}_{t-1}\pi_t - r_{t-1}^*\right) + \theta x_{t-1} + (1-\theta)\mathbb{E}_{t-1}(x_{t+1}), \tag{3}$$

where i_t is the nominal interest rate, and the parameter φ captures the rate-sensitivity of output.[3] Expressing the term $i_t - \mathbb{E}_t\pi_{t+1}$ as the real rate r_t, and using the identity $g_t - g_t^* = x_t - x_{t-1}$ yields the equation for output growth:

$$g_t = g_t^* - \varphi(r_{t-1} - r_{t-1}^*) - (1-\theta)x_{t-1} + (1-\theta)\mathbb{E}_{t-1}(x_{t+1}). \tag{4}$$

The key parameters governing Equation (4) are the potential growth rate g_t^*, the natural real interest rate r_t^*, and the interest-rate sensitivity φ.[4]

[3] In basic formulations, φ corresponds to the inverse of the elasticity of intertemporal substitution. FOMC discussions also tend to emphasize the role of interest rate transmission to output through financial conditions and housing, which is also captured in modern IS curve formulations (Cúrdia and Woodford, 2016; Caballero and Simsek, 2020; Kekre and Lenel, 2022; Beraja et al., 2019; Berger et al., 2021; Eichenbaum, Rebelo and Wong, 2022). We take the more inclusive view in our measurement.

[4] The equations contain additional coefficients, including ϕ and θ, which govern the dynamic propagation of inflation and output gaps. These coefficients affect the model's dynamics, but we do not measure potential cross-member heterogeneity in them.

Together, Equations (2) and (4) provide a common framework that nests different beliefs about the structure of the macroeconomy. Larger values of κ and φ imply stronger effects of the output gap on inflation and of the real interest rate on the output gap, consistent with a more Keynesian view. Smaller values correspond more closely to a New Classical view, in which output gaps play a limited role in inflation dynamics, and real interest rates have weaker effects on real activity.

Equations (2) and (4) also illustrate a feature central to our analysis: model heterogeneity can generate long-lasting differences in beliefs even when members are Bayesian. For example, in Equation (2), after observing low realized inflation, a member who perceives a steeper Phillips curve slope (higher κ) and a lower neutral unemployment rate u^* may attribute the low inflation reading to labor market slack. A member perceiving a flatter Phillips curve slope may attribute the same observation to low inflation expectations $\mathbb{E}_t(\pi_{t+1})$ and deflationary cost-push shocks s_t, an interpretation that implies a different outlook for future inflation. That is, based on their model parameters, members may perceive different macroeconomic shocks, which can sustain persistent differences in beliefs and forecasts.[5]

2.2 Measuring Parameter Beliefs

We measure model parameter beliefs for each member of the FOMC using transcripts and data released in conjunction with regularly scheduled meetings. The committee consists of 19 members: seven Governors of the Federal Reserve Board, including a Chair, and presidents of twelve regional Federal Reserve banks. At any given meeting, up to twelve of these FOMC members vote on policy action: the seven Governors, the President of the New York Fed, who acts as Vice Chair of the FOMC, and a rotating cast of four regional Fed presidents.[6] We construct model parameter belief measures at the member-meeting level.

2.2.1 Parameter Beliefs from the Summary of Economic Projections

To measure beliefs about u^*, r^*, and g^*, we use members' self-reported numerical forecasts from the Summary of Economic Projections (the SEP). Since late 2007, as part of the SEP, the FOMC has released detailed forecasts about macroeconomic variables at

[5]This logic follows prior work emphasizing slow-moving model parameters when agents simultaneously learn about both model parameters and macroeconomic states (Johannes, Lochstoer and Mou, 2016; Farmer, Nakamura and Steinsson, 2024; Li, Van Nieuwerburgh and Renxuan, 2025; Crump et al., 2025).

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

four of the eight scheduled annual meetings, including members' projections for unemployment, growth, and inflation. Since 2009, the SEP has included long-run projections for these variables, defined as the rates "to which a policymaker expects the economy to converge over time—maybe in five or six years—in the absence of further shocks and under appropriate monetary policy."[7] Accordingly, we use members' long-run forecasts of unemployment and real GDP growth as their beliefs about u^* and g^*. Since 2012, the SEP has also included participants' assessments of the appropriate federal funds rate, including the long-run interest rate, which we combine with their forecasts of long-run inflation to obtain their beliefs about r^*. The identities of the FOMC members making the forecasts are revealed with a five-year lag, meaning we have data through December 2020; this corresponds to 48 meetings for g^* and u^*, and 36 meetings for r^*.

2.2.2 Parameter Belief Arguments from FOMC Transcripts

Members do not directly report beliefs about the Phillips curve slope and interest-rate sensitivity. To measure these parameters, we use transcripts from 373 regularly scheduled FOMC meetings, from April 1976 through December 2020. These transcripts contain verbatim exchanges from the meetings, which generally involve an economic discussion followed by a monetary policy discussion. We identify statements in which a member takes a stance on the Phillips curve slope or interest-rate sensitivity and score the strength of the relationship each expresses. We then use these statement-level scores to estimate parameter beliefs.

Specifically, we use Claude Sonnet 4.6 ("the LLM") to identify and categorize statements about the Phillips curve and interest-rate sensitivity. We first provide the LLM with all members' speeches within a given meeting, segmented into distinct arguments, each an exact quotation making a claim about unemployment, inflation, or growth.[8] We then evaluate whether an argument expresses a view about the parameter of interest, which requires that it take a stance on the strength of a structural relationship between variables. For example, to be scored as expressing a view about the slope of the Phillips curve, simply discussing unemployment or inflation is insufficient: the argument must take an explicit stance on how economic slack affects inflation. To be scored as expressing a view on interest-rate sensitivity, an argument must take an explicit stance on how interest rates affect real activity.

From this set of arguments, we extract whether each Phillips curve argument expresses a steep, moderate, or flat slope, and whether each IS curve argument expresses

[7] See https://www.federalreserve.gov/monetarypolicy/fomc_projectionsfaqs.htm.

[8] Appendix IA.E reports the prompts, implementation details, and validation against human coders.

strong, moderate, or weak sensitivity. Table I presents example quotations and classifications. Janet Yellen's statement that inflation will rise as resource utilization tightens implies slack is a meaningful driver of inflation and, according to our prompting instructions, is classified as "steep." In contrast, James Bullard's statement that low unemployment is not a harbinger of higher inflation is classified "flat" because it suggests a weak relationship. Between these extremes, Stanley Fischer's quote is classified as "moderate" as it accepts the channel in principle, but qualifies its importance. With respect to the IS curve, Alan Blinder's statement frames rate policy as a meaningful driver of output and is classified as "strong." Richard Fisher's statement treats the channel as largely absent and is classified as "weak," while Jeremy Stein's argument is classified as "moderate" because it accepts the channel in concept but identifies high uncertainty as temporarily attenuating its force.

Appendix Table IA.1 reports summary statistics on parameter belief arguments. We identify an average of 0.380 Phillips curve and 0.481 interest-rate sensitivity arguments per member-meeting. The majority of arguments (70%+) are scored as "steep" or "strong," consistent with the importance of these channels in workhorse macroeconomic models. However, there is substantial time-series and across-member variation that passes through into our estimates of members' parameter beliefs, as discussed in Section 2.3.

2.2.3 Mapping Arguments to Parameter Beliefs

We construct a quantitative measure of parameter beliefs from these scored arguments using a simple state-space model. The model treats each argument as a noisy signal of a member's latent parameter belief and estimates all members' beliefs jointly. This approach pools information across our unbalanced panel of members and meetings, filters out potential measurement error, and separates common shifts in the committee's views from disagreement across its members.

For a given parameter, we assume that member i's belief in meeting t, denoted $\theta_{i,t}$, is unobserved and consists of two parts: a common component, μ_t, which follows a random walk; and a member-specific deviation, $b_{i,t}$, which follows an AR(1) process. The equations governing the evolution of these beliefs are

$$\begin{aligned}
\theta_{i,t} &= \mu_t + b_{i,t} && && \text{(Member } i\text{'s parameter belief)} \\
\mu_t &= \mu_{t-1} + \eta_t, && \eta_t \sim N(0, \sigma_\mu^2), && \text{(Committee component)} \\
b_{i,t} &= \rho b_{i,t-1} + \xi_{i,t}, && \xi_{i,t} \sim N(0, \sigma_b^2). && \text{(Member-specific component)}
\end{aligned}$$

For argument j made by member i in meeting t, we assume that the member draws a

conditionally independent signal, $s^*_{i,j,t}$, based on their parameter belief, which determines whether they discuss the parameter and the valence of that discussion:

$$s^*_{i,j,t} = \theta_{i,t} + \varepsilon_{i,j,t}, \qquad \varepsilon_{i,j,t} \sim N(0,1).$$

If the member chooses to mention the parameter in their argument, the observed argument classification score, $Y_{i,j,t}$, provides a discrete measurement of the signal:

$$Y_{i,j,t} = \begin{cases} -1, & s^*_{i,j,t} < -\zeta, \\ 0, & -\zeta \le s^*_{i,j,t} \le \zeta, \\ 1, & s^*_{i,j,t} > \zeta, \end{cases} \qquad \text{(Argument Scoring Equation)}$$

where Flat/Weak maps to -1, Moderate maps to 0, and Steep/Strong maps to 1 for the Phillips curve and interest-rate sensitivity arguments, respectively.

A member's choice to mention the parameter in their argument may also provide information about their parameter belief. To capture this idea, we assume that the member draws an additional conditionally independent signal based on their parameter belief, denoted $d^*_{i,j,t}$, which determines whether they mention the parameter:

$$d^*_{i,j,t} = a_t + \psi_t b_{i,t} + \nu_{i,j,t}, \qquad \nu_{i,j,t} \sim N(0,1), \qquad \text{(Selection Equation)}$$

where the member mentions the parameter when $d^*_{i,j,t} > 0$. Accordingly,

$$\Pr(\text{member } i \text{ mentions the parameter in argument } j \mid b_{i,t}) = \Phi(a_t + \psi_t b_{i,t}),$$

where a_t absorbs meeting-level variation in discussion propensity and ψ_t governs how the member's parameter belief affects selection into discussion.

When $\psi_t > 0$, members with stronger or steeper beliefs are more likely to select into discussion; we expect this given the skewed distribution toward strong or steep arguments. We estimate ψ_t for each meeting, identified jointly by the pattern of parameter mentions across all arguments and the valence of the arguments that mention the parameter, which the score and state equations relate to member-specific deviations. We allow for time-variation in ψ_t to capture time-variation in the selection margin, e.g., the fact that Phillips curve arguments are overwhelmingly classified as steep early in the sample, but more balanced later on.

Estimating the Model and Extracting Parameter Beliefs. The parameters estimated in each sample are the common component of beliefs μ_t, member-specific deviations $b_{i,t}$, the persistence parameter ρ, innovation variances for the common and member-specific components (σ_μ^2 and σ_b^2), the ordered probit threshold ζ, and the selection parameters a_t and ψ_t governing the probability that an argument mentions the parameter. We normalize the argument-signal and selection-shock variances to one. As a result, a one-unit difference in μ_t, $b_{i,t}$, or $\theta_{i,t} = \mu_t + b_{i,t}$ corresponds to one standard deviation of idiosyncratic variation in the argument signal. In our main analysis, we use parameter estimates from 1992 onward, though the model is estimated for the full sample for which transcripts are available. We use weakly informative priors on the persistence and innovation variances, the ordered probit threshold, and the meeting-level selection parameters, and our estimation uses eight independent Gibbs chains of 100,000 iterations each. In addition to the main specification, we also estimate the model without the selection equation and show that our main results persist. Appendix IA.C provides the implementation details.

2.3 Dynamics and Determinants of Members' Models

Figure I plots the time series of the committee median of members' parameter beliefs, with the pre-SEP sample of u^* and g^* spliced with Federal Reserve staff estimates, and of r^* spliced with the historically assumed 2% real interest rate, as used in later analyses. Table II reports parameter belief summary statistics.

The time series of parameter beliefs reads as an abbreviated history of U.S. monetary policy. In the late 1970s and 1980s, beliefs about the Phillips curve slope are steep and volatile, and interest-rate sensitivity is perceived as strong, consistent with Paul Volcker's use of monetary restraint to bring runaway inflation under control.[9] The perception of a steep Phillips curve persists into the early 1990s, reflecting the 1990–1991 recession and the weak, low-inflation recovery that followed, while perceived interest-rate sensitivity rises through the late 1990s; the 1994–1995 soft landing and the 1998 easings made policy-rate changes appear effective in modulating aggregate demand, and perceived interest-rate sensitivity peaks at the end of the decade. From the late 1990s through the end of our sample, perceived Phillips curve steepness trends downward, as labor market tightness did not translate into high inflation. Perceived interest-rate sensitivity falls sharply during the Global Financial Crisis and zero lower bound period, when low inflation and growth persisted despite interest rates at zero, before rising as discussion of liftoff began. In the post-crisis SEP sample, median beliefs about u^* and r^* decline steadily, as inflation

[9] Notably, the Volcker Fed emphasized reserve and money-growth control as policy instruments rather than the federal funds rate target.

repeatedly fell short of target despite a tightening labor market and output grew only moderately despite sustained policy accommodation. These patterns also reflect the time-series variation in the measured arguments used to estimate parameter beliefs (Appendix Table IA.1).[10]

Members' model parameters exhibit substantial and persistent cross-sectional heterogeneity. Table II shows that the average cross-sectional standard deviation of members' beliefs is about 50% of the corresponding time-series standard deviation for all parameters. Figure II illustrates that this dispersion is due to persistent differences in parameter beliefs. For each parameter, we assign members to terciles based on their belief within a quarter, and plot the average percentile rank for each tercile for sixteen quarters. Members initially in the top and bottom terciles remain separated four years later for all parameters. The quarterly AR(1) persistence of the member-specific components of members' estimated Phillips curve slope and interest-rate sensitivity beliefs from the state-space model are 0.933 and 0.846, and the quarterly AR(1) persistence of members' reported u^*, g^*, and r^* beliefs, after controlling for meeting fixed effects, are 0.983, 0.922, and 0.988. As discussed in Section 2.1, persistence in parameter belief disagreements is an expected feature of model heterogeneity, as parameter learning can be slow in the presence of macroeconomic shocks.[11]

Model parameters also exhibit systematic and stable cross-member correlations. Figure III plots meeting-level cross-sectional correlations of parameters. Phillips curve beliefs and interest-rate sensitivity have an average cross-sectional correlation of 0.22. In the SEP sample (2009–2020), beliefs in a steeper Phillips curve slope are associated with lower u^* and lower r^*, with average correlations of -0.31 and -0.44. Beliefs in higher interest-rate sensitivity are likewise associated with lower u^* and lower r^* beliefs, with average correlations of -0.26 and -0.38. Beliefs about u^* and r^* have an average correlation of 0.60. By contrast, g^* exhibits an average correlation of 0.25 with r^*, but has low correlations with other parameters (-0.03 with u^* and -0.17 with interest-rate sensitivity).

Panel A of Table III reports a principal component analysis summarizing joint variation in the Phillips curve slope, interest-rate sensitivity, g^*, and u^*, scaling each parameter

[10]The Phillips curve evidence also accords with the account of the Fed's evolving beliefs in Cogley and Sargent (2005), who model the Fed as a Bayesian learner choosing among Phillips curve specifications.

[11]We find beliefs about steady-state values to be particularly persistent, consistent with Farmer, Nakamura and Steinsson (2024), who characterize the slow pace of Bayesian learning about such values when inference is complicated by shocks. The lower persistence of beliefs about interest-rate sensitivity may reflect the greater state dependence of the underlying transmission relationship, as documented, for example, in recent work on the mortgage refinancing channel (Beraja et al., 2019; Berger et al., 2021; Eichenbaum, Rebelo and Wong, 2022).

in-sample to zero mean and unit standard deviation.[12] In the pre-SEP sample, the first principal component, an average of the Phillips curve slope and interest-rate sensitivity measures, explains about 58% of their variation. In the SEP sample, the first component loads positively on the Phillips curve slope and interest-rate sensitivity and negatively on u^* and g^*, with respective loadings of 0.557, 0.604, -0.479, and -0.309; it explains 39% of the joint variation. Members' positions along this first component are highly persistent: member fixed effects explain 58% (pre-SEP) and 75% (SEP) of its variation, as shown in Panel B. The table also reports the second principal component: in the pre-SEP sample it reflects a Phillips curve slope versus interest-rate sensitivity factor, explaining 42% of parameter variation (member fixed effects account for 37% of the component); in the SEP sample it contrasts u^* and g^*, with loadings of 0.566 and -0.819, explaining 26% of parameter variation (member fixed effects account for 56% of the component).

Taken together, the cross-member correlations and the principal components analysis point to a dominant axis underlying members' models. Models fall along a spectrum between the Keynesian and New Classical traditions, distinguished by beliefs about the effectiveness of aggregate demand stabilization. Views closer to the Keynesian tradition hold that output gaps generate inflationary pressures via the Phillips curve, and that policy can close them by adjusting its stance. Views closer to the New Classical tradition give output gaps a more limited role in inflation and the policy stance less traction on activity. In the post-GFC period, these views align with members' assessments of u^* and r^*: more Keynesian members perceive persistently low inflation and growth as driven by a low u^* (implying substantial output gaps) and a low r^*, which at the effective lower bound limits the ability of policy to provide sufficient accommodation. More New Classical members assign greater importance to inflation expectations and supply-side forces, such as productivity growth and structural constraints. These two perspectives are consistent with characterizations provided by FOMC members, which we discuss in Section 4.1 in the context of the committee's post-GFC policy.[13]

We examine potential sources of persistent member differences in model parameters by relating them to separate blocks of observables. For institution and career background, we construct new measures; for macroeconomic experience, demographics, and education, we adopt the measures of Malmendier, Nagel and Yan (2021) and Bordo and Istrefi (2023). Panel B of Table III reports the explanatory power of each block after control-

[12] We exclude r^* here due to the shorter sample, but report results with it in Appendix Tables IA.8 and IA.9.

[13] The weak correlations of g^* with other parameters are consistent with long-run monetary neutrality, under which beliefs about g^* need not be tied to views about cyclical mechanisms. Dispersion in g^* therefore captures disagreement over supply-side fundamentals outside the primary dimension we study.

ling for meeting fixed effects; Appendix Tables IA.6 and IA.7 report exact variable definitions and coefficients. The strongest individual explanatory variable is an institution peer measure—the average first principal component value of all other members of the same Federal Reserve Bank (or the Board)—with coefficients of 0.304 (R^2 of 9.2%) in the pre-SEP sample and 0.432 (R^2 of 18.6%) in the SEP sample, suggesting that institutional staff and internal reasoning are important determinants of members' models. Career background explains 9.4% and 20.8% in the two samples; the relationship is strongest in the SEP sample, where experience in the financial sector or financial regulation is associated with a higher first principal component score. Following Malmendier, Nagel and Yan (2021) and Bordo and Istrefi (2023), we also find a role for experienced inflation and unemployment (R^2 values of 4.3% and 7.5% using the variables of Bordo and Istrefi (2023)), though the power of individual variables differs across the two samples. Finally, in the pre-SEP sample, obtaining a PhD from a "saltwater" school is associated with a higher first principal component value, consistent with the Keynesian tradition of those departments, while being a supply-side (Reagan) Republican appointee is associated with a lower one. These relationships weaken in the SEP sample, consistent with convergence of modern macroeconomics across schools and the dwindling number of Reagan appointees.

Despite these patterns, no individual block of observables approaches the explanatory power of member fixed effects, indicating that the views represented by committee members are not reducible to a readily observable set of characteristics. In all, our results accord with evidence that professional and macroeconomic experiences shape policymakers' views, while documenting a persistent configuration of parameter beliefs that captures contrasting perspectives on how the macroeconomy works.

3 Testing Model Heterogeneity

If members hold different models of the macroeconomy, their forecasts can differ even when they observe the same data, with these differences systematically determined by their model parameters. We test this prediction by examining whether members' parameters govern how they map labor market slack and policy tightness into forecasts of real GDP growth and inflation. We then show that alternative explanations—perceived shocks, fixed hawk–dove outlooks, and private information—do not account for these patterns.

3.1 Regression Framework for Phillips- and IS-Curve Based Forecasts

We relate members' forecasts to their model parameters using panel regressions with meeting fixed effects. The fixed effects absorb variation common to members within a meeting, so the coefficients are identified from cross-member differences in parameter beliefs and perceived conditions.

For the Phillips curve, we estimate regressions of the form

$$\begin{aligned} F_t^i[\pi_{t+1}] = \alpha_t^\pi + \beta_1 \left(u_t - u_{i,t}^*\right) \\ + \beta_2 \big(\text{PCSlope}_{i,t} \times (u_t - u_{i,t}^*)\big) + \delta_\pi \, \text{Supply}_{i,t} + \varepsilon_{i,t}^\pi, \end{aligned} \tag{5}$$

where F_t^i is the forecast operator for member i in meeting t, $u_t - u_{i,t}^*$ is member i's perceived labor market slack, $\text{PCSlope}_{i,t}$ is the member-specific component of i's estimated Phillips curve slope, and $\text{Supply}_{i,t}$ measures i's perception of supply conditions, including forecasted inflation expectations, separate from labor market slack and the slope, with more positive values indicating a more inflationary outlook on those conditions.

To interpret this regression, we iterate Equation (2) forward one period and apply member i's forecast operator, writing member i's inflation forecast as

$$F_t^i[\pi_{t+1}] = -\kappa_{i,t}\chi \left(u_t - u_{i,t}^*\right) + \phi \pi_t + F_t^i[(1-\phi)\beta \, \mathbb{E}_{t+1}(\pi_{t+2}) + s_{t+1}], \tag{6}$$

where $\mathbb{E}$ is the expectations operator for agents in the economy, potentially distinct from i's own.

We write member i's Phillips curve slope as

$$\kappa_{i,t}\chi \equiv \overline{\kappa\chi} + \lambda_{PC} \, \text{PCSlope}_{i,t}, \qquad \lambda_{PC} > 0, \tag{7}$$

where $\overline{\kappa\chi}$ is the average slope, $\text{PCSlope}_{i,t}$ is the standardized member-specific component of i's slope belief, and λ_{PC} converts it to the units of the slope. We likewise express the member-specific part of the forward-looking supply term in Equation (6) as $\lambda_S \, \text{Supply}_{i,t}$, with λ_S the corresponding scaling; its component common to all members in meeting t is absorbed by α_t^π.

Substituting Equation (7) into Equation (6) and collecting the terms that vary across

members within a meeting yields

$$F_t^i[\pi_{t+1}] = \alpha_t^\pi + \underbrace{(-\overline{\kappa\chi})}_{=\beta_1}(u_t - u_{i,t}^*) + \underbrace{(-\lambda_{PC})}_{=\beta_2}\big(\text{PCSlope}_{i,t}\,(u_t - u_{i,t}^*)\big) + \underbrace{\lambda_S}_{=\delta_\pi}\,\text{Supply}_{i,t}, \tag{8}$$

where α_t^π absorbs every component common across members in meeting t.

Equation (8) maps the coefficients in Equation (5) to the model parameters. The main coefficient of interest is β_2, which equals $-\lambda_{PC}$ and which we expect to be negative. For the same positive level of perceived labor market slack, a member perceiving a steeper Phillips curve forecasts lower inflation. The coefficient β_1 equals the negative of the average perceived slope, $-\overline{\kappa\chi}$, and we likewise expect it to be negative. It measures how many percentage points lower the average member forecasts inflation for each additional percentage point of perceived slack. The coefficient δ_π equals λ_S and should be positive, as members who perceive more inflationary supply conditions, for example from unanchored inflation expectations or cost-push shocks, forecast higher inflation.

For the IS curve, we estimate the analogous regressions of the form

$$F_t^i[g_{t+1} - g_{t+1}^*] = \alpha_t^g + \gamma_1\,(r_{i,t} - r_{i,t}^*) + \gamma_2\big(\text{ISSensitivity}_{i,t} \times (r_{i,t} - r_{i,t}^*)\big) + \delta_g\,\text{Demand}_{i,t} + \varepsilon_{i,t}^g, \tag{9}$$

where $r_{i,t} - r_{i,t}^*$ is member i's perceived policy tightness, $\text{ISSensitivity}_{i,t}$ is the member-specific component of i's estimated IS curve sensitivity, and $\text{Demand}_{i,t}$ measures i's perception of demand-side forces separate from both, with more positive values indicating a more expansionary outlook. Decomposing interest-rate sensitivity as $\varphi_{i,t} \equiv \bar{\varphi} + \lambda_g\,\text{ISSensitivity}_{i,t}$ and applying the same steps as above to Equation (4) yields an analogous mapping:

$$F_t^i[g_{t+1} - g_{t+1}^*] = \alpha_t^g + \underbrace{(-\bar{\varphi})}_{=\gamma_1}(r_{i,t} - r_{i,t}^*) + \underbrace{(-\lambda_g)}_{=\gamma_2}\big(\text{ISSensitivity}_{i,t}\,(r_{i,t} - r_{i,t}^*)\big) + \underbrace{\lambda_D}_{=\delta_g}\,\text{Demand}_{i,t}, \tag{10}$$

where λ_D scales the member-specific demand term.

The predictions are direct analogues of those for the Phillips curve. The main coefficient of interest is γ_2, which equals $-\lambda_g$ and which we expect to be negative. For the

same positive level of perceived policy tightness, a member who perceives output as more interest-rate sensitive forecasts lower growth. We likewise expect $\gamma_1 < 0$, measuring how many percentage points lower the average member forecasts growth relative to potential for each additional percentage point of perceived tightness; and $\delta_g > 0$, as members perceiving more expansionary demand conditions should forecast higher growth.

3.2 Data and Variable Construction

To estimate our regressions, we supplement our model parameter estimates with members' forecasts and macroeconomic data. We also construct measures of each member's perceived supply and demand conditions from the FOMC transcripts.

We obtain members' forecasts from the Monetary Policy Report to Congress (MPR) from 1992 to 2007 and from the SEP thereafter.[14] Our regressions use forecasts made at meeting t for the calendar year following the meeting.

We compute member i's perceived labor market slack as $u_t - u^*_{i,t}$, where u_t is the most recent monthly unemployment rate released before the meeting and $u^*_{i,t}$ is the member's long-run unemployment projection. We compute the real interest rate perceived by member i, $r_{i,t}$, as their projected nominal federal funds rate at the end of the meeting year minus their forecasted inflation over the following year. A member's perceived policy tightness is then $r_{i,t} - r^*_{i,t}$, where $r^*_{i,t}$ is the member's long-run real rate in the SEP, constructed as described in Section 2.2.[15] We measure members' forecasts of growth relative to potential as their GDP growth forecast minus their long-run growth projection. When a member's perceived labor market slack, policy tightness, or projection of potential growth is unavailable, we use the average among members with available values in the meeting; when no member value is available, we use the staff's Greenbook counterpart, with an assumed r^* of 2%.

Finally, we construct members' perceived supply and demand conditions following the same approach as in our transcript-based parameter belief estimation. For each argument in our dataset, we evaluate whether it identifies a force moving inflation outside

[14]The MPR is the Federal Reserve's semiannual report to Congress, submitted in February and July. Member-level forecasts of inflation, GDP growth, and unemployment prepared in conjunction with these meetings are available from 1992 through 2007 and have been studied in prior work (Romer and Romer, 2008; Romer, 2010; Malmendier, Nagel and Yan, 2021). The data are available from the Federal Reserve Bank of Philadelphia; additional details are provided by Romer (2010). As discussed in Section 2.2, the FOMC began releasing the SEP in late 2007, reporting individual members' forecasts at four of the eight scheduled meetings each year.

[15]Before the SEP, perceived slack does not vary across members within a meeting and is absorbed by the meeting fixed effects; the same holds for perceived tightness before 2012. The slack and tightness coefficients are therefore identified from the later samples, while the interaction coefficients are identified throughout from cross-member variation in parameter beliefs.

the slack channel, such as inflation expectations, energy prices, or supply disruptions, and if so, sign it as inflationary (+) or deflationary (−); and whether it identifies a force moving output outside the policy rate channel, such as fiscal policy, external demand, or inventories, and if so, sign it as expansionary (+) or contractionary (−). We then apply the state-space model of Section 2.2.3 to these signed arguments to estimate $\text{Supply}_{i,t}$ and $\text{Demand}_{i,t}$.

3.3 Testing Model Heterogeneity

Table IV presents the main results from estimating Equations (5) and (9). To facilitate interpretation, we express the dependent variables, along with labor market slack, $u_t - u^*_{i,t}$, and policy tightness, $r_{i,t} - r^*_{i,t}$, in percentage points. We also scale $\text{PCSlope}_{i,t}$, its interaction with $u - u^*_{i,t}$, $\text{ISSensitivity}_{i,t}$, its interaction with $r_{i,t} - r^*_{i,t}$, and the inflationary supply and expansionary demand controls to be mean zero with unit standard deviation within the sample, pooling observations across members.

Panel A reports the Phillips curve specifications, with the main estimates in columns (1) and (2). Members who perceive a steeper Phillips curve exhibit inflation forecasts that are more sensitive to perceived labor market slack. The interaction between Phillips curve beliefs and perceived labor market slack enters with a coefficient of -0.064 without controls and -0.059 after controlling for inflationary supply conditions. These state-dependent forecast differences are large relative to overall disagreement: a one-standard-deviation difference in the interaction between Phillips-curve beliefs and perceived slack corresponds to 23% to 25% of the 0.25% average within-meeting dispersion in annual inflation forecasts. The coefficient on labor market slack ranges from -0.20 to -0.24, implying that at the average Phillips curve slope perceived by committee members, a one-percentage-point increase in slack lowers annual inflation forecasts by 20–24 basis points.[16]

Panel B reports the IS-curve specifications, again with the main estimates in columns (1) and (2). Growth forecasts of members who perceive output as more interest-rate sensitive respond more strongly to perceived policy tightness. The interaction between perceived interest-rate sensitivity and policy tightness enters with a coefficient of -0.099 both without controls and after controlling for perceived expansionary demand conditions. These state-dependent forecast differences are similarly large relative to overall

[16]Our regression provides a forecast-based estimate of the Phillips curve slope, which is challenging to identify in practice. Hazell et al. (2022) estimate a slope of about -0.34 by mapping regional estimates into an aggregate coefficient, and Barnichon and Mesters (2020) estimate unemployment-gap Phillips-curve slopes of about -0.45 and -0.24, using different measures of monetary policy shocks.

disagreement, with a one-standard-deviation difference across members in the interaction accounting for 27–28% of the 0.36% average within-meeting dispersion in forecasts. The coefficient on perceived policy tightness ranges from -0.33 to -0.32, implying that, at the average interest-rate sensitivity perceived by committee members, a one-percentage-point increase in policy tightness lowers forecast growth relative to potential in the following year by 0.32 to 0.33%.

Columns (3) and (4) re-estimate the regressions using measures of Phillips curve slopes, interest-rate sensitivity, and perceived supply and demand conditions constructed without the selection component of the state-space model. In these specifications, transcript-based estimates only use the valence of members' discussion, rather than differences in their propensity to mention a parameter. The results are broadly similar, indicating that our inferences are robust to how transcript arguments are mapped into parameter estimates.

In the appendix, we estimate and report results for alternative specifications, and find similar relationships between members' inflation and growth forecasts and their model parameters. Table IA.2 reports results where members' Phillips curve and interest-rate sensitivity parameters are estimated as simple averages of scored arguments from all prior meetings and from the immediately prior 8 meetings. Table IA.3 reports results where we use common measures of perceived conditions, $u_t - u_t^*$ and $r_t - r_t^*$, rather than member-specific versions, and reports similar relationships between members' perceived Phillips curve slopes, interest-rate sensitivities, and inflation and growth forecasts. Table IA.4 reports versions of the IS curve regressions where the dependent variable is one-year-ahead GDP forecasts and g^* is included as a control, rather than the dependent variable being growth relative to potential. Table IA.5 includes members' Phillips curve slopes and interest-rate sensitivity parameters as standalone independent variables. Across the specifications, we find similar qualitative and quantitative results to our main specification.

3.4 Alternative Explanations

The interaction effects in both panels confirm the key prediction of model heterogeneity: for a given level of perceived labor market slack or policy tightness, members' forecasts vary systematically with their perceived Phillips curve slopes and interest-rate sensitivities. We next consider three alternative explanations for members' forecast heterogeneity: differences in perceived shocks, persistent hawkish or dovish outlooks, and private information. None appears to account for the results, although each may coexist

with model heterogeneity.

Different assessments of macroeconomic shocks. As illustrated by our framework in Section 3.1, an alternative reason that members may have different forecasts is that they perceive different supply and demand conditions acting upon the economy. For example, members may differentially perceive unanchored inflation expectations or cost-push shocks that could drive inflation. A potential concern is that our measures of Phillips curve slopes and interest-rate sensitivity partly proxy for these perceived shocks: a member whom we classify as perceiving a steep Phillips curve may instead perceive a flat one but infer deflationary shocks precisely when labor markets are slack, producing the state-dependent forecast pattern without any difference in slope beliefs.

The specifications in columns (2) and (4) of Table IV include controls for members' perceived supply and demand conditions, estimated from their discussions in the transcripts. Perceived shocks carry substantial explanatory power for members' forecasts, but do so largely independently of estimated parameter beliefs. A one-standard-deviation increase in perceived inflationary supply conditions corresponds to a 7.6 basis point higher inflation forecast, while a one-standard-deviation increase in perceived expansionary demand conditions corresponds to a 2.4 basis point higher forecast of growth. Yet, as discussed above, including these measures produces only modest changes in the interaction estimates. Perceived supply and demand conditions, therefore, explain meaningful forecast variation without accounting for the state-dependent relationship between model parameters and forecasts.

The two mechanisms are also not mutually exclusive. As discussed in Section 2.1, differences in perceived shocks are compatible with model heterogeneity, and partly a product of it. When members hold different model parameters, they observe the same macroeconomic realizations and draw different inferences about the shocks driving them.

Persistent and dogmatic belief disagreements. Prior work and anecdotal accounts characterize members as hawks, who place greater weight on controlling inflation, or doves, who place greater weight on maintaining strong output (Istrefi, 2019; Malmendier, Nagel and Yan, 2021; Bordo and Istrefi, 2023). One possible extension of this characterization is that forecast disagreements reflect fixed member outlooks: hawkish members systematically forecast higher inflation and dovish members lower inflation, regardless of the state of the economy. More generally, such an account implies a stable forecast ranking across members.

The interaction estimates reported above imply a different pattern: the forecast differ-

ences associated with members' model beliefs vary with labor market slack and policy tightness rather than entering as fixed differences across members. Figure IV provides a descriptive illustration of this state dependence.

Using meeting-level measures of labor market and policy conditions, we classify meetings as having slack labor markets when $u_t - u_t^* > 0$ and tight labor markets otherwise, and as having a tight policy environment when $r_t - r_t^* > 0$ and a loose one otherwise. The u^* measure is the meeting average of member projections when available and the staff NAIRU forecast otherwise; the $r - r^*$ measure is the meeting average of member values when available and the staff measure with an assumed r^* of 2% otherwise. This classification imposes a common assessment of conditions and therefore abstracts from differences across members in perceived slack and policy tightness. Panel A divides members within each meeting into those with below-median and above-median Phillips curve slopes and plots their mean cross-sectionally demeaned inflation forecasts. When labor markets are slack, members who perceive a steeper Phillips curve forecast lower inflation than those who perceive a flatter curve; when labor markets are tight, this gap closes. Panel B similarly divides members according to their perceived interest-rate sensitivity and plots their mean cross-sectionally demeaned forecasts of growth relative to potential. Members who perceive greater interest-rate sensitivity forecast higher growth when policy is loose, while the gap closes when policy is tight.

Janet Yellen provides a concrete illustration of state-dependent forecast differences and their policy implications. Our estimates characterize her Phillips curve belief as consistently steep throughout the sample. In July 1996, facing a modestly tight labor market, Yellen's year-ahead inflation forecast was 0.77 standard deviations above the committee average. As labor markets continued to tighten by the September meeting, Yellen argued that "the probability of an increase in inflation is clearly higher when labor market slack is lower" and that there was "a very solid case" to increase the interest rate by 25 basis points, as the committee adopted an unchanged policy. In contrast, in June 2010, Yellen's year-ahead inflation forecast was 0.73 standard deviations below the committee average, when there was substantial slack in the labor market. She argued that "the economy's substantial slack will continue to weigh on inflation" and stated her belief that "macroeconomic considerations continue to call for accommodative policy and, ideally, even more accommodative policy than we have now." The same persistent belief in a steep Phillips curve provides a unified interpretation of both episodes: an above-average inflation forecast and willingness to tighten policy when labor markets were tight, but a below-average forecast and support for accommodation when slack was substantial. The flip in her relative forecast, together with the corresponding shift in policy stance, is difficult to reconcile

with a fixed hawk–dove view.

We do note, however, that the state-dependent pattern we document is compatible with some persistence in members' forecasts, for two reasons. First, the macroeconomic state is itself persistent in our sample: we identify 45 meetings with slack labor markets versus 26 with tight, and 59 with loose policy versus 12 with tight. Such persistence may lead even purely state-dependent forecast differences to appear persistent. Second, members' beliefs about u^* and r^*, which we find to be highly persistent, lead members to perceive the same unemployment and policy conditions differently across meetings, which may feed through to persistent forecast differences.

Differential private information. Finally, we consider whether the patterns we attribute to model heterogeneity instead reflect differences in the information available to committee members, a traditional rationale for committee decision-making. We prompt the LLM to classify the information sources cited in each argument in our dataset, and to assess whether those sources were likely commonly observed by other committee members or private to the individual member.

Figure V reports the composition of sources cited in Phillips curve and IS curve arguments by classification. We identify approximately 82% of cited sources as commonly observed, 14% as private, and 4% as ambiguous. Official statistics, Board staff forecasts and analysis, and financial market indicators account for 65% of cited sources and are almost entirely common across members. Private information is concentrated in reports from business and regional contacts and in Reserve Bank forecasts and analysis.

Excluding arguments based on private information has little effect on our measures of members' model beliefs. The within-meeting correlation between the average valence scores constructed from all arguments and those constructed using only commonly observed information is 0.98 for Phillips curve slopes and 0.99 for IS-curve sensitivity (Appendix Table IA.19). The predominance of common versus private information extends to the broader set of arguments that do not concern these two parameters, as shown in Appendix Table IA.18 and Figure IA.3.

The evidence implies that, while some private information surfaces during meetings, primarily from regional or specialized sources, the vast majority of information discussed is commonly observed. Moreover, private information appears unlikely to drive our measurement of members' models.

4 Models and Committee Decisions

Sections 2 and 3 document that FOMC members hold heterogeneous models of the macroeconomy, broadly organized along a spectrum between the Keynesian and New Classical traditions, and that these differences generate systematic forecast disagreement. We now study how the committee aggregates the policy positions associated with these models into collective decisions.

The committee's December 2012 adoption of state-contingent forward guidance provides a concrete illustration. The proposal was initially supported by a minority, but gained overwhelming support as incoming data during 2012 appeared to validate the more Keynesian models held by its early advocates. Building on this example, we construct measures of each member's alignment with, and contribution to, the decisions adopted at FOMC meetings. We find that committee decisions tilt toward the policy positions of members whose models better fit recent data. This form of aggregation can improve decisions when macroeconomic data are informative, but it can also lead the committee to overfit policy to the prevailing macroeconomic regime, as may have been the case when Flexible Average Inflation Targeting (FAIT) was adopted in August 2020.

4.1 The Evans Rule: Persistent Models and Evolving Support

4.1.1 Adoption of the Evans Rule

In December 2012, the FOMC took a significant step in its communication strategy by adopting state-contingent forward guidance. The committee committed to keeping the federal funds rate near zero at least as long as unemployment remained above 6.5 percent, projected inflation one to two years ahead stayed no more than half a percentage point above the 2 percent objective, and longer-run inflation expectations remained anchored. The policy became known as the "Evans rule," after Chicago Fed President Charles Evans, who had advocated numerical thresholds since 2011.

When the committee discussed an early version at its November 2011 meeting, Evans framed the debate as a choice between two explanations for the weak post-crisis recovery at the ZLB:

> "There are essentially two story lines that are used to account for the aberrantly slow growth and high unemployment following the Great Recession... The first story line I refer to as the 'structural impediments' scenario... We are up against a supply constraint that monetary policy can't fix... [and] it's

> probably time to begin considering removing excess accommodation before inflation rises above target... The second story line I referred to as the 'liquidity trap' scenario... Actual real rates are modestly negative, but the real natural rate of interest is strikingly negative... drags on demand are exceedingly large and persistent... [and] there are monetary policy prescriptions that can vastly improve outcomes in such an event."

These two interpretations map closely onto the primary axis of model disagreement documented in Section 2. Under the demand-centric interpretation in the Keynesian tradition—Evans's "liquidity trap" view—a low r^* left policy tight even at the ZLB, generating weak growth and substantial slack; a low u^* increased the perceived degree of slack, and a steep Phillips curve translated it into low inflation. Under the supply-centric "structural impediments" interpretation, structural constraints, rather than deficient aggregate demand, explained the weak recovery, leaving less scope for monetary policy. Evans aligned with the demand-centric interpretation: his u^* belief, well below the committee's, implied substantial slack, and combined with his relatively steep Phillips curve led him to forecast low inflation—1.5% for 2013 as of January 2012 versus the committee median of 1.8%.

4.1.2 Evolution of Support for the Evans Rule

Committee members' positions on the Evans rule were closely aligned with their models. Table V and Figure VI document this pattern and trace the rule's progression from a minority position to adoption as incoming data became more consistent with the liquidity trap interpretation. Panel A of Table V groups participants by their eventual degree of support. The groups differ primarily in whether members held a demand- or supply-centric view of the recovery. Strong supporters perceived a steeper Phillips curve, greater interest-rate sensitivity, and lower u^* and r^*. They also projected a substantially later liftoff than opponents (Panel A of Figure VI), consistent with an aggregate demand shortfall that called for continued policy accommodation.

Panel B of Table V and Panels B and C of Figure VI trace how support broadened during 2012 as incoming data led members to revise their outlooks for inflation and growth downward, strengthening the case for the demand-centric interpretation. By September 2012, twelve members supported state-contingent forward guidance. Chairman Bernanke described a "pretty clear consensus... that there's a lot of hunger for moving away from the date toward some kind of state-contingent guidance," while noting that the committee was "not quite ready for prime time." Between October and Decem-

ber, as inflation expectations declined further, four additional members came around, and the committee adopted the rule with sixteen participants in support and three opposed.

The rule's strongest supporters were generally also its earliest supporters. Later support, however, did not necessarily reflect convergence toward the models held by the early advocates. Esther George ultimately supported the policy but warned that the committee's inflation forecasts "may be based on misestimates of the amount of slack in the economy," and she indicated that she saw "lower longer-term rates as unlikely to provide sufficient benefits to the economy." Consistent with these views, her r^* stayed above, and her perceived interest-rate sensitivity below, those of the rule's strong supporters. More broadly, Appendix Figure IA.1 shows that members' loadings on the first principal component of model parameters remained stable throughout the period, indicating that eventual support did not stem from convergence in model parameter beliefs.

Taken together, the episode illustrates how a policy can gain support as incoming data strengthen its proponents' case while weakening the position of opponents. Members retained their distinct models, consistent with the persistence documented in Section 2, but as the outlook evolved, the committee shifted toward policy positions implied by models that better fit recent developments. This example motivates our next set of analyses where we take a more systematic approach using the full set of committee decisions.

4.2 Measuring Decisions and Member Alignment

We begin by constructing a comprehensive set of policy decisions. For each meeting, we provide the LLM with the minutes, the policy alternatives circulated before the meeting, and the policy discussion from the transcript. We prompt the LLM to identify the decisions reported in the minutes and cross-reference them against the transcript, capturing decisions that emerged or changed during deliberation. The resulting set includes target-rate choices, statement language, forward guidance, asset purchases, and other policy actions. Appendix Table IA.10 presents examples of the different types of identified decisions. Appendix Table IA.11 reports summary statistics on their number, distribution, and types. We identify about five decisions per meeting; about 22% of them are interest rate decisions, 43% are communications decisions, and 35% are other types of decisions, e.g., asset purchases. In Appendix Section IA.A.1, as a validation exercise, we classify the stance of each decision from -3 (highly accommodative) to $+3$ (highly restrictive), and find that the average decision score in each meeting is strongly related to its monetary policy surprise, measured using the asset price response in the 30 minutes following the committee's announcement. Appendix Section IA.E reports the prompts

and human-coding validation.

To measure the extent to which these decisions reflect individual policymakers' views, we construct three outcomes describing each member's relationship to the decisions adopted at a meeting. *Alignment* measures the degree of support for or opposition to a decision. *Adoption contribution* measures the member's observable, decision-specific role in developing and advancing it. *Dissent* records whether a voting member formally opposes the committee's action.

Concretely, for each member-decision pair, we construct an alignment score ranging from -2 (clear opposition) to $+2$ (clear support), and an adoption contribution score ranging from 0, when the member makes no positive decision-specific contribution, to 3, when the member offers a concrete proposal or specific language that appears in the adopted decision. We average these decision-level scores within each meeting to construct $Alignment_{i,t}$ and $AdoptionContribution_{i,t}$ for each member i.

The two measures capture complementary dimensions of the committee's decision-making. Alignment captures stated agreement with the decision. Because deliberations may also occur outside of the formal meeting (Schonhardt-Bailey, 2013), alignment may reflect influence that is not directly visible in the transcripts. Adoption contribution, on the other hand, isolates the member's observable role in developing or building support for the decision during the meeting.

Table VI presents examples from the December 2012 meeting, and illustrates the distinction between the two measures. For state-dependent forward guidance adopted at the meeting, William Dudley received an alignment of 2 and an adoption contribution of 2 ($2/2$ hereafter) for actively favoring threshold guidance as a data-contingent "automatic stabilizer." Sandra Pianalto received a $1/2$: she preferred a 7 percent unemployment threshold, but accepted 6.5 percent provided the Committee remained vigilant on inflation, and supported the language conditional on a press-conference explanation of its qualifications. Charles Plosser received a $-2/0$ for opposing a proposal he viewed as premature.

The Treasury-purchase rows describe a separate decision: whether to replace the expiring maturity extension program with $45 billion per month in outright purchases of longer-term Treasury securities. Eric Rosengren received a $2/2$ for advocating continued purchases, citing the fragile recovery, fiscal headwinds, limits on expanding MBS purchases, and continued shortfalls from the dual mandate. Jerome Powell received a $-1/1$ for accepting the decision despite reservations about balance-sheet expansion, exit risks and fiscal risks, and the marginal benefits of Treasury purchases. Charles Plosser received a $-2/0$ for opposing additional Treasury purchases as non-neutral balance-sheet expan-

sion that risked being confused with fiscal policy.

4.3 Decision Measure Summary Statistics

Table VII reports summary statistics for $Alignment_{i,t}$ and $AdoptionContribution_{i,t}$ and validates them against recorded dissents. Panel A reports statistics for voting members. Among voters with $Alignment_{i,t}$ between -2 and -1, 90.5% dissented. The dissent rate falls to 24.7% for alignment in $(-1, 0]$, 1.6% for alignment in $(0, 1]$, and zero in the highest category. Low values of $Alignment_{i,t}$ map closely to formal opposition. At the same time, because dissents occur in only 5.1% of voter-observations, the alignment measure captures substantial variation in agreement that is not reflected in recorded votes. As expected, $AdoptionContribution_{i,t}$ also increases with alignment.

Panel B reports the corresponding statistics for non-voters. Non-voters have lower alignment and adoption contribution than voters, consistent with their more limited role. Panel C breaks down the measures by committee role. The chair has an average alignment of 1.49 and an average adoption contribution of 1.12. These exceed the corresponding averages of 1.09 and 0.64 for other governors, 0.92 and 0.55 for voting Reserve Bank presidents, and 0.73 and 0.47 for non-voting presidents. This pattern is consistent with the chair's role in managing the meeting, shaping the policy round, and building consensus (Blinder, 1999; Chappell Jr, McGregor and Vermilyea, 2004).

Panel D shows that chair-member gaps vary substantially across chair tenures. They are largest under Greenspan and Bernanke. Greenspan's average $Alignment_{i,t}$ and $AdoptionContribution_{i,t}$ were 1.48 and 1.38, compared to 0.65 and 0.52 for other members during his tenure. Under Bernanke, the corresponding averages were 1.74 and 1.21 for the chair and 1.08 and 0.58 for other members. By contrast, the gaps narrow under Yellen and reverse under Powell. These patterns are consistent with anecdotal accounts of the chairs' leadership styles. Narayana Kocherlakota, for example, describes a shift from the chair-centered Greenspan FOMC to a format in which Bernanke and Yellen spoke later in the policy round, with Yellen placing particular emphasis on facilitating agreement within the committee (Bowmaker, 2025).

4.4 Measuring Model Fit

Having constructed measures of members' relationships to adopted decisions, we next measure how well their models fit recent data. Forecasts provide observable implications of a member's model for inflation and growth, so comparing them with realized outcomes indicates which models have fit the data more closely. Because members' mod-

els may extend well beyond the parameters we estimate, we construct two complementary measures. *Realized forecast fit* evaluates forecast performance as a whole, whereas *parameter-implied forecast fit* isolates the component associated with the parameter beliefs estimated in Section 2.

We construct *Realized forecast fit* from half-year forecast errors for inflation and real GDP growth. In the SEP sample, members report forecasts for the first and second halves of the year. At each meeting t, we identify the most recent half-year whose outcome is available and compare it with each member's latest forecast for that period. We take the absolute value of each forecast error and, within each meeting, standardize these errors across members separately for inflation and growth. We then compute an eight-observation rolling average by member and variable, average the two rolling-average values, and multiply by -1. The resulting measure, $\text{Fit}^{\text{RF}}_{i,t}$, is therefore higher for members whose recent forecasts more closely match realized outcomes.

To extend realized forecast fit to the pre-SEP period, we construct analogous half-year errors from the MPR forecasts. Because the MPR contains annual rather than half-year forecasts, we impute each member's half-year error by adding the member's annual forecast difference relative to the Federal Reserve staff to the staff's half-year forecast error. The full-sample measure combines the directly observed SEP half-year errors with the MPR-imputed errors and applies the same within-meeting standardization, rolling-average, and aggregation procedure.

Our second measure, *parameter-implied forecast fit*, isolates the component of relative forecast performance predicted by the parameter beliefs we measure. At each meeting t, we estimate separate rolling panel regressions for inflation and growth using the preceding eight FOMC meetings. The dependent variable is each member's forecast accuracy, defined as the negative of the cross-sectionally standardized absolute forecast error. The inflation regression includes perceived labor market slack $u_\tau - u^*_{i,\tau}$ and its interaction with $\text{PCSlope}_{i,\tau}$. The growth regression includes perceived policy tightness $r_{i,\tau} - r^*_{i,\tau}$, its interaction with $\text{ISSensitivity}_{i,\tau}$, and $g^*_{i,\tau}$. These variables capture the same model components underlying the state-dependent forecast tests in Section 3. Each regression includes meeting fixed effects, and all variables are standardized using means and standard deviations from the corresponding rolling window.

We use the resulting coefficient estimates, denoted $\widehat{\Gamma}^{\pi}_t$ and $\widehat{\Gamma}^{g}_t$, to predict each member's forecast accuracy at each meeting, compute an eight-observation rolling average of each prediction by member, and average the inflation and growth predictions:

$$\text{Fit}^{\text{PI}}_{i,t} = \frac{1}{2}\left[\overline{(X^{\pi})'\widehat{\Gamma}^{\pi}}_{i,t} + \overline{(X^{g})'\widehat{\Gamma}^{g}}_{i,t}\right],$$

where the overlines denote eight-observation rolling averages by member, $X^{\pi}_{i,t}$ and $X^{g}_{i,t}$ collect the inflation and growth regressors, and $\widehat{\Gamma}^{\pi}_{t}$ and $\widehat{\Gamma}^{g}_{t}$ are estimated using the preceding eight meetings. Realized forecast fit measures whose forecasts were closest to realized inflation and growth; parameter-implied forecast fit measures the extent to which the Phillips- and IS-curve components of a member's model, evaluated at prevailing conditions going into the meeting, predict that performance.

4.5 Model Fit and Committee Decisions

We test whether adopted decisions tilt toward the policy positions of members whose models better fit recent data. Better-fitting members would therefore be more closely aligned with adopted decisions, contribute more to their development, and dissent less often.

We estimate regressions of the form

$$Y_{i,t} = \alpha_t + \beta \, \text{Fit}_{i,t} + \varepsilon_{i,t}, \tag{11}$$

where $Y_{i,t}$ is alignment, adoption contribution, or dissent of member i in meeting t; $\text{Fit}_{i,t}$ is realized or parameter-implied forecast fit going into meeting t; and α_t is a meeting fixed effect. The outcomes are scored at the member-decision level and averaged within each member-meeting. Fit, alignment, and adoption contribution are standardized within the estimation sample; dissent is a $0/1$ indicator, so its coefficient reports the change in dissent probability associated with a one-standard-deviation increase in fit. The meeting fixed effects absorb variation common to all members at a meeting, so β is identified from differences in model fit across members facing the same set of adopted decisions.

Table VIII reports standardized estimates from Equation (11). Panel A uses *realized forecast fit*. In the full sample of voting members (columns 1–3), a one-standard-deviation increase in realized fit is associated with 0.117 standard deviations greater alignment and 0.058 standard deviations greater adoption contribution. Better-fitting voters also dissent less often, with a coefficient of -0.033, implying a 3.3 percentage point decline in the probability of dissent for a one-standard-deviation increase in fit. This is substantial relative to the 5.1% unconditional dissent probability in our sample. Among all participants, the alignment and adoption-contribution coefficients are 0.117 and 0.059. The estimates are larger in the SEP sample and retain the same signs.

These results establish a broad relationship between forecast performance and committee decisions. Adopted decisions more closely reflect the positions of members whose inflation and growth forecasts were close to realized outcomes: these members exhibit

greater alignment and adoption contribution and are less likely to dissent.

Panel B uses *parameter-implied forecast fit*, restricting attention to the component of forecast performance predicted by the Phillips curve and interest-rate sensitivity beliefs we measure. In the full voter sample, a one-standard-deviation increase in parameter-implied fit is associated with 0.206 standard deviations greater alignment and 0.142 standard deviations greater adoption contribution; the dissent coefficient is -0.048, implying a 4.8 percentage point decline in the probability of dissent. Among all participants, the alignment and adoption-contribution coefficients are 0.224 and 0.167. In the SEP sample, the corresponding estimates among all participants rise to 0.359 and 0.259. The estimates retain the same sign across both member samples and both sample periods.[17]

Taken together with the persistence of members' model parameters documented in Section 2.3, we interpret the relationship between fit and adopted decisions as evidence that members acquiesce when recent data do not support their positions, even as they continue to disagree about the economy. Appendix Table IA.20 provides concrete examples: members voice outlooks and policy preferences that differ from those of members whose models fit the data better, but ultimately go along with the adopted decision.

Appendix Tables IA.13 and IA.14 report results for $Alignment_{i,t}$ and $AdoptionContribution_{i,t}$, separately considering non-rate and rate decisions; the main relationships hold for both non-rate and rate decisions. The non-rate results provide evidence that the relationship between model fit and alignment cannot be explained simply by the committee following a mechanical Taylor rule, since these decisions, reflecting communication and balance sheet decisions, are typically discretionary.

Appendix Tables IA.15 and IA.16 report results separately using growth- and inflation-based fit measures. We find that both contribute to the relationship between members' forecast errors and their views being reflected in the committee's decisions; members' Phillips curve model parameters appear to have slightly more explanatory power than their IS curve model parameters, consistent with their higher R^2 values in explaining cross-sectional variation in members' forecasts from Section 3.3.

Lastly, Appendix Section IA.A.1 further shows that the average stance of the decisions adopted at a meeting is strongly related to the high-frequency monetary policy surprise at the announcement, indicating that the relationship between model fit and committee decisions concerns policy actions that changed market expectations.

Taken together, the results show a consistent pattern across outcomes, samples, and

[17]The difference in coefficient magnitudes between the two fit measures partly reflects differences in the variation remaining after meeting fixed effects are absorbed. The magnitudes should therefore not be interpreted as a decomposition of forecast performance into components.

decision types: members whose models better fit recent data are more closely aligned with adopted decisions, contribute more to their development, and dissent less often. The relationship appears both in realized forecast performance and in the component predicted by the Phillips- and IS-curve beliefs we measure, consistent with the committee placing greater weight on models that have performed well under recent conditions.

4.6 Model Fit and Dynamics of Major Policy Innovations

The Evans rule episode illustrates how support for a major policy decision can develop over many meetings, whereas our baseline measures are constructed from the meeting in which the policy is adopted. In this section, we examine whether these measures, based on the meeting at which the decision was adopted, retain information about the preceding deliberative process. Among major policy decisions with traceable run-ups, members who are more aligned with and contribute more to the adopted decision tend to have been the decision's earliest advocates.

Starting from our decisions dataset, we identify a sample of major adopted policy innovations, drawing on the St. Louis Fed's FRASER monetary policy timeline.[18] We retain the decisions that mark substantive changes in policy, split broad episodes such as large-scale asset purchases and forward guidance into their constituent decisions, and exclude emergency facilities, implementation arrangements, information releases and governance or transparency changes. We further require that pre-adoption deliberation be observable in the FOMC transcripts. These restrictions yield 16 decisions, concentrated in forward guidance and balance sheet policy, which are listed in Appendix Table IA.21.[19]

For each decision, we trace the pre-adoption discussion and record the first instance at which each participant expresses affirmative support for the policy ultimately adopted. We define the voiced-support window as the period from the first observed expression of support through adoption and classify members into three groups: those who first express support in the first half of the window, those who first express support in the second half, and those who express no support before adoption.

Figure VII relates the timing of support to members' positions at adoption. Panel A shows that participants who first support a decision in the beginning half of the voiced-

[18] https://fraser.stlouisfed.org/timeline/monetary-policy-history.

[19] We focus on major innovations because their pre-adoption histories can be identified. A decision such as the Evans rule leaves a distinct trail in the transcripts, which allows us to date each participant's first expression of support. Although all decisions may follow a similar process, less prominent decisions blend into the routine discussion of alternatives, and their run-ups cannot be reliably traced. Along these lines, our decision sample restriction reflects a constraint of measurement rather than a feature of the underlying decision process.

support window have the greatest alignment with the eventual decision, while those expressing support later on or not at all prior to the decision have lower alignment. The pattern appears both in raw means and after removing decision-level fixed effects. Panel B shows a similar relationship for adoption contribution, which captures active participation at the final meeting. A one-standard-deviation increase in the support-timing rank is associated with a 0.41-standard-deviation increase in final alignment and a 0.29-standard-deviation increase in final adoption contribution.[20] Finally, in Appendix Figure IA.2, we show that differences in members' model parameters are stable from the first meeting at which these decisions are discussed until their adoption. That is, the adoption of the decisions is not driven by members' model parameters shifting towards the initial advocates and proposers.

These patterns indicate that the final-meeting measures retain information about positions developed over a policy's run-up, rather than merely recording assent at the moment of adoption. Members who support a proposal while it is still taking shape remain more aligned with the eventual decision and contribute more actively when it is adopted. The length of these run-ups also weighs against interpretations in which alignment captures only *ex post* agreement or in which final-meeting statements are opportunistic endorsements of an outcome already settled: the relationship is already visible in positions expressed while the policy remains under development. Persistent model fit may help explain why some members support a proposal early and remain aligned at adoption, but it does not by itself explain the subsequent broadening of support. This interpretation is also compatible with a central role for the chair. As the Evans rule episode illustrates, policies can originate with individual members and be advanced through repeated advocacy, while the chair coordinates deliberation and consolidates support.

Viewed alongside the baseline aggregation results and the persistence documented in Section 2, the run-up evidence is consistent with a committee in which underlying models change slowly while support for the policies they imply varies with incoming data. One interpretation is that incoming macroeconomic data make it easier for members to defend those policies when the evidence accords with their view of the economy, and harder when it does not, allowing support to shift without convergence in structural beliefs. Committee policy may therefore tilt toward better-fitting models even as disagreement over the structure of the economy persists.

[20] For the regression analysis we standardize the ordinal measure, with higher values indicating earlier support.

4.7 Adaptation and Overfitting

The preceding analysis shows that committee decisions tilt toward members whose models better fit recent macroeconomic outcomes. The Evans rule illustrates how this process can benefit the committee. The decision reflected the preferences of members whose models fit the incoming data, and the low-inflation, high-slack conditions that favored those models persisted for years after adoption, vindicating the policy tilt. Yet there is no guarantee that recent model performance identifies the model best suited to future conditions. The committee's 2020 adoption of Flexible Average Inflation Targeting (FAIT) provides a cautionary example.

Under FAIT, the committee changed how it pursued its 2% objective, committing to inflation that "averages 2 percent over time," aiming for inflation "moderately above 2 percent for some time" following persistent shortfalls, and recasting its employment objective to respond to "shortfalls," rather than deviations, from maximum employment. The framework's rationale combined two features of the preceding decade: a low-r^* environment that made policy more likely to encounter the effective lower bound, and a flat Phillips curve that implied little inflationary pressure from labor market tightness. Over this period, inflation had fallen short of the committee's target in nearly every year, and members had repeatedly over-predicted it. President Williams, who ranked high on both our model fit and alignment measures, drew the lesson explicitly: this "pattern of consistently…overpredicting inflation and being overly confident about the return of inflation to its goal…should teach us…that the world has changed in important ways that should affect our decisions."

This conditioning of the framework on the prevailing macroeconomic regime drew a contemporaneous warning. As the committee deliberated over the draft statement at the July 2020 meeting, Atlanta Fed President Raphael Bostic cautioned:

> My concern about the draft statement...is that it seems to exist uncomfortably in a zone between a general statement of long-run strategy and an operational statement that is strongly conditioned on current and medium-term economic conditions...this may be appropriate, as the most important of those conditions—low r^*, below-target inflation, and persistent overestimation of the NAIRU—have been with us for a long time. But these are facts until they're not, and at times the statement reads like we are unable or unwilling to contemplate any alternative configuration of economic circumstances...

Within two years, the configuration of economic circumstances had changed sharply: inflation rose over 2021 and 2022 to levels well above the committee's target and beyond

any path its members had projected. In the 2025 framework review, Chair Jerome Powell placed the low-rate environment and flat Phillips curve at the center of both the original framework and its subsequent revision:

> At the time of the last review, we were living in a new normal, characterized by the proximity of interest rates to the effective lower bound (ELB), along with low growth, low inflation, and a very flat Phillips curve—meaning that inflation was not very responsive to slack in the economy...The economic conditions that...drove the 2020 framework changes were thought to be rooted in slow-moving global factors that would persist for an extended period...the post-pandemic reopening brought the highest inflation in 40 years...[W]e eliminated [FAIT]... the idea of an intentional, moderate inflation overshoot had proved irrelevant.

Vice Chair for Supervision Michelle Bowman drew the broader lesson for the committee's use of recent data:

> Our experience during and following the pandemic highlights the difficulty in assessing the current state of the economy and predicting how it will evolve in the presence of major supply- and demand-side shocks, possible structural changes in the economy, and real-time data and measurement uncertainty...An inflexible and dogmatic view of data dependence gives an inherently backward-looking view of the economy and would guarantee that we remain behind the curve, requiring us to catch up in the future...[W]e should consider shifting our focus from overweighting the latest data points to a proactive forward-looking approach.

Evans articulated the adaptive case for data-based model weighting in August 2012. Discussing how the Committee should compare competing models, he argued that "[I]f you were going to bring these other models to the table and say a little bit about how the data speak to that, you would want to...weight those that are most informative about the data and down-weight others." Read alongside Bowman's warning, the remark captures the central trade-off: tilting toward models with better recent fit may improve decisions when the relationships underlying that fit persist, but may make policy excessively backward-looking when those relationships change. In Appendix IA.D, we develop a model of committee aggregation that illustrates that weighting members by model fit may improve performance when members have diversifying models, but can generate excess sensitivity to recent conditions. It also indicates that some level of stubbornness by members in holding to their views may be useful to prevent such excess sensitivity.

5 Conclusion

Our findings indicate that FOMC members hold distinct and persistent models of the macroeconomy, reflected in their beliefs about structural relationships and steady-state values. These differences govern how members interpret the same economic conditions and generate state-dependent forecast disagreement. In aggregating these views, the committee tilts adopted decisions toward the policy positions of members whose models better fit recent data. These results extend theories of group decision-making beyond private-information aggregation by showing that committees also aggregate competing models. Such aggregation can be adaptive when recent model performance is informative, but it also creates a risk of overfitting when the economic relationships underlying that performance change.

References

Aina, Chiara. 2025. "Tailored stories."

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

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

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

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

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

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

Barnichon, Régis, and Geert Mesters. 2020. "Identifying Modern Macro Equations with Old Shocks." *Quarterly Journal of Economics*, 135(4): 2255–2298.

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

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

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

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra. 2019. "Regional Heterogeneity and the Refinancing Channel of Monetary Policy." *Quarterly Journal of Economics*, 134(1): 109–183.

Berger, David, Konstantin Milbradt, Fabrice Tourre, and Joseph Vavra. 2021. "Mortgage Prepayment and Path-Dependent Effects of Monetary Policy." *American Economic Review*, 111(9): 2829–2878.

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.

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

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

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

Bowmaker, Simon W. 2025. *Fed Reckoning: Conversations on America's Central Bank.* Yale University Press. Manuscript in preparation.

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

Caballero, Ricardo J., and Alp Simsek. 2020. "A Risk-Centric Model of Demand Recessions and Speculation." *Quarterly Journal of Economics*, 135(3): 1493–1566.

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, 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.

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

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

Cúrdia, Vasco, and Michael Woodford. 2016. "Credit Frictions and Optimal Monetary Policy." *Journal of Monetary Economics*, 84: 30–65.

Eichenbaum, Martin, Sergio Rebelo, and Arlene Wong. 2022. "State-Dependent Effects of Monetary Policy: The Refinancing Channel." *American Economic Review*, 112(3): 721–761.

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.

Galí, Jordi. 2008. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework.* Princeton, NJ:Princeton University Press.

Gerardi, Dino, and Leeat Yariv. 2008. "Information acquisition in committees." *Games and Economic Behavior*, 62(2): 436–459.

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

Gibbons, Robert, and John Roberts. 2013. *The handbook of organizational economics.* Princeton University Press.

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

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

Hansen, 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.

Hazell, Jonathon, Juan Herreño, Emi Nakamura, and Jón Steinsson. 2022. "The Slope of the Phillips Curve: Evidence from U.S. States." *Quarterly Journal of Economics*, 137(3): 1299–1344.

Howes, Cooper, Marc Dordal i Carreras, Olivier Coibion, and Yuriy Gorodnichenko. 2026. "How Monetary Policy Is Made: Lessons from Historical FOMC Discussions." National Bureau of Economic Research.

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.

Kakhbod, Ali, Amir Kermani, and Bernardo Maciel. 2025. "In the Fed's Mind."

Kekre, Rohan, and Moritz Lenel. 2022. "Monetary Policy, Redistribution, and Risk Premia." *Econometrica*, 90(5): 2249–2282.

Ke, Shikun. 2025. "Analysts' Belief Formation in Their Own Words."

Li, Hao, Sherwin Rosen, and Wing Suen. 2001. "Conflicts and common interests in committees." *American Economic Review*, 91(5): 1478–1497.

Li, Zigang, Stijn Van Nieuwerburgh, and Wang Renxuan. 2025. "Understanding Rationality and Disagreement in House Price Expectations." *Review of Financial Studies, Forthcoming*.

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

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

Marschak, Jacob, and Roy Radner. 1972. *Economic Theory of Teams. Cowles Foundation for Research in Economics at Yale University. Monograph Series*, New Haven and London:Yale University Press.

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

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

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

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

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

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.

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

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

Schwartzstein, Joshua, and Adi Sunderam. 2025. "Sharing Models to Interpret Data." Mimeo.

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

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

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

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton, NJ:Princeton University Press.

Tables

TABLE I
PARAMETER BELIEF ARGUMENTS AND CLASSIFICATIONS

Speaker	Date	Parameter	Value	Quote
Janet Yellen	Mar-2015	Phillips Curve	Steep	"Experience here and abroad teaches us that, as resource utilization tightens, eventually inflation will begin to rise. ... the more labor and product markets tighten, the more confident I'll become in the inflation outlook."
Stanley Fischer	Dec-2015	Phillips Curve	Moderate	"the Phillips-curve relationship will reassert itself, and I think we are beginning to see some signs of that in the behavior of labor compensation, but only faintly so."
James Bullard	Jun-2017	Phillips Curve	Flat	"The unemployment rate has fallen to 4.3 percent nationally. I do not regard this as a harbinger of higher inflation ahead. The most recent research ... suggests that any linkage between relatively low unemployment and inflation is empirically very minor in recent years."
Alan Blinder	Nov-1994	IS Curve	Strong	"higher interest rates are going to hurt autos ... They are going to hurt housing ... put a damper on business fixed investment."
Jeremy Stein	Jun-2012	IS Curve	Moderate	"when you're in an environment with extremely high uncertainty, investment is going to be less responsive to interest rates than it usually is. ... the interest rate is not going to play a big role in that."
Richard Fisher	Aug-2012	IS Curve	Weak	"I asked ... 'is there anything further the Federal Reserve can do that would impact your desire to expand domestic cap-ex or payroll?' ... the answer was no."

Note: The table presents examples of members' arguments about the Phillips curve slope and interest-rate sensitivity and their classifications. Phillips curve slope is classified as steep, moderate, or flat. Interest-rate sensitivity is classified as strong, moderate, or weak.

TABLE II
MODEL PARAMETER SUMMARY STATISTICS

PANEL A: OVERALL SAMPLE

	Summary Statistics			Estimate Distribution			
	Member-meetings	Members	Meetings	Mean	TS σ	XS σ	P90–P10
Argument-implied parameter beliefs							
PC slope	6,661	107	362	0.987	0.544	0.340	0.777
IS sensitivity	6,661	107	362	0.964	0.470	0.289	0.661
Longer-run equilibrium parameter beliefs							
u^*	787	36	48	4.979	0.458	0.268	0.566
g^*	787	36	48	2.202	0.304	0.172	0.349
r^*	584	31	36	1.311	0.602	0.294	0.644

PANEL B: BY COMMON HISTORICAL REGIME

	Summary Statistics			Estimate Distribution			
	Member-meetings	Members	Meetings	Mean	TS σ	XS σ	P90–P10
PC slope							
Inflation/Disinflation (1977–1989)	2,107	51	115	1.416	0.440	0.073	0.134
Early Greenspan (1990–1995)	952	35	48	1.379	0.300	0.382	0.806
Great Moderation (1996–2005)	1,568	34	80	0.911	0.238	0.468	1.090
GFC/ZLB (2006–2013)	1,137	39	64	0.669	0.261	0.505	1.209
Liftoff/Pandemic (2014–2020)	897	30	55	0.227	0.182	0.486	1.140
IS sensitivity							
Inflation/Disinflation (1977–1989)	2,107	51	115	1.015	0.270	0.250	0.559
Early Greenspan (1990–1995)	952	35	48	0.608	0.349	0.384	0.901
Great Moderation (1996–2005)	1,568	34	80	1.308	0.523	0.269	0.596
GFC/ZLB (2006–2013)	1,137	39	64	0.647	0.484	0.320	0.749
Liftoff/Pandemic (2014–2020)	897	30	55	1.039	0.296	0.280	0.661
u^*							
GFC/ZLB (2006–2013)	366	24	21	5.373	0.214	0.334	0.716
Liftoff/Pandemic (2014–2020)	421	28	27	4.674	0.391	0.217	0.450
g^*							
GFC/ZLB (2006–2013)	366	24	21	2.521	0.102	0.178	0.335
Liftoff/Pandemic (2014–2020)	421	28	27	1.954	0.158	0.168	0.360
r^*							
GFC/ZLB (2006–2013)	163	19	9	2.055	0.100	0.324	0.763
Liftoff/Pandemic (2014–2020)	421	28	27	1.063	0.460	0.284	0.604

Note: The table summarizes member-meeting estimates of parameter beliefs. Panel A reports statistics over the full sample and Panel B reports statistics by sub-samples. Argument-implied beliefs are estimated from members' meeting-level arguments. Longer-run equilibrium beliefs are from the Summary of Economic Projections and are reported in percentage-point units. For each row, the mean column reports the average meeting-level cross-sectional mean, the TS σ column reports the time-series standard deviation of the meeting-level committee median, the XS σ column reports the average within-meeting cross-sectional standard deviation, and the P90–P10 column reports the average within-meeting 90th-minus-10th percentile spread.

TABLE III

FACTOR STRUCTURE IN PARAMETER BELIEFS AND EXPLANATORY POWER

PANEL A: PRINCIPAL COMPONENT LOADINGS				
	Principal component 1		Principal component 2	
	Pre-SEP (1992–2008)	SEP (2009–2020)	Pre-SEP (1992–2008)	SEP (2009–2020)
PC slope	0.707 ($= 1/\sqrt{2}$)	0.557	0.707 ($= 1/\sqrt{2}$)	0.080
IS sensitivity	0.707 ($= 1/\sqrt{2}$)	0.604	-0.707 ($= -1/\sqrt{2}$)	-0.042
u^*		-0.479		0.566
g^*		-0.309		-0.819
Variance explained	0.581	0.389	0.419	0.255
PANEL B: EXPLANATORY POWER				
	Principal component 1		Principal component 2	
	Pre-SEP (1992–2008)	SEP (2009–2020)	Pre-SEP (1992–2008)	SEP (2009–2020)
Member fixed effects				
Member fixed-effect R^2	0.579	0.748	0.374	0.558
Determinant blocks				
Institution fixed effect excluding member	0.092	0.186	0.000	0.026
Career background	0.094	0.208	0.009	0.153
Bordo-Istrefi macro experience	0.043	0.075	0.025	0.149
Malmendier-Nagel macro experience	0.014	0.065	0.008	0.036
Demographics and education	0.219	0.060	0.060	0.122
Observations	2,653	784	2,653	784
Members	52	35	52	35
Meetings	136	48	136	48
Meeting fixed effects	Yes	Yes	Yes	Yes

Note: Panel A reports principal component loadings and variance explained by the principal components of members' parameter beliefs. Panel B reports R^2 values from regressions of the principal components on member fixed effects after absorbing meeting fixed effects, and separately, on blocks of independent variables after absorbing meeting fixed effects. Pre-SEP principal components use PC slope and IS sensitivity; SEP principal components additionally include u^* and g^* from the Summary of Economic Projections. The independent variables used are defined in Table IA.6.

TABLE IV
PARAMETER BELIEFS AND ONE-YEAR-AHEAD FORECASTS

PANEL A: PHILLIPS CURVE AND INFLATION FORECASTS

	Main specification		No-selection specification	
	(1)	(2)	(3)	(4)
$(u_t - u^*_{i,t})$	-0.242	-0.204	-0.220	-0.204
	(-5.66)	(-5.19)	(-5.26)	(-4.18)
$\text{PCSlope}_{i,t} \times (u_t - u^*_{i,t})$	-0.064	-0.059	-0.081	-0.069
	(-3.31)	(-3.84)	(-4.79)	(-4.16)
$\text{Supply}_{i,t}$		0.076		0.057
		(6.13)		(4.95)
Avg. cross-sectional forecast σ	0.250	0.250	0.250	0.250
Interaction effect / forecast σ	-25.4%	-23.4%	-32.4%	-27.6%
Observations	1,197	1,197	1,197	1,197
R^2	0.127	0.205	0.161	0.203
Meeting fixed effects	Yes	Yes	Yes	Yes

PANEL B: INTEREST RATE SENSITIVITY AND $g - g^*$

	Main specification		No-selection specification	
	(1)	(2)	(3)	(4)
$(r_{i,t} - r^*_{i,t})$	-0.331	-0.322	-0.324	-0.302
	(-1.93)	(-1.90)	(-1.96)	(-1.83)
$\text{IS sensitivity}_{i,t} \times (r_{i,t} - r^*_{i,t})$	-0.099	-0.099	-0.096	-0.095
	(-4.69)	(-4.73)	(-5.07)	(-5.05)
$\text{Demand}_{i,t}$		0.024		0.031
		(1.86)		(2.58)
Avg. cross-sectional forecast σ	0.361	0.361	0.361	0.361
Interaction effect / forecast σ	-27.5%	-27.5%	-26.6%	-26.3%
Observations	1,197	1,197	1,197	1,197
R^2	0.064	0.067	0.061	0.066
Meeting fixed effects	Yes	Yes	Yes	Yes

Note: Panel A reports regressions of one-year-ahead inflation forecasts on observed unemployment less the member's perceived natural rate, $(u_t - u^*_{i,t})$, its interaction with perceived Phillips curve slope, and perceived supply conditions. Panel B reports regressions of one-year-ahead real GDP growth relative to potential growth (g^*) on the forecasted real policy rate less the member's perceived neutral rate, $(r_{i,t} - r^*_{i,t})$, its interaction with perceived interest-rate sensitivity, and perceived demand conditions. The sample spans 1992–2020. Perceived Phillips curve slope and interest-rate sensitivity measures are based on transcripts from meetings prior to the forecasts. The u^*, g^*, and policy-tightness inputs use member-specific SEP values when available; missing member values are filled with the within-meeting mean among members with available values. Meetings without any member-specific values use the corresponding Board staff measure, assuming $r^* = 2\%$. Columns 1 and 2 use transcript-based parameter estimates that model both argument scores and selection into discussion; Columns 3 and 4 use estimates from argument scores alone. Transcript-based parameters, interaction terms, and perceived supply and demand conditions are standardized; forecasts and economic conditions are in percentage-point units. Standard errors are two-way clustered by member and meeting; t-statistics are in parentheses.

TABLE V
FOMC PARTICIPANTS AND THE DECEMBER 2012 THRESHOLD DECISION

PANEL A: ALIGNMENT WITH THE DECEMBER 2012 DECISION

	Model (beliefs)					Proj. liftoff		Position	
Group (n)	PC	IS	u^*	r^*	g^*	Jan '12	Dec '12	Align.	Adopt.
Strong supporters (12)	−0.02	+0.10	5.40	1.93	2.42	2014.4	2015.1	+2.0	2.00
Reluctant supporters (4)	−0.31	+0.04	5.88	2.26	2.35	2013.5	2014.2	+1.0	0.75
Opposers (3)	−0.25	−0.49	5.83	2.25	2.43	2012.3	2013.7	−1.7	0.00

PANEL B: FIRST SUPPORT FOR NUMERICAL THRESHOLDS

	Model (beliefs)					Proj. liftoff		Position	
Group (n)	PC	IS	u^*	r^*	g^*	Jan '12	Dec '12	Align.	Adopt.
Early supporters (2011) (7)	+0.03	+0.18	5.39	1.77	2.49	2014.7	2015.0	+1.9	1.71
2012 converts (Mar.–Sep.) (5)	+0.04	+0.00	5.58	2.14	2.34	2013.8	2015.2	+2.0	2.00
Late adopters (Oct.–Dec.) (4)	−0.47	+0.01	5.67	2.26	2.35	2013.3	2014.2	+1.2	1.25
Holdouts (never) (3)	−0.25	−0.49	5.83	2.25	2.43	2012.3	2013.7	−1.7	0.00

Note: The table groups the nineteen participants in the December 2012 threshold-guidance decision two ways. In Panel A, participants are grouped by final alignment with the adopted decision; in Panel B, participants are grouped by the first meeting at which each member supported numerical thresholds. Columns report group means. PC and IS are transcript-based model-belief deviations as of the December 2012 meeting. The u^*, r^*, and g^* columns use members' December 2012 long-run projections. Projected liftoff is the average SEP-implied year in which the group expected the federal funds rate to leave the zero lower bound in January and December 2012. Align. and Adopt. report alignment and adoption contribution, which range from −2 to +2 and 0 to 3, respectively.

TABLE VI
POLICY DECISIONS AND POLICY ALIGNMENT: EXAMPLES

PANEL A: EXAMPLE POLICY DECISIONS

Decision	Change
State-Dependent Forward Guidance	Replace calendar-date funds-rate guidance with economic thresholds: 0 to 1/4 percent remains appropriate while unemployment is above 6.5 percent, projected inflation is no more than 2.5 percent, and inflation expectations remain anchored.
Treasury Purchases	Begin $45 billion per month in longer-term Treasury purchases after the Maturity Extension Program ends.

PANEL B: EXAMPLE SPEAKER ALIGNMENT AND ADOPTION SCORES

Decision	Speaker	Alignment	Adoption	Evidence
State-Dependent Forward Guidance	Dudley	2	2	Supports thresholds as dynamic automatic stabilizers that improve on date-based guidance.
State-Dependent Forward Guidance	Pianalto	1	2	Supports threshold language as a compromise and asks the Chair to explain key qualifications.
State-Dependent Forward Guidance	Plosser	-2	0	Opposes thresholds as not "yet ready for prime time" and prefers simpler language.
Treasury Purchases	Rosengren	2	2	Supports continuing Treasury purchases, tying $45 billion to a fragile recovery and labor-market improvement.
Treasury Purchases	Powell	-1	1	Supports the package despite misgivings, but favors fewer purchases and more MBS.
Treasury Purchases	Plosser	-2	0	Opposes adding Treasury purchases as confused with fiscal policy.

Note: Panel A describes the state-dependent forward-guidance and Treasury-purchase decisions made at the December 2012 meeting. Panel B reports examples of LLM-scored alignment and adoption contribution at the speaker level for each decision. Alignment ranges from -2 to $+2$, where negative values indicate opposition and positive values indicate support. Adoption contribution ranges from 0 to 3, where larger values indicate more direct contributions to adopted language or decision-specific reasoning. Panel B also reports the rationale for each assigned score.

TABLE VII
DECISION ALIGNMENT, ADOPTION CONTRIBUTION, AND DISSENT

PANEL A: VOTERS

	$Alignment_{i,t}$ range			
	[-2, -1]	(-1, 0]	(0, 1]	(1, 2]
Average $Alignment_{i,t}$	-1.17	-0.16	0.65	1.68
Average adoption contribution	0.10	0.10	0.51	0.90
Percentage dissenting votes	90.5%	24.7%	1.6%	0.0%
Number of voter-meetings	21	381	749	1,280

PANEL B: NON-VOTERS

	$Alignment_{i,t}$ range			
	[-2, -1]	(-1, 0]	(0, 1]	(1, 2]
Average $Alignment_{i,t}$	-1.09	-0.17	0.59	1.66
Average adoption contribution	0.05	0.11	0.48	0.81
Number of nonvoter-meetings	45	452	567	608

PANEL C: CROSS-MEMBER STATISTICS

	Chair	Governors	Presidents	
			Voting	Non-voting
Average $Alignment_{i,t}$	1.49	1.09	0.92	0.73
Average adoption contribution	1.12	0.64	0.55	0.47
Percentage dissenting votes	0.0%	1.4%	9.6%	–
N	231	1,050	1,150	1,584

PANEL D: CHAIR TENURES

	Greenspan	Bernanke	Yellen	Powell
Chair average $Alignment_{i,t}$	1.48	1.74	1.48	0.83
Other members' average $Alignment_{i,t}$	0.65	1.08	1.15	1.27
Chair average adoption contribution	1.38	1.21	0.56	0.36
Other members' average adoption contribution	0.52	0.58	0.57	0.62
Number of meetings	113	64	32	22

Note: Panels A and B group member-meeting observations into bins based on $Alignment_{i,t}$, the average of member-decision alignment scores in the meeting. In this measure, $Alignment_{i,t}$ ranges from -2 (clear opposition) to +2 (clear support). Adoption contribution is the average member-decision score, where 0 indicates no positive decision-specific contribution and 3 indicates that a concrete item proposed by the member appears in the adopted decision. Panel C reports averages by member role. Panel D reports each chair's average alignment and adoption contribution during that chair's tenure, alongside the corresponding averages for non-chair committee members in meetings chaired by that person. The table is restricted to the MPR and SEP forecast eras, beginning with the first MPR forecast-panel meeting on February 5, 1992.

TABLE VIII

REALIZED AND PARAMETER-IMPLIED FORECAST FIT, AND COMMITTEE DECISIONS

PANEL A: REALIZED FORECAST FIT

	Full sample					SEP sample				
	Voters			All participants		Voters			All participants	
	Align. (1)	Adopt. (2)	Dissent (3)	Align. (4)	Adopt. (5)	Align. (6)	Adopt. (7)	Dissent (8)	Align. (9)	Adopt. (10)
Forecast fit	0.117	0.058	-0.033	0.117	0.059	0.173	0.098	-0.041	0.163	0.092
	(3.33)	(1.88)	(-3.15)	(3.11)	(2.55)	(2.33)	(1.98)	(-2.01)	(2.44)	(2.79)
Observations	2,122	2,122	2,131	3,583	3,583	1,017	1,017	1,017	1,702	1,702
Meetings	227	227	227	227	227	105	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

PANEL B: PARAMETER-IMPLIED FORECAST FIT

	Full sample					SEP sample				
	Voters			All participants		Voters			All participants	
	Align. (1)	Adopt. (2)	Dissent (3)	Align. (4)	Adopt. (5)	Align. (6)	Adopt. (7)	Dissent (8)	Align. (9)	Adopt. (10)
Parameter-implied fit	0.206	0.142	-0.048	0.224	0.167	0.355	0.222	-0.077	0.359	0.259
	(2.70)	(2.90)	(-2.17)	(3.19)	(4.01)	(2.88)	(3.01)	(-2.16)	(3.74)	(4.23)
Observations	2,039	2,039	2,048	3,449	3,449	1,017	1,017	1,017	1,702	1,702
Meetings	219	219	219	219	219	105	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Panel A reports results from regressions of decision agreement measures on realized forecast fit, including meeting fixed effects. Panel B reports the corresponding regressions using parameter-implied forecast fit. Realized forecast fit is the negative mean of eight-observation rolling averages of within-meeting standardized inflation and GDP absolute forecast errors. Parameter-implied forecast fit projects inflation and GDP fit on members' model parameters and perceived economic conditions, then computes eight-observation rolling averages of the fitted values. Policy tightness is the forecasted current-year federal funds rate less one-year-ahead inflation and r^*. The u^*, g^*, and policy-tightness inputs use member-specific values when available, meeting means for missing members within a meeting, and Board staff measures otherwise; the Board staff policy-tightness measure uses an assumed 2-percent r^*. Fit, alignment, and adoption contribution are standardized within their estimation samples. Dissent is a 0/1 outcome. Standard errors are two-way clustered by member and meeting and for parameter-implied rows include a stacked sandwich correction for first-stage estimation. t-statistics are in parentheses. Observation and meeting counts are reported by column.

Figures

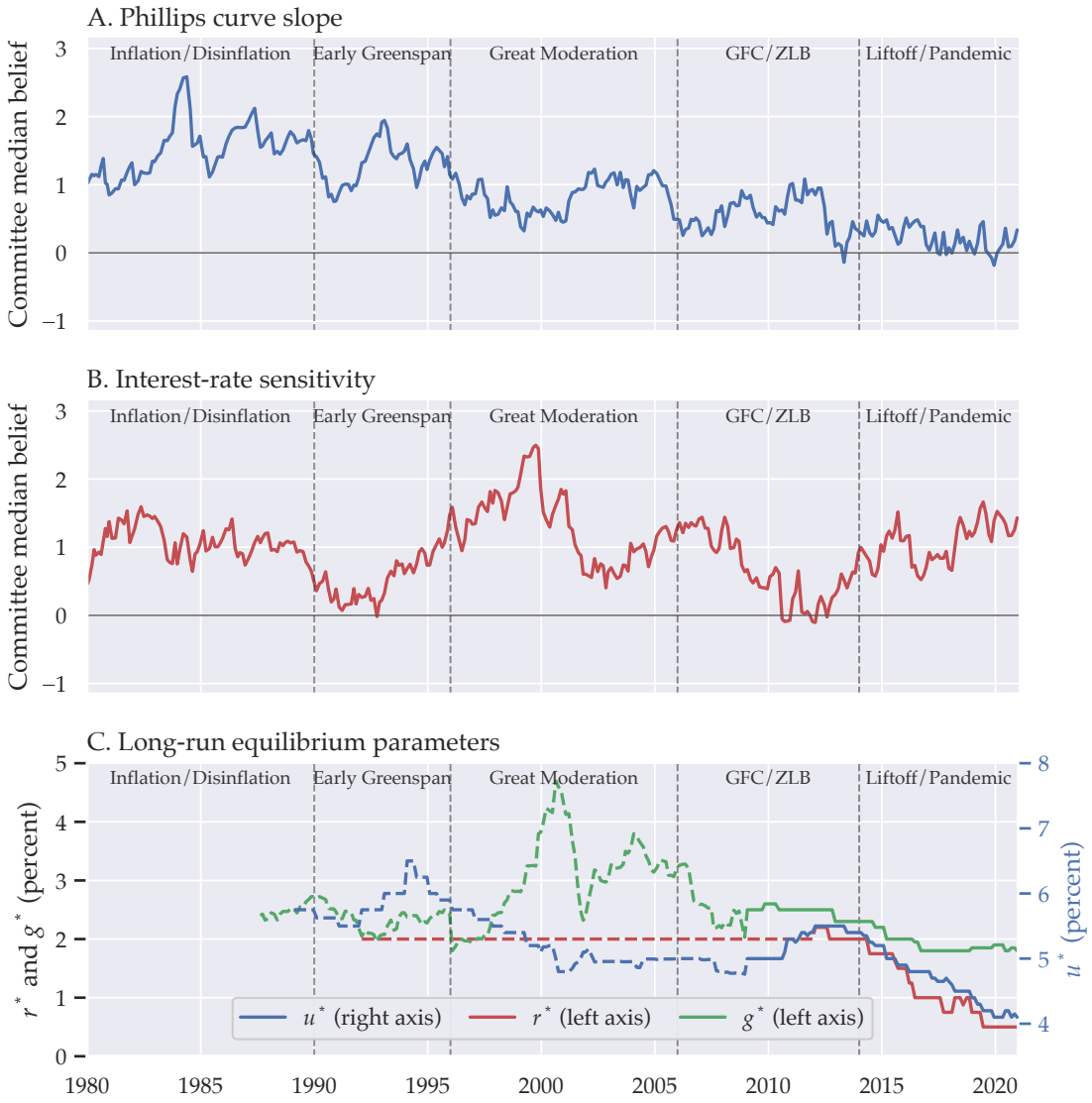


FIGURE I
COMMITTEE MEDIAN PARAMETER BELIEFS

Note: Panel A plots the within-meeting committee median of members' beliefs about the Phillips curve slope. Panel B plots the corresponding median of members' beliefs about interest-rate sensitivity. Panel C plots the long-run equilibrium parameters used in the analysis. In Panel C, solid lines plot the meeting median of member-specific SEP beliefs when available; dashed lines extend the series to the pre-SEP period with common inputs. For u^*, the dashed pre-SEP line uses the real-time NAIRU series beginning in 1989. For g^*, it uses Greenbook-implied one-year potential growth. For r^*, it uses the canonical 2-percent Taylor rule benchmark. Vertical dashed lines mark the non-overlapping historical regimes.

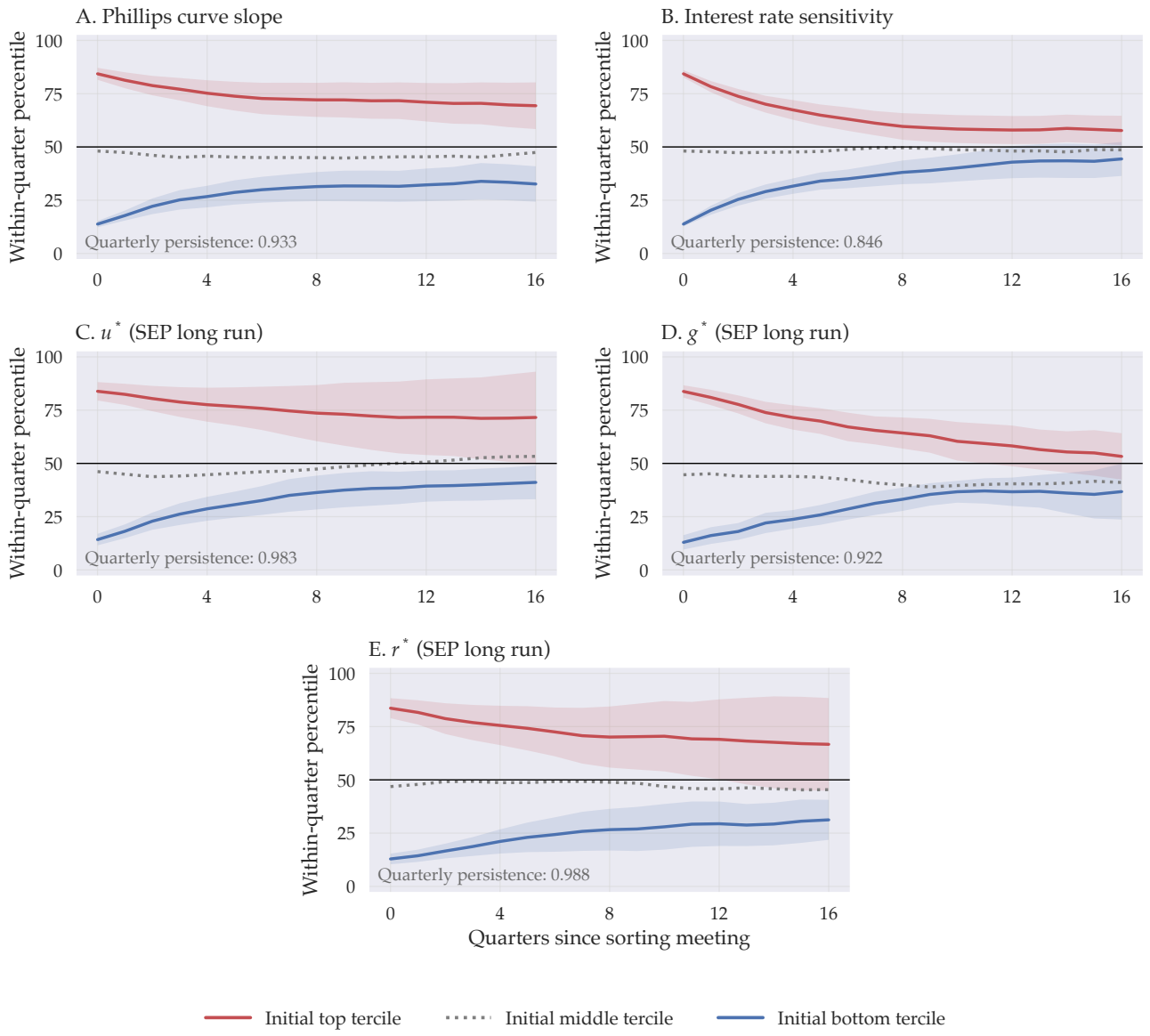


FIGURE II
PARAMETER BELIEFS' PERSISTENCE

Note: This figure displays the evolution of parameter beliefs by initial tercile. Members are assigned to terciles based on their belief in a sorting quarter ($h = 0$). The figure tracks each tercile's mean within-quarter percentile rank from the sorting quarter through the following sixteen calendar quarters ($h = 0, \ldots, 16$) for the Phillips curve slope, interest-rate sensitivity, u^*, g^*, and r^*. Panels A and B use real-time member components for Phillips curve slope and interest-rate sensitivity, and estimated quarterly persistence values are from the state-space model described in Section 2.2.3. Panels C, D, and E use long-run SEP values for u^*, g^*, and r^*, and persistence values are estimated from an AR(1) model including meeting fixed effects. Shaded regions show 95 percent confidence intervals. Standard errors are two-way clustered by quarter and member.

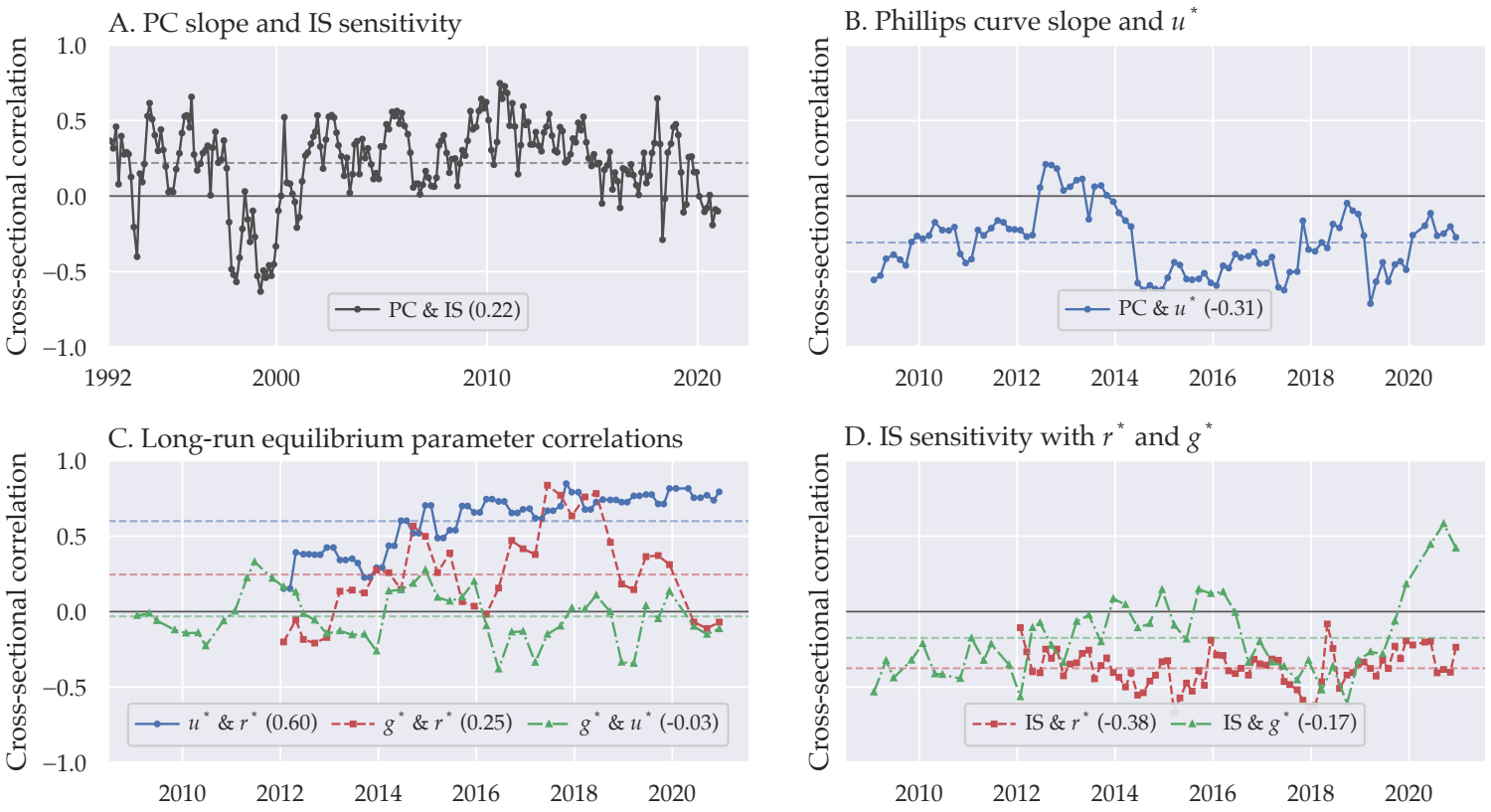


FIGURE III
CROSS-PARAMETER CORRELATIONS

Note: The figure plots meeting-level cross-sectional correlations of members' parameter beliefs. Panel A reports the correlation between perceived Phillips curve slope and perceived interest-rate sensitivity. Panel B reports the correlation between perceived Phillips curve slope and u^*. Panel C reports the pairwise correlations among u^*, r^*, and g^*. Panel D reports the correlations of perceived interest-rate sensitivity with r^* and g^*. Phillips curve slope and interest-rate sensitivity use estimated member-specific deviations from the meeting-level common component. The u^*, r^*, and g^* measures use member-specific SEP long-run projections. In each panel, dashed horizontal lines report the full sample average correlation.

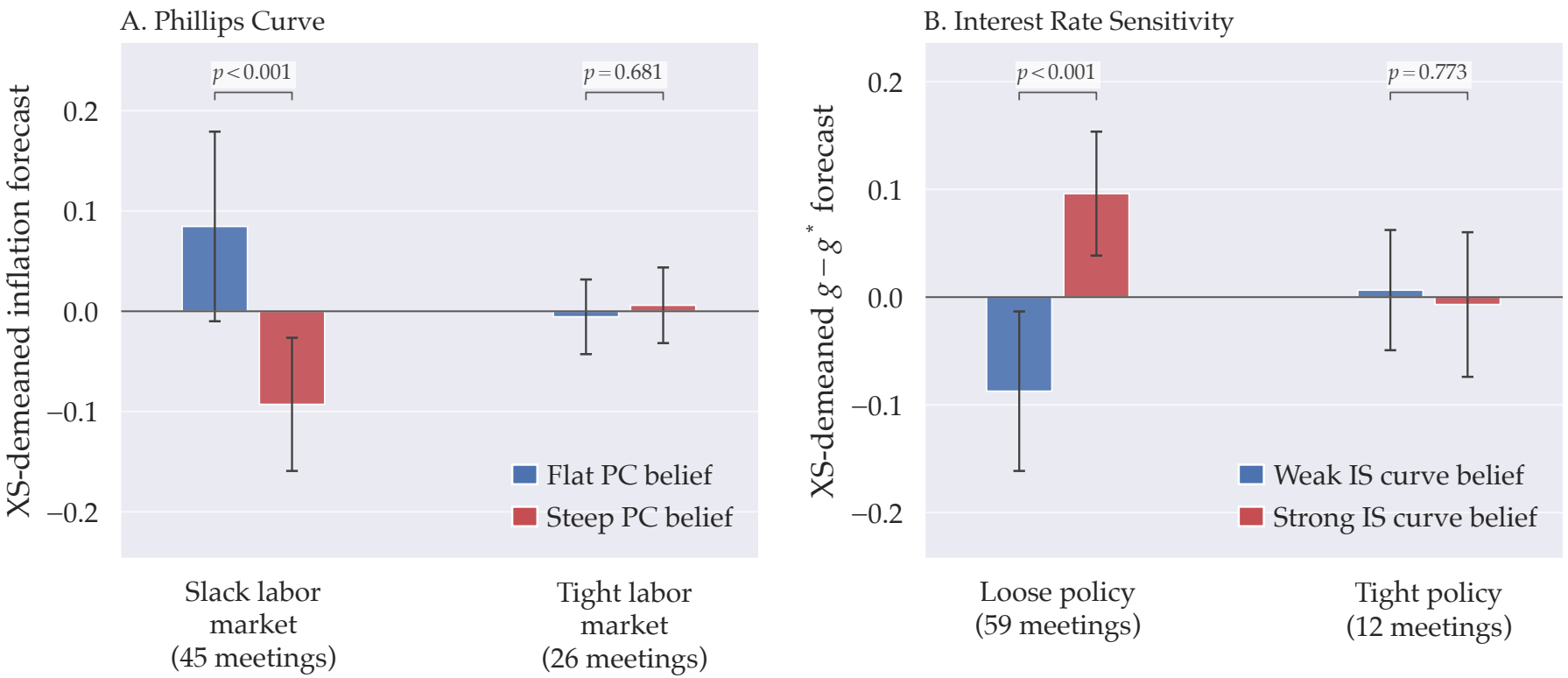


FIGURE IV
STATE DEPENDENCE AND FORECASTS

Note: Panel A reports the cross-sectionally demeaned average one-year-ahead inflation forecasts for members with below- and above-median Phillips curve slope beliefs in slack and tight labor market conditions. Labor market tightness is $u_t^* - u_t$, where u_t is the most recent unemployment rate as of the meeting and u_t^* is the meeting mean of members' $u_{i,t}^*$ beliefs; meetings without member-specific SEP projections use the Greenbook NAIRU forecast for u_t^*. Slack labor market conditions correspond to $u_t - u_t^* > 0$ and tight labor market conditions correspond to $u_t - u_t^* < 0$. Panel B reports the cross-sectionally demeaned average one-year-ahead forecasts of GDP growth relative to potential for members with below- and above-median interest-rate sensitivity beliefs in loose and tight policy conditions. Policy conditions use the meeting mean of member-specific $(r_{i,t} - r_{i,t}^*)$, where $r_{i,t}$ is the end-of-year Federal funds rate forecast less one-year-ahead inflation; meetings without member-specific inputs use the Greenbook Federal funds rate forecast with an assumed 2% r^*. Loose policy conditions correspond to $r_t - r_t^* < 0$ and tight policy conditions correspond to $r_t - r_t^* > 0$. Standard errors are two-way clustered by member and meeting. Error bars show 95 percent confidence intervals.

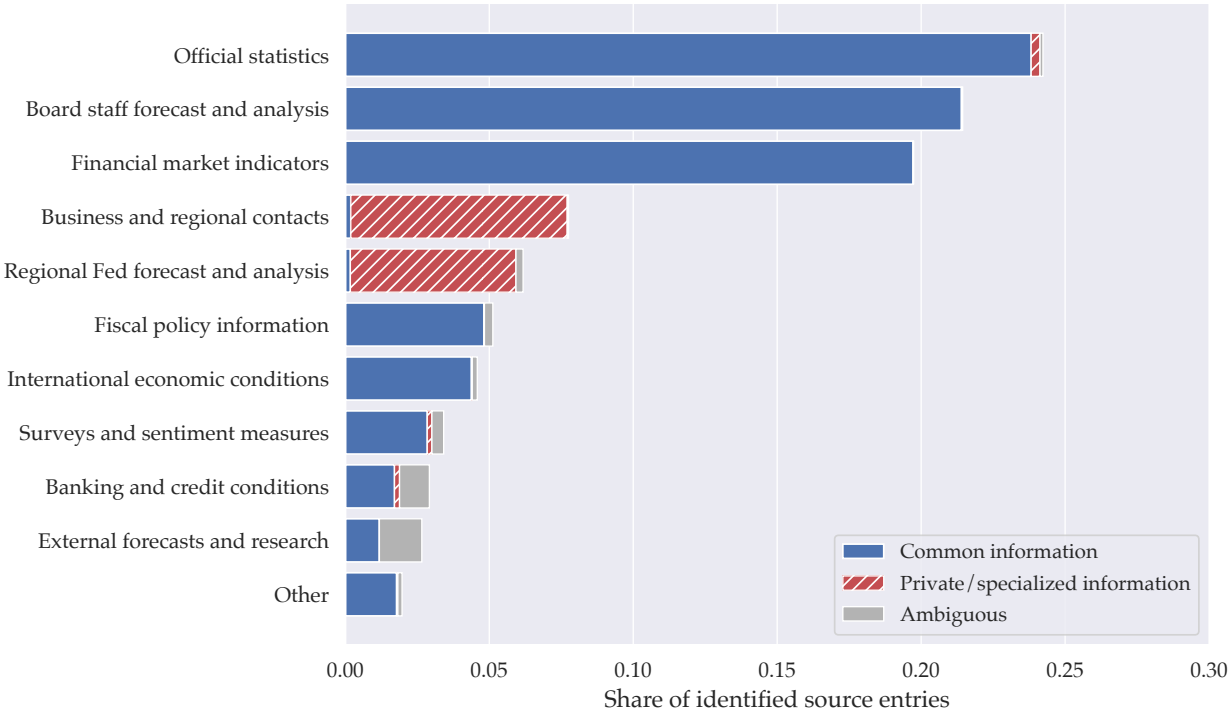


FIGURE V
COMMON AND PRIVATE INFORMATION BY SOURCE

Note: The figure plots the share of identified information sources by source category, for arguments used to construct the Phillips curve slope and interest-rate sensitivity parameter values. Bars distinguish sources classified as common to the Committee from sources classified as private or specialized. Gray bars indicate identified sources whose common observability is ambiguous.

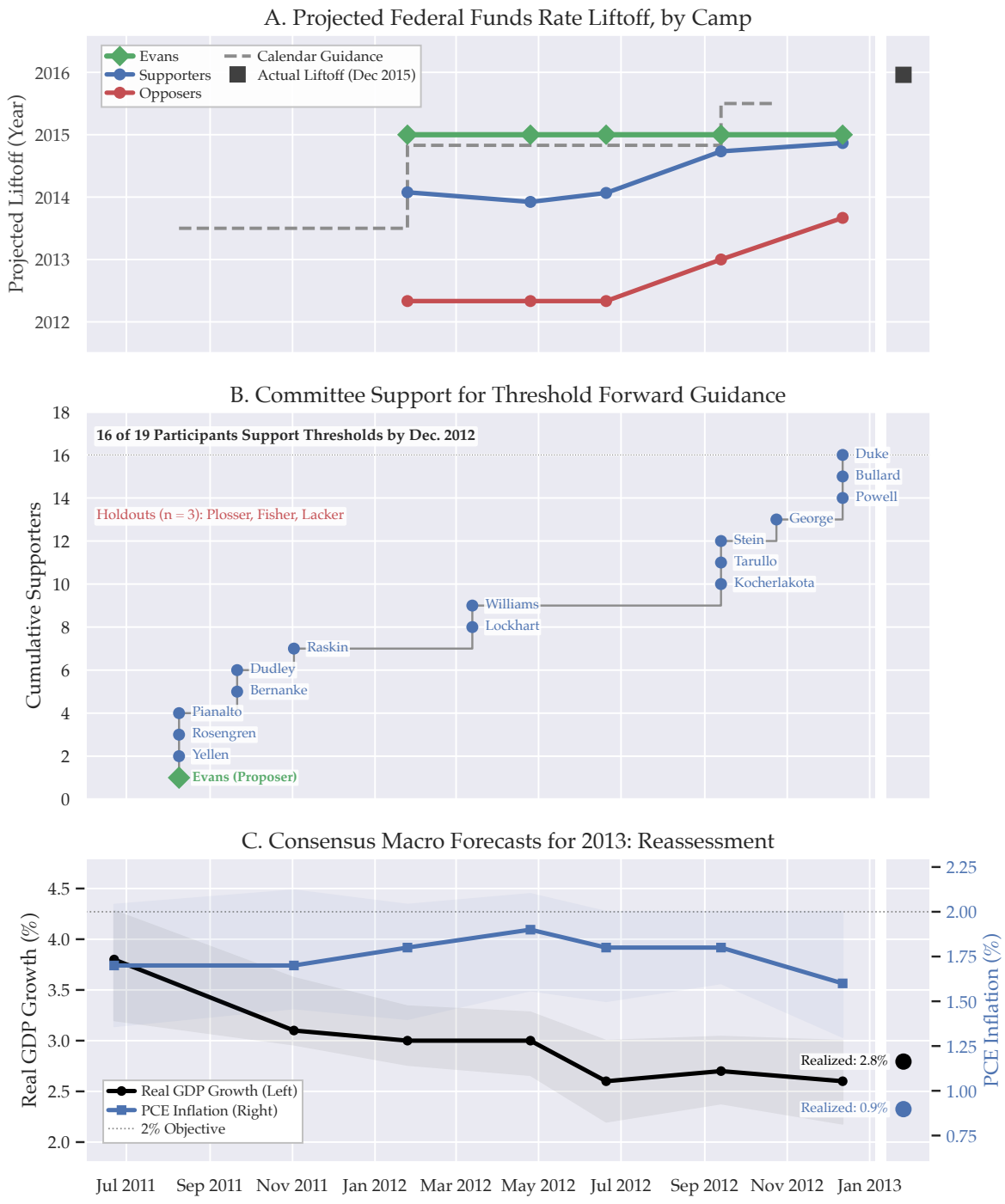


FIGURE VI

THE PATH TO THE DECEMBER 2012 THRESHOLD DECISION

Note: Panel A plots the projected year of federal-funds-rate liftoff by Evans-rule camp, along with the Committee's stated calendar guidance and the realized December 2015 liftoff. Panel B labels participants by the first meeting at which they supported numerical threshold guidance, with holdouts listed separately. Panel C plots the Committee median SEP projection for 2013 real GDP growth and PCE inflation during the run-up, with forecast dispersion bands. Realized outcomes appear to the right in the top and bottom panels.

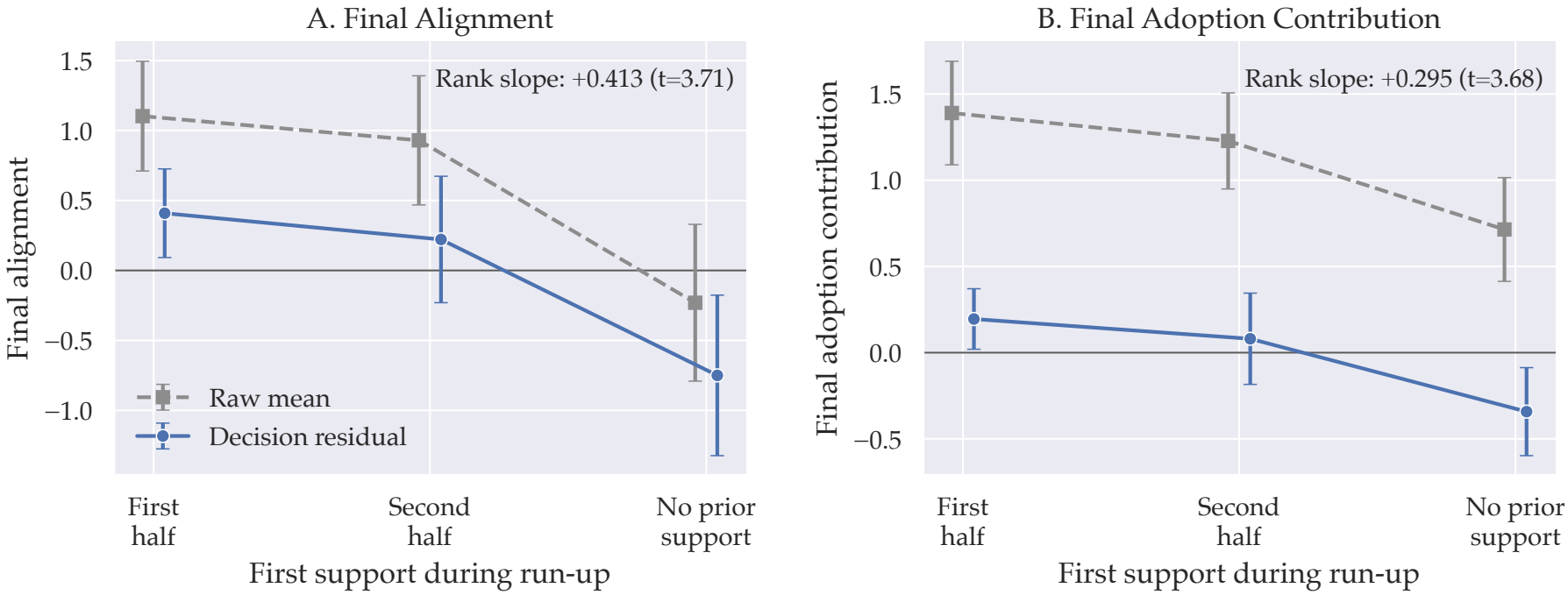


FIGURE VII
RUN-UP SUPPORT AND FINAL DECISION POSITIONS

Note: The figure reports final decision alignment and adoption contribution for the sixteen major policy-innovation decisions with traced pre-adoption discussion and within-decision support-timing variation. Members are grouped by whether their first positive support prior to the adoption meeting appears in the first half of the pre-adoption window, the second half, or is not observed before adoption. Bars show 95 percent confidence intervals. The rank slope reports the coefficient from a regression of the standardized outcome on rank of timing of support, including decision fixed effects. Standard errors are two-way clustered by member and decision.

Internet Appendix for

Policy by Committee

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

The Internet Appendix consists of five sections. Section IA.A presents additional empirical analyses. Section IA.B is a data appendix presenting details on the processing of FOMC transcripts. Section IA.C provides details on the state space estimation used to estimate members' parameter beliefs from their arguments. Section IA.D presents a simple model of committee decision-making based on our results. Section IA.E provides LLM prompts used in our analysis and presents comparisons of LLM-based coding with human coding.

IA.A Appendix Empirical Analyses

TABLE IA.1
ARGUMENT SUMMARY STATISTICS

PANEL A: OVERALL SAMPLE

	Summary Statistics			Score Distribution		
	Relevant arguments	Share of total arguments	Per member-meeting	Flat/weak	Moderate	Steep/strong
Phillips Curve						
All members	2,603	4.98%	0.380	16.2%	9.0%	74.8%
Voting members	1,536	4.98%	0.380	16.3%	10.2%	73.6%
Non-voting members	1,067	5.00%	0.380	16.1%	7.4%	76.5%
IS Curve						
All members	3,292	6.30%	0.481	11.3%	16.9%	71.9%
Voting members	1,955	6.33%	0.484	10.4%	18.1%	71.5%
Non-voting members	1,337	6.26%	0.476	12.5%	15.1%	72.4%

PANEL B: BY COMMON HISTORICAL REGIME

	Summary Statistics			Score Distribution		
	Relevant arguments	Share of total arguments	Per member-meeting	Flat/weak	Moderate	Steep/strong
Phillips Curve						
Inflation/Disinflation (1976–1989)	295	3.29%	0.129	7.5%	3.1%	89.5%
Early Greenspan (1990–1995)	304	4.78%	0.319	7.2%	5.3%	87.5%
Great Moderation (1996–2005)	819	6.13%	0.522	11.6%	7.4%	81.0%
GFC/ZLB (2006–2013)	519	4.24%	0.456	14.8%	8.9%	76.3%
Liftoff/Pandemic (2014–2020)	666	5.89%	0.742	30.9%	15.5%	53.6%
IS Curve						
Inflation/Disinflation (1976–1989)	896	10.00%	0.390	9.7%	13.8%	76.5%
Early Greenspan (1990–1995)	564	8.87%	0.592	9.6%	18.8%	71.6%
Great Moderation (1996–2005)	732	5.48%	0.467	6.6%	15.3%	78.1%
GFC/ZLB (2006–2013)	669	5.47%	0.588	20.2%	22.9%	57.0%
Liftoff/Pandemic (2014–2020)	431	3.81%	0.480	10.9%	13.9%	75.2%

Note: The table summarizes arguments used to construct members' Phillips curve slope and interest-rate sensitivity estimates. Scores equal 1 for steep Phillips curve or strong interest-rate sensitivity statements, 0 for moderate statements, and -1 for flat Phillips curve or weak interest-rate sensitivity statements. The first column reports the number of arguments deemed relevant for parameter beliefs. The second column, Share of total arguments, reports the percentage of extracted committee-member arguments that correspond to the row's model-belief measure. The third column, Arguments per member-meeting, divides relevant arguments by the number of member-meetings. The fourth, fifth, and sixth columns report percentages of arguments by classification. For Phillips curve arguments, flat/weak and steep/strong refer to flat and steep statements; for IS arguments, they refer to weak and strong statements. Panel A reports the full sample for all members, voters, and non-voters. Panel B reports the same statistics across the historical regimes.

TABLE IA.2
PARAMETER BELIEFS AND ONE-YEAR-AHEAD FORECASTS: PRIOR ARGUMENT AVERAGES

PANEL A: PHILLIPS CURVE AND INFLATION FORECASTS				
	Expanding average		Trailing eight-meeting average	
	(1)	(2)	(3)	(4)
$(u_t - u^*_{i,t})$	-0.231	-0.202	-0.252	-0.240
	(-5.40)	(-4.51)	(-5.97)	(-6.17)
$\text{PCSlope}_{i,t} \times (u_t - u^*_{i,t})$	-0.092	-0.095	-0.086	-0.083
	(-4.60)	(-4.70)	(-3.80)	(-3.58)
$\text{Supply}_{i,t}$		0.061		0.048
		(3.33)		(3.20)
Avg. cross-sectional forecast σ	0.252	0.252	0.244	0.245
Interaction effect / forecast σ	-36.5%	-37.7%	-35.3%	-33.8%
Observations	1,070	1,060	937	909
R^2	0.190	0.217	0.170	0.185
Meeting fixed effects	Yes	Yes	Yes	Yes

PANEL B: INTEREST RATE SENSITIVITY AND $g - g^*$				
	Expanding average		Trailing eight-meeting average	
	(1)	(2)	(3)	(4)
$(r_{i,t} - r^*_{i,t})$	-0.314	-0.310	-0.327	-0.328
	(-1.80)	(-1.76)	(-1.83)	(-1.80)
$\text{IS sensitivity}_{i,t} \times (r_{i,t} - r^*_{i,t})$	-0.085	-0.085	-0.111	-0.112
	(-2.66)	(-2.66)	(-4.08)	(-4.10)
$\text{Demand}_{i,t}$		0.008		0.002
		(0.32)		(0.10)
Avg. cross-sectional forecast σ	0.358	0.358	0.361	0.361
Interaction effect / forecast σ	-23.8%	-23.7%	-30.8%	-31.0%
Observations	1,120	1,120	1,034	1,032
R^2	0.048	0.048	0.066	0.066
Meeting fixed effects	Yes	Yes	Yes	Yes

Note: The table repeats the regressions from Table IV using simple averages to aggregate members' scored arguments to proxy for their parameter beliefs. Columns 1 and 2 use averages from all past meetings; columns 3 and 4 use trailing eight-meeting averages. Supply and demand controls use the corresponding directional argument-score averages.

TABLE IA.3
PARAMETER BELIEFS AND ONE-YEAR-AHEAD FORECASTS: COMMON ECONOMIC CONDITIONS

PANEL A: PHILLIPS CURVE AND INFLATION FORECASTS				
	Main specification		No-selection specification	
	(1)	(2)	(3)	(4)
$\text{PCSlope}_{i,t} \times (u - u^*)_t$	-0.071	-0.065	-0.095	-0.083
	(-3.22)	(-3.62)	(-5.30)	(-4.56)
$\text{Supply}_{i,t}$		0.084		0.061
		(6.28)		(5.09)
Avg. cross-sectional forecast σ	0.250	0.250	0.250	0.250
Interaction effect / forecast σ	-28.6%	-26.2%	-38.0%	-33.1%
Observations	1,197	1,197	1,197	1,197
R^2	0.071	0.169	0.127	0.177
Meeting fixed effects	Yes	Yes	Yes	Yes

PANEL B: INTEREST RATE SENSITIVITY AND $g - g^*$				
	Main specification		No-selection specification	
	(1)	(2)	(3)	(4)
$\text{IS sensitivity}_{i,t} \times (r - r^*)_t$	-0.081	-0.081	-0.075	-0.075
	(-4.65)	(-4.74)	(-4.36)	(-4.40)
$\text{Demand}_{i,t}$		0.030		0.044
		(2.01)		(2.98)
Avg. cross-sectional forecast σ	0.361	0.361	0.361	0.361
Interaction effect / forecast σ	-22.4%	-22.5%	-20.6%	-20.8%
Observations	1,197	1,197	1,197	1,197
R^2	0.034	0.039	0.029	0.039
Meeting fixed effects	Yes	Yes	Yes	Yes

Note: The table repeats Table IV, replacing member-specific labor market slack and policy tightness with common measures of conditions, $(u - u^*)_t$ and $(r - r^*)_t$. Each common condition is the within-meeting mean of completed member conditions. Meetings without member-specific conditions use the Board staff unemployment gap in Panel A and Board staff year-end federal funds and one-year-ahead inflation forecasts with an assumed 2-percent r^* in Panel B. Meeting fixed effects absorb the common-condition main effects, so the table reports only the parameter-by-condition interactions.

TABLE IA.4
INTEREST-RATE SENSITIVITY AND GDP FORECASTS: g^* CONTROL

INTEREST RATE SENSITIVITY AND GDP GROWTH FORECASTS				
	Main specification		No-selection specification	
	(1)	(2)	(3)	(4)
$(r_{i,t} - r^*_{i,t})$	-0.360	-0.350	-0.353	-0.334
	(-2.08)	(-2.05)	(-2.11)	(-1.99)
IS sensitivity$_{i,t} \times (r_{i,t} - r^*_{i,t})$	-0.090	-0.089	-0.086	-0.086
	(-3.97)	(-3.99)	(-4.00)	(-3.99)
g^*	0.443	0.426	0.442	0.452
	(2.31)	(2.24)	(2.44)	(2.50)
Demand$_{i,t}$		0.030		0.027
		(2.45)		(2.32)
Avg. cross-sectional forecast σ	0.357	0.357	0.357	0.357
Interaction effect / forecast σ	-25.3%	-25.1%	-24.3%	-24.1%
Observations	1,197	1,197	1,197	1,197
R^2	0.082	0.087	0.079	0.083
Meeting fixed effects	Yes	Yes	Yes	Yes

Note: The table repeats Panel B of Table IV, using GDP forecasts as the dependent variable, and adding members' g^* values as controls. The g^* terms use member-specific SEP long-run projections when available and Greenbook-based measures otherwise.

TABLE IA.5
PARAMETER BELIEFS AND ONE-YEAR-AHEAD FORECASTS: PARAMETER LEVEL INCLUDED

PANEL A: PHILLIPS CURVE AND INFLATION FORECASTS				
	Main specification		No-selection specification	
	(1)	(2)	(3)	(4)
$\text{PCSlope}_{i,t}$	-0.027	-0.021	-0.017	-0.019
	(-2.44)	(-1.92)	(-1.35)	(-1.66)
$(u_t - u^*_{i,t})$	-0.222	-0.189	-0.203	-0.185
	(-5.50)	(-5.37)	(-5.04)	(-4.31)
$\text{PCSlope}_{i,t} \times (u_t - u^*_{i,t})$	-0.050	-0.048	-0.073	-0.060
	(-2.56)	(-3.20)	(-4.50)	(-3.75)
$\text{Supply}_{i,t}$		0.075		0.057
		(5.91)		(4.93)
Avg. cross-sectional forecast σ	0.250	0.250	0.250	0.250
Interaction effect / forecast σ	-20.1%	-19.3%	-29.3%	-24.0%
Observations	1,197	1,197	1,197	1,197
R^2	0.134	0.209	0.164	0.207
Meeting fixed effects	Yes	Yes	Yes	Yes

PANEL B: INTEREST RATE SENSITIVITY AND $g - g^*$				
	Main specification		No-selection specification	
	(1)	(2)	(3)	(4)
$\text{IS sensitivity}_{i,t}$	0.048	0.051	0.039	0.037
	(1.51)	(1.60)	(1.54)	(1.46)
$(r_{i,t} - r^*_{i,t})$	-0.339	-0.330	-0.332	-0.310
	(-1.93)	(-1.91)	(-1.99)	(-1.85)
$\text{IS sensitivity}_{i,t} \times (r_{i,t} - r^*_{i,t})$	-0.062	-0.059	-0.063	-0.064
	(-2.58)	(-2.54)	(-2.26)	(-2.30)
$\text{Demand}_{i,t}$		0.026		0.030
		(2.01)		(2.49)
Avg. cross-sectional forecast σ	0.361	0.361	0.361	0.361
Interaction effect / forecast σ	-17.0%	-16.4%	-17.5%	-17.7%
Observations	1,197	1,197	1,197	1,197
R^2	0.068	0.072	0.063	0.068
Meeting fixed effects	Yes	Yes	Yes	Yes

Note: The table repeats Table IV, including the standalone parameter level as an independent variable, in addition to the parameter-by-condition interaction.

TABLE IA.6
OBSERVABLE DETERMINANTS OF PARAMETER BELIEFS

PANEL A: INSTITUTION ENVIRONMENT

Variable	Original source	Construction
Institution fixed effect excluding member		For each displayed outcome, standardize the outcome in the relevant sample and remove meeting fixed effects. Average the member's meeting-residualized value within institution, and assign the leave-one-out institution mean that excludes the member.

PANEL B: CAREER BACKGROUND

Variable	Original source	Construction
Research career		Equals one for an academic research background or Federal Reserve research/economist background; equals zero otherwise.
Fed system staff career		Equals one for prior Federal Reserve staff-side experience in research, supervision, markets, or operations; equals zero otherwise.
Private financial-sector career		Equals one for prior private finance, financial markets, or commercial banking experience; equals zero otherwise.
Financial regulation career		Equals one for prior supervision, markets, operations, legal, or regulatory experience; equals zero otherwise.
Business management career		Equals one for prior consulting, corporate executive, CFO, chief risk officer, management, or business leadership experience; equals zero otherwise.
Public policy career		Equals one for prior domestic government-policy or international-policy experience; equals zero otherwise.

PANEL C: MACRO EXPERIENCE

Variable	Original source	Construction
Bordo-Istrefi macro experience		
Max inflation before FOMC	Bordo-Istrefi	Maximum quarterly annualized CPI inflation from the member's birth through the quarter before the member first joins the FOMC.
Max unemployment before FOMC	Bordo-Istrefi	Maximum unemployment rate from the member's birth through the quarter before the member first joins the FOMC.
Rolling max inflation since FOMC entry (as of meeting)	Bordo-Istrefi	At the member's first FOMC meeting, the entry baseline is the latest complete quarter whose observation end and availability date are no later than that meeting. At each meeting, the measure is the maximum from that entry-baseline quarter through the complete quarters available by the meeting. Quarterly inflation is 400 times the quarter-over-quarter log change in quarter-average monthly CPI; a quarter becomes available on the 15th of the month after quarter end.
Rolling max unemployment since FOMC entry (as of meeting)	Bordo-Istrefi	At the member's first FOMC meeting, the entry baseline is the latest complete quarter available by that meeting. At each meeting, the measure uses the local ALFRED monthly vintage valid on that date, selecting the latest real-time start when validity windows overlap; it averages only quarters with all three monthly observations and takes the maximum from the entry-baseline quarter through complete quarters available by the meeting. Historical revisions are allowed, but future vintages and future quarters are excluded.
Malmendier-Nagel macro experience		
M/N inflation experience	Malmendier-Nagel	Lifetime weighted average of quarterly annualized CPI inflation from birth through the meeting quarter, with weights increasing with age using $\theta = 3.044$ and exponent $\lambda = \theta - 1$.
M/N unemployment experience	Malmendier-Nagel	Lifetime weighted average of quarterly unemployment from birth through the meeting quarter, using the same Malmendier-Nagel weights.
M/N real-rate experience	Malmendier-Nagel	Lifetime weighted average of the real short rate from birth through the meeting quarter, using the same Malmendier-Nagel weights.

PANEL D: DEMOGRAPHICS AND EDUCATION

Variable	Original source	Construction
Reserve Bank president	Bordo-Istrefi	Equals one if the member's title at the meeting is Federal Reserve Bank president; equals zero otherwise.
Saltwater PhD	Bordo-Istrefi	Equals one if the member has a PhD from a saltwater institution using the Bordo-Istrefi school classification; equals zero otherwise.
Freshwater PhD	Bordo-Istrefi	Equals one if the member has a PhD from a freshwater institution using the Bordo-Istrefi school classification; equals zero otherwise.
Other PhD	Bordo-Istrefi	Equals one if the member has a PhD in the Bordo-Istrefi omitted PhD-school category or a PhD school in the extended sample that does not appear in the Bordo-Istrefi appendix list; equals zero otherwise.
Appointed by Democrat	Bordo-Istrefi	Equals one if the appointing president is a Democrat; equals zero otherwise.
Appointed by supply-side Republican	Bordo-Istrefi	Equals one if the appointing president is classified as a supply-side Republican, corresponding to Reagan in the Bordo-Istrefi classification; equals zero otherwise.

Note: The table describes the independent variables used in the regressions of parameter beliefs on their determinants, reported in Table III. Continuous regressors are standardized in the regressions; indicator variables remain in levels.

TABLE IA.7
MULTIVARIATE BLOCK COEFFICIENTS: FIRST AND SECOND PRINCIPAL COMPONENTS

	Inst. FE				Career				B-I macro				M-N macro				Dem./educ.			
	PC1		PC2		PC1		PC2		PC1		PC2		PC1		PC2		PC1		PC2	
	Pre-SEP (1)	SEP (2)	Pre-SEP (3)	SEP (4)	Pre-SEP (5)	SEP (6)	Pre-SEP (7)	SEP (8)	Pre-SEP (9)	SEP (10)	Pre-SEP (11)	SEP (12)	Pre-SEP (13)	SEP (14)	Pre-SEP (15)	SEP (16)	Pre-SEP (17)	SEP (18)	Pre-SEP (19)	SEP (20)
Institution fixed effect excluding member																				
Institution fixed effect excluding member	0.304	0.432	-0.013	0.164																
	(3.54)	(3.63)	(-0.10)	(1.37)																
Career background																				
Research career					0.364	0.410	-0.033	-0.142												
					(1.69)	(1.30)	(-0.13)	(-0.62)												
Fed system staff career					-0.192	0.569	0.033	0.419												
					(-0.69)	(1.74)	(0.13)	(1.73)												
Private financial-sector career					0.009	0.325	-0.005	-0.323												
					(0.04)	(1.48)	(-0.02)	(-1.31)												
Financial regulation career					-0.050	0.581	-0.216	-0.159												
					(-0.19)	(2.06)	(-0.84)	(-0.72)												
Business management career					-0.432	-0.262	-0.095	-0.564												
					(-1.39)	(-0.89)	(-0.48)	(-1.90)												
Public policy career					-0.461	-0.092	-0.093	-0.004												
					(-2.13)	(-0.27)	(-0.47)	(-0.02)												
Bordo-Istrefi macro experience																				
Max inflation before FOMC									0.086	0.247	-0.052	-0.094								
									(0.81)	(1.97)	(-0.47)	(-0.99)								
Max unemployment before FOMC									-0.151	-0.028	-0.108	0.068								
									(-1.24)	(-0.70)	(-1.57)	(2.22)								
Rolling max inflation since FOMC entry									-0.024	-0.006	0.044	0.389								
									(-0.18)	(-0.05)	(0.38)	(2.74)								
Rolling max unemployment since FOMC entry									0.145	0.167	0.097	0.090								
									(0.76)	(1.29)	(0.55)	(0.67)								
Malmendier-Nagel macro experience																				
M/N inflation experience													0.246	1.838	-0.132	-1.170				
													(0.49)	(1.86)	(-0.30)	(-1.11)				
M/N unemployment experience													0.813	0.615	1.234	0.879				
													(1.18)	(0.65)	(1.27)	(0.94)				
M/N real-rate experience													-0.545	-3.278	-1.015	1.530				
													(-0.99)	(-2.48)	(-1.17)	(0.97)				
Demographics and education																				
Reserve Bank president																	-0.147	-0.174	0.221	0.478
																	(-0.69)	(-0.66)	(1.05)	(1.81)
Saltwater PhD																	0.817	0.560	0.264	0.126
																	(3.14)	(1.68)	(1.03)	(0.50)
Freshwater PhD																	-0.089	0.447	-0.364	-0.348
																	(-0.50)	(0.85)	(-1.77)	(-1.39)
Other PhD																	0.242	0.442	-0.197	0.442
																	(1.01)	(0.59)	(-0.79)	(1.39)
Appointed by Democrat																	-0.231	0.068	-0.176	-0.232
																	(-1.02)	(0.24)	(-0.72)	(-0.96)
Appointed by supply-side Republican																	-0.436	0.140	-0.255	0.948
																	(-1.99)	(0.29)	(-1.18)	(4.63)
Block R^2	0.092	0.186	0.000	0.026	0.094	0.208	0.009	0.153	0.043	0.075	0.025	0.149	0.014	0.065	0.008	0.036	0.219	0.060	0.060	0.122
Observations	2,517	620	2,517	620	2,653	784	2,653	784	2,653	784	2,653	784	2,653	784	2,653	784	2,653	784	2,653	784
Members	51	31	51	31	52	35	52	35	52	35	52	35	52	35	52	35	52	35	52	35
Meetings	136	48	136	48	136	48	136	48	136	48	136	48	136	48	136	48	136	48	136	48
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The columns of the table report the pre-SEP and SEP regressions for principal component 1 followed by the corresponding regressions for principal component 2, corresponding to the results displayed in Table III. Continuous regressors are standardized and indicator variables remain in levels. Variable definitions are provided in Appendix Table IA.6. Parentheses report *t*-statistics computed from standard errors two-way clustered by member and meeting.

TABLE IA.8
PARAMETER-BELIEF BUNDLES IN THE r^* SAMPLE

Member-specific r^* sample (2012–2020)					
Explanatory power			Principal component loadings		
Determinant or benchmark	PC1	PC2	Component	PC1	PC2
Member fixed-effect R^2	0.787	0.669	PC slope	0.429	0.203
Institution fixed effect excluding member	0.333	0.049	IS sensitivity	0.481	-0.094
Career background	0.141	0.218	u^*	-0.422	0.540
Bordo-Istrefi macro experience	0.049	0.159	r^*	-0.592	0.010
Malmendier-Nagel macro experience	0.053	0.068	g^*	-0.237	-0.811
Demographics and education	0.058	0.151	Variance explained	0.382	0.207
Observations			583		
Members			31		
Meetings			36		
Meeting fixed effects			Yes		

Note: The table repeats the exercise in Table III, for the sample in which members' r^* estimates are available. PC1 and PC2 are the first and second principal components of PC slope, IS sensitivity, u^*, r^*, and g^*. The left three columns report the explanatory power of member fixed effects and determinant-block R^2 after absorbing meeting fixed effects. The right three columns report the principal-component loadings. The institution fixed-effect variable is constructed from the average meeting-level outcome within institution, excluding the member. Continuous regressors are standardized and indicator variables remain in levels. Coefficient-table t-statistics are computed from standard errors two-way clustered by member and meeting.

TABLE IA.9
MULTIVARIATE BLOCK COEFFICIENTS IN THE r^* SAMPLE

	Inst. FE		Career		B-I macro		M-N macro		Dem./educ.	
	PC1 (1)	PC2 (2)	PC1 (3)	PC2 (4)	PC1 (5)	PC2 (6)	PC1 (7)	PC2 (8)	PC1 (9)	PC2 (10)
Institution fixed effect excluding member										
Institution fixed effect excluding member	0.577	0.224								
	(6.13)	(1.56)								
Career background										
Research career			0.509	-0.311						
			(1.12)	(-1.38)						
Fed system staff career			0.595	0.399						
			(1.79)	(1.30)						
Private financial-sector career			0.624	-0.571						
			(2.16)	(-2.22)						
Financial regulation career			0.367	-0.135						
			(1.09)	(-0.56)						
Business management career			-0.134	-0.492						
			(-0.35)	(-1.55)						
Public policy career			0.082	-0.194						
			(0.23)	(-0.62)						
Bordo-Istrefi macro experience										
Max inflation before FOMC					0.199	-0.146				
					(1.20)	(-1.29)				
Rolling max inflation since FOMC entry					0.011	0.324				
					(0.06)	(2.03)				
Rolling max unemployment since FOMC entry					0.146	0.149				
					(0.68)	(0.98)				
Malmendier-Nagel macro experience										
M/N inflation experience							2.575	0.320		
							(1.10)	(0.12)		
M/N unemployment experience							0.101	1.805		
							(0.09)	(1.66)		
M/N real-rate experience							-3.438	-0.399		
							(-1.35)	(-0.13)		
Demographics and education										
Reserve Bank president									-0.298	0.604
									(-0.95)	(2.12)
Saltwater PhD									0.372	0.159
									(1.06)	(0.54)
Freshwater PhD									0.595	-0.153
									(1.10)	(-0.39)
Other PhD									0.599	0.732
									(0.88)	(1.82)
Appointed by Democrat									0.261	0.067
									(0.79)	(0.23)
Block R^2	0.333	0.049	0.141	0.218	0.049	0.159	0.053	0.068	0.058	0.151
Observations	430	430	583	583	583	583	583	583	583	583
Members	26	26	31	31	31	31	31	31	31	31
Meetings	36	36	36	36	36	36	36	36	36	36
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table repeats the regressions in Appendix Table IA.7, where the principal components are constructed including members' r^* estimates. Variable definitions are provided in Appendix Table IA.6.

TABLE IA.10
FOMC DECISION TYPES: EXAMPLES

Decision type	Date	Quotation
Rate decision	2001-08-21	"decided to lower its target for the federal funds rate by 25 basis points to 3-1/2 percent"
	2008-12-16	"The Federal Open Market Committee decided today to establish a target range for the federal funds rate of 0 to 1/4 percent."
	2015-12-16	"the Committee decided to raise the target range for the federal funds rate to 1/4 to 1/2 percent"
Communication (Economic assessment)	2001-08-21	"Household demand has been sustained, but business profits and capital spending continue to weaken and growth abroad is slowing, weighing on the U.S. economy."
Communication (Forward guidance)	2003-08-12	"In these circumstances, the Committee believes that policy accommodation can be maintained for a considerable period."
Communication (Reaction function)	2013-09-18	"the Committee decided to await more evidence that progress will be sustained before adjusting the pace of its purchases"
Other	1992-02-05	"the Committee at this meeting established ranges for growth of M2 and M3 of 2-1/2 to 6-1/2 percent and 1 to 5 percent, respectively"
	2009-03-18	"the Committee decided today to increase the size of the Federal Reserve's balance sheet further by purchasing up to an additional $750 billion of agency mortgage-backed securities"
	2011-09-21	"The Committee intends to purchase, by the end of June 2012, $400 billion of Treasury securities with remaining maturities of 6 years to 30 years and to sell an equal amount of Treasury securities with remaining maturities of 3 years or less."

Note: The table presents examples of the three decision types in the production FOMC decision dataset. Communication decisions are additionally labeled by subtype.

TABLE IA.11
FOMC DECISION SUMMARY STATISTICS

PANEL A: DECISIONS PER MEETING

Sample	Period	Meetings	Decisions	Mean	SD	Min.	p25	Median	p75	Max.
All meetings	1992–2020	231	1,160	5.02	1.95	1	3.0	5.0	6.0	10
Greenspan	1992–2006	113	476	4.21	1.85	2	3.0	4.0	4.0	10
Bernanke	2006–2014	64	364	5.69	2.03	1	3.8	6.0	7.0	10
Yellen	2014–2018	32	186	5.81	1.06	5	5.0	5.0	7.0	8
Powell	2018–2020	22	134	6.09	1.44	4	5.0	5.5	7.0	9

PANEL B: DISTRIBUTION ACROSS DECISION TYPES

		Rate decision		Communication		Other	
Sample	Period	*N*	%	*N*	%	*N*	%
All meetings	1992–2020	260	22.4	499	43.0	401	34.6
Greenspan	1992–2006	132	27.7	135	28.4	209	43.9
Bernanke	2006–2014	64	17.6	193	53.0	107	29.4
Yellen	2014–2018	37	19.9	104	55.9	45	24.2
Powell	2018–2020	27	20.1	67	50.0	40	29.9

PANEL C: MEETING-LEVEL COMPOSITION

Sample	Period	Multiple decisions (%)	Mean distinct types	Rate (%)	Communication (%)	Other (%)
All meetings	1992–2020	99.6	2.52	100.0	80.1	71.9
Greenspan	1992–2006	100.0	2.19	100.0	60.2	58.4
Bernanke	2006–2014	98.4	2.70	100.0	98.4	71.9
Yellen	2014–2018	100.0	3.00	100.0	100.0	100.0
Powell	2018–2020	100.0	3.00	100.0	100.0	100.0

Note: Panel A summarizes the meeting-level decision count. Panel B reports shares among all decisions. Panel C reports the fraction of meetings containing each decision type. The sample begins with the first 1992 meeting and ends with the last meeting available in those data (December 16, 2020).

TABLE IA.12
FORECAST FIT, PARAMETER FIT, AND COMMITTEE DECISIONS: PARAMETER LEVEL INCLUDED

PANEL A: REALIZED FORECAST FIT

	Full sample					SEP sample				
	Voters			All participants		Voters			All participants	
	Align. (1)	Adopt. (2)	Dissent (3)	Align. (4)	Adopt. (5)	Align. (6)	Adopt. (7)	Dissent (8)	Align. (9)	Adopt. (10)
Forecast fit	0.117	0.058	-0.033	0.117	0.059	0.173	0.098	-0.041	0.163	0.092
	(3.33)	(1.88)	(-3.15)	(3.11)	(2.55)	(2.33)	(1.98)	(-2.01)	(2.44)	(2.79)
Observations	2,122	2,122	2,131	3,583	3,583	1,017	1,017	1,017	1,702	1,702
Meetings	227	227	227	227	227	105	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

PANEL B: PARAMETER-IMPLIED FORECAST FIT

	Full sample					SEP sample				
	Voters			All participants		Voters			All participants	
	Align. (1)	Adopt. (2)	Dissent (3)	Align. (4)	Adopt. (5)	Align. (6)	Adopt. (7)	Dissent (8)	Align. (9)	Adopt. (10)
Parameter-implied fit	0.213	0.158	-0.051	0.180	0.139	0.358	0.212	-0.083	0.295	0.206
	(2.98)	(3.76)	(-2.40)	(2.60)	(3.82)	(3.06)	(3.26)	(-2.30)	(2.97)	(3.71)
Observations	2,039	2,039	2,048	3,449	3,449	1,017	1,017	1,017	1,702	1,702
Meetings	219	219	219	219	219	105	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parameter levels included in projected fit	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table repeats the analysis in Table VIII, including the standalone Phillips curve slope and interest rate sensitivity parameter levels in the first-stage projection of fit on members' model parameters.

TABLE IA.13
FORECAST FIT, PARAMETER FIT, AND NON-RATE COMMITTEE DECISIONS

PANEL A: REALIZED FORECAST FIT								
	Full sample				SEP sample			
	Voters		All participants		Voters		All participants	
	Align. (1)	Adopt. (2)	Align. (3)	Adopt. (4)	Align. (5)	Adopt. (6)	Align. (7)	Adopt. (8)
Forecast fit	0.097	0.041	0.111	0.054	0.160	0.086	0.162	0.096
	(3.00)	(1.36)	(3.03)	(2.44)	(2.33)	(1.76)	(2.55)	(3.13)
Observations	2,113	2,113	3,568	3,568	1,017	1,017	1,702	1,702
Meetings	226	226	226	226	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
PANEL B: PARAMETER-IMPLIED FORECAST FIT								
	Full sample				SEP sample			
	Voters		All participants		Voters		All participants	
	Align. (1)	Adopt. (2)	Align. (3)	Adopt. (4)	Align. (5)	Adopt. (6)	Align. (7)	Adopt. (8)
Parameter-implied fit	0.188	0.132	0.217	0.163	0.350	0.217	0.364	0.258
	(2.63)	(2.57)	(3.17)	(3.69)	(3.04)	(2.78)	(3.88)	(4.14)
Observations	2,030	2,030	3,434	3,434	1,017	1,017	1,702	1,702
Meetings	218	218	218	218	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table reproduces Table VIII for decisions coded as communication or other. Panel A reports results using realized forecast fit, and Panel B reports results using parameter-implied forecast fit.

TABLE IA.14
FORECAST FIT, PARAMETER FIT, AND RATE DECISIONS

PANEL A: REALIZED FORECAST FIT								
	Full sample				SEP sample			
	Voters		All participants		Voters		All participants	
	Align. (1)	Adopt. (2)	Align. (3)	Adopt. (4)	Align. (5)	Adopt. (6)	Align. (7)	Adopt. (8)
Forecast fit	0.118	0.069	0.093	0.041	0.169	0.072	0.121	0.024
	(3.39)	(3.06)	(3.13)	(2.30)	(2.33)	(1.95)	(2.11)	(0.92)
Observations	2,122	2,122	3,583	3,583	1,017	1,017	1,702	1,702
Meetings	227	227	227	227	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
PANEL B: PARAMETER-IMPLIED FORECAST FIT								
	Full sample				SEP sample			
	Voters		All participants		Voters		All participants	
	Align. (1)	Adopt. (2)	Align. (3)	Adopt. (4)	Align. (5)	Adopt. (6)	Align. (7)	Adopt. (8)
Parameter-implied fit	0.156	0.071	0.146	0.067	0.255	0.092	0.237	0.088
	(2.15)	(2.22)	(2.38)	(2.24)	(2.02)	(1.96)	(2.56)	(2.03)
Observations	2,039	2,039	3,449	3,449	1,017	1,017	1,702	1,702
Meetings	219	219	219	219	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table reproduces Table VIII for decisions coded as rate decisions. Panel A reports results using realized forecast fit, and Panel B reports results using parameter-implied forecast fit.

TABLE IA.15
REALIZED FORECAST FIT AND COMMITTEE DECISIONS: INFLATION AND GROWTH COMPONENTS

	Full sample					SEP sample				
	Voters			All participants		Voters			All participants	
	Align.	Adopt.	Dissent	Align.	Adopt.	Align.	Adopt.	Dissent	Align.	Adopt.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PANEL A: INFLATION										
Inflation forecast fit	0.086	0.025	-0.027	0.062	0.019	0.160	0.051	-0.037	0.111	0.040
	(2.29)	(0.89)	(-2.83)	(1.65)	(0.82)	(2.42)	(1.12)	(-2.43)	(1.72)	(1.06)
Observations	2,120	2,120	2,129	3,581	3,581	1,015	1,015	1,015	1,700	1,700
Meetings	227	227	227	227	227	105	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
PANEL B: GDP GROWTH										
	Full sample					SEP sample				
	Voters			All participants		Voters			All participants	
	Align.	Adopt.	Dissent	Align.	Adopt.	Align.	Adopt.	Dissent	Align.	Adopt.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Growth forecast fit	0.080	0.057	-0.021	0.107	0.066	0.084	0.088	-0.021	0.138	0.102
	(2.53)	(1.74)	(-2.25)	(3.62)	(2.65)	(1.49)	(1.93)	(-1.16)	(2.53)	(3.10)
Observations	2,120	2,120	2,129	3,581	3,581	1,015	1,015	1,015	1,700	1,700
Meetings	227	227	227	227	227	105	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table repeats the analysis in Table VIII separately for inflation and GDP forecast fit. Both panels use observations with nonmissing inflation and GDP component fit scores. Fit, alignment, and adoption contribution are standardized within their estimation samples. Dissent is a 0/1 outcome. Standard errors are two-way clustered by member and meeting. t-statistics are in parentheses.

TABLE IA.16

PARAMETER-IMPLIED FORECAST FIT AND COMMITTEE DECISIONS: INFLATION AND GROWTH COMPONENTS

	Full sample					SEP sample				
	Voters			All participants		Voters			All participants	
	Align. (1)	Adopt. (2)	Dissent (3)	Align. (4)	Adopt. (5)	Align. (6)	Adopt. (7)	Dissent (8)	Align. (9)	Adopt. (10)
Panel A: Inflation										
Inflation parameter-implied fit	0.217	0.162	-0.051	0.233	0.172	0.363	0.242	-0.079	0.378	0.264
	(2.63)	(2.92)	(-2.44)	(3.10)	(3.56)	(2.74)	(3.18)	(-2.29)	(3.49)	(3.98)
Observations	2,037	2,037	2,046	3,447	3,447	1,015	1,015	1,015	1,700	1,700
Meetings	219	219	219	219	219	105	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: GDP Growth										
	Full sample					SEP sample				
	Voters			All participants		Voters			All participants	
	Align. (1)	Adopt. (2)	Dissent (3)	Align. (4)	Adopt. (5)	Align. (6)	Adopt. (7)	Dissent (8)	Align. (9)	Adopt. (10)
Growth parameter-implied fit	0.091	0.050	-0.021	0.101	0.078	0.170	0.091	-0.037	0.164	0.127
	(1.50)	(0.98)	(-1.19)	(1.60)	(1.84)	(1.70)	(1.10)	(-1.30)	(1.71)	(1.85)
Observations	2,037	2,037	2,046	3,447	3,447	1,015	1,015	1,015	1,700	1,700
Meetings	219	219	219	219	219	105	105	105	105	105
Meeting fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table repeats the analysis in Table VIII separately for inflation and GDP forecast fit. Both panels use observations with nonmissing inflation and GDP component fit scores. Fit, alignment, and adoption contribution are standardized within their estimation samples. Dissent is a 0/1 outcome. Standard errors are two-way clustered by member and meeting and include a stacked sandwich correction for first-stage estimation. t-statistics are in parentheses.

TABLE IA.17
INFORMATION SOURCES IN PARAMETER-VALUE ARGUMENTS

Information source	Source entries	Share of identified entries	Common information	Private/ specialized information	Ambiguous
Official statistics	1,660	24.2%	98.3%	1.3%	0.4%
Board staff forecast and analysis	1,469	21.4%	99.8%	0.1%	0.1%
Financial market indicators	1,352	19.7%	99.9%	0.1%	0.0%
Business and regional contacts	531	7.8%	2.4%	97.0%	0.6%
Regional Fed forecast and analysis	423	6.2%	2.8%	93.1%	4.0%
Fiscal policy information	351	5.1%	94.0%	0.3%	5.7%
International economic conditions	314	4.6%	95.5%	0.6%	3.8%
Surveys and sentiment measures	234	3.4%	83.3%	4.7%	12.0%
Banking and credit conditions	200	2.9%	58.5%	6.0%	35.5%
External forecasts and research	182	2.7%	44.5%	0.0%	55.5%
Other	135	2.0%	90.4%	2.2%	7.4%
Total	**6,851**	**100.0%**	**82.0%**	**14.0%**	**3.9%**

Note: The table summarizes identified information sources in arguments used to construct the Phillips curve slope and interest-rate sensitivity estimates. Common information denotes sources classified as common to the Committee; private or specialized information denotes sources classified as not common to the Committee; ambiguous denotes sources whose common observability is unclear. The total row reports the composition across identified source entries. The sample contains 6,851 identified source entries from 5,733 arguments.

TABLE IA.18
INFORMATION SOURCES IN ALL ARGUMENTS

Information source	Source entries	Share of identified entries	Common information	Private/ specialized information	Ambiguous
Official statistics	17,103	27.2%	94.1%	4.1%	1.7%
Board staff forecast and analysis	10,802	17.2%	99.8%	0.1%	0.1%
Business and regional contacts	9,042	14.4%	1.8%	97.8%	0.4%
Financial market indicators	8,294	13.2%	99.7%	0.1%	0.2%
Regional Fed forecast and analysis	4,030	6.4%	5.1%	91.7%	3.1%
International economic conditions	3,337	5.3%	91.7%	0.7%	7.6%
Surveys and sentiment measures	3,202	5.1%	69.0%	21.3%	9.7%
Fiscal policy information	2,793	4.4%	93.5%	0.4%	6.1%
Other	1,525	2.4%	84.1%	6.6%	9.4%
External forecasts and research	1,474	2.3%	47.5%	2.0%	50.5%
Banking and credit conditions	1,315	2.1%	46.5%	16.0%	37.5%
Total	**62,917**	**100.0%**	**73.1%**	**22.8%**	**4.1%**

Note: The table repeats Table IA.17, using all extracted arguments. Source categories and classifications match Table IA.17. The sample contains 62,917 identified source entries from 52,224 arguments.

TABLE IA.19
INFORMATION SOURCES AND PARAMETER ARGUMENT SCORES

	Phillips curve slope	IS sensitivity
Scored arguments	2,603	3,292
Scored arguments excluding private/specialized	2,135	2,847
Member-meetings	1,727	2,173
Member-meetings excluding private/specialized	1,443	1,901
Overlap member-meetings	1,443	1,901
Average within-meeting correlation	0.983	0.986
Pooled member-meeting correlation	0.982	0.988
Meetings used for average correlation	171	245

Note: The table compares member-meeting average parameter-argument scores using all scored arguments with averages that exclude arguments citing private or specialized information. The within-meeting correlation is the average cross-sectional correlation across meetings with at least two overlapping member observations. The pooled correlation uses all overlapping member-meeting observations.

TABLE IA.20
EXAMPLES OF MODEL DISAGREEMENT AND ACQUIESCENCE IN FOMC DECISIONS

Adopted and preferred policy	Measured model and recent fit	Member's stated outlook	Acceptance of the adopted decision
DECEMBER 1995: HOENIG *Adopted:* A 25 basis point funds-rate cut without a discount-rate cut. *Preferred:* No additional easing.	*Model (SD):* PC +0.93; IS -1.04. *Model favors:* A steeper PC and weaker IS favored leaving rates unchanged. *Recent data:* Falling inflation and softer activity favored modest easing. *Recent fit:* Overall, 32nd percentile.	"Growth for the next several quarters will be about 2 to 2-1/2 percent and inflation will be 3 percent. ... Lower yields in long-term debt markets have been stimulative. That suggests to me that ... our policy has been neutral and not slightly tight."	"I would prefer that we not change the funds rate right now ... However, given that we are talking about a small adjustment and ... no discount rate change, I can accept your proposal at this time."
JANUARY 2004: BIES *Adopted:* Replace "considerable period" with "can be patient." *Preferred:* Less commitment to keeping rates low.	*Model (SD):* IS -1.15. *Model favors:* A weaker IS favored removing forward guidance. *Recent data:* Low inflation and remaining slack favored patience. *Recent fit:* Growth, 5th percentile.	"Having more-balanced growth going forward gives me much more confidence that this recovery has legs"	"I can support the recommended wording change to 'patience,' but ... I would prefer not to use the 'patience' language."
NOVEMBER 2010: WARSH *Adopted:* $600 billion of Treasury purchases under QE2. *Preferred:* No new asset purchases at that meeting.	*Model (SD):* PC -0.79; IS -0.95; u^* +2.03; g^* -0.67. *Model favors:* A flatter PC, weaker IS, higher u^*, and lower g^* favored waiting rather than starting QE2. *Recent data:* High unemployment and falling inflation favored QE2. *Recent fit:* Overall, 6th percentile.	"Harm is being done to the supply side of the economy, and I share the staff's view that the NAIRU has moved up. ... The Fed is capable of a lot of things, but it's not capable, in my view, of moving the dial tremendously on economic growth from here."	"The version of alternative B that I can, with the greatest reluctance, not dissent from ... [is] subject to serious regular review."
SEPTEMBER 2012: LOCKHART *Adopted:* $40 billion in monthly open-ended MBS purchases. *Preferred:* A smaller monthly purchase pace.	*Model (SD):* PC -1.29; IS -1.03; u^* -1.06. *Model favors:* A flatter PC and weaker IS favored a $30 billion pace; lower u^* supported some accommodation. *Recent data:* Weak hiring favored a larger monthly purchase pace. *Recent fit:* Growth, 11th percentile.	"This time around, it seems to me that we face a more conventional problem of inadequate demand, and I am not convinced that lowering general market rates will stimulate much credit expansion and spending."	Voted for the adopted $40 billion pace after preferring $30 billion: "I will support alternative B's modest pace of $30 billion per month"; roll call: "Yes."
DECEMBER 2013: LACKER *Adopted:* A $10 billion taper with stronger forward guidance. *Preferred:* A larger taper with less forward guidance.	*Model (SD):* IS -1.25; u^* +1.22; g^* -2.97. *Model favors:* A weaker IS, higher u^*, and low g^* implied less slack and favored a larger, clean taper. *Recent data:* Labor gains favored taper; weak growth and inflation favored a smaller step. *Recent fit:* Overall, 12th percentile.	"Overall, the U.S. economy seems to be showing modest fluctuations around a modest average growth rate. ... [My forecast] has GDP growing just a little above 2 percent for next year and beyond"	"If left to my own devices ... I would prefer a clean taper. ... If we are going to push back ... alternative B is a statement I can support."
DECEMBER 2015: TARULLO *Adopted:* A 25 basis point increase in the funds-rate target range. *Preferred:* No rate increase.	*Model (SD):* r^* -0.75. *Model favors:* A lower r^* made near-zero policy less accommodative and favored delay. *Recent data:* Strong employment gains supported liftoff, while inflation remained below target. *Recent fit:* Growth, 44th percentile.	"[T]oday seems an odd moment to declare reasonable confidence that inflation will return to the Committee's stated 2 percent target . . . [I]t [is] more prudent to act on evidence than on faith that traditional economic patterns will reassert themselves."	"Despite all of that, I will reluctantly go along ... [given] the institutional reality."

Note: The table reports six examples, accompanied by meeting quotations, in which a member's model parameters and stated economic interpretation favored a policy different from the adopted decision, while the member accepted the Committee's choice. Model (SD) reports deviations from the meeting mean in cross-sectional standard deviations; positive PC values indicate steeper Phillips curve beliefs and positive IS values indicate stronger interest-rate sensitivity. Recent fit reports the member's percentile in the meeting cross-section of parameter-implied forecast fit over the previous eight meetings; lower percentiles indicate worse fit. Quotations are from the corresponding FOMC transcript.

TABLE IA.21
MAJOR POLICY-INNOVATION DECISIONS

Date	Decision and inclusion basis	Type	Support window	Pre-adoption support timing
1994-02-04	**First public post-meeting policy statement**: first statement tied to a firming	Comm.	8 mtgs.	9 early; 4 late; 5 none
2003-08-12	**"Considerable period" forward guidance**: forward-guidance spell begins	F. guidance	5 mtgs.	1 early; 9 late; 9 none
2004-05-04	**"Measured pace" forward guidance**: exit-guidance turning point	F. guidance	5 mtgs.	13 early; 5 late; 0 none
2008-12-16	**ZLB target range 0–0.25%**: ZLB regime begins	Rate	8 mtgs.	10 early; 1 late; 7 none
2009-03-18	**QE1 MBS expansion to $1.25T**: agency/MBS scale expansion	Balance sheet	2 mtgs.	11 early; 1 late; 5 none
2009-03-18	**QE1 long-term Treasury purchases**: Treasury-purchase pivot	Balance sheet	2 mtgs.	11 early; 1 late; 5 none
2010-11-03	**QE2 ($600B Treasuries)**: Treasury-purchase spell begins	Balance sheet	7 mtgs.	1 early; 7 late; 9 none
2011-08-09	**Calendar forward guidance**: "through mid-2013" guidance begins	F. guidance	8 mtgs.	5 early; 1 late; 11 none
2011-09-21	**Operation Twist**: maturity-extension tool begins	Balance sheet	10 mtgs.	4 early; 2 late; 11 none
2012-09-13	**QE3 outcome-based guidance**: purchases tied to labor-market improvement	F. guidance	10 mtgs.	10 early; 2 late; 7 none
2012-09-13	**QE3 open-ended MBS purchases**: no fixed terminal quantity	Balance sheet	10 mtgs.	11 early; 1 late; 7 none
2012-12-12	**Evans-rule threshold guidance**: economic-contingency rate guidance begins	F. guidance	11 mtgs.	9 early; 4 late; 6 none
2012-12-12	**QE3 Treasury purchases**: $45B/month added after Operation Twist	Balance sheet	10 mtgs.	7 early; 4 late; 8 none
2013-12-18	**Taper of asset purchases**: first QE3 purchase reduction	Balance sheet	10 mtgs.	13 early; 4 late; 0 none
2015-12-16	**Liftoff**: first federal-funds-rate hike off the ZLB	Rate	12 mtgs.	11 early; 5 late; 0 none
2020-09-16	**FAIT guidance**: flexible average-inflation targeting translated into rate guidance	F. guidance	11 mtgs.	10 early; 6 late; 1 none

Note: The table lists the sixteen major policy-innovation decisions used in the timing of support analysis. Each retained decision supplies within-decision variation in support timing. The decision column identifies the policy change and its inclusion in the sample. Voiced-support window reports the number of transcript meetings from the first observed positive support through adoption. Pre-adoption support timing classifies participants by whether their first positive support appears in the first half of that window, the second half, or is not observed prior to the meeting in which the decision is adopted.

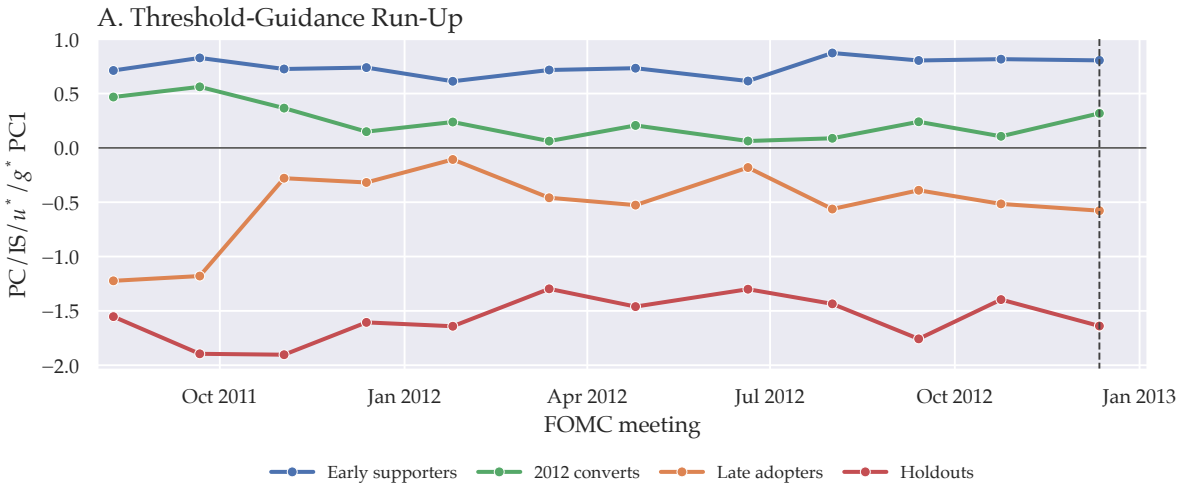


FIGURE IA.1
MODEL-PARAMETER PC1 DURING THE THRESHOLD-GUIDANCE RUN-UP

Note: The figure repeats the December 2012 threshold-guidance run-up in Figure VI, reporting the first principal component of Phillips curve slope, interest-rate sensitivity, u^*, and g^*. Members are grouped by the first transcript meeting at which they support numerical threshold guidance. Holdouts do not support the decision before adoption. The dashed vertical line marks the adoption meeting.

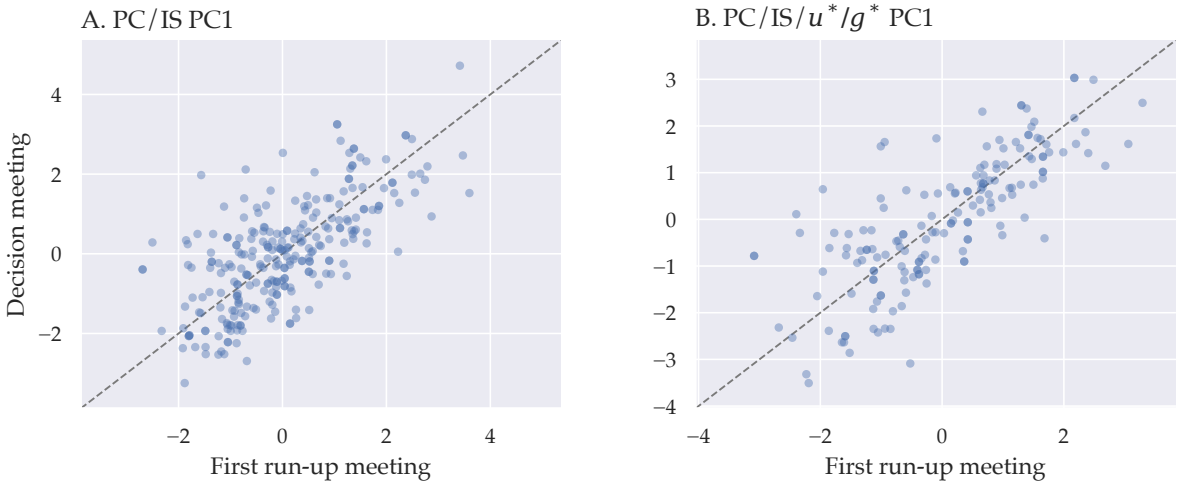


FIGURE IA.2
PERSISTENCE OF MODEL-PARAMETER PC1 DURING DYNAMIC DECISIONS

Note: The figure reports the persistence of members' first principal component during the run-ups to the 16 major policy innovations in Table IA.21. Each point is a member-decision observation. The horizontal axis reports the first meeting prior to adoption in the decision's run-up window, and the vertical axis reports the adoption meeting. The left panel uses Phillips curve slope and interest-rate sensitivity. The right panel also includes u^* and g^* when available. Values are demeaned by decision and meeting date. The dashed line is the 45-degree line.

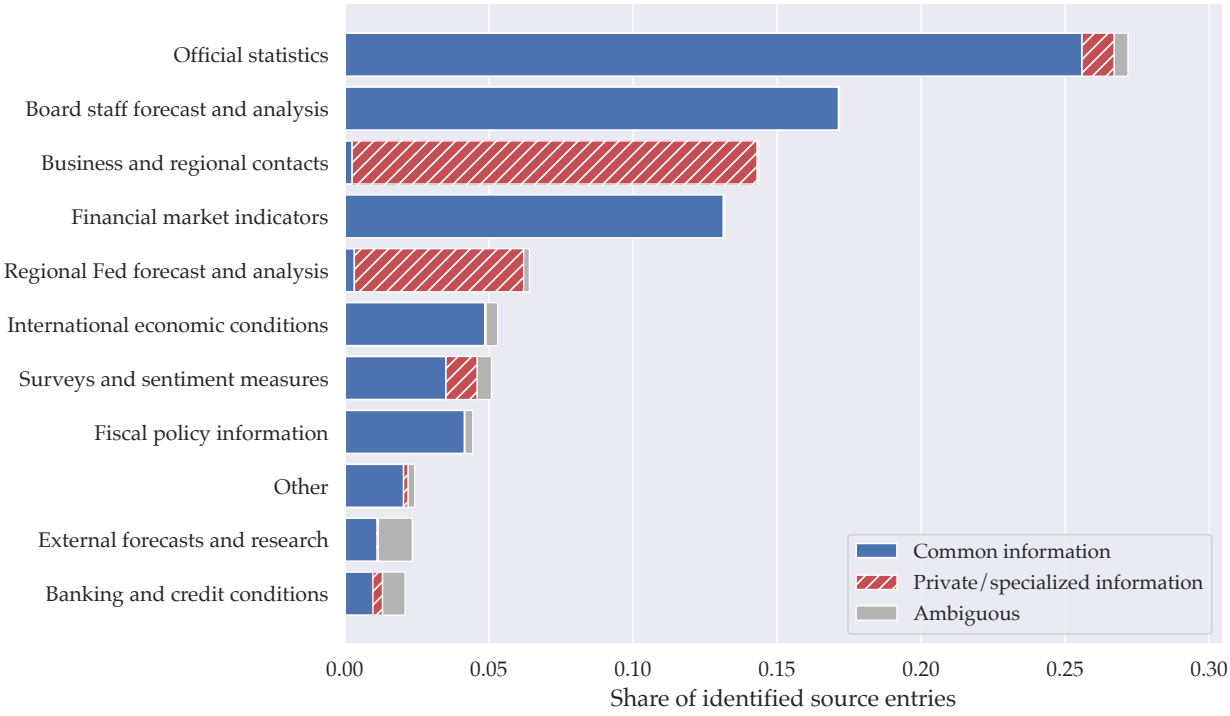


FIGURE IA.3

COMMON AND PRIVATE INFORMATION BY SOURCE: ALL ARGUMENTS

Note: The figure repeats Figure V, using all extracted arguments rather than arguments used to construct parameter estimates. Bars distinguish sources classified as common to the Committee from sources classified as private or specialized. Gray bars indicate identified sources whose common observability is ambiguous. Entries with no identifiable source are omitted.

IA.A.1 Decisions and Monetary Policy Surprises

In this section, we link the policy decisions we identify from the transcripts to high-frequency monetary policy surprises measured from asset price responses to FOMC announcements. This provides additional validation of our decisions dataset.[21]

Following prior work (Gürkaynak, Sack and Swanson, 2005; Nakamura and Steinsson, 2018), we measure monetary policy surprises using the high-frequency responses of interest rate futures in a 30-minute window following the FOMC's announcement of its policy decision. Our main measure of monetary policy surprises is mps_t, constructed as the first principal component of changes in the one- to four-quarter ahead Eurodollar futures in the 30-minute window following FOMC announcements, from Swanson and Jayawickrema (2024) and Bauer and Swanson (2023*b*). The sample runs from 1988 to 2019. We also analyze the changes in n-quarter ahead Eurodollar futures (labeled EDn), as well as the near- and 3-month maturity Fed Funds futures in the 30-minute window following the FOMC announcement (labeled FF and $FF3m$).[22] The sample of Fed Funds futures runs from 1995 to 2019.

For each meeting, we assign a score from -3 (highly accommodative) to +3 (highly restrictive) to each of the identified changes to the policy directive. We then compute two measures: $Avg(\{ExRatePolicyScores_t\})$ as the average score across all changes excluding the target rate; and $Avg(\{PolicyScores_t\})$ as the average score across all changes.[23] 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.22 report results from univariate regressions of mps_t on $Avg(\{PolicyScores_t\})$ and $Avg(\{ExRatePolicyScores_t\})$. The coefficients from these regressions are 1.32 and 1.21, respectively, indicating that a one standard deviation change in the $Avg(\{PolicyScores\})$ measures captures a 1.2 to 1.3 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 policy surprises stemming from both the target rate decision and other elements of the committee's decision, most importantly forward guidance about the future path of interest rates. In columns 3-6 of Table IA.22, we explicitly separate the surprise of the target rate decision from forward guidance, in the spirit of Gürkaynak, Sack and Swanson (2005). Columns 3 and 4 report results from regressions of the high-frequency change in four quarter-ahead Eurodollar futures on the $Avg(\{PolicyScores\})$ variables, controlling for the contemporaneous

[21] 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.

[22] 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.

[23] 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.

change in one quarter-ahead Eurodollar futures. The coefficients on $Avg(\{PolicyScores_t\})$ and $Avg(\{ExRatePolicyScores_t\})$ are 0.82 and 0.65 in these regressions, compared to the 5.7 bps standard deviation of $ED4_t$. Columns 5 and 6 report results from regressions of the high-frequency change in 3-month ahead Fed Funds futures on the $Avg(\{PolicyScores\})$ variables, controlling for the contemporaneous change in the near-maturity Fed Funds futures contract. The coefficients on $Avg(\{PolicyScores_t\})$ and $Avg(\{ExRatePolicyScores_t\})$ are 0.64 and 0.75 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.

TABLE IA.22
POLICY DECISION SCORES AND MONETARY POLICY SURPRISES

	$y = mps_t$		$y = ED4_t$		$y = FF3m_t$		$y = mps_t^{orth}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Avg(\{PolicyScores_t\})$	1.32		0.82		0.64		0.54	
	(3.36)		(2.85)		(3.11)		(1.74)	
$Avg(\{ExRatePolicyScores_t\})$		1.21		0.65		0.75		0.39
		(2.90)		(3.68)		(3.55)		(1.16)
$ED1_t$			0.91	1.01				
			(8.00)	(16.68)				
FF_t					0.69	0.71		
					(7.35)	(7.77)		
σ_y	4.95	4.95	5.70	5.70	3.74	3.74	4.63	4.63
N	262	261	262	261	190	189	262	261
R^2	0.07	0.06	0.51	0.80	0.53	0.54	0.01	0.01

Note: The table reports regressions of high-frequency asset-price measures of monetary policy surprises on the restrictive or accommodative tone of adopted FOMC decisions. The dependent variables are: mps_t, the first principal component of changes in one- to four-quarter ahead Eurodollar contracts within a 30-minute post-announcement window; $ED4_t$, the change in the four-quarter ahead Eurodollar contract; $FF3m_t$, the change in 3-month ahead federal funds futures yields; and mps_t^{orth}, residuals from regressing mps_t on controls for incoming macroeconomic and financial news. The independent variables of interest are the standardized meeting-level average scores of adopted policy decisions. Scores range from -3 for strongly accommodative decisions to +3 for strongly restrictive decisions. Controls are the high-frequency change in one-quarter ahead Eurodollar futures, $ED1_t$, and near-maturity federal funds futures, FF_t. The table reports the standard deviation of the dependent variable as σ_y; t-statistics computed using HC1 robust standard errors are reported in parentheses.

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 contribution to the adoption of policy decisions to how well their models match recent economic data. The relationship with monetary policy surprises indicates that our prior evidence is not just measuring internal deliberation but is capturing behavior reflected in financial market movements to the resulting policy.

Second, taken in conjunction with the evidence on the committee's tilt towards different members' models based on the data, the results also suggest that the committee structure may play an important role in generating monetary policy surprises. The committee's tilting behavior is consistent with it not using a fixed policy rule to make a decision, as argued by Hack, Istrefi and Meier (2025), with shifts in the policy stance arising from the shifting models 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.22 report results from regressions of mps_t^{orth}, measured as the residual of mps after controlling for macroeconomic and financial news arriving prior to the FOMC meeting, on the $Avg(\{PolicyScores\})$ variables. These results indicate a negligible relationship, suggesting that the component of the $Avg(\{PolicyScores\})$ variables that is related to monetary policy surprises stems from investor surprise at the committee's response to macroeconomic and financial news.

IA.B 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 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

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

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

Chair for Supervision (since 2017), or Governor. Regional bank presidents are identified as such. On rare occasions, a regional Fed is represented at the meeting by the First Vice President and the title is adjusted accordingly. Separately, we use the information in the header of the transcript file or in the vote roll call to identify the meeting-by-meeting list of voting members of the FOMC. We verify the total number of voters with data from Thornton, Wheelock et al. (2014), updated by the Federal Reserve Bank of St. Louis, and also merge the identities of dissenters at each meeting.

In the fourth step, we assign speaker identifiers that stay constant across meetings regardless of the role the individual may take in a given meeting. For instance, Donald Kohn is part of 194 distinct transcripts, first as economist, then as Governor, and finally as Vice Chair of the Board of Governors. In the rare cases where two participants at a given meeting share a last name, the transcript indicates one of the names with an initial. In those instances, we assign the speakers their respective full names as the stable identifier across meetings. We also confirm that the last names of speakers across meetings do not incidentally coincide. Recall that for members of the Board of Governors and for regional Fed presidents, we separately collect their names and roles, which allows us to confirm we are not identifying distinct people who happen to share a name.

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

In the sixth and final step, we use Anthropic Claude to construct a dataset of demographics for all participants who ever vote in any of the meetings in our sample. In 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.C Details on Parameter Estimation

Below we provide details on our estimation of model parameters based on members' arguments from the FOMC transcripts. Section IA.C.1 discusses the choice of priors. We select weakly informative priors, so that the posterior estimates of members' model parameters are primarily informed by data. Section IA.C.2 discusses the estimation and its stability and convergence properties. We run eight independent chains for each time-period and verify that the estimates produced are stable across different starting conditions.

IA.C.1 Choice of Priors

As noted in the main text, we normalize the argument level shocks to have variance 1:

$$\mathrm{Var}(\varepsilon_{i,j,t}) = 1, \qquad \mathrm{Var}(\nu_{i,j,t}) = 1.$$

To estimate the model, we choose prior distributions over the parameters to estimate:

$$\rho \sim N(0.5, 0.60^2) \quad \text{truncated to } [-0.995, 0.995],$$

$$\sigma_b^2 \sim IG(9.0, 0.60), \qquad \sigma_\mu^2 \sim IG(0.5, 0.0133), \qquad \sigma_a^2 \sim IG(0.5, 0.08),$$

$$\psi_t \sim N(0, 0.5),$$

where the inverse gamma distribution is parameterized by shape and scale, and 0.5 is the prior variance of ψ_t. We use a flat prior for ζ over its admissible support, with the lower bound $\zeta \geq 0.05$, and we initialize the common-component variance to $\mathrm{Var}(\mu_0) = 4.0$.

For the inverse gamma variance parameters, we choose the priors so that they contribute roughly the amount of conditional Gaussian information in one meeting at the start of the post-1992 analysis. For an inverse gamma prior with shape a_0, the conditional posterior shape is $a_0 + n/2$, where n is the number of Gaussian terms whose variance is the parameter being sampled. We therefore summarize the contribution of the prior by its share of posterior shape,

$$\frac{a_0}{a_0 + n/2}.$$

For σ_b^2, the innovation variance governing member-specific deviations, n is the number of member-level AR(1) residuals, $\sum_i (T_i - 1)$. We set $a_0 = 9$, corresponding to the number of member level innovations in one typical meeting of approximately 18. For σ_μ^2, the innovation variance governing the common component, $n = T - 1$, the number of meeting innovations. For σ_a^2, the variance of the meeting selection intercepts, $n = T$, the number of draws of a_t under $a_t \sim N(0, \sigma_a^2)$. For these latter two variances, we set $a_0 = 0.5$, corresponding to one pseudo innovation or intercept draw.

We consider the dynamic variance priors to be weakly informative for our sample, given that by the time the first forecast-sample meeting is reached, the estimator has already accumulated 144 meetings, 2,656 member-meetings, and 11,287 argument opportunities; each dynamic variance prior then accounts for less than one percent of the

conditional posterior shape.

The other important choice of prior is the prior precision for the meeting specific selection coefficient ψ_t. Because ψ_t is estimated only using meeting-level information, we calibrate it separately. Its prior variance of 0.5 contributes a non-vanishing amount of shrinkage toward zero. At the 1992 and 2020 reference cutoffs, it accounts for 11.9 and 7.0 percent of the conditional precision for the Phillips curve selection coefficient, and 7.5 and 13.9 percent for the IS curve selection coefficient. Equivalently, the selection data contribute precisions of 14.8 and 26.6 for the Phillips curve, and 24.7 and 12.4 for the IS curve, compared with prior precision 2.

These quantities should be interpreted relative to the amount of meeting level selection data. A typical meeting contains roughly 179 arguments and about ten parameter mentions. Both mentions and nonmentions enter the selection equation, but each argument's information about ψ_t is weighted by the squared member deviation, $\bar{b}^2_{i(j),t}$. The relevant conditional information per meeting is therefore $\sum_{j \in t} \bar{b}^2_{i(j),t}$, measured in unit-deviation argument equivalents. The prior precision of 2 is equivalent to two arguments from members whose deviations equal one latent score standard deviation; it provides modest stabilization when the observed arguments contain limited dispersion in member beliefs.

IA.C.2 Estimation and Multi-Chain Stability

For each post-1992 meeting, we estimate the model eight times using different random seeds, producing eight independently simulated chains. Each chain runs for 100,000 Gibbs iterations. At the first cutoff in a segment, each chain begins from a rough, data-based starting guess based on the observed argument scores and mention rate. We discard the first 10,000 iterations of these initial runs. At later cutoffs, each chain begins from its own final state at the preceding period's cutoff and discards its first 5,000 iterations. These warm starts reduce burn-in time but do not add posterior draws from the earlier cutoff. We retain every fifth remaining draw, calculate a posterior mean for $b_{i,t}$ within each chain, and report the average of these eight chain-specific posterior means as the member's parameter deviation in our main analysis.

To assess whether the Gibbs sampler produces stable posterior estimates at each cutoff, we compare the posterior summaries from the eight chains. Denoting the retained posterior mean and variance for quantity x in chain m by $\bar{x}_m$ and s^2_m, respectively, the number of retained draws per chain by n, the sample variance of the chain means by B, and the average within-chain variance by W, we calculate

$$\widehat{R}_x = \max\left\{1, \sqrt{1 - \frac{1}{n} + \frac{B}{W}}\right\}.$$

Thus, $\widehat{R}$ compares variation in posterior means across chains with average posterior variation within chains. Values close to one indicate that the chains are exploring the same posterior distribution. As a complementary check, at each cutoff we calculate all pairwise correlations of the current vector of member deviations, $b_{i,t}$, across the eight chains

and take their median. We then summarize these cutoff-level medians across the post-1992 sample; correlations close to one indicate that the chains produce nearly the same cross-section of member beliefs.

For the Phillips-curve parameter, across 4,193 post-1992 member-meeting states, the mean $\widehat{R}$ is 1.002 for both $\theta_{i,t}$ and $b_{i,t}$, the median is 1.000, and the 95th percentile is 1.006. The median pairwise correlation of the current $b_{i,t}$ vector is 0.999, and its fifth percentile across cutoffs is 0.996. For interest-rate sensitivity, the mean $\widehat{R}$ is 1.016 for $\theta_{i,t}$ and 1.024 for $b_{i,t}$; the corresponding medians are 1.005 and 1.004, and the 95th percentiles are 1.056 and 1.042. The median pairwise correlation of the current $b_{i,t}$ vector is 0.983, and its fifth percentile across cutoffs is 0.885.

IA.D A Model of Committee Aggregation

Our results indicate that members hold heterogeneous models of the macroeconomy, and we provide evidence that the committee tilts in the direction of members whose models better fit recent data. As we discuss in the main text, such tilting might improve decision-quality. It may also contribute to suboptimal decisions, if the committee tilts too far in the direction of better fitting models, as may have been the case with the committee's adoption of Flexible Average Inflation Targeting (FAIT) in 2020. Below, we present a simple theoretical framework that formalizes these points.

IA.D.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 M is a probability measure on $(\Omega, \mathcal{F})$, where Ω can be taken to represent all possible macroeconomic realizations. We consider "structured models," in the language of Hansen and Sargent (2022), where each model M_θ is induced by a parameter $\theta \in \Theta$. Throughout, we use uppercase letters to denote probability measures and the corresponding lowercase letters to denote their densities.[26]

Each member i has their own personal, fixed structured model M_i. In particular, $M_i = M_{\theta_i}$ for some $\theta_i \in \Theta$. When the models are dominated by a common reference measure ν, we write $m_i = dM_i/d\nu$ for the density of member i's model. We make three assumptions about members' models.

Assumption 1 (Absolute Continuity) *Members' models agree with the true model on zero probability events (models are mutually absolutely continuous with respect to μ).*

Assumption 2 (Partial Misspecification) *Members' models are partially misspecified ($\forall i$, $\exists S_i \in \mathcal{F}$ 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 ($\exists S \in \mathcal{F}$ such that $\mu(S) > 0$ and $M_i(S) \neq M_j(S)$).*

The committee makes its decision as follows. Both members observe data D_t from a prior period, which is a realization of a random variable, realized CPI and GDP growth, for example, whose distribution depends on the state of the world. For any state ω, the probability of observing D_t is $\pi(D_t|\omega)$. Given a model P, with density 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,

$$v_i = \alpha_i m_i + (1 - \alpha_i)m_j, \text{ where } j \neq i \text{ and } \alpha_i \in [0, 1]. \tag{12}$$

[26] For expositional ease, we focus on a static setting and abstract from the model-updating process.

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

$$m = \frac{1}{2}(v_A + v_B) = w_A m_A + (1 - w_A) m_B, \tag{13}$$

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

$$U_i(\alpha_i; \alpha_j) = \log \mathcal{L}(D_t|m) - \lambda KL(m_i \| v_i), \tag{14}$$

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 $\lambda \geq 0$.

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 reflects the difficulty of advocating for one's own model when it is misaligned with the observed data, and induces the committee to tilt its decision toward models that better fit D_t, consistent with the empirical evidence. On its own, the fact that this term is based on the committee's model can be interpreted as indicating that the strength of members' arguments for their own models, captured by α_i, depends on the strength of the other member's arguments. For example, in the case of a committee decision to put 50% weight on each model, this term enters similarly regardless of whether both members agree on the 50% weight, or if both members vote for 100% weight on their own model and 0% weight on the other member's model.[27]

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, or a committee chair that encourages open disagreements.

IA.D.2 Equilibrium

We study pure strategy Nash equilibria.

Definition IA.D.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.

[27] This interaction is plausible in a deliberative committee: members advance their views before colleagues who observe the same evidence and respond to one another's arguments, so the strength of an argument depends on how well the resulting committee position accounts for both the data and the other members' arguments.

Proposition 1 (Equilibrium Existence and Uniqueness) *If $\lambda > 0$, there is a unique equilibrium.*

Proof. All proofs are provided at the end of the section in Appendix Section IA.D.5. □

IA.D.3 Committee Performance

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

Proposition 2 (Potential Gains from 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. The result characterizes when there are potential gains from combining members' models: there exists a set of weights $\mathbf{w}^*$ under which the weighted average of members' models outperforms either individual model. In the context of model heterogeneity of the sort we document in the paper, where members exhibit persistent belief disagreements, this proposition indicates that committee design is useful to keep in mind. A committee designer who knows what members' models are and how they will respond to different macroeconomic settings should choose members with different models that diversify one another.

However, the condition in Proposition 2 is not sufficient for the committee to outperform its best member in equilibrium, because the committee need not choose the optimal weights $\mathbf{w}^*$. Instead, the committee's equilibrium weights depend on the realized data and on members' willingness to advocate for models different from their own. Thus, 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 sensitive to noise in the data, or to missing hard-to-detect tail risks captured in members' models that did not manifest. As $\lambda \to \infty$, each member advocates for their own model, and in equilibrium, the chosen model is always $m = \frac{1}{2}m_A + \frac{1}{2}m_B$. That is, there is no variability, but the weights on members' models are biased relative to the optimum (except when $w_A^* = \frac{1}{2}$). For intermediate values, as λ increases, the equilibrium model weights exhibit more bias and less variance relative to the optimum.

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

Proposition 3 (Committee Outperformance) *Under the conditions of Proposition 2, denote $w_A(\lambda)$ as the equilibrium weight on model A, which depends on the realized data. Define the expected performance of a committee model with weight w_A on member A's model as*

$$\phi(w_A) \equiv \mathbb{E}_\mu \left[\log\left(w_A m_A + (1 - w_A)m_B\right)\right].$$

Let

$$\Delta \equiv \phi(w_A^*) - \max\{\phi(1), \phi(0)\}, \tag{15}$$
$$\beta(\lambda) \equiv \mathbb{E}(w_A(\lambda) - w_A^*), \tag{16}$$
$$\sigma^2(\lambda) \equiv \mathbb{V}(w_A(\lambda)), \text{ and} \tag{17}$$
$$c \equiv -\phi''(w_A^*), \tag{18}$$

where w_A^* is the weight that maximizes $\phi(w_A)$, 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 expected performance evaluated at the optimal model weights.[28] *Under a second-order approximation, if*

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

then

$$\mathbb{E}\left[\phi(w_A(\lambda))\right] > \max\{\phi(1), \phi(0)\}.$$

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.

[28] Expectations and variances without a subscript are taken over the realized data D_t.

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

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

$$\mathbb{E}\left[(w_A(0) - w_A^*)^2\right] > \left(\frac{1}{2} - w_A^*\right)^2,$$

there is a $\lambda > 0$ at which the expected log likelihood of the committee model exceeds its value at $\lambda = 0$.

The logic of Corollary 1 is that, to second order, maximizing the expected performance of the committee's model is equivalent to minimizing the expected mean-squared error of the chosen weights with respect to w^*,

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

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.

IA.D.4 Implications for Committee Design

Our framework illustrates novel tradeoffs and practical implications for committee decision-making of the form we document at the FOMC, where the committee's decisions tilt toward the preferences of members whose models better fit recent data. Committees may improve decision-making. However, member selection is important, as potential benefits require members with a diversity of models who are sufficiently willing to entertain one another's views. One way to push members in such a direction is via structured deliberation, for example, as in FOMC meetings, where members discuss their interpretation of economic conditions and policy views before voting. At the same time, it is possible for the committee to tilt too far in the direction of models that better fit recent data, because limited data can only imperfectly reveal the true model. For example, a model may capture recent data well but do a poor job of capturing unobserved tails of the distribution. Tilting too far in the direction of such a model is likely to be suboptimal.[29]

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

Indeed, the committee's adoption of Flexible Average Inflation Targeting may be taken as a case where the committee leaned too far in the direction of models that better fit recent data. The policy was rooted in the more Keynesian view that the economy was facing a large aggregate demand shortfall in the post-GFC era, contributing to low inflation, with a concern that this could anchor stubbornly low inflation expectations. Implicitly, the policy gave little weight to the possibility of inflationary supply shocks, which hit the economy following the COVID pandemic. The committee then reverted from the policy in its framework review in 2025.

Finally, 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.

IA.D.5 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 the fact that the strategy space is convex and compact; this maximizer corresponds 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

$$\phi(w_A) = \mathbb{E}_\mu \left[\log\left(w_A m_A + (1 - w_A) m_B\right)\right]$$

around w_A^*, observe that $\phi(w_A) \approx \phi(w_A^*) - \frac{c}{2}(w_A - w_A^*)^2$, where $c \equiv -\phi''(w_A^*) > 0$, and the linear term disappears because $\phi'(w_A^*) = 0$. Taking expectations, we have

$$\mathbb{E}\left[\phi(w_A(\lambda))\right] \approx \phi(w_A^*) - \frac{c}{2}\mathbb{E}\left[(w_A(\lambda) - w_A^*)^2\right]. \tag{IA.D.2}$$

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

$$\mathbb{E}\left[\phi(w_A(\lambda))\right] \approx \phi(w_A^*) - \frac{c}{2}(\beta(\lambda)^2 + \sigma^2(\lambda)). \tag{IA.D.3}$$

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. From Equation (IA.D.2), maximizing the expected log likelihood is equivalent, to a second-order approximation, to minimizing

$$MSE(\lambda) = \mathbb{E}\left[(w_A(\lambda) - w_A^*)^2\right].$$

When $\lambda = 0$, members place no cost on advocating for models different from their own, so $w_A(0)$ is the fully data-driven committee weight. As $\lambda \to \infty$, deviations from members' own models become arbitrarily costly. Hence each member advocates for their own model, and

$$w_A(\lambda) \to \frac{1}{2}.$$

It follows that

$$MSE(0) = \mathbb{E}\left[(w_A(0) - w_A^*)^2\right], \qquad \lim_{\lambda\to\infty} MSE(\lambda) = \left(\frac{1}{2} - w_A^*\right)^2.$$

Thus, when

$$\mathbb{E}\left[(w_A(0) - w_A^*)^2\right] > \left(\frac{1}{2} - w_A^*\right)^2,$$

the fully data-driven committee weight does not maximize expected performance. Therefore, there is some $\lambda > 0$ such that $MSE(\lambda) < MSE(0)$, so the expected log likelihood of the committee model exceeds its value at $\lambda = 0$. □

IA.E LLM Prompts and Validation

Section IA.E.1 presents the prompts used to construct transcript-based measures for analysis. Section IA.E.2 presents comparisons of LLM-based coding with human coding.

IA.E.1 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.

- The prompt to identify committee members' arguments is in Section IA.E.1.1.
- Given their arguments, prompts to extract beliefs about model parameters are presented in Section IA.E.1.2.
- Given their arguments, prompts to identify and categorize the information sources cited by committee members are in Section IA.E.1.3.
- The prompts to identify each policy decision made during the meeting are in Section IA.E.1.4.
- Given the policy decisions made, the prompts to identify each member's alignment with the adopted decisions and their contributions to those decisions' adoption are in Section IA.E.1.5.

IA.E.1.1 Prompts to Identify Committee Members' Arguments

For each speaker-meeting, we take all of the member's speeches and input them into the LLM. We prompt the LLM to identify all economically distinct arguments concerning GDP growth or output, inflation, and unemployment. The prompt 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 and extract economically distinct arguments about GDP
 growth/output, inflation, and/or unemployment.

## Task

Read through the speeches and identify every economically distinct argument that pertains to
  GDP growth/output, inflation, and/or unemployment. An argument qualifies for inclusion if
 it makes a claim about the current level, recent or past changes, or expected future
 evolution of at least one of these three variables. Arguments that discuss other economic
 forces (e.g., credit conditions, asset prices, business investment) qualify as long as the
 speaker connects them --- explicitly or through clear economic reasoning --- to GDP growth/
 output, inflation, or unemployment.

Where the speaker provides reasoning, cause-and-effect logic, or an explanation for why a
 variable is at its current level, has changed, or is expected to evolve in a particular way
 , that reasoning is part of the argument and should be included in the quotation. For
 example, "inflation remains elevated because supply chains have not fully recovered" is a
 single argument --- the causal reasoning is inseparable from the claim about inflation.
 Bare factual statements without accompanying reasoning (e.g., "unemployment has fallen to
 3.6 percent") still qualify as arguments and should be extracted.
```

```
The boundary between arguments is defined by the economics, not by the text. Two claims are
 separate arguments if they involve distinct economic or causal content --- for example, a
 claim about the current level of inflation and a claim about the cause of rising
 unemployment are separate arguments even if the speaker makes both in the same sentence.
 Conversely, a claim and the causal reasoning the speaker offers in support of it are part
 of the same argument even if they span multiple sentences.

Do not extract the same argument twice. For example, "GDP growth was strong in Q3" and "the
 economy grew at 4 percent in Q3" are the same argument and should be extracted only once.
 But "GDP growth was strong in Q3" and "I expect growth to slow next year" are separate
 arguments.

## Guidelines

- Include only the speaker's own views or views they clearly endorse.
- Exclude arguments the speaker cites to refute or disagree with.
- Each argument must pertain to the current level, recent or past changes, or expected
 future evolution of GDP growth/output, inflation, and/or unemployment.
- When the speaker provides reasoning or causal logic behind a statement about one of these
 variables, include that reasoning as part of the same argument and quotation --- do not
 strip the reasoning away from the claim it supports.
- Quotations must be extracted verbatim from the speeches --- do not paraphrase, summarize,
 or alter the wording in any way.
- Each quotation should include enough surrounding context to make the argument intelligible
  on its own. It is acceptable for quotations to overlap if two economically distinct
 arguments are made in the same passage.
- If the speaker makes no qualifying arguments about GDP growth/output, inflation, or
 unemployment, return an empty output block: <output></output>

## Output Format

Provide your results in this exact XML structure. Each argument should be contained within
 an `<argument>` block:

'''xml
<output>
<argument>
    <description>[One-sentence summary stating the claim, its direction, the reasoning if
     any, and which variable(s) it pertains to]</description>
    <quotation>[Verbatim quotation from the speeches]</quotation>
</argument>
</output>
'''

Provide only the XML output above with no additional commentary.

Here are the FOMC member speeches you need to analyze:

<speeches>
{speeches}
</speeches>
"""
```

IA.E.1.2 Prompts to Extract Model Parameters.

Phillips curve: To extract FOMC members' beliefs about the Phillips curve relationship, we pass each argument through the LLM with a classification prompt. We classify beliefs as *steep*, *moderate*, *flat*, or *null*. The prompt also identifies whether the speaker cites a separate non-slack force affecting inflation. The model returns an XML object containing the classification and supporting evidence.

```
"""
```

```
Classify an FOMC speaker's belief about the Phillips curve: how strongly economic slack
 affects price inflation. The main object is the structural slope of the slack-inflation
 relationship, holding fixed non-slack shocks and the level or measurement of slack.
 Classify `phillips_slope` as `steep`, `moderate`, `flat`, or `null`. Separately, record
 whether the speaker explicitly identifies a separate non-slack inflation force.

Use only the quotation as evidence. Do not rely on external summaries, metadata, or inferred
  context that is not stated in the quotation itself.

## Step 1: Identify a slack-to-inflation claim

Classify only statements that connect economic slack to price inflation. Slack may be
 expressed through unemployment, labor market tightness, wage pressure, the output gap,
 growth relative to potential, or capacity utilization.

A statement is admissible only if the quotation explicitly says or clearly entails that more
  or less slack changes inflation or inflation pressure, or that this relationship itself
 has weakened or strengthened. Wages alone are not enough: the statement must reach price
 inflation.

Policy or forecast shorthand is not enough unless the mechanism is stated. Examples such as
 "further easing risks overheating," "strong labor market conditions warrant higher rates,"
 or "inflation may rise as the economy strengthens" are `null` unless the quotation explains
  that reduced slack will push inflation up. However, a conditional forecast in which the
 speaker attributes their inflation path to a specific degree of slack or tightness is
 admissible: "with the economy slightly tighter, I have inflation getting to 2 percent"
 treats tightness as the reason inflation rises and identifies the slack-to-inflation
 channel.

## Step 2: Separate slope beliefs from gap measurement and shocks

The object here is the structural parameter: how strongly slack affects inflation. Do not
 classify the current level of slack, the current inflation outcome, the preferred policy
 response, or the claim that the output gap is larger or smaller than commonly measured.

Use this test: **could a speaker with the opposite view of the Phillips curve plausibly make
  the same statement?** If yes, the statement does not identify the slope and should be `
 null`.

Claims that measured slack or the output gap is uncertain, mismeasured, smaller or larger
 than conventional estimates, or only inferable from inflation outcomes are not slope
 beliefs by themselves. Code `null` unless the quotation explicitly says the slack-to-
 inflation coefficient is strong, weak, unreliable, or otherwise identifies the relationship
  itself.

Statements that inflation did not move as a measured gap predicted are admissible only if
 the speaker draws a structural inference about the link itself. "Inflation has not fallen
 as much as the estimated output gap predicted, so the true gap may be smaller" is `null`. "
 Output-gap measures have little predictive value for inflation because the link itself is
 weak" is `flat`.

Use shock or other non-slack-force language only insofar as it helps identify the slope. If
 the quotation only says what else is moving inflation in the current episode --- energy,
 commodity prices, supply disruptions, import prices, exchange rates, markups, global
 pricing forces, or inflation expectations treated as an autonomous force --- then `
 phillips_slope` is `null`.

Explicit pass-through statements can identify the slope. If the speaker says that tight
 labor markets, higher labor costs, or reduced slack are not feeding through to prices as
 they typically did, or says the relationship itself has changed or weakened, that is
 evidence about the slope rather than a mere current-episode observation. If the speaker
 says slack is preventing a named shock from passing through to inflation, that also
 identifies the slack-to-inflation mechanism.

If a sentence contains both a non-slack-force clause and an independent structural claim
 about slack, classify `phillips_slope` from the structural claim. For example, "slack
 should exert disinflationary pressure, but energy prices are pushing in the other direction
 " is admissible because the first clause states the channel.
```

```
Common ‘null‘ cases:
- pure shock attribution: "inflation is being driven by energy prices"
- shock offset without an independently stated slack channel: "commodity prices may offset
 disinflationary forces"
- ambiguous net outcome: "inflation has remained subdued despite low unemployment"
- level of slack only: "the output gap has closed"
- gap measurement only: "the output gap may be much smaller than standard estimates imply"
- wages without prices
- forecasts, policy views, or historical references without Phillips-curve reasoning
- labor market listed as one factor among several in a forecast without isolating its effect
  on inflation
- current-episode observations such as "firms lack pricing power despite tight labor markets
 " unless the quotation explicitly treats that as evidence the relationship itself is weak
- reporting what other people believe about inflation drivers rather than the speaker’s own
 structural view

Admissible examples:
- "we need more slack to bring inflation down"
- "tight labor markets are feeding into higher wages and prices"
- "tight labor markets have not fed through to inflation as they typically have"
- "despite low unemployment, inflation remains subdued, suggesting the Phillips curve has
 flattened"
- "slack should reduce inflation, but energy prices are offsetting that pressure"

## Step 3: Classify ‘phillips_slope‘

- ‘steep‘: the speaker treats slack as a meaningful driver of inflation. Ordinary FOMC
 hedging words such as "should," "may," "expected," or "likely" do not by themselves move a
 statement to ‘moderate‘. If the speaker makes an unqualified directional claim that slack
 is putting upward or downward pressure on inflation, code ‘steep‘ even if other forces
 offset that pressure in the current episode. Likewise, if the speaker says slack is
 preventing a non-slack shock from passing through to inflation and does not qualify that
 channel as weak or uncertain, code ‘steep‘.
- ‘moderate‘: the speaker accepts the channel but explicitly qualifies it as weaker, partial
 , less reliable, or uncertain. Examples: "the Phillips curve has flattened somewhat"; "
 limited passthrough"; "depends on whether the relationship still holds." Statements about
 difficulty measuring the output gap, detecting slack in real time, or forecasting with
 current gap estimates are ‘null‘ unless they clearly qualify the slope itself rather than
 the gap measure.
- ‘flat‘: the speaker concludes the channel is very weak or largely absent. Examples: "the
 link has broken down"; "tight labor markets have simply not fed through to inflation"; "the
  Phillips curve has weakened considerably"; "domestic slack no longer does much work for
 inflation." Statements that slack has little reliable predictive content for inflation
 because the link itself is weak, or that the labor market is not susceptible to setting off
  a wage-price spiral, belong here. Reserve ‘flat‘ for language that treats the channel as
 near zero, irrelevant, or no longer operative. Do not use ‘flat‘ for language that merely
 questions the current level of slack or the measurement of the gap.
- ‘null‘: no slope belief is identified.

Important distinctions:
- "tight labor markets should push inflation higher" -> ‘steep‘
- "the Phillips curve exists but has flattened" -> ‘moderate‘
- "inflation is low despite tight labor markets" -> ‘null‘
- "inflation is low despite tight labor markets, suggesting the link has weakened" -> ‘flat‘
- "inflation did not fall as much as the measured output gap predicted, so the true output
 gap may be smaller" -> ‘null‘
- "I am skeptical of output-gap estimates; depending on the gap, policy implications differ"
  -> ‘null‘
- "output gaps are unreliable predictors of inflation because the link itself is weak" -> ‘
 flat‘
- "tight labor markets are not feeding through to prices" -> ‘flat‘
- "slack should lower inflation, but energy prices are offsetting it" -> ‘steep‘
- "energy prices are raising inflation despite slack" -> ‘null‘

## Step 4: Classify ‘non_slack_inflation_force‘

After classifying ‘phillips_slope‘, independently record whether the quotation explicitly
```

```
 identifies a separate non-slack inflation force that is actively moving inflation. This
 field is auxiliary and does NOT determine `phillips_slope`.

Use:
- `inflationary`: a separate non-slack inflation force is described as pushing inflation up
- `deflationary`: a separate non-slack inflation force is described as pushing inflation
 down or holding it down
- `unclear`: a separate non-slack inflation force is identified, but its direction is not
 clear
- `na`: no separate non-slack inflation force is clearly identified

Decision order:
1. First strip away named shocks and current inflation outcomes, and ask what the quotation
 says about the slack-to-inflation parameter itself.
2. Then separately record whether some other force is described as currently moving
 inflation.

Examples of non-slack forces include energy or commodity prices, supply bottlenecks, import
 prices, tariffs, exchange-rate pass-through, markups, global pricing forces, or inflation
 expectations when the speaker treats them as actively moving inflation.

Code only forces described as actively moving inflation in the current episode. Do NOT infer
  a separate non-slack force merely because inflation moved despite slack; if no distinct
 driver is named, use `na`.

If the text mentions only a hypothetical, contingent, or future risk ("could raise inflation
 ", "might put upward pressure") rather than a force already operating, do not code it as
 active; use `na` unless the passage separately says it is currently affecting inflation.

If a non-slack shock is mentioned only as pressure that slack is preventing from passing
 through to inflation, use `na`; the shock does not count as active unless the quotation
 says it is currently moving inflation.

Do NOT code labor market tightness, unemployment, the output gap, overheating, strong
 spending, fiscal stimulus, credit conditions, or the slack-inflation channel itself as `
 non_slack_inflation_force`. Code expectations only when the speaker treats them as an
 active force moving inflation, such as expectations having recently risen and pushing
 inflation up, or having recently fallen and holding inflation down or undermining return to
  target. Anchor language such as "anchored," "stable," "firmer anchor," or "reduced tail
 risks" is `na` unless the passage goes further and says expectations are currently pushing
 inflation up or down.

## Output format

Respond with only the following XML block and nothing else.

<classification>
<phillips_slope>[steep|moderate|flat|null]</phillips_slope>
<non_slack_inflation_force>[inflationary|deflationary|unclear|na]</non_slack_inflation_force
 >
<reasoning>[1-2 sentences citing specific phrases from the quotation]</reasoning>
</classification>

Here is the economic argument you need to analyze:
<economic_argument>
<quotation>{quotation}</quotation>
</economic_argument>

Remember: classify only from the quotation. If the quotation does not itself identify a
 Phillips-curve slope, return `null`.
"""
```

IS curve: To extract FOMC members' beliefs about the IS curve relationship, we pass each argument through the LLM with a classification prompt. We classify the strength of monetary-policy transmission to real activity as *strong*, *moderate*, *weak*, or *null*. The prompt

also identifies the transmission channel and whether the speaker cites a separate non-rate force affecting real activity. The model returns an XML object containing the classification and supporting evidence.

```
"""
Classify an FOMC speaker's belief about the IS curve: how strongly monetary policy (interest
  rates) affects real economic output. The main object is the structural strength of the
 rate-to-output relationship, holding fixed non-rate shocks and the level or stance of rates
 . Classify `is_transmission` as `strong`, `moderate`, `weak`, or `null`. Separately, record
  whether the speaker reasons through the direct (Euler / cost-of-capital) channel, the
 indirect (downstream financial-conditions) channel, or both, and whether the speaker
 explicitly identifies a separate non-rate force actively moving real activity.

Use only the quotation as evidence. Do not rely on external summaries, metadata, or inferred
  context that is not stated in the quotation itself.

## Step 1: Identify a rate-to-output claim

Classify only statements that connect monetary policy to real economic activity. Monetary
 policy may be expressed through the federal funds rate, real rates, the policy stance (
 tight/restrictive/easy/accommodative), rate hikes or cuts, monetary conditions, or monetary
  aggregates. Real activity includes GDP, output, growth, aggregate demand, consumption,
 investment, employment, net exports, or interest-sensitive sectors such as housing,
 durables, business capex, and trade.

A statement is admissible only if the quotation explicitly says or clearly entails that
 monetary policy, operating directly or through its rate-driven downstream financial
 conditions, changes real activity, or that this relationship itself has weakened or
 strengthened. Mention of the rate stance alone is not enough: the statement must reach real
  activity. Forecasts, policy preferences, and policy-reaction statements are not enough
 unless the rate-to-output mechanism is stated. "We should cut rates," "I forecast slower
 growth," or "with economic growth slowing, we pushed back our estimate for the timing of
 liftoff" are `null` without transmission reasoning. Do not infer the channel from a Taylor-
 rule-style reaction to economic conditions; the quotation must say or clearly entail that
 the rate path itself would change real activity.

This prompt covers rate-driven monetary-policy transmission to output, including downstream
 pass-through. The rate channel admits both direct effects (the rate operating on the cost
 of capital and intertemporal substitution between current and future spending) and indirect
  effects (rate-driven changes in mortgage rates, household borrowing costs, refinancing,
 asset prices including housing prices and equity prices, household wealth, and the exchange
  rate). Statements about financial conditions not downstream of the policy stance, such as
 credit spreads from risk sentiment, lending standards from bank losses, house prices not
 linked to rate policy, or balance-sheet policies that operate independently of the rate
 path, are outside scope and `null`. The test: does the speaker trace the mechanism back to
 the policy rate, the policy stance, or rate-sensitive prices that reflect the stance?

Forward guidance about the rate path operates through the same channel: output depends on
 the entire expected path of future real rates. "Extending our forward guidance would
 provide additional stimulus to the economy" is admissible. Forward guidance about
 communication strategy without an output link ("we should communicate our rate path more
 clearly") is `null`. Beliefs about balance-sheet policies, asset purchases, or QE that
 operate independently of the rate path are `null`; if the speaker discusses both rate
 policy and balance-sheet policy, classify based on the rate-policy component only.

## Step 2: Separate channel-strength beliefs from level-of-rates and shocks

The object here is the structural parameter: how strongly monetary policy affects output. Do
  not classify the current level of rates, the current output outcome, the preferred policy
 response, or the claim that the policy stance is more or less restrictive than commonly
 measured.

Use this test: could a speaker with the opposite view of rate-policy-to-output transmission
 strength plausibly make the same statement? If yes, the statement does not identify `
 is_transmission` and should be `null`.

Claims that measured rates or the policy stance is uncertain, mismeasured, more or less
```

```
 restrictive than conventional estimates, or identifiable only from observed output
 movements are not channel-strength beliefs by themselves. Code ‘null‘ unless the quotation
 explicitly says the rate-to-output coefficient is strong, weak, unreliable, or otherwise
 identifies the relationship itself.

Statements about how output moved relative to the policy stance, whether output moved
 together with policy, opposite to policy, or did not move as expected, are admissible only
 if the speaker draws a structural inference about the rate-policy-to-output transmission
 relationship itself. Pure observation of co-movement or anti-movement without causal
 attribution is ‘null‘. "Growth has remained solid despite tighter policy" is ‘null‘ because
  it is equally consistent with strong transmission plus offsetting shocks or weak
 transmission. "Growth is slowing as we tighten" alone is ‘null‘ because it is co-movement
 without attribution. Both become admissible when the speaker draws a structural conclusion:
  "growth has remained solid despite tighter policy, suggesting a smaller-than-usual rate
 effect" is ‘moderate‘; "growth has remained solid despite tighter policy, suggesting little
  or no rate effect" is ‘weak‘; "growth has slowed as our tightening works through" is ‘
 strong‘.

Use shock or other non-rate-force language only insofar as it helps identify ‘
 is_transmission‘: the structural transmission of rate policy to real output. If the
 quotation only says what else is moving output in the current episode, such as fiscal
 policy, supply shocks, terms of trade, demographics, or productivity, then ‘is_transmission
 ‘ is ‘null‘ unless the speaker makes a parallel structural claim about rate-policy-to-
 output transmission. If a sentence contains both a non-rate-force clause and an independent
  structural claim about rate-policy-to-output transmission, classify ‘is_transmission‘ from
  the structural claim. "Fiscal policy is the main driver now" alone is ‘null‘. "Fiscal
 stimulus is offsetting our tightening" is admissible because it presupposes tightening is
 contractionary. "Fiscal policy is the main driver and monetary policy has almost no
 traction" is ‘weak‘.

Both direct and indirect channel claims identify the parameter. Direct: the speaker reasons
 through the policy rate operating on the cost of capital, intertemporal substitution, or
 the consumption Euler equation. "Higher rates are raising the cost of capital and
 discouraging investment" is admissible. "Tight policy should slow growth" is admissible.
 Indirect: the speaker traces a chain from the policy stance through a downstream financial-
 conditions variable (mortgage rates, household borrowing costs, refinancing, asset prices,
 equity wealth, exchange rate) to output. "Lower mortgage rates from our easing are
 supporting housing" is admissible. "Our rate hikes have strengthened the dollar and that is
  hurting exporters" is admissible. "Our low rates have boosted house prices, supporting
 consumption through wealth effects" is admissible. "Lower rates have allowed mortgage
 refinancing, freeing up household cash flow for spending" is admissible. The downstream
 variable must be connected to both the policy stance and real activity; observing
 downstream variables alone is not admissible. For indirect-channel claims, the speaker must
  explicitly attribute the downstream variable’s movement to monetary policy or the rate
 stance. "Falling mortgage rates are supporting housing" alone is ‘null‘; "Falling mortgage
 rates from our easing are supporting housing" is admissible.

Common ‘null‘ cases:
- pure shock attribution: "growth is being driven by fiscal stimulus"
- ambiguous resilience: "growth has remained solid despite tighter policy"
- level of rates only: "policy is restrictive"
- generic financial conditions without rate attribution: "tighter financial conditions might
  slow the economy"
- downstream observations without output link: "I would like to see mortgage rates come down
 "
- wealth or asset-price effects without rate linkage: "falling house prices are hurting
 spending"
- forecasts or policy views without rate-to-output reasoning: "I forecast slower growth," "
 we should cut"
- forward guidance without output link: "we should communicate our rate path more clearly"
- balance-sheet / QE policy operating independently of the rate path

Admissible examples:
- "tight policy should slow growth"
- "rate hikes are slowing investment by raising the cost of capital"
- "lower mortgage rates from our easing are supporting housing"
- "the restrictive stance has strengthened the dollar, hurting exports"
- "our low rates have boosted equity prices, providing a wealth effect that supports
```

```
 consumption"
- "lower rates have allowed refinancing, freeing up household cash flow for spending"
- "growth has remained solid despite tighter policy, suggesting a smaller-than-usual rate
 effect"

## Step 3: Classify `is_transmission`

- `strong`: the speaker treats rate policy as a meaningful driver of output. Ordinary FOMC
 hedging words such as "should," "may," "expected," or "likely" do not by themselves move a
 statement to `moderate`. If the speaker makes an unqualified directional claim that rate
 policy is restraining, supporting, stimulating, or slowing output, code `strong` even if
 other forces offset that pressure in the current episode. Likewise, if the speaker says a
 non-rate force is offsetting rate-policy effects and does not qualify the rate channel as
 weak or uncertain, code `strong`.
- `moderate`: the speaker accepts the channel but explicitly qualifies its strength as
 weaker, partial, sector-specific, less reliable, limited in magnitude, unusually delayed in
  a way that limits current force, or uncertain. Examples: attenuated pass-through, one rate
 -sensitive sector responding while another does not, one channel being limited while
 another may still operate, or uncertainty about whether the usual rate effect is operating.
  If the speaker rejects one rate channel or says one sector is unresponsive while leaving
 other rate channels operative or possible, code `moderate`, not `strong`. Statements about
 difficulty measuring the stance of policy, identifying neutral rates, or forecasting with
 current rate estimates are `null` unless they clearly qualify the rate-to-output channel
 itself rather than the stance measure.
- `weak`: the speaker concludes the channel is very weak or largely absent. Examples: broken
  transmission, pushing-on-a-string language, an explicit claim that rate policy is
 ineffective for real demand, or a claim that the relevant spending or investment margin is
 not materially responsive to realistic rate changes. Statements that rate policy has little
  reliable effect on output because the channel itself is weak belong here. Reserve `weak`
 for language that treats the channel as near zero, irrelevant, no longer operative, or not
 material at the relevant margin. If the relevant real-activity margin in the quotation
 would not respond to a realistic rate or cost-of-capital change, classify `weak` even if
 the speaker does not separately discuss every other possible channel. Do not use `weak` for
  language that merely says the channel is attenuated, partial, sector-specific, or
 uncertain while still operative.
- `null`: no channel-strength belief is identified.

Important distinctions:
- "tight policy should slow growth" -> `strong` (standard hedging)
- "tight policy may slow growth, but the rate effect appears attenuated" -> `moderate`
- "one direct rate channel is not binding, but other rate channels may still help activity"
 -> `moderate`
- "the rate-to-output link has broken down" -> `weak`
- "policy is restrictive" -> `null` (no output claim)
- "policy is restrictive and is weighing on growth" -> `strong`
- "we should pause because the economy is fragile" -> `null` (policy reaction without
 transmission reasoning)
- "we should pause because further tightening would slow a fragile economy" -> `strong`
- "growth is solid despite tighter policy" -> `null` (ambiguous net outcome)
- "growth is solid despite tighter policy, suggesting a smaller-than-usual effect" -> `
 moderate`
- "growth is solid despite tighter policy, suggesting little or no effect" -> `weak`
- "I forecast slower growth because policy is restrictive" -> `strong`
- "the full effects will arrive in 2-3 quarters" -> `strong` (presupposes the channel)
- "monetary policy cannot fix this because the problems are structural" -> `weak`
- "rising mortgage rates from our tightening are cooling housing" -> `strong` (indirect
 channel)
- "our easing has boosted equity prices, providing a wealth effect for consumption" -> `
 strong` (indirect channel)
- "lower rates allowed refinancing that has supported consumer spending" -> `strong` (
 indirect channel)
- "the restrictive stance has strengthened the dollar, hurting exports" -> `strong` (
 indirect channel)

When in doubt, `null`.

## Step 4: Classify `is_subchannel`
```

```
After classifying `is_transmission`, independently record which sub-channel of rate-to-
 output transmission the speaker invokes. This field is auxiliary and does not determine `
 is_transmission`. The distinction follows the HANK literature (Kaplan-Moll-Violante 2018;
 Auclert 2019), which decomposes monetary transmission into a direct (intertemporal-
 substitution) channel and indirect channels operating through downstream financial-
 conditions variables.

Use:
- `direct`: the speaker reasons through the policy rate (or expected rate path) operating on
  the cost of capital, firm borrowing rates, intertemporal substitution, or the consumption
 Euler equation.
- `indirect`: the speaker reasons through specific rate-driven downstream financial-
 conditions variables, such as mortgage rates, asset prices (housing or equity), household
 wealth, refinancing or cash-flow effects, the exchange rate, household borrowing costs, or
 other named downstream variables.
- `both`: the speaker explicitly invokes both direct and indirect channels in the same
 argument.
- `null`: applies whenever `is_transmission` is `null`.

Decision order:
1. First classify `is_transmission` from the rate-to-output claim.
2. Then ask: does the speaker name a specific downstream financial-conditions variable as
 the operative mechanism (mortgage rates, asset prices, refinancing, exchange rate, etc.)?

If yes, classify as `indirect`; if the speaker also invokes direct rate effects on cost of
 capital or intertemporal substitution in the same argument, classify as `both`. Count a
 channel as invoked when it is accepted, rejected, or explicitly qualified as part of the
 rate-to-output argument; a negated direct-channel claim can still make the subchannel `both
 ` when paired with an indirect-channel claim. If no specific downstream variable is named,
 classify as `direct`. Business borrowing rates, private-business rates, and the cost of
 capital for investment count as `direct` unless the speaker routes the mechanism through a
 named downstream variable such as the exchange rate, mortgage rates, refinancing, asset
 prices, or household wealth. Generic rate-stance language ("policy is restrictive and is
 weighing on activity," "tight policy should slow growth") defaults to `direct` because rate
 -stance reasoning operates through the textbook discount-rate / Euler-equation channel.

Code `both` only when the speaker explicitly invokes both mechanisms ("our easing has
 lowered the cost of capital for businesses and lowered mortgage rates for households,
 supporting both investment and consumption"). Discussion of multiple downstream variables
 alone is `indirect`, not `both`.

## Step 5: Classify `non_rate_output_force`

After classifying `is_transmission` and `is_subchannel`, independently record whether the
 quotation explicitly identifies a separate non-rate force that is actively moving real
 activity. This field is auxiliary and does not determine `is_transmission`.

Use:
- `expansionary`: a separate non-rate force is described as pushing real activity up
- `contractionary`: a separate non-rate force is described as pushing real activity down
- `unclear`: a separate non-rate force is identified, but its direction on real activity is
 not clear
- `na`: no separate non-rate force is clearly identified

Decision order:
1. First strip away named non-rate shocks and current output outcomes, and ask what the
 quotation says about the rate-to-output parameter itself.
2. Then separately record whether some other force is described as currently moving real
 activity.

Examples of non-rate forces include fiscal policy, supply shocks, trade and external-demand
 shocks not driven by rate policy, demand shocks, pandemics or other disruptions,
 demographics, structural shifts, inventory dynamics, and global growth conditions when
 treated as autonomous of the rate stance.

Code only forces described as actively moving real activity in the current episode. Do not
 infer a separate non-rate force merely because output moved despite the rate stance; if no
 distinct driver is named, use `na`. Do not code generic demand language as a separate force
```

```
  unless the quotation says autonomous demand conditions are currently pushing real activity
  up or down. Conditional threshold language for future spending, hiring, or investment is
 not by itself an active non-rate output force; use `na`, not `unclear`, unless the text
 identifies a current autonomous demand force.

Examples: "firms are waiting for demand to improve before investing" -> `na`; "weak
 autonomous demand is holding down investment" -> `contractionary`.

Do not code monetary policy or rate-stance effects as `non_rate_output_force`. Do not code
 rate-driven downstream financial conditions when the speaker has tied them to rate policy;
 those are part of the rate-to-output channel itself, not separate non-rate forces.

## Output format

Respond with only the following XML block and nothing else.

<classification>
<is_transmission>[strong|moderate|weak|null]</is_transmission>
<is_subchannel>[direct|indirect|both|null]</is_subchannel>
<non_rate_output_force>[expansionary|contractionary|unclear|na]</non_rate_output_force>
<reasoning>[1-2 sentences citing specific phrases from the quotation]</reasoning>
</classification>

Here is the economic argument you need to analyze:
<economic_argument>
<quotation>{quotation}</quotation>
</economic_argument>

Remember: classify only from the quotation. If the quotation does not itself identify an IS
 curve belief, return NULL.
"""
```

IA.E.1.3 Prompts to Identify and Categorize Information Sources

Having identified the arguments made, we next classify the information sources cited by each of these arguments. We first use a seeded sample of 1,000 arguments, pass them into the LLM in groups of five, and prompt it to construct categories based on the prompt below.

```
"""
You are identifying information-source category candidates from sampled FOMC economic
 arguments.

Identify source category candidates that can be aggregated into a source menu for the full
 argument set. A source is the information input the speaker appears to rely on for the
 economic belief in the argument.

Return raw XML only. The first non-whitespace character must be `<`. Return one <
 source_candidate> block per source category represented in the provided arguments.

# What counts as an information source

An information source is the observable input behind the speaker's belief: an official
 statistic, market indicator, staff material, regional report, contact evidence, survey,
 financial condition measure, or other information object used to support the claim.

The category should describe the information input, not the economic variable by itself. For
   example, GDP, payrolls, CPI, and industrial production may belong to an official-
 statistics source category if the sampled arguments support that grouping.

# What is not a source
```

```
A speaker's own forecast, model, or judgment is not a source by itself. If the argument says
  "my forecast" but does not identify the underlying information input, the source is not
 identified.

Do not create categories for broad speaker reasoning, policy preferences, theoretical
 mechanisms, or variables with no identifiable information input.

# Category construction

Use categories that actually appear in the provided arguments. Do not start from a preset
 taxonomy.

Keep categories broad enough to classify the full argument set. Do not create one category
 per variable, one category per wording variant, or one category per quotation.

Use `unclear` only as the residual category for arguments with no identifiable information
 source.

Do not classify whether the information is common to the Committee in this discovery step.
 This step defines source categories only.

For each source category candidate, include a short machine-readable name, a definition,
 cues for when to use it, boundary cases where it should not be used, one example argument
 id, and short exact evidence from that argument.

# Output

Return XML only. Escape XML special characters in evidence.

<source_discovery>
  <source_candidate>
    <source>short_source_name</source>
    <definition>definition</definition>
    <include_when>cues for using this source</include_when>
    <exclude_when>boundary cases that should use another source or unclear</exclude_when>
    <example_argument_id>argument_id</example_argument_id>
    <evidence>short exact evidence</evidence>
  </source_candidate>
</source_discovery>

Source discovery sample chunk: {chunk_id}

<arguments>
<argument id="{argument_id}">
  <meeting_date>{YYYYMMDD}</meeting_date>
  <speaker>{speaker}</speaker>
  <description>{description}</description>
  <quotation>{quotation}</quotation>
</argument>
{additional argument blocks}
</arguments>
"""
```

We then prompt the LLM to consolidate the information-source category names and descriptions into one combined list. We use the following prompt:

```
"""
You are consolidating information-source candidates into a small source menu for FOMC
 economic arguments.

The input contains candidate source names, definitions, inclusion cues, exclusion cues,
 counts, and example evidence. Many candidates are duplicates or near-duplicates. Build a
 compact menu that can classify the same kind of arguments without preserving wording
 variants.

Return raw XML only. The first non-whitespace character must be `<`. Return one <
 source_category> block per final source category.
```

```
# What counts as a source category

A source category is a recurring type of information input behind a speaker's economic
 belief: an official statistic, market indicator, staff material, regional report, contact
 evidence, survey, financial condition measure, or other information object used to support
 the claim.

The category should describe the information input, not the economic variable by itself, not
  the speaker's policy preference, and not a forecast or model label with no identifiable
 underlying input.

# Consolidation rules

Merge candidate names that describe the same information input even if they use different
 wording, specificity, or examples.

Keep meaningful distinctions when they would matter for classification. Do not collapse
 contact-based regional intelligence into official statistics; do not collapse Federal
 Reserve staff materials into public data releases; do not collapse financial market prices
 into broad macro statistics.

Use a small menu. Target roughly 8 to 15 substantive categories plus `unclear`, but use the
 evidence rather than a fixed count. Do not create one category per variable, one category
 per wording variant, or one category per rare example.

Keep source names short, lower-case, and machine-readable with underscores.

Include `unclear` as the residual category for arguments with no identifiable information
 source.

Do not classify whether the information is common to the Committee. Define source categories
  only.

# Evidence requirement

For each final source category, include one short exact evidence span copied from the
 candidate input. The evidence should illustrate the category, not justify every merged
 candidate.

Before writing the menu, check that each merged candidate belongs in exactly one final
 source category and that each final category has clear inclusion and exclusion boundaries.

# Output

Return XML only. Escape XML special characters in evidence.

<source_menu>
  <source_category>
    <source>short_source_name</source>
    <definition>definition</definition>
    <include_when>cues for using this source</include_when>
    <exclude_when>boundary cases that should use another source or unclear</exclude_when>
    <merged_candidates>comma-separated candidate source names</merged_candidates>
    <evidence>short exact evidence</evidence>
  </source_category>
</source_menu>

<source_candidates>
  <candidate>
    <source>{source}</source>
    <count>{count}</count>
    <definition>{definition}</definition>
    <include_when>{include_when}</include_when>
    <exclude_when>{exclude_when}</exclude_when>
    <evidence>{evidence}</evidence>
  </candidate>
  {additional candidate blocks}
```

```
</source_candidates>
"""
```

We refine and freeze the consolidated list into the source menu used for classification. We then individually input each argument into the LLM, alongside the information-source category names and descriptions, and prompt Claude to categorize each type of information source cited by the speaker in the argument, as well as whether the source was commonly observable to the Committee.

```
"""
You are labeling the information source for one FOMC economic argument.

Use the numbered source list supplied in the prompt. In the `source` field, write the
 category name, not the number.

# Source label

Label each distinct information input behind the speaker's economic belief. Do not label the
  policy preference, economic variable, forecast label, or reasoning style as the source.

If the argument uses multiple distinct source categories, return one <source_entry> for each
  category.

Use `Other` when an identifiable information source is present but none of the listed
 substantive categories fits. Use `Unclear` only when no information source is identifiable.

A speaker's own forecast, model, or judgment is not a source by itself. If the underlying
 input is not identifiable, use `Unclear`.

# Boundary rules

Return a separate source only when the quotation uses that source as direct evidence. Do not
  add background inputs or assumptions merely embedded inside a forecast, model, or staff
 document. Add the underlying input only when the speaker separately cites it as evidence
 outside the forecast or document.

Use Board staff forecast and analysis for Greenbook, Tealbook, Bluebook, Board staff
 projections, and Board staff models. Use Regional Fed forecast and analysis for Reserve
 Bank or District staff forecasts, models, or research.

Use FOMC participant projections and SEP only for explicit SEP, central tendency, range, dot
  plot, or participant-submission references. Generic statements that "participants expect"
 or "people project" are forecast descriptions, not sources by themselves.

Historical comparisons are not a separate source. Classify the underlying information input
 when it is identifiable, such as official statistics, staff analysis, external research, or
  another listed source. If the quotation only makes a broad historical analogy with no
 identifiable information input, use `Unclear`.

Use Business and regional contacts for human contact reports, anecdotes, directors, firms,
 bankers, or Beige Book-style reports. Use Other for identifiable non-contact data objects
 or source institutions that do not fit the list, including private or speaker-only datasets
 . A private source is Other, not Unclear, if the information object is identifiable.

# Common to the Committee

Set `common_to_committee` to `yes` if the source would have been observable by all Committee
  members at the meeting, including public data, market data, and shared FOMC or Board staff
  materials.

Set `common_to_committee` to `no` if the source appears private, local to a Reserve Bank or
 speaker, contact-based, supervisory, or otherwise not commonly observable by all Committee
 members.

Use `unclear` if the source is identifiable but common observability is not clear. For `
```

```
 Unclear‘, set ‘common_to_committee‘ to ‘unclear‘.

# Evidence

Evidence must be a short exact span from the quotation. If no source is identifiable, use ‘
 No source identified.‘

# Output

Return raw XML only. The first non-whitespace character must be ‘<‘.

<source_classification>
  <sources>
    <source_entry>
      <source>category name</source>
      <common_to_committee>yes|no|unclear</common_to_committee>
      <evidence>short exact evidence</evidence>
    </source_entry>
  </sources>
</source_classification>

Frozen source menu:
1. {category name}: {definition}
{additional categories}

Argument to classify:
<meeting_date>{YYYYMMDD}</meeting_date>
<speaker>{speaker}</speaker>
<description>{description}</description>
<quotation>{quotation}</quotation>
"""
```

IA.E.1.4 Prompts to Identify Policy Decisions Made

To identify the policy decisions made, we first compare the released statement for each meeting with the statement released after the preceding meeting. The prompt to identify these statement changes is below.

```
"""
You are comparing two FOMC statements.

Unit: one full sentence. Use only the sentence IDs shown in the input.

Rules:
1. Return content-bearing sentence additions, deletions, replacements, or revisions.
2. Include changes to policy actions, guidance, reaction-function language, economic
 assessments, implementation details, rationale, and tool-readiness language.
3. Pair current and previous sentences only when they are the same statement element: policy
  action, guidance, reaction function, economic assessment, implementation detail, rationale
 , or tool-readiness language.
4. If an old sentence disappeared and a new sentence covers different content, use separate
 deletion/addition rows rather than a forced match.
5. Do not return rows for pure renumbering, reordering, sentence splits/merges that preserve
  content, moved text, or trivial wording changes, including determiner/pronoun-only
 substitutions such as the/this/its/these.
6. If content is split or merged, record only the net substantive change: pair the best
 substantive successor, do not add rows for unchanged fragments or duplicated rationale, and
  do not reuse the same sentence ID in multiple rows.
7. For additions use ‘NA‘ for ‘id_before‘; for deletions use ‘NA‘ for ‘id_now‘.
8. Ignore web navigation, headers, footers, and release mechanics if any remain in the input
 .
9. Output XML only. The first character of your response must be ‘<‘; begin exactly with ‘<
 output>‘ and end exactly with ‘</output>‘. Return no markdown, no triple-backtick code
 fences, and no prose.
```

```
Provide your results in this exact XML structure:

<output>
  <statement_changes>
    <change>
      <id_now>S0007</id_now>
      <id_before>S0006</id_before>
    </change>
  </statement_changes>
</output>

Prompt version: v4_statement_changes_xml_unique_policy_guard

PREVIOUS STATEMENT:
<<<PREVIOUS
{previous_statement}
PREVIOUS>>>

CURRENT STATEMENT:
<<<CURRENT
{current_statement}
CURRENT>>>
"""
```

We then input the minutes from each meeting into the LLM and prompt it to extract the adopted policy language. The instructions differ between the directive era and the public-statement era. For meetings before February 4, 1994, we extract the adopted directive or policy record using the following prompt:

```
"""
You are an expert economic analyst specializing in Federal Reserve monetary policy records.
 Analyze FOMC minutes and extract the adopted directive or policy record.

Measurement object:
- Extract the adopted policy directive or policy record.
- The relevant policy language usually appears near the end of the minutes, often in "
 Committee Policy Action," "Policy Statement," "Directive," or similar sections.
- Use exact language from the minutes. Do not paraphrase, summarize, or add interpretation
 inside extracted fields.
- Extract complete text. Do not omit sentences, clauses, implementation instructions, or
 relevant vote-adopted language from the adopted statement/directive.

Document-structure rules:
- Output only `<policy_directive>`.
- In the pre-1994 era there is no regular released public statement comparable to the post
 -1994 statement series. Do not extract current-condition summaries, objective preambles, or
  narrative rationale as a separate statement field.
- If the directive contains both background and operative instructions, include the complete
  adopted directive text inside `<policy_directive>`.
- Do not include rejected alternatives, member preferences, discussion summaries, economic
 rationale, or other sections that were not adopted policy language.

Before the final XML, you may include a `<systematic_extraction>` block that verifies where
 the adopted language appears and checks verbatim completeness. The final parseable output
 must be:

<adopted_policy>
<policy_directive>Full verbatim directive or policy record text</policy_directive>
</adopted_policy>

Use exact text from the minutes without modification.

Meeting date: {YYYYMMDD}
Extraction scope: directive_record
```

```
Here are the FOMC minutes to analyze:

<fomc_minutes>
{minutes}
</fomc_minutes>

Extract the adopted policy language for the requested scope and return the XML output.
"""
```

For meetings beginning February 4, 1994, we instead extract the adopted directive or implementation record using the following prompt:

```
"""
You are an expert economic analyst specializing in Federal Reserve monetary policy
 implementation records. Analyze FOMC minutes and extract only the adopted directive or
 implementation language.

Measurement object:
- Extract the adopted Desk directive and implementation instructions.
- Include open market operation instructions, federal funds target/range implementation
 language, administered-rate implementation if part of the adopted policy record, balance-
 sheet purchase/sale/reinvestment/rollover/runoff instructions, facility/liquidity
 implementation language, and formal operating instructions.
- Do not extract the released public statement. In the public-statement era, released
 statement text comes from statement files and statement-change rows downstream.
- Do not include discussion summaries, rejected alternatives, member preferences, economic
 rationale, or communication-debate narrative unless the exact text is part of the adopted
 directive/implementation language.
- Use exact language from the minutes. Do not paraphrase, summarize, or add interpretation
 inside extracted fields.
- Extract complete directive/implementation text. Do not omit sentences, clauses,
 implementation instructions, or relevant vote-adopted language.

Document-structure rules:
- The relevant language usually appears near the end of the minutes in "Committee Policy
 Action," "Domestic Policy Directive," "Implementation Note," or similar sections.
- Output only `<policy_directive>`.
- Do not output a separate statement field even when the minutes quote the released
 statement.
- If no adopted directive or implementation language is present, return an empty `<
 policy_directive></policy_directive>` rather than extracting the public statement.

Before the final XML, you may include a `<systematic_extraction>` block that verifies where
 the adopted directive/implementation language appears and checks verbatim completeness. The
  final parseable output must be:

<adopted_policy>
<policy_directive>Full verbatim adopted directive or implementation text</policy_directive>
</adopted_policy>

Use exact text from the minutes without modification.

Meeting date: {YYYYMMDD}
Extraction scope: implementation_directive_only

Here are the FOMC minutes to analyze:

<fomc_minutes>
{minutes}
</fomc_minutes>

Extract the adopted policy language for the requested scope and return the XML output.
"""
```

We next input the Bluebook or Tealbook B for each meeting into the LLM and prompt it to extract the policy alternatives presented for the Committee's consideration. For all

meetings except February 9, 1983 and July 1, 1998, we use the following prompt.

```
"""
You are an expert economic analyst specializing in Federal Reserve communications. Analyze
 the FOMC Bluebook or Tealbook-B text and extract all alternative policy statements and
 proposed directives.

Task:
- Find every policy alternative presented for Committee consideration, usually labeled
 Alternative A, Alternative B, Alternative C, etc.
- For each alternative, extract the policy statement if one is present and distinct, and the
  proposed directive or directive-like policy language.
- Extract all alternatives, even if there are more than three.
- Use exact document language. Do not paraphrase, summarize, or add interpretation inside
 extracted fields.
- Clearly distinguish policy statements from proposed directives when they are separate.
- If the document gives alternative-specific public statements but no alternative-specific
 directive, omit `<proposed_directive>` for those alternatives; do not copy a common
 implementation note into each alternative.
- If the document presents alternatives in a nonstandard narrative form, do not output
 IGNORE. Set `<non_standard>true</non_standard>` and represent the major policy alternatives
  using the document's own language.
- Output IGNORE only if the document truly contains no alternative policy approaches or
 policy differences.

Before the final XML, you may include a `<systematic_analysis>` block that verifies labels,
 boundaries, and completeness. The final parseable output must be an `<
 alternative_statements>` block with this structure:

<alternative_statements>
<non_standard>false</non_standard>
<alternative>
    <label>Alternative A</label>
    <description>Brief document-based description or context</description>
    <policy_statement>Full policy statement text, if present and distinct</policy_statement>
    <proposed_directive>Full proposed directive or directive-like policy language</
     proposed_directive>
</alternative>
</alternative_statements>

Omit `<policy_statement>` when there is no distinct policy statement. Omit `<
 proposed_directive>` when there is no alternative-specific directive. Provide only exact
 text from the document inside policy/directive fields.

Meeting date: {YYYYMMDD}

Here is the Bluebook or Tealbook-B text to analyze:

<bluebook_text>
{bluebook_text}
</bluebook_text>

Extract the policy alternatives and return the XML output.
"""
```

For the February 9, 1983 and July 1, 1998 meetings, we use the following prompt.

```
"""
You are an expert economic analyst specializing in Federal Reserve communications. Analyze
 the FOMC Bluebook text and extract all alternative policy approaches and their technical
 specifications.

In this document, alternatives may be presented mainly through technical specifications
 rather than formal directive statements. Extract every alternative and all quantitative
 parameters, rates, targets, timing, operational details, balance-sheet details, and forward
 -guidance language that define it.
```

```
Requirements:
- Extract every alternative present, even if there are more than three.
- Use exact document language. Do not paraphrase extracted policy text.
- Capture all technical parameters completely.
- If alternatives are discussed narratively or through embedded parameter choices, set `<
 non_standard>true</non_standard>` and represent the alternatives using the document's own
 language.
- Output IGNORE only if the document truly contains no alternative policy approaches or
 policy differences.

Before the final XML, you may include a `<systematic_analysis>` block that verifies labels,
 boundaries, and completeness. The final parseable output must be an `<
 alternative_statements>` block with this structure:

<alternative_statements>
<non_standard>true</non_standard>
<alternative>
    <label>Alternative A</label>
    <description>Brief document-based description or context</description>
    <policy_statement>Policy characterization, if present and distinct</policy_statement>
    <proposed_directive>All technical specifications and directive-like language defining
     this alternative</proposed_directive>
</alternative>
</alternative_statements>

Omit `<policy_statement>` when there is no distinct policy statement. Provide only exact
 text from the document inside policy/directive fields.

Meeting date: {YYYYMMDD}

Here is the Bluebook or Tealbook-B text to analyze:

<bluebook_text>
{bluebook_text}
</bluebook_text>

Extract the policy alternatives and return the XML output.
"""
```

We then input the adopted policy language, the released statement and its changes when available, and the policy alternatives into the LLM and prompt it to identify each adopted monetary policy decision. The instructions differ between the directive era and the public-statement era. For meetings before February 4, 1994, the prompt is below.

```
"""
You are constructing the directive-era FOMC monetary-policy output dataset.

Use only the evidence in the meeting packet. Statement-change rows are a search aid; do not
 let extracted change text override the adopted directive or policy record.
Policy alternatives are proposed options, not adopted-policy evidence; use them only to
 contextualize rows whose adoption is evidenced by the adopted directive or policy record.
 When an adopted numeric target/range exactly matches an alternative, use that alternative's
  description to contextualize the score. Do not treat the merely closest alternative as
 adopted evidence; if adopted values are above, below, or between alternatives, describe
 that relationship in the justification. Do not use alternatives to create implementation-
 only rows unless the alternatives materially vary on that item and the adopted policy
 record independently supports the row.

Measurement object:
- A row is a current-meeting monetary-policy output adopted in the directive or policy
 record.
- Include the main funds-rate range/constraint when operative, formal funds-rate
 consultation/tolerance ranges, reserve-pressure/restraint or borrowed-reserve/discount-
 window-borrowing stance, discount-rate actions only when adopted/approved as current-
 meeting policy, formal M1/M1-A/M1-B/M2/M3/reserve/credit/debt targets or ranges, operating
```

```
 instructions, technical measurement/implementation provisos that condition reserve paths,
 and directive bias/flexibility.
- Exclude routine public communication: before February 1994 there is no regular public
 statement. Exclude directive objective preambles, current-condition or rationale background
 , prior Board discount-rate changes recited as background, expected aggregate paths,
 rejected alternatives, member preferences, SOMA/FX ratifications, stale recitations,
 routine administration, and settlement plumbing.

Types:
- `rate decision`: direct funds-rate range/constraint, formal funds-rate consultation/
 tolerance range, reserve-pressure/restraint stance, borrowed-reserve or discount-window-
 borrowing stance, or discount-rate action/hold when separately adopted as a current-meeting
  policy action. Do not code discount-rate changes that appear only as prior Board actions
 in background text, such as "On [earlier date] the Board announced/approved..."
- `other`: aggregate/credit/debt targets or ranges, operating instructions, directive bias/
 flexibility, and other adopted monetary-policy outputs.
- `communication`: use only if the packet separately evidences outward release or a distinct
  adopted communication object. Otherwise all rows should have `communication_subtype: "none
 "`.

Granularity:
- Split rows when instruments or horizons differ. When separate numeric M1/M1-A/M1-B/M2/M3
 ranges are adopted, return one row per aggregate and horizon; never output one bundled
 aggregate-ranges row. Keep bank-credit/debt ranges separate.
- Do not add a row for the whole directive sentence when separate rate, aggregate, or
 operating-instruction rows capture its operative pieces.
- M1/M2/M3 or reserve-aggregate targets are `other`, not `rate decision`, even if the
 directive implements them through reserve aggregates.
- Merge technical measurement provisos, such as NOW-account distortion language, into the
 related short-run aggregate or operating row unless the proviso is independently adopted
 and could move separately.
- Keep formal annual ranges consistently when the adopted directive presents them as current
  policy record; drop only historical background that is not maintained, reaffirmed, changed
 , or otherwise operative at the meeting.

Scoring:
- Use one row-level `score` from -3 to +3. Score the directional policy signal in the row,
 not the meeting overall and not a realized market surprise. Sign convention: negative means
  more accommodative/easier/supportive; positive means more restrictive/tighter/less
 supportive.
- Restrictive signals: firming, increased reserve pressure/restraint, higher borrowed
 reserves when used as restraint, restrained aggregate growth, inflation pressure, or
 readiness to withdraw support. Accommodative signals: easing, reduced pressure/restraint,
 lower borrowing assumption, tolerance for faster money growth, weak demand, low inflation,
 or readiness to provide support.
- Reserve-targeting polarity check: in 1979-1982 especially, slower target growth for
 reserves or money, reserve-aggregate deceleration, higher funds-rate/discount-rate/
 borrowing bands, reserve-requirement increases, or explicit restraint are restrictive and
 must score positive. Faster/tolerated money growth, reserve expansion, lower funds-rate/
 discount-rate/borrowing bands, reserve-requirement relief, or liquidity support are
 accommodative and must score negative.
- For formal funds-rate consultation bands, use exact numeric matches to alternatives for
 directional context: a matched alternative describing lower or declining rates is
 accommodative, while one describing higher rates, reserve restraint, or firmer borrowing
 assumptions is restrictive. Direction is relative to recent/intermeeting rates; do not call
  a band neutral merely because it straddles the current discount rate if the matched
 alternative says rates would likely decline from recent levels. If no directional context
 is available, an unchanged formal band can score 0.
- If your justification says "tightening", "restraint", "firming", "restrictive", or "anti-
 inflationary", the score should not be negative. If it says "easing", "accommodative", "
 support", "liquidity provision", or "reserve expansion", the score should not be positive
 unless the row contains a clearly offsetting stronger signal.
- Score 0 for neutral, procedural, stale, balanced, unchanged formal consultation/tolerance
 bands, or monitoring rows. Do not neutralize a funds-rate consultation/tolerance band when
 it is newly raised, newly lowered, unusually high, or paired with explicit reserve
 restraint or declining-rate language; score that directional signal. Use +/-1 for routine/
 moderate directive signals, +/-2 for strong actions or forceful support/restraint, and +/-3
  for maximum row-level policy signals: crisis-scale actions, ELB-style maximum
```

```
 accommodation, or extraordinary commitments.

Evidence discipline:
- Every row needs statement-change ids, a short exact quote, or both. Do not invent endpoint
  target levels. Keep justifications tied to the row evidence.
- Final self-check: a funds-rate consultation/tolerance row whose matched alternative says
 short-term rates would likely decline from recent levels must have a negative score, not 0.
  A funds-rate consultation/tolerance row whose evidence says restraint, higher rates, or
 firming must have a positive score, not 0 or negative.
- Output XML only. The first non-whitespace character must be `<`. Return no markdown, code
 fences, or prose.

Meeting date: {YYYYMMDD}
Prompt version: v4_regime_split_v93_lean_decision_schema

Attention notes:
{meeting-specific instructions, where applicable}

Return XML only, following this structure:
<adopted_policies ymd="{YYYYMMDD}">
  <adopted_policy>
    <description>concise adopted decision or communication description</description>
    <type>communication|rate decision|other</type>
    <communication_subtype>economic_assessment|forward_guidance|reaction_function|none</
     communication_subtype>
    <score>SCORE_INTEGER</score>
    <justification>1-2 sentences tied to row evidence</justification>
    <source_statement_change_ids>statement_change_id; statement_change_id</
     source_statement_change_ids>
    <adopted_policy_evidence>short exact quote</adopted_policy_evidence>
  </adopted_policy>
</adopted_policies>

Output discipline: first non-whitespace character must be `<`; return no markdown, code
 fences, or prose. Keep every XML field clean and complete. Replace `SCORE_INTEGER` with
 exactly one integer from -3 to 3. Never leave `communication_subtype` blank; use `none` for
  every non-communication row. Never leave `adopted_policy_evidence` blank; use the shortest
  exact quote that supports the row.

Minutes directive / adopted policy text:
{adopted_policy_text}

Released statement text from FOMCtidy:
{released_statement}

Statement changes:
{statement_changes}

Policy alternatives:
{policy_alternatives}
"""
```

For meetings beginning February 4, 1994, the prompt to identify the adopted policy decisions is below.

```
"""
You are constructing the public-statement-era FOMC monetary-policy output dataset.

Use only the evidence in the meeting packet. Statement-change rows are a search aid; do not
 let extracted change text override the adopted/released policy evidence.
Use the released statement text from FOMCtidy as the source for public communication. Use
 minutes/adopted-policy text only for directive or implementation evidence, not as a
 duplicate source for communication rows.
Policy alternatives are proposed options, not adopted/released evidence; use them only to
 contextualize rows whose adoption is evidenced by released statement text, statement
 changes, or adopted-policy text. Do not use alternatives to create implementation-only rows
```

```
  unless the alternatives materially vary on that item and adopted/released evidence
 independently supports the row.

Measurement object:
- A row is a current-meeting monetary-policy output that was adopted, released, or otherwise
  made part of the policy record.
- Include rate actions and holds, coordinated or separately evidenced administered-rate
 actions/holds, balance-sheet/facility/liquidity actions, formal aggregate/debt targets or
 ranges, operating instructions, and substantive adopted/released policy communication.
 During crisis periods, include named swap lines, auction facilities, liquidity facilities,
 TALF/PDCF/TSLF/AMLF/CPFF/MMIFF-type programs, and material facility extensions/expansions
 when they are adopted or announced as current policy; do not drop them merely because they
 are not the main statement sentence.
- Exclude rejected alternatives, member preferences, routine administration, SOMA/FX
 ratifications, settlement plumbing, stale recitations, trivial wording polish, and evidence
 -free inferences.
- For 1994-1999 transition meetings, code the action itself, but do not emit a standalone
 communication row for a merely mechanical public release. Keep immediate-announcement or
 released-statement communication when exact text gives a policy-relevant purpose,
 substantive rationale, economic/risk assessment, reaction-function language, or forward
 guidance. Do not code internal minutes/directive current-condition reviews as public
 economic-assessment communication unless the evidence supplies exact released statement or
 announcement text. Keep adopted directive operating choices: directive bias/asymmetric/
 symmetric intermeeting flexibility is `other`, separate from the rate action.

Types:
- `rate decision`: main funds-rate target/range action or hold; coordinated administered-
 rate implementation; or separately announced/evidenced administered-rate action or hold.
- `other`: adopted monetary-policy output that is not a rate decision or communication.
- `communication`: substantive outward-facing/adopted statement or announcement message. Use
  one subtype: `economic_assessment`, `forward_guidance`, or `reaction_function`. Non-
 communication rows use `communication_subtype: "none"`.

Communication subtypes:
- `economic_assessment`: one message-level row for current economic and financial conditions
 , outlook, inflation, labor markets, growth risks, inflation risks, and balance-of-risks
 language. Balance-of-risks language is not `reaction_function` unless it states how policy
 will react; do not output it as a separate row when it can be merged into the meeting's
 economic assessment. Monitoring/evaluation or reaction-function commitments belong in `
 reaction_function`.
- `forward_guidance`: expected policy path, timing, calendar, threshold, state-contingent
 continuation/exit commitment, or specific instrument scale/duration/use commitment,
 including funds-rate and asset-purchase guidance. Specific readiness to expand the quantity
 /duration of named purchases, or to purchase a named instrument if conditions warrant, is
 forward guidance.
- `reaction_function`: data-dependence, monitoring, risk-management, preparedness/readiness/
 tools, policy-focus, current-action rationale, or other language explaining how the
 Committee will react to incoming information or risks, when it does not itself specify an
 instrument path, scale, duration, timing, or threshold.
- Bare language that the Committee will assess, review, or monitor the appropriate pace/size
 /use of an instrument at future meetings is `reaction_function`, not `forward_guidance`,
 unless it states an expected direction, threshold, duration, or commitment. Never create a
 `forward_guidance` row whose only future-looking content is "assess the appropriate pace"
 or continued monitoring.

Granularity:
- Split rows only when instruments, horizons, or communication functions differ. Consolidate
  sentence-level fragments into message-level rows.
- Hard cap communication granularity: at most one `economic_assessment` and at most one `
 reaction_function` row per meeting. This is a merge rule, not an exclusion rule: merged
 rows must name all material clauses they subsume unless a clause is distinct instrument-
 path `forward_guidance`.
- Merge bare monitoring or assessment clauses, including language that the Committee will "
 assess the appropriate pace" at future meetings, into the one `reaction_function` row; do
 not emit them as separate `forward_guidance` rows, even when the instrument is asset
 purchases. When this language is only reaction-function evidence, do not describe it as
 forward guidance.
- When balance-of-risks language is merged into `economic_assessment`, the description
```

```
 should explicitly mention risks or balance of risks.
- Do not hide forward guidance inside a rate or action row. A rate/action row should state
 the current action or hold; threshold, path, timing, or state-contingent continuation/exit
 language in the same sentence belongs in the separate merged `forward_guidance` row.
- Do not merge federal-funds-rate guidance with balance-sheet, asset-purchase, or facility
 guidance; they are different instruments and should be separate `forward_guidance` rows
 when both are present.
- Do not hide specific instrument-readiness guidance inside a general all-tools reaction-
 function row. General "employ tools" readiness is `reaction_function`; named-instrument
 quantity, duration, purchase, continuation, or exit guidance is `forward_guidance`.
- For one instrument, merge same-direction threshold, timing, expected-path, below-normal,
 and data-dependence clauses into one `forward_guidance` row. Split same-instrument guidance
  only when clauses send separable opposite directional signals, such as likely further
 purchase reductions versus not-on-a-preset-course contingency language.
- Do not duplicate the same object as both action and communication. Balance-sheet
 transmission or stock-of-holdings language tied to an active or maintained balance-sheet
 action belongs in that action row's justification unless it states a separable future
 reaction rule. A maintained reinvestment/rollover policy remains an `other` balance-sheet
 action even when it includes an "until normalization" timing condition.
- Open-ended current purchase actions phrased as "will continue to purchase ... in the
 amounts needed" are `other` action rows; do not add a duplicate `forward_guidance` row for
 the same purchase commitment unless the text also gives a separable future path, threshold,
  exit, or reaction rule.
- Do not bundle a separate crisis expansion of eligible purchased assets, such as adding
 agency CMBS, into the main purchase row when it appears as a separate sentence or statement
 -change row.
- Do not bundle formal M1/M2/M3 ranges when separate numeric ranges are given. Do not bundle
  bank-credit/domestic-nonfinancial-debt ranges into aggregate rows. Exclude expected
 aggregate paths unless they are formal numeric current-meeting target/tolerance ranges.
- If ambiguous but economically meaningful, keep the best defensible row and note the
 ambiguity in the justification.

Scoring:
- Use one row-level `score` from -3 to +3. It is not a meeting-level stance, not a change
 score, and not a realized surprise. It is the directional policy signal in the announcement
  that could plausibly matter for asset-price monetary-policy surprises. Sign convention:
 negative means more accommodative/easier/supportive; positive means more restrictive/
 tighter/less supportive.
- Restrictive signals: hikes, firming, increased reserve pressure/restraint, tapers/
 reductions/runoff, inflation pressure, overheating, upside inflation risk, or readiness to
 withdraw support. Accommodative signals: cuts, easing, purchases/reinvestment/facilities/
 liquidity support, weak growth, downside risks, low/subdued inflation, easier expected
 policy path, or readiness to provide support.
- Score 0 for neutral, procedural, stale, balanced, or two-sided signals. Use +/-1 for
 routine/moderate signals, +/-2 for strong signals, and +/-3 for maximum row-level policy
 signals. Do not make the tails unreachable: a `-3` is appropriate for ELB-style maximum
 accommodation, crisis-scale LSAP or market-support actions, or extraordinary whatever-it-
 takes commitments; a `+3` is appropriate for comparably forceful maximum restraint.
- A literal 0 to 1/4 percent or 0 to 0.25 percent federal-funds target/range normally scores
  -3 as maximum accommodative rate stance. A 1-1/4 percent hold is not a zero-bound hold and
  normally scores 0 unless its own row evidence gives a directional signal.
- Repeated ZLB/ELB holds: score -3 for maintaining an explicit 0 to 1/4 percent or 0 to 0.25
  percent target/range, but do not use the rate-hold row to absorb separate guidance,
 purchases, or facility decisions. If a non-ELB hold simply maintains the prior target/range
  without new directional content, score 0.
- Ordinary 25 basis point moves usually score +/-1, but liftoff from the ELB/ZLB after an
 extended zero-bound period, or a coordinated target-range and administered-rate package
 that begins normalization, normally scores +2.
- Large new or expanded crisis-era asset purchases and emergency market-support programs can
  score -3. Bounded major facilities, crisis-support extensions, and strong low-rate
 guidance normally score -2 unless the row evidence is clearly maximum-scale or open-ended.
 Continuations, reinvestments, or supportive maintenance rows without expansion normally
 score -1 or -2 by force. Material named crisis facilities and swap-line decisions should
 still appear as separate `other` rows even when their directional score is only -1 or -2.
- Preparedness/readiness/tools language tied to stronger recovery, support, or additional
 accommodation scores by direction and force; two-sided flexibility can score 0. Forceful
 crisis-era all-tools language with balance-sheet support can be -3 when it communicates a
 whatever-it-takes reaction function. Named-instrument readiness guidance is scored as
```

```
 forward guidance by its direction and force.
- If the row justification says support, liquidity provision, accommodation, easing,
 purchases, or facilities, the score should not be positive unless the same row contains a
 stronger withdrawal/tightening signal.
- Facility exits score by announcement signal: +1 for material withdrawal, 0 for procedural/
 stale/inactive expirations, and -1/-2 for crisis-support extensions. Mixed liftoff
 rationale plus "stance remains accommodative" usually scores 0 unless one direction
 dominates.

Evidence discipline:
- Every row needs statement-change ids, a short exact quote, or both. `
 adopted_policy_evidence` must be a short exact quote from adopted/released policy evidence
 in this packet; exact `phrase_now` or `smallest_change_now` text may be quoted when
 statement-change evidence supplies released text.
- Do not invent endpoint target levels, neutral-rate claims, concrete facility extensions,
 or overall-stance language. Keep justifications tied to the row evidence. If a statement-
 change row and its exact text disagree, code the exact text. A concrete action row needs
 exact text for that action; if the exact text only says assess/evaluate/consider/stand
 ready, code the communication as `reaction_function`, not as an adopted facility extension
 or implementation. If named crisis facilities or swap lines appear only in alternatives or
 historical/background text, do not invent adopted rows; if they appear in adopted/released
 evidence, include them. If removed language appears in adopted policy evidence, the adopted
  text is binding.
- Final self-check: if a candidate `forward_guidance` row's evidence only says the Committee
  will "assess the appropriate pace" or monitor conditions, delete that row and merge the
 clause into the meeting's `reaction_function` row. Such a row is invalid even if the
 instrument is securities purchases.
- Output XML only. The first non-whitespace character must be `<`. Return no markdown, code
 fences, or prose.

Meeting date: {YYYYMMDD}
Prompt version: v4_regime_split_v93_lean_decision_schema

Attention notes:
{meeting-specific instructions, where applicable}

Return XML only, following this structure:
<adopted_policies ymd="{YYYYMMDD}">
  <adopted_policy>
    <description>concise adopted decision or communication description</description>
    <type>communication|rate decision|other</type>
    <communication_subtype>economic_assessment|forward_guidance|reaction_function|none</
     communication_subtype>
    <score>SCORE_INTEGER</score>
    <justification>1-2 sentences tied to row evidence</justification>
    <source_statement_change_ids>statement_change_id; statement_change_id</
     source_statement_change_ids>
    <adopted_policy_evidence>short exact quote</adopted_policy_evidence>
  </adopted_policy>
</adopted_policies>

Output discipline: first non-whitespace character must be `<`; return no markdown, code
 fences, or prose. Keep every XML field clean and complete. Replace `SCORE_INTEGER` with
 exactly one integer from -3 to 3. Never leave `communication_subtype` blank; use `none` for
  every non-communication row. Never leave `adopted_policy_evidence` blank; use the shortest
  exact quote that supports the row.

Minutes directive / adopted policy text:
{adopted_policy_text}

Released statement text from FOMCtidy:
{released_statement}

Statement changes:
{statement_changes}

Policy alternatives:
{policy_alternatives}
```

```
"""
```

Finally, for each adopted decision, we prompt the LLM to identify the same-meeting policy alternatives that contain that decision. The prompt is below.

```
"""
You are linking adopted FOMC policy decisions to same-meeting policy alternatives. Use only
 the meeting packet. Do not create, delete, merge, split, reclassify, or rescore decisions.

Task:
- For each adopted decision, identify the alternatives that explicitly contain that same
 decision.
- This is a choice-set link, not a claim about which alternative was expected, preferred, or
  adopted.
- Do not identify a default, baseline, expected, central, or status-quo alternative.
- `choice_set_alternatives` lists the same-meeting policy alternative records supplied for
 evaluating the meeting, even when none match the decision. It is not a claim that each
 record is an option for that decision's instrument. Use `None` only when no same-meeting
 policy alternatives are provided.

Matching criteria:
- Match only when the alternative contains the same policy object: instrument/action,
 direction, amount or range, timing/horizon, and conditionality.
- Use the adopted-policy evidence as binding. If it contains a distinctive phrase or
 condition that is absent from the short description, the alternative still must match it.
- Numeric matches must be exact. Equivalent notation is the same value: 6-1/2 = 6.5 = 6 1/2.
  For ranges, both endpoints and the time period must match. A point estimate is not a match
  for a range.
- For multi-part decisions, every component must match. If one component differs or is
 missing, including a modifier or sub-clause, the alternative is not a match.
- Extra separable text does not block a match if all distinctive components of the adopted
 decision are present. Extra or missing qualifiers inside the action, trigger, timing,
 horizon, amount, or commitment do block a match.
- Evaluate each alternative individually. Do not combine alternatives to reconstruct an
 adopted decision.
- Approximate words do not block a match when the number itself is the same: "about 3
 percent" matches "3 percent." They do not make nearby numbers match.
- For directive language, direction, intensity, and modal verbs matter. Do not treat "
 slightly" and "somewhat" as exact synonyms, or "would" and "might" as exact synonyms.
- Blanks and unselected options in directive templates are not evidence. Fill blanks only
 from exact values printed in the same named alternative record, never from another
 alternative or from the adopted decision.
- Unresolved slash or parenthetical choices such as "would/might" or "might (WOULD)" are
 templates, not selected wording. A template containing both possible words is not an exact
 match to either selected word.
- Do not match against memo/reference/prior ranges included only for comparison.
- Do not match narrative analysis of rationale, consequences, or expected market reactions;
 match only proposed decision text, ranges, or actions.
- For communication decisions, match only when the alternative contains the distinctive
 adopted communication content. Shared background paragraphs, generic macro assessments,
 bracketed risk templates, and broad topical similarity are not enough. If the adopted
 decision combines multiple communication elements, all distinctive elements must match.
- For released public-statement decisions, match against proposed announcement text in the
 alternative's `policy_statement`, not `proposed_directive`. If `policy_statement` is blank
 or is only narrative/analytical rationale rather than proposed announcement text, use `None
 `.
- Communication modality and commitment strength matter: different uncertainty, condition,
 or commitment language is not an exact match even when the broad policy theme is similar.
- For forward guidance and reaction-function decisions, the trigger/condition and promised
 action must both match. Do not match variants with different triggers, horizons, action
 sets, or program designs.
- Event anchors are part of forward-guidance timing: "after the economic recovery
 strengthens" is not the same as "after the conclusion of this asset-purchase program."

Pre-1994 Bluebook notes:
- Values in an alternative description or table can supply matchable evidence for that same
 alternative.
```

```
- When a meeting has A/B/C policy alternatives plus Directive Variant wording records, use A
 /B/C only for matching numeric operating paths, consultation ranges, and reserve-stance
 choices, even if directive variants mention blanks or cross-reference A/B/C. Use Directive
 Variants for matching only decisions about directive wording.
- Standing, reaffirmed, or tentative ranges should remain unmatched unless the alternatives
 actually print those ranges as part of the choice set.
- Communication language embedded only as a blank or bracketed template is not a match
 unless the same alternative prints the adopted choices and wording.
- Do not include prior-directive or reference records in `choice_set_alternatives` unless
 the adopted decision is itself about that prior/reference language.

Output:
- Return XML only. The first non-whitespace character must be `<`.
- Fill the XML skeleton below. Do not add, remove, reorder, rename, repeat, or restart any
 `<decision>` element. Each `id` must appear exactly once.
- Fill `justification` before `matched_alternatives`.
- If alternatives match, explain briefly why the listed alternatives match.
- If none match, explain briefly why no alternative matches and use `None` for `
 matched_alternatives`.
- Do not output any default, baseline, expected, central, or status-quo field.

Example format:
<alternative_origin ymd="YYYYMMDD">
  <decision id="d001">
    <choice_set_alternatives>Alternative A; Alternative B; Alternative C</
     choice_set_alternatives>
    <justification>Concise explanation.</justification>
    <matched_alternatives>Alternative A; Alternative B</matched_alternatives>
  </decision>
</alternative_origin>

Meeting date: {YYYYMMDD}

Adopted decisions:
{decisions}

Policy alternatives:
{policy_alternatives}

Output XML skeleton to complete:
<alternative_origin ymd="{YYYYMMDD}">
  <decision id="{decision_id}">
    <choice_set_alternatives></choice_set_alternatives>
    <justification></justification>
    <matched_alternatives></matched_alternatives>
  </decision>
  {one decision block per adopted decision}
</alternative_origin>

Return the completed XML skeleton only.
"""
```

IA.E.1.5 Prompts to Identify Alignment and Adoption Contribution

With the policy decisions identified, we next run prompts to identify each member's alignment with each of the policy decisions made and their contribution to the adoption of those decisions.

The prompt to identify a member's alignment with decisions adopted at the current meeting is given below.

```
"""
You are scoring one FOMC speaker's position alignment with each adopted policy decision
 listed in `policy_decisions`.
```

```
For each decision, score how strongly the focus speaker's current stated position supports
 or opposes that adopted decision. Do not score whether the speaker shaped the outcome, and
 do not score votes.

Default to `0`. Use a nonzero score only when the focus speaker's own words show support for
  or opposition to that decision.

Return raw XML only. The first non-whitespace character must be `<`. Return one <decision>
 block per decision listed in `policy_decisions`, carrying its `id`.

# Evidence

Use only current-position evidence from the focus speaker:

- direct: explicit statements about this decision's action, component, number, wording,
 instrument, or communication content;
- alternative: explicit support for or opposition to a same-meeting alternative, statement,
 directive, recommendation, or package that contains or selects this decision;
- stance: a clear view on a policy stance or direction when the adopted decision clearly
 embodies that stance.

Prioritize direct evidence over alternative evidence, and alternative evidence over stance
 evidence. Component-specific evidence controls broader package evidence.

Ignore biography, ideology, silence, votes, chair status, procedural consensus, clarifying
 questions, retrospective narration, and willingness to join the result unless the same
 passage states a current policy position. If the usable evidence is only a yes vote,
 procedural acceptance, or "I can live with it," score `0`.

# Alternatives and Packages

Alternative, statement, directive, recommendation, or package support can establish
 alignment, but only when the quoted same-meeting policy choice contains or selects the
 adopted decision. Use the decision description, adopted-policy evidence, and alternatives
 in the packet as authoritative for the adopted number, range, wording, instrument, and
 content.

If the focus speaker clearly supports a policy choice that contains the decision, score
 support for that decision even if the speaker does not repeat the decision separately. If
 the focus speaker clearly opposes a policy choice that contains the decision, score
 opposition for that decision. If the mapping is approximate, unclear, or based only on a
 nearby compromise or directionally similar value, score `0`.

Do not infer support for a component merely because it is adjacent to a supported component
 in the same package. If the speaker names or describes a target action, paragraph, phrase,
 instrument, rate, amount, program, rationale, or guidance, score that component directly
 and do not let broader package evidence override it. Component-specific support can be
 positive even if the speaker opposes the package overall on another dimension; component-
 specific opposition can be negative even if the speaker supports the package overall.

# Communication and Implementation Rows

For communication decisions, score the message itself, not just the underlying rate or asset
  action. Support for the rate decision, asset action, broad policy stance, or chair
 recommendation is not enough for a specific guidance, reaction-function, rationale, or
 economic-assessment row unless the speaker identifies that message, statement, paragraph,
 phrase, wording, or communication approach, or clearly supports the same statement or
 policy choice whose content selects that message. Clear support for the statement as
 written can support a statement-level decision or a communication row selected by that
 statement, but it should not automatically score every adjacent paragraph or unrelated
 component.

For routine implementation details, require component-specific evidence. Repo or reverse-
 repo operations, administered rates, settlement mechanics, reserve-management details,
 small operational deviations, and similar mechanics should not receive nonzero scores from
 generic package support alone. This implementation guardrail does not block whole-choice
 support for central substantive policy decisions such as the funds-rate target or path,
```

```
 asset-purchase stance, balance-sheet stance, liquidity facility, or swap action when the
 endorsed policy choice validly selects that decision.

# Scores

- `2`: clear support. Use `2` for explicit support for this decision, explicit support for a
  policy choice that validly contains this decision, or a clear stance that this decision
 embodies. Simple support such as "I support this decision," "I support Alternative B," or "
 I support the statement as written" can be `2` when it validly maps to the decision; no
 extended reasoning is required. Also use `2` when the speaker names or clearly describes
 the target sentence, paragraph, phrase, instrument, rate, path, program, condition, or
 component and gives substantive reasoning for supporting it. Minor reservations about
 secondary aspects stay `2`.
- `1`: qualified support. Use `1` when the speaker supports the decision but with
 significant reservations, supports only part of it, gives conditional or second-best
 support, or the evidence is indirect and only loosely connects to the decision.
- `0`: no clear position, ambiguous mapping, procedural acceptance only, or balanced support
  and opposition.
- `-1`: qualified opposition. Use `-1` when the speaker opposes the decision but with
 significant qualifications, opposes only part of it, states an unmet condition for support,
  or the evidence is indirect and only loosely connects to the decision.
- `-2`: clear opposition. Use `-2` for explicit opposition to this decision, explicit
 opposition to a policy choice that validly contains this decision, or a clear stance that
 this decision contradicts. Minor favorable comments about secondary aspects stay `-2`.

# Evidence Requirement

Every nonzero score requires an exact quote from the focus speaker stated at this meeting.
 The quote can be direct evidence about the decision, alternative evidence about a policy
 choice that contains the decision, or stance evidence that clearly maps to the decision. If
  there is no usable quote, score `0` and set `evidence` to `No explicit statement found.`

Before writing each score, check that its sign matches the quote and that the quote concerns
  this decision through direct, alternative, or stance evidence.

# Output

Return XML only. Escape XML special characters in evidence.

<position_alignment_analysis>
  <decisions>
    <decision>
      <id>Decision id from policy_decisions</id>
      <position_alignment>-2|-1|0|1|2</position_alignment>
      <evidence>Exact quote from the focus speaker, or No explicit statement found.</
       evidence>
    </decision>
  </decisions>
</position_alignment_analysis>

Meeting packet:

<policy_decisions>
{policy_decision_blocks}
</policy_decisions>

<alternatives>
{alternatives}
</alternatives>

meeting_date: {ymd}

Speaker packet:

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

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

Score only this focus speaker and return the XML block only.
"""
```

To identify a member's alignment with a decision before it was adopted, we input the later-adopted decision and all of the member's speeches from the earlier meeting into the LLM. The prompt is given below.

```
"""
You are scoring one FOMC speaker's position in this meeting on each policy decision listed
 in `policy_decisions`. Each listed decision was adopted at a later meeting.

Score whether the focus speaker explicitly supports or opposes the decision's substantive
 action and instrument at this meeting. The final adopted wording or calibration may not
 exist yet, so match the action and instrument using the decision description and adopted
 evidence provided for the decision.

Return raw XML only. The first non-whitespace character must be `<`. Return one <decision>
 block per decision listed in `policy_decisions`, carrying its `id`.

# What counts as a position on this decision

Match on the decision's substantive ACTION and INSTRUMENT. The instrument is whatever lever
 the decision pulls -- the asset-purchase program, the funds-rate target or path, the
 balance-sheet operation, a liquidity facility or swap arrangement, or the statement's
 directional signal/approach.

Score a position (support `+` or opposition `-`), symmetrically, when the speaker explicitly
  engages that lever and takes a side on the target action:

- advocating, or arguing against, the SAME instrument the decision operates -- including a
 stronger/larger/faster or weaker/smaller/slower version of the target action;
- backing or rejecting the predecessor/interim wording step the adopted guidance extends, or
  the policy approach the decision embodies, so long as the speaker states their own stance;
- score the sign relative to the later adopted action, not relative to the old policy object
 . Favoring the same direction as the adopted action is support; favoring the opposite
 direction on the same lever is opposition.

# What is `0`

- general economic outlook, forecast, or risk discussion;
- data-readiness or timing judgments without a stance on the decision's action;
- keeping the existing stance, wording, pace, or target range, or preferring no action yet,
 unless the speaker directly argues against the specific change the decision later makes;
- future-agenda or process framing with no stance of the speaker's own;
- a position only on a different instrument or different action, including a current-meeting
  move or alternative that does not describe the later target action;
- reluctant acceptance, procedural consensus, clarifying questions, biography, ideology, or
 routine language.

# Direction and form

Use the listed decision information to identify the relevant action and instrument. Match
 the same lever and direction, not only the final wording, amount, date, or target range. A
 stronger, larger, earlier, later, or faster version of the same target action is support
 unless the speaker rejects the adopted content. If the speaker opposes acting today but
 clearly endorses the same action later, score that later endorsement as support. The
 opposite direction on the same lever is opposition. A different lever is `0`.

For rate decisions, the lever is the funds-rate path: support for cuts, further cuts, or
 more aggressive easing supports a later lower-rate target; support for hikes, tightening,
 or avoiding further cuts opposes it. Holding or pausing is `0` unless the speaker also
 states a clear easing or tightening view.
```

```
For asset-purchase and balance-sheet decisions, match purchases versus sales, expansion
 versus contraction, and the broad asset class. Exact dollar amounts and program labels need
  not match, but simply keeping an option open, leaving something on the table, or saying "
 not yet" is not support by itself.

For communication decisions, score the message itself, not just the underlying rate or asset
  action. A quote must identify the guidance, announcement, rationale, or predecessor signal
 . A predecessor stance carries over only when it is about the same signal, same direction,
 or same communication approach; do not carry over a stance that is only about the old
 signal's opposite policy direction.

# Evidence

Use only explicit current-position evidence from the focus speaker: direct statements about
 the decision's action, instrument, wording, number, or communication content, or explicit
 agreement/disagreement with another speaker about the same policy object. A quote must come
  from the focus speaker; a question that only repeats proposed wording is not support.

Alternative labels are meeting-specific. A label, vote, or bare "I support Alternative X" by
  itself is not evidence for a later adopted decision; use the quote only if the same
 speaking turn directly describes the relevant action, wording, rate path, or purchase
 program.

# Scores

- `2`: explicit support for the action or a stronger version. Simple support such as "I
 support this decision" is `2`; no extended reasoning is required.
- `1`: qualified, hedged, conditional, reluctant, or second-best support.
- `0`: engages the object but is genuinely undecided/balanced, OR takes no explicit position
  on this decision's action.
- `-1`: qualified/hedged/conditional opposition, including "not now," conditions-not-met, or
  the opposite direction on the same instrument. If the speaker clearly says the same target
  action should be taken later and treats it as desirable, score qualified support;
 preserving the option is not enough.
- `-2`: explicit opposition to the action or the opposite direction on the same instrument.
 Simple opposition such as "I oppose this decision" is `-2`; no extended reasoning is
 required.

Apply the same evidentiary standard to positive and negative scores: a `2` requires no more
 explicit a statement than a `-2`.

Every nonzero score requires an exact quote stating the speaker's current position ON THIS
 DECISION'S ACTION. If there is none, score `0` with evidence `No explicit statement found.`
  Before scoring, check the sign matches the evidence and the quote is about the target
 action.

# Output

Return XML only. Escape XML special characters in evidence.

<position_alignment>
  <decisions>
    <decision>
      <id>Decision id from policy_decisions</id>
      <evidence>Exact quote from the focus speaker, or No explicit statement found.</
       evidence>
      <reasoning>One phrase: does the speaker take an explicit current position on THIS
       decision's specific action/instrument, or is it only outlook/timing/process or a
       different action?</reasoning>
      <alignment>-2|-1|0|1|2</alignment>
    </decision>
  </decisions>
</position_alignment>

Meeting packet:

<policy_decisions>
```

```
  <decision>
    <id>{decision_id}</id>
    <adoption_date>{decision_ymd}</adoption_date>
    <policy_direction>{policy_direction}</policy_direction>
    <type>{type}</type>
    <communication_subtype>{communication_subtype}</communication_subtype>
    <description>{description}</description>
    <adopted_policy_evidence>{adopted_policy_evidence}</adopted_policy_evidence>
  </decision>
</policy_decisions>

meeting_date: {meeting_ymd}   (this is {meetings_back} meeting(s) BEFORE adoption_date)

Speaker packet:

<speeches>
{speeches}
</speeches>

<focus_speaker>
{stablespeaker}
</focus_speaker>

Score only this focus speaker and return the XML block only.
"""
```

The prompt to identify a member's contribution to the adoption of each decision is given below.

```
"""
You are scoring one FOMC speaker's contribution to each adopted policy decision listed in `
 policy_decisions`.

For each decision, score whether the focus speaker supplied positive decision-specific
 support, reasoning, or content that appears in the adopted policy. Support alone can count
 only when it addresses this decision's action, number, wording, instrument, or
 communication content. Higher scores require reasoning or adopted content. Do not score
 votes or whether the speaker shaped the outcome.

Default to `0`. Use a positive score only when the focus speaker's own words supply a
 positive decision-specific contribution.

Return raw XML only. The first non-whitespace character must be `<`. Return one <decision>
 block per decision listed in `policy_decisions`, carrying its `id`.

# Positive evidence

A positive score requires a focus-speaker quote that directly concerns this adopted decision
 's component, number, wording, instrument, or communication content. A quote that only
 names or endorses an alternative, package, statement, or vote choice as a whole is not
 direct evidence for any listed component and must be `0`.

- proposing an adopted number, range, instrument, phrase, communication language, or
 refinement;
- explicitly supporting this adopted decision/component;
- giving reasoning, evidence, or defense for this adopted decision/component.

Explicit support for THIS adopted decision's substantive action is decision-specific
 contribution, including support for the specific rate move, hold, hike, asset-purchase
 action, or guidance the Committee adopted. This can count even if others also voted for it
 and even if the speaker would have preferred a somewhat different size or calibration. With
  the speaker's own reasoning it is a 2; without reasoning it is a 1. For communication or
 guidance decisions, support for the underlying rate action is not by itself a contribution
 to the guidance; the quote must address the adopted guidance or communication content.
 Whole-alternative agreement such as "I agree with Alternative B" does not become explicit
 support for a listed component unless the quote itself names or describes that component.
```

```
Do not count: support voiced ONLY at the alternative/package level with nothing said about
 this decision's action or rationale; a yes vote; lack of objection; reluctant acceptance;
 generic agreement with the broad direction of policy that is not tied to this decision's
 specific action; implementation questions; reasoning for a rejected component; or criticism
  of the adopted mechanism. Phrases like "not opposed" or "no objection" are lack of
 objection, not positive contribution, unless paired with affirmative support or reasoning
 for this component. If the speaker opposes the adopted decision overall but separately
 proposes wording or a subcomponent that was adopted, score that adopted trace.

# Alternatives

Endorsing an alternative, package, or statement as a whole is not enough by itself. This
 remains `0` even when that alternative is matched to the decision or contains the component
 . If the only quote is like "I support Alternative B," "I agree with Alternative B," or "
 people have focused on Alternative B, which I agree with," score `0` for each component
 unless the quote also addresses that component's action, number, wording, instrument, or
 communication content. This is especially important for components common to several
 alternatives. A positive score requires the quote itself to address this decision's
 component, number, wording, instrument, or communication content. For a `3`, the proposed
 number or wording must match the adopted content exactly or near-verbatim. A proposed
 midpoint or objective can qualify for a point-objective decision when the same number is
 the adopted target. Nearby, compromise, directionally similar, or rejected values do not
 qualify. Backing the adopted action while preferring a different calibration can still
 count as support at the `1` or `2` level. Use the decision description and adopted evidence
  in the packet as authoritative for the adopted number, range, wording, and content.

# Scores

- `3`: proposed a concrete item copied exactly or near-verbatim into the adopted decision or
  adopted communication. Use sparingly: similar themes, structural overlap, broad policy
 direction, partial wording matches, rejected wording, and approximate matches do not
 qualify.
- `2`: affirmatively supported this adopted decision/component and gave decision-specific
 reasoning.
- `1`: explicitly supported this adopted decision/component without added reasoning.
- `0`: no positive decision-specific contribution.

# Evidence requirement

Every positive score requires an exact quote from the focus speaker stated at this meeting.
 If there is no usable evidence, score `0` and set `evidence` to `No explicit statement
 found.`

Before writing the score, check it against the evidence: a `3` requires adopted content
 traceable to the quote; a `2` requires reasoning in the quote, not just support; package-
 only, alternative-only, vote-only, or no-objection-only evidence must be `0`. If the same
 evidence quote could be copied unchanged to several different decisions in the package, it
 is not decision-specific and must be `0`.

# Output

Return XML only. Escape XML special characters in evidence.

<adoption_contribution_analysis>
  <decisions>
    <decision>
      <id>Decision id from policy_decisions</id>
      <adoption_contribution>0|1|2|3</adoption_contribution>
      <evidence>Exact quote from the focus speaker, or No explicit statement found.</
       evidence>
    </decision>
  </decisions>
</adoption_contribution_analysis>

Meeting packet:

<policy_decisions>
{policy_decision_blocks}
```

```
</policy_decisions>

<alternatives>
{alternatives}
</alternatives>

meeting_date: {ymd}   (this is the ADOPTION meeting for all decisions listed)

Speaker packet:

<speeches>
{speeches}
</speeches>

<focus_speaker>
{stablespeaker}
</focus_speaker>

Score only this focus speaker and return the XML block only.
"""
```

IA.E.2 LLM Validation

We validate our automated extraction of measures from FOMC transcripts by comparing against trained human coders. This analysis helps to characterize the potential impact of measurement error arising from using LLMs to label our data.

We use Anthropic's Claude to extract arguments and classify parameter beliefs from the transcripts. Here we compare Claude's classifications against those of trained human coders. We validate three types of measures that we constructed using Claude:

(i) *Parameter Beliefs:* beliefs about Phillips curve slope and IS curve sensitivity.

(ii) *Arguments:* types of information and data sources cited.

(iii) *Decisions:* measures of adoption contribution and alignment of committee members regarding each decision.

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

We report Pearson correlations for ordinal variables (parameter beliefs, adoption contribution, and alignment), and percentage agreement for classifications without meaningful ordinal structure (information type and data source indicators). For data source indicators, we also report classification rates. We find agreement between LLM and human coding that is on the order of agreement between two trained human coders.

IA.E.2.1 Human Coding Protocol

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

For each parameter belief (Phillips curve and IS curve), we construct stratified random samples of 200 arguments with 50 arguments per classification category. This stratified approach ensures sufficient observations in each category to estimate reliability, including for less common classifications. For information-source validation, we selected 200 arguments for independent coding. For decision validation, we selected three FOMC meetings representing distinct policy contexts: August 1994 (tightening cycle), December 2008 (financial crisis emergency measures), and August 2011 (calendar-based forward guidance). The dynamic-alignment sample covers positions expressed before four later decisions.

Coders received a detailed written manual defining each classification scheme with exact category definitions, illustrative examples, and guidance on edge cases. To ensure standardization, we developed custom web-based validation interfaces using Streamlit, an open-source platform. For parameter beliefs and information sources, the interface presented the quotation and description; Claude's classifications were stored in separate files never provided to coders. For decision measures, the interface presented Claude's extracted decisions alongside quoted evidence and justification, and the coder assessed whether the evidence supported Claude's interpretation and provided their own scores. Before coding the full samples of 200 observations, human coders were given a training sample of 20 observations for each classification category, with solutions provided.

IA.E.2.2 Summary of Results

Table IA.23 presents consolidated validation results. Here, we report the concordance between the measures constructed by Claude and the alternatives from human coders.

Overall, Claude achieves high concordance with human classifiers. Claude achieves high correlations with human coders on parameter measures ($r = 0.74$–0.76). To put these numbers into context, we also report the correlation between the two human classifiers as "Human Ceiling." Claude's correlations are on the order of this human inter-coder reliability.

Moving to the middle panel, information type classification is on the order of human reliability: Claude agrees with the human coders in 86.9 percent of cases on average, compared with 85.6 percent agreement between the two human coders. The primary limitation is systematic over-identification of private information. Finally, decision-related measures show near-perfect agreement: adoption contribution correlates at $r = 0.93$, position alignment at $r = 0.96$, and dynamic position alignment at $r = 0.98$, though as noted above, these validations used an audit-based approach without an independent human ceiling.

Measure Class	Measure	Metric	Claude	Human Ceiling
Parameter Beliefs	Phillips Curve Slope	r	0.735	0.760
	IS Curve Sensitivity	r	0.762	0.754
Arguments	Information Type	Agreement	86.9%	85.6%
Decisions	Adoption Contribution	r	0.927	—
	Position Alignment	r	0.957	—
	Dynamic Position Alignment	r	0.979	—

TABLE IA.23
SUMMARY OF VALIDATION RESULTS

Note: The Claude column reports correlation with human average for parameter beliefs, average percentage agreement with the human coders for information type, and correlation with the human coder for decision measures. Human ceiling is inter-coder agreement among human raters. Information type uses the 118 arguments for which Claude and both human coders provide definite common/private classifications. Decision measures used audit-based validation without independent human coding, so no human ceiling is available.

IA.E.2.3 Parameter Beliefs

We examine the measures of parameter beliefs. We validate two measures capturing FOMC members' beliefs about economic relationships: the Phillips curve slope (the sensitivity of inflation to economic slack) and IS curve sensitivity (the strength of monetary-policy transmission to real activity). The Phillips curve measure uses categories for steep, moderate, and flat relationships, while the IS curve measure uses categories for strong, moderate, and weak transmission. Both measures include a null category for arguments not expressing a pertinent belief.

Table IA.24 reports detection rates for parameter beliefs. By construction from the random sampling, Claude identifies parameter beliefs in 75% of arguments for both parameters. Coder A detects beliefs in 73–75% of arguments, while Coder B detects beliefs in 65–71%. Notably, Claude rarely misses beliefs that humans detect. Across all comparisons, Claude fails to identify a belief when a human coder does in only 1–7 cases per parameter. This pattern indicates that Claude's slightly higher detection rate reflects a lower threshold for identifying beliefs rather than a systematic error.

	Phillips Curve	IS Curve
Non-null classifications (out of 200 arguments)		
Claude	150	150
Coder A	146	149
Coder B	129	142
Human non-null, Claude null		
Coder A	7	7
Coder B	2	1

TABLE IA.24
PARAMETER MEASURES: DETECTION RATES

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

Table IA.25 presents classification performance. Claude achieves high correlations with human coders across both measures ($r = 0.74$–0.76). For the IS curve, Claude's correlation with the human average exceeds the human ceiling, the correlation between the two human coders. For the Phillips curve, Claude maintains strong agreement.

	Phillips Curve	IS Curve
Human Average	0.735	0.762
	($n = 145$)	($n = 149$)
Coder A	0.653	0.643
	($n = 139$)	($n = 142$)
Coder B	0.825	0.809
	($n = 127$)	($n = 141$)
Human Ceiling	0.760	0.754

TABLE IA.25
PARAMETER MEASURES: CLASSIFICATION PERFORMANCE

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

IA.E.2.4 Arguments Dataset

The arguments dataset captures individual points made by FOMC members during policy deliberations. We validate two dimensions: information type and data source categories. The validation sample consists of 200 arguments coded independently by two human coders.

Information Type. Information type classifies whether arguments rely on common information or private/specialized information. Claude provides a definite classification as common or private for 162 of the 200 arguments, Coder A for 145, and Coder B for 171. Conditional on providing a definite classification, Claude classifies 37 of 162 arguments (22.8 percent) as private, compared with 13 of 145 (9.0 percent) for Coder A and 21 of 171 (12.3 percent) for Coder B. This pattern suggests Claude applies overly broad criteria for private information.

Claude's classification shows agreement with human coding that is on the order of agreement between the two human coders. Among the 118 arguments for which all three provide a definite classification, Claude agrees with Coder A in 97 cases (82.2 percent) and with Coder B in 108 cases (91.5 percent). The two human coders agree in 101 cases (85.6 percent). The classifications follow a systematic pattern: among the 129 arguments for which both humans provide a definite classification, they disagree in 19 cases. Fourteen disagreements involve Coder B classifying the argument as private when Coder A classifies it as common, compared with five disagreements in the opposite direction.

This systematic over-identification is conservative for our analysis. Our finding that private information plays a limited role in FOMC deliberations is robust; Claude's classification rate for private information is higher than that of either human coder.

Data Source Categories. Table IA.26 presents validation results for the ten binary data source indicators. Performance varies substantially across categories. Claude's average agreement with the two human coders meets or exceeds human inter-coder agreement for eight of the ten reported categories. The relatively low agreement rates for some categories indicate that data source identification is inherently difficult, likely because arguments often cite multiple sources implicitly. The largest difference in classification rates is for Official Statistics: Coder A identifies this category in 17.6 percent of arguments, compared with 45.7 percent for Coder B and 36.2 percent for Claude.

	Classification Rate			Agreement		
Data Category	Claude	Coder A	Coder B	Claude–Coder A	Claude–Coder B	Human Inter-Coder
Banking and Credit Conditions	5.5%	4.0%	6.5%	97.5%	98.0%	96.5%
International Economic Conditions	11.1%	11.1%	11.6%	97.0%	96.5%	96.5%
Board Staff Forecast and Analysis	31.2%	33.2%	27.1%	88.9%	92.0%	88.9%
Financial Market Indicators	23.1%	24.1%	26.6%	86.9%	90.5%	87.4%
Business and Regional Contacts	10.6%	6.0%	8.5%	93.5%	98.0%	95.5%
Official Statistics	36.2%	17.6%	45.7%	75.4%	81.4%	67.8%
Surveys and Sentiment Measures	5.5%	1.5%	4.5%	96.0%	99.0%	97.0%
External Forecasts and Research	4.5%	3.0%	3.0%	96.5%	98.5%	98.0%
Regional Fed Forecast and Analysis	9.0%	2.0%	7.0%	91.0%	97.0%	93.0%
FOMC Participant Projections / SEP	0.0%	1.0%	1.5%	99.0%	98.5%	99.5%

TABLE IA.26
DATA CATEGORY VALIDATION

Note: The table reports classification rates and percentage agreement for Claude and the two human coders. Classification Rate is the percentage of arguments assigned to each data category by each coder. Agreement is the percentage of arguments for which the indicated pair gives the same binary classification. All columns use the same 199 arguments for which both human coders provide a valid response. "Fiscal Policy Information" is excluded because it is rarely identified; "Other" and "No Identifiable Source" are residual categories and are not reported.

IA.E.2.5 Decision Measures

We validate Claude's scoring of FOMC meeting decisions. The validation sample is composed of speaker–decision pairs across 18 decisions from three meetings spanning different policy contexts. As noted above, decision validation uses an audit-based approach in which a human coder reviewed Claude's extractions rather than coding independently, so no human ceiling is available for comparison.

Adoption Contribution and Position Alignment. Table IA.27 presents validation results for member-level adoption contribution, position alignment, and dynamic position alignment. Claude achieves near-perfect correlations with the human coder for adoption contribution ($r = 0.93$), position alignment ($r = 0.96$), and dynamic position alignment ($r = 0.98$). Performance is consistent across all three contemporaneous meetings despite their different policy contexts.

	Adoption Contribution (0–3)	Position Alignment (-2 to $+2$)	Dynamic Position Alignment (-2 to $+2$)
1994-08-16	0.955	0.931	—
2008-12-16	0.897	0.958	—
2011-08-09	0.931	0.972	—
Overall	0.927	0.957	0.979

TABLE IA.27
DECISION SCORING VALIDATION

Note: The table reports Pearson correlations between Claude's scores and the human coder's scores. The contemporaneous measures use $N = 324$ speaker–decision pairs; dynamic alignment uses $N = 153$ observations.

Six adoption contribution scores, two position alignment scores, and one dynamic alignment score differ by at least two scale points. These cases are rare, indicating that the scoring criteria are consistently identifiable in the transcripts.

IA.E.2.6 Discussion

Overall, Claude's automated extraction achieves reliability comparable to trained human coding across most measures. For parameter beliefs, Claude achieves high correlations ($r = 0.74$–0.76), matching human inter-coder reliability. For decision-related measures, Claude achieves near-perfect correlations ($r = 0.93$–0.98), though the audit-based methodology provides a less stringent test than blind independent coding. Information type and data source categories validate at human-level reliability. The primary difference in these classifications relative to humans is that Claude exhibits over-classification of private information.

References

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

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

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

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

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

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

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

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

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

Hansen, Lars Peter, and Thomas J Sargent. 2022. "Structured ambiguity and model misspecification." *Journal of Economic Theory*, 199: 105165.

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

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

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