Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

Optimal Pooling in Taylor Rule Estimation with Multiple-Horizon Forecast Panels

Edward Herbst and Karen Page

2026-064

Please cite this paper as:
Herbst, Edward, and Karen Page (2026). "Optimal Pooling in Taylor Rule Estimation with Multiple-Horizon Forecast Panels," Finance and Economics Discussion Series 2026-064. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2026.064.

Optimal Pooling in Taylor Rule Estimation with Multiple-Horizon Forecast Panels

Edward Herbst
Federal Reserve Board

Karen Page
UT Austin

August 27, 2026

Abstract

Multiple-horizon forecast panels are increasingly used to infer perceived monetary policy rules, but inference depends on how coefficients are pooled across forecasters, dates, and horizons. We treat this pooling structure as the object of inference. In a participant-date-horizon Taylor-rule regression model, we compare pooling patterns using Bayesian marginal likelihoods, applying the framework to the Blue Chip Financial Forecasts, Survey of Professional Forecasters, and the Summary of Economic Projections. The preferred specifications place much of the systematic variation in policy-rate forecasts in intercepts that vary across forecast horizons and survey dates. Evidence of "changing perceptions" of monetary policy via economically meaningful time-varying response coefficients is weak overall.

Keywords: Taylor rules; monetary policy expectations; survey forecasts; Bayesian model selection; coefficient heterogeneity

JEL classification: C11, C23, E47, E52

Correspondence: edward.p.herbst@frb.gov; karen.page@my.utexas.edu. We thank Mark Bognanni, Manuel González-Astudillo, Christopher Gust, Benjamin Johannsen, and Matthias Paustian for helpful comments and discussions. All errors are our own. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Board or the Federal Reserve System.

1 Introduction

Economists increasingly use macroeconomic surveys, including central bank projections, to estimate monetary policy rules. These estimates can inform us about how forecasters, and perhaps policymakers themselves, perceive a central bank's reaction function. Because monetary policy works in part through expectations, beliefs about the reaction function can shape how private agents and financial markets interpret policy communications and macroeconomic news; see, for example, Woodford (2005) and Hamilton et al. (2011). Engen et al. (2015) and Bundick (2015) use consensus forecasts to argue in favor of time variation in perceived policy rules. This is important because it suggests that changes in macroeconomic outcomes may reflect shifts in beliefs about policy behavior—rather than, or in addition to, changes in policy itself—complicating the transmission of monetary policy to the economy. Another strand of the literature uses the median paths of the Federal Reserve's own policy projections in the Summary of Economic Projections (SEP) to infer an implicit policy rule underlying policymakers' forecasts. Knotek (2019) and González-Astudillo and Tanvir (2023) study how this implied rule varies over time.

More recently, researchers have turned to panel data on individual forecasts, rather than private-sector consensus or central bank median projections. These data provide projected paths for policy rates, inflation, and real activity at multiple horizons for each forecaster or participant, allowing the econometrician to exploit cross-sectional and cross-horizon heterogeneity in expectations. For the United States, the most popular of these surveys—all used in this paper—are the Blue Chip Financial Forecasts and the Survey of Professional Forecasters (SPF) for private-sector forecasts and the Federal Reserve's SEP. In influential work, Bauer et al. (2024) estimate Taylor rules using the Blue Chip panel. They find substantial time variation in the estimated coefficient governing the Federal Reserve's response to fluctuations in real activity and argue that this represents the market's changing perception of monetary policy, which importantly affects the transmission of policy.[1]

It is the wealth of information in these multiple-horizon forecast panels that supports inference on (for example) time variation in parameters, but this richness is both a blessing and a burden. In principle, policy rule coefficients can vary by forecaster, horizon, and time, but in practice, estimation of a completely unrestricted rule is infeasible because this specification is not identified. Consequently, estimation necessarily requires *pooling* observations by imposing restrictions on how coefficients vary across forecasters, forecast horizons, or time. Existing work typically adopts a particular pooling specification—often allowing coefficients to vary along one dimension while holding them fixed along others—and then interprets the resulting estimates as evidence of heterogeneity or time variation in the perception of monetary policy, treating alternative specifications as robustness checks.

In this paper, we focus on the specification—how observations are pooled together—

[1] In earlier work, Carlstrom and Jacobson (2015) use SPF microdata to estimate one-year-ahead forecaster-specific inertial Taylor rules and document substantial dispersion in the implied coefficients. Arai (2023) performs a related analysis on the SEP.

as the central object of interest in these Taylor-rule regressions. Starting from a general forecaster-date-horizon regression, we enumerate all possible pooling patterns, including the familiar fixed-effects specification in Bauer et al. (2024). We use a Bayesian framework to quantify the evidence for each. This allows the data to determine where pooling is appropriate and where heterogeneity is supported, providing inference that is robust to specification uncertainty. The Bayesian approach is particularly useful here—as argued by Leamer (1978)—since classical model selection is awkward when the model space is large and its models are not nested. The standard Bayesian calculus using marginal likelihoods, on the other hand, gives clear guidance on how to trade off fit against parsimony.

We show how the pooling-specification problem can be cast as a more familiar variable-selection problem. In doing so, we can draw on the large literature on Bayesian model uncertainty and variable selection beginning with George and McCulloch (1993), Madigan and Raftery (1994), and Kass and Raftery (1995). The same framework also accommodates structured extensions, including inertial rules, structural breaks, calendar effects, and group-based heterogeneity, some of which we consider here. Our approach relies heavily on the analytical tractability of the g-prior of Zellner (1986), using the more modern hierarchical framework of Liang et al. (2008) to guard against pathologies when comparing pooling patterns.

While our framework is designed to engage with the empirical literature emerging around the estimation of perceived monetary policy reaction functions, it also fits within a broader literature on grouped heterogeneity in panel data. Bonhomme and Manresa (2015) and Su et al. (2016) propose approaches to generalizing grouped fixed and slope effects. Our work builds on these ideas, but cannot be written directly in either of these forms. Instead, we treat each Taylor-rule pooling specification as a set of restrictions on a latent grouping structure and use marginal data densities to determine which candidate groupings are best supported by the data. Bayesian approaches to grouped heterogeneity are not new; see, for a macroeconomic application, Smith (2023). But to the best of our knowledge, this paper is the first to cast coefficient heterogeneity as a Bayesian variable-selection problem.

We estimate the optimal posterior pooling patterns for the Blue Chip, SPF, and SEP datasets. These datasets feature different time periods, forecast horizons, and forecasters, but a few broad conclusions emerge. First, while all three datasets support considerable heterogeneity in the Taylor rules, the favored specifications never coincide with the influential one in which the intercept of the rule varies by forecaster and time, while the response coefficients vary only by time. In the baseline static specifications, all three panels instead select intercepts that vary by forecast horizon and survey date, rather than by participant. Thus, the preferred specifications place much of the systematic variation in policy-rate forecasts in these intercepts rather than in heterogeneous response coefficients. Second, even when these response coefficients vary by time in the optimal specification, the extent of their variation is often small. This is because—across all three datasets—the projections for real activity and inflation explain very little of the variation in the policy-rate forecast. Third, forecast horizon is an important source of heterogeneity in our estimated Taylor-rule coefficients.

Such behavior might arise, for example, if forecasters or participants write conventional Taylor-rule-consistent projections at further-out horizons, with idiosyncratic considerations dominating their near-term projections.

When we extend our framework to inertial rules and specifications with partial pooling—enlarging the model space from hundreds to thousands to billions—the main conclusions survive, with some additional nuance. Inertial rules shift much of the forecast-path variation into persistence, leaving the direct responses to activity and inflation even less influential. Allowing adjacent dates and horizons to form data-selected groups similarly compresses variation in response coefficients into a small number of regimes and horizon groups. Thus the richer specifications uncover structured departures from a common rule, but not the highly local response-coefficient heterogeneity implied by conventional period-by-period specifications.

Ultimately, the exercise measures how much these forecast panels can say about perceived monetary policy rules once the pooling structure is not taken as fixed, echoing the recommendation in Leamer (1983). While all three panels clearly reject a common rule and contain information about broad forms of heterogeneity, the preferred specifications place much of the systematic variation in policy-rate forecasts in intercept objects, while heterogeneity in response coefficients is much more limited. Claims about economically meaningful time variation in policy-rule reaction coefficients rely on restrictions that the data themselves do not strongly favor.

The rest of the paper proceeds as follows. Section 2 describes the econometric framework. Section 3 describes the Blue Chip, SPF, and SEP panels and reports the exact restriction-search results. Section 4 considers inertial rules and partial pooling. Section 5 concludes.

2 Econometric Framework

The data consist of forecaster-level projections

$$\big(R_{jht}, x_{jht}, \pi_{jht}\big), \qquad j = 1, \ldots, J,\ h = 0, \ldots, H-1,\ t = 1, \ldots, T,$$

where j indexes forecasters or FOMC participants, h indexes forecast horizons (including the nowcast $h = 0$), and t indexes survey dates.[2] Here R_{jht} denotes the projected policy rate, x_{jht} the activity measure (an output or unemployment-rate gap), and π_{jht} projected inflation. For exposition, we take the panel to be balanced.[3]

Our baseline specification is a linear relationship between projected policy rates, inflation, and the activity measure, allowing for unrestricted heterogeneity across forecasters, horizons,

[2]Strictly speaking, h and t do not always measure time in the same units. This matters most for the SEP, where t indexes projection rounds and h indexes forecast years. This detail only complicates the exposition, so it is ignored here.

[3]Appendix A gives the corresponding definitions and marginal-likelihood formulas for the unbalanced panels used in the applications.

and time:

$$R_{jht} = \alpha_{jht} + \gamma_{jht} x_{jht} + \beta_{jht} \pi_{jht} + \varepsilon_{jht}, \qquad \varepsilon_{jht} \overset{\text{iid}}{\sim} N(0, \sigma^2). \tag{1}$$

When $\alpha_{jht} = \alpha$, $\gamma_{jht} = \gamma$, and $\beta_{jht} = \beta$ for all (j, h, t), (1) corresponds to a constant-coefficient Taylor rule. For example, the Taylor (1993) rule uses $\alpha = 4.0$, $\gamma = 0.5$, and $\beta = 1.5$, with activity measured by the output gap and inflation measured as the annual percentage change in the GDP deflator minus 2 percentage points.

The coefficients in (1) are not identified, since the regression model assigns three free parameters to each observation. Identification requires pooling, that is, some observations must share coefficients. The empirical question this paper asks is: which pooling structure is favored by the data? Existing applications typically impose this choice before estimation and treat alternatives as robustness checks. Bauer et al. (2024), for example, allow the intercept to vary by participant and date (α_{jt}) while restricting both response coefficients, γ_t and β_t, to vary only by date.

We instead treat the pooling pattern as unknown and use Bayesian model-comparison techniques to quantify the nature of coefficient heterogeneity across participants, horizons, and time. Formally, for each coefficient $c \in \{\alpha, \gamma, \beta\}$, let $S_c \subseteq \{j, h, t\}$ denote the set of dimensions along which that coefficient varies. We call S_c the pooling pattern for coefficient c. For example, $S_c = \emptyset$ means the coefficient is constant across forecasters, horizons, and time; $S_c = \{h\}$ means the coefficient is horizon specific; and $S_c = \{j, t\}$ indicates that the coefficient is participant-date specific. For a single coefficient, the *a priori* admissible set of pooling patterns is

$$\mathcal{S} = \mathcal{P}(\{j, h, t\}) \setminus \{\{j, h, t\}\}$$

where $\mathcal{P}(\cdot)$ denotes the power set. As discussed above, we exclude the completely unrestricted pattern $\{j, h, t\}$ because it is not identified. In the empirical exercises, we further restrict this set to patterns for which the corresponding regression is point identified in the realized sample. A *pooling pattern* is then a triple

$$M = (S_\alpha, S_\gamma, S_\beta) \in \mathcal{M},$$

where $\mathcal{M} = \mathcal{S}^3$. Since $|\mathcal{S}| = 7$, the prior admissible model space contains $7^3 = 343$ restriction patterns. For a given restriction pattern M, the pooling pattern induces a partition of the observations into groups that share a coefficient. Let D_{S_c} be an $n \times k(S_c)$ dummy matrix. The number of columns is given by the number of unique coefficients associated with a pooling pattern:

$$k(S_c) = \prod_{i \in S_c} N_i \quad \text{with} \quad N_j = J, \quad N_h = H, \text{ and } \quad N_t = T, \quad \text{and} \quad k(\emptyset) = 1. \tag{2}$$

The matrix D_{S_c} has entries of 1s and 0s, where 1s in a given column indicates that those entries are pooled together for coefficient estimation. Let x and π be $n \times 1$ vectors that stack the activity and inflation observations, respectively, and let ι be an $n \times 1$ vector of 1s. For

a given model M, the matrix of regressors can be written as

$$X_M = \left[\text{diag}(\iota) D_{S_\alpha},\quad \text{diag}(x) D_{S_\gamma},\quad \text{diag}(\pi) D_{S_\beta}\right].$$

The matrix X_M has dimensions $n \times k_M$, where the number of columns is given by

$$k_M = k(S_\alpha) + k(S_\gamma) + k(S_\beta).$$

Let Y be the $n \times 1$ vector of policy-rate projections. The regression model can be written as

$$Y = X_M \theta_M + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2 I_n), \tag{3}$$

where θ_M collects all coefficients implied by the restriction pattern M. For a realized sample and candidate pooling pattern M, point identification requires $\text{rank}(X_M) = k_M$. In the empirical analysis, we restrict attention to pooling patterns satisfying this condition.[4]

Before describing the econometric model further, we decompose each pooling-pattern specification into a component that is common across all models and a component that captures coefficient heterogeneity. This separation distinguishes two conceptually different questions: whether the Taylor rule contains intercept, activity, and inflation terms, and how the associated coefficients are allowed to vary across participants, horizons, and time. This distinction allows us to cleanly introduce prior beliefs about the degree of heterogeneity in the Taylor rule regressions.[5]

Let $X_\emptyset = [\iota, x, \pi]$ denote the regressor matrix corresponding to the constant-coefficient Taylor rule, and let $k_\emptyset = \text{rank}(X_\emptyset) = 3$ denote the number of linearly independent regressors common to every candidate model. Define $Q_\emptyset = I - X_\emptyset (X_\emptyset' X_\emptyset)^{-1} X_\emptyset'$ as the annihilator matrix associated with the column space of $X_\emptyset$. For a pooling pattern M, the matrix X_M contains $X_\emptyset$ as a subspace. Applying $Q_\emptyset$ removes the component of X_M already spanned by the constant-coefficient specification, leaving only directions associated with coefficient heterogeneity. Let $X_{M\perp\emptyset}$ denote any full-column-rank basis for the column space of $Q_\emptyset X_M$.[6] Under our identification restriction, the rank of $X_{M\perp\emptyset}$ is $r_M = k_M - k_\emptyset = k_M - 3$. Using this decomposition, the likelihood implicit in (3) can be rewritten as

$$Y \mid \theta_\emptyset, \theta_{M\perp\emptyset}, \sigma^2, M \sim N\left(X_\emptyset \theta_\emptyset + X_{M\perp\emptyset} \theta_{M\perp\emptyset},\, \sigma^2 I_n\right), \tag{4}$$

where $\theta_\emptyset$ collects the coefficients of the constant-coefficient Taylor rule and $\theta_{M\perp\emptyset}$ parameterizes departures from that benchmark.[7] The original coefficient vector θ_M can be recovered

[4] In practice, some participant-horizon-date combinations are missing. We simply leave out their dummy columns, so a model's coefficient count reflects the groups we actually observe rather than the count in an ideal balanced panel; see Appendix A.

[5] This is analogous to the way the constant is typically treated in Bayesian variable selection problems.

[6] We calculate $X_{M\perp\emptyset}$ from the QR decomposition of $Q_\emptyset X_M$.

[7] Throughout, we suppress conditioning on X_M for notational simplicity.

from

$$\theta_M = (X_M' X_M)^{-1} X_M' \big(X_\emptyset \theta_\emptyset + X_{M \perp \emptyset} \theta_{M \perp \emptyset} \big).$$

The likelihood implied by (4) forms the basis of our Bayesian analysis. We next combine this likelihood with priors over the coefficients and, subsequently, priors over restriction patterns to obtain posterior probabilities for each pattern.

2.1 Posterior Distribution over Pooling Patterns

Our main object of interest is the posterior probability of M,

$$p(M \mid Y) \propto p(Y \mid M)\, p(M),$$

where $p(Y|M)$ is the marginal likelihood under model M and $p(M)$ is the prior for model M. The marginal likelihood is a central object in Bayesian analysis.[8] It is obtained by integrating the likelihood over the parameter space of model M. Thus, it measures how likely the observed data are under the model after accounting for uncertainty about its parameters. This integration implicitly penalizes richly parameterized models, trading off fit and complexity. We next derive the marginal likelihood for our setting and highlight how these two forces shape posterior model probabilities.

Prior Distribution. A given model M has parameters $\theta_\emptyset$, $\theta_{M \perp \emptyset}$, and σ^2. For the common coefficients and error variance, we use the improper prior

$$p(\theta_\emptyset, \sigma^2 | M) \propto \sigma^{-2}. \tag{5}$$

Since these parameters are common across all models, the corresponding prior terms cancel in posterior model comparisons. The prior distribution for $\theta_{M \perp \emptyset}$ is more sophisticated. We introduce an additional parameter g, following Zellner (1986), as a way to describe prior beliefs about the degree of heterogeneity in the Taylor rule. The so-called g-prior allows one to compare models that may feature very different pooling patterns using closed-form marginal likelihoods. The prior is

$$\theta_{M \perp \emptyset} \mid \sigma^2, g, M \sim N\big(0, g\sigma^2 (X_{M \perp \emptyset}' X_{M \perp \emptyset})^{-1}\big)\,. \tag{6}$$

As $g \to 0$, the prior describes dogmatic beliefs that there is no coefficient heterogeneity, whereas as $g \to \infty$, the prior is diffuse. The term in $X_{M \perp \emptyset}' X_{M \perp \emptyset}$ reflects the scaling and correlation structure of the regressors, ensuring that model comparisons are not driven by arbitrary differences in parameterization.[9]

[8] See Kass and Raftery (1995) for more discussion of the marginal likelihood and its role in Bayesian model comparison.

[9] For the QR decomposition used in the paper $X_{M \perp \emptyset}' X_{M \perp \emptyset} = I$. As mentioned above, any full-rank basis for the column space of $Q_\emptyset X_M$ may be used. Different choices do not affect the resulting marginal likelihood.

Marginal Likelihoods. The marginal likelihood conditional on g is computed by integrating the product of the likelihood and the prior distribution of the model parameters:

$$p(Y \mid g, M) = \int p(Y|\theta_\emptyset, \theta_{M\perp\emptyset}, \sigma^2, g, M)p(\theta_{M\perp\emptyset}|\sigma^2, g, M)p(\theta_\emptyset, \sigma^2 \mid M)d\theta_{M\perp\emptyset}d\theta_\emptyset d\sigma^2. \quad (7)$$

Define the $R^2_{M\perp\emptyset}$ as the fraction of variation in Y explained by model M above the constant coefficient model:

$$R^2_{M\perp\emptyset} = \frac{\text{RSS}_\emptyset - \text{RSS}_M}{\text{RSS}_\emptyset} = 1 - \frac{\text{RSS}_M}{\text{RSS}_\emptyset}.$$

Here, RSS_M is the residual sum of squares of model M. Note that $R^2_{M\perp\emptyset} \in [0, 1]$ since $\text{RSS}_M \leq \text{RSS}_\emptyset$, because each model nests the constant parameter specification. This measure equals 0 if coefficient heterogeneity adds no explanatory power to the regression, and 1 if the model perfectly explains the residual variation left by the common rule. Taking the logs, the marginal likelihood in (7) can be written as:

$$\log p(Y \mid g, M) = \text{constant} - \frac{r_M}{2}\log(1+g) - \frac{\nu}{2}\log\left(1 - \frac{g}{1+g}R^2_{M\perp\emptyset}\right), \qquad \nu = n - k_\emptyset. \quad (8)$$

The marginal likelihood highlights the two forces determining posterior model probabilities under the g-prior. The first term, $r_M/2\log(1+g)$, penalizes models with more heterogeneity, i.e., less pooling. The size of this penalty also increases with g, reflecting the larger parameter space associated with a more diffuse prior. The second term rewards the incremental goodness of fit through $R^2_{M\perp\emptyset}$. As g increases, the factor $g/(1+g)$ approaches one, so models that achieve larger reductions in the residual sum of squares receive greater support. The choice of g therefore governs the balance between fit and parsimony.[10]

Various choices of g have been proposed in the literature. Setting $g = n$, for instance, yields the so-called unit information prior, leading to model comparisons in the spirit of the Bayesian information criterion. Still, using a single choice for g usually leads to several unsatisfactory characteristics of posterior model probabilities. While setting g arbitrarily large—the unit information prior in large samples—ought to reflect greater prior uncertainty about the nature of pooling in the Taylor-rule regression, this choice is problematic. As $g \to \infty$, the complexity penalty diverges, causing posterior mass to concentrate on the most restrictive models regardless of fit, a phenomenon known as Bartlett's paradox. More generally, posterior model probabilities can be highly sensitive to the choice of g, making it difficult to select a single fixed value that appropriately reflects prior beliefs across a broad

[10]The same factor controls shrinkage of the coefficient point estimates. Let $\widehat{\theta}_M$ denote the least-squares estimate under model M, and let $\widehat{\theta}^{(M)}_\emptyset$ denote the common-rule least-squares estimate expressed in the coefficient space of M. Conditional on g and M,

$$\text{E}[\theta_M \mid Y, g, M] = \widehat{\theta}^{(M)}_\emptyset + \frac{g}{1+g}\left(\widehat{\theta}_M - \widehat{\theta}^{(M)}_\emptyset\right).$$

After integrating out g, the shrinkage factor is replaced by $\text{E}[g/(1+g) \mid Y, M]$.

class of restriction patterns.

Rather than fixing g, we place a prior on it. The marginal likelihood then becomes a weighted average of fixed-g marginal likelihoods:

$$p(Y \mid M) = \int p(Y \mid g, M) p(g \mid M)\, dg.$$

We use the hyper-g/n prior of Liang et al. (2008),

$$p(g \mid M) = \frac{a-2}{2n}\left(1 + \frac{g}{n}\right)^{-a/2}, \qquad a > 2,$$

where a is a shape parameter. The prior can be derived as the prior induced on g when $g/(g+n) \sim \text{Beta}\,(1, a/2-1)$. Then, up to constants common across restriction patterns,

$$p(Y|M) \propto \frac{a-2}{n(a+r_M-2)} F_1\left(1; \frac{a}{2}, \frac{\nu}{2}; \frac{a+r_M}{2}; 1-\frac{1}{n}, R^2_{M\perp\emptyset}\right), \tag{9}$$

where F_1 is Appell's hypergeometric function; see Bayarri et al. (2012) for the corresponding derivation. The marginal likelihood depends on the same model-specific objects as the fixed-g expression, namely the heterogeneity rank r_M and the partial fit $R^2_{M\perp\emptyset}$.

Prior over restriction patterns. We assign prior probability to restriction patterns according to their heterogeneity complexity:

$$p(M) \propto \exp\{-\lambda r_M\}. \tag{10}$$

The complexity penalty $\lambda \geq 0$ controls the tradeoff between fit and parsimony on top of the marginal-likelihood penalty. When $\lambda = 0$, all restriction patterns receive equal prior weight; as λ increases, models that add fewer independent heterogeneity directions are favored *a priori*. In the baseline specification we set $\lambda = 0$.

Posterior model probabilities. Posterior probabilities over restriction patterns are given by

$$p(M \mid Y) = \frac{p(Y \mid M) p(M)}{\sum_{M' \in \mathcal{M}} p(Y \mid M') p(M')}. \tag{11}$$

Because the marginal likelihood has the closed-form representation above and can also be evaluated as a stable one-dimensional integral over g, these probabilities are straightforward to compute over the full model space.

In our applications, the posterior distribution over restriction patterns is often highly concentrated, with one or two models receiving most of the posterior weight. This concentration is informative about which kinds of heterogeneity receive posterior support, so we report the dominant restriction pattern and use it to organize the discussion. We also

report marginal posterior probabilities for specific parameters—for example, the posterior probability that β varies by horizon—by summing $p(M \mid Y)$ over all models.

3 Optimal Pooling in Three Datasets

This section reports the posterior distribution $p(M|Y)$ for three multiple-horizon forecast panels: Blue Chip, SPF, and SEP. Each contains projections of an interest rate, inflation, and activity, but the panels differ in survey frequency, horizon structure, sample length, and respondent composition. In the baseline specifications, inflation is measured as the forecast rate minus 2 percentage points.[11] The Blue Chip and SPF datasets are long-running private-sector panels of quarterly forecast paths. The SEP contains annual projections from FOMC participants; its sample is much shorter because interest rate projections began only in 2012 and because participant-level data are released with a roughly five-year lag. Despite these differences, a common result emerges. All three panels reject a constant-coefficient rule and favor horizon-date rather than respondent-date intercepts. The evidence for economically meaningful time variation in the response coefficients is weak: the posterior modal models often pool these coefficients across dates, and when pooling specifications with date-specific coefficients receive meaningful probability, their estimated paths vary only modestly over time.

As discussed above, the model space contains 343 pooling patterns but for each dataset, we retain only those patterns whose realized design matrix has full column rank. We use the hyper-g/n prior with $a = 3$ and assign equal prior probability to the retained patterns ($\lambda = 0$). A central reference is the period-by-period panel regression used by Bauer et al. (2024), corresponding to

$$(S_\alpha, S_\gamma, S_\beta) = (\{j, t\}, \{t\}, \{t\}).$$

This specification assigns participant-date intercepts, α_{jt}, while allowing the activity and inflation response coefficients, γ_t and β_t, to vary by date but remain common across participants and forecast horizons. Equivalently, the parameters can be recovered using a separate participant fixed-effects regression at each survey date. We refer to it as the *period-by-period benchmark*.

3.1 Optimal Pooling in the Blue Chip Panel

The Blue Chip Financial Forecasts survey provides monthly private-sector financial and macroeconomic forecasts. Our policy-rate variable is each forecaster's federal funds rate projection. Following Bauer et al. (2024), we construct annual CPI inflation forecasts by combining quarterly forecasts with previous realizations where necessary, and we construct the projected output gap by combining forecasters' real GDP growth forecasts with the

[11] The Blue Chip and SPF panels report CPI inflation, whereas the SEP reports PCE inflation. Appendix C.2 repeats the analysis with inflation in levels.

CBO output gap. The estimation sample begins in July 1984 and ends in June 2026. After applying the sample filters described below, it contains $N = 119{,}435$ observations across $T = 489$ survey dates, 300 distinct forecasters, and six quarterly horizon points from the current quarter, $h = 0$, through five quarters ahead, $h = 5$.[12]

The posterior distribution puts essentially all mass on a single pooling model.

$$(S_\alpha, S_\gamma, S_\beta) = (\{h, t\}, \{t\}, \{t\}),$$

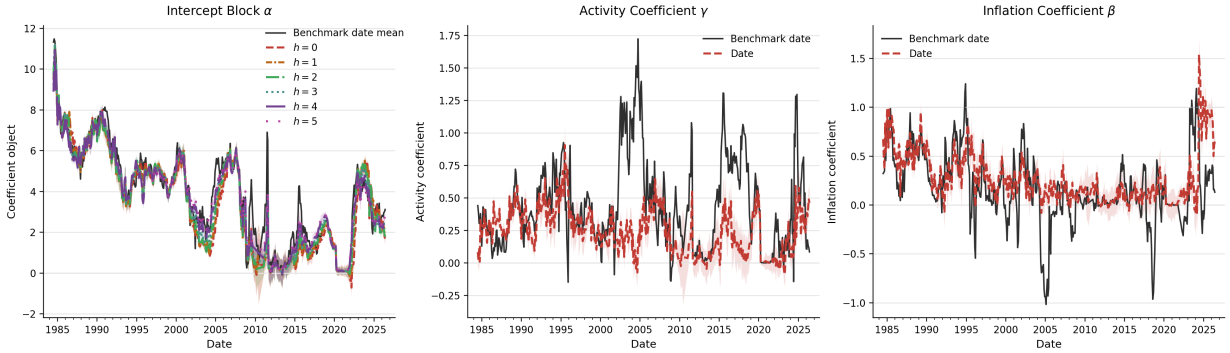
that is, horizon-date intercepts together with date-specific activity and inflation response coefficients. This specification is similar to the period-by-period benchmark, but the intercept term depends on horizon and time rather than forecaster and time. Even this small change can lead to markedly different paths for γ_t and β_t. Figure 1 plots the model-averaged posterior paths, organized according to the coefficient objects in the modal model, alongside estimates from the period-by-period specification. The horizon-specific intercepts all track, roughly, the mean of the forecaster-specific ones from the period-by-period benchmark (left panel). But there are important periods where these optimal coefficients diverge from the forecaster-specific average.

The different intercept structure is associated with paths for γ_t and β_t that are much less volatile than those from the benchmark. The standard deviation of the selected-model γ_t path is about half that of the period-by-period estimate. In particular, the large benchmark increases in the early 2000s and during parts of the 2010s are substantially attenuated. The selected-model β_t path is similarly smoother, and the large positive and negative benchmark swings in the early 2000s are absent under the optimal pooling structure.

The selected model uses 3,732 coefficients. The posterior comparison strongly rejects the common rule: the selected model's log marginal likelihood is about 190,648 points higher than that of the specification in which all three coefficients are constant. The period-by-period benchmark provides a better in-sample fit, but its much larger participant-date intercept block places it about 25,665 log marginal likelihood points below the selected model and gives it no meaningful posterior weight. Under the hyper-g/n prior, the posterior mean $\mathrm{E}[g \mid M^\star, Y]$ is about 935, far below the sample size. Fixing the prior scale at the conventional value $g = n$ selects the same pooling pattern. Thus the benchmark's additional fit is too expensive to overcome its complexity under either prior treatment.

[12] Several Blue Chip vintages around the 2009–2010 financial-crisis period lack the full-horizon inflation forecast. July–September 2010 are missing all but the nearest horizon and are dropped directly; the realized-rank screen also removes October–December 2010. We exclude other vintages only when the realized design cannot identify the participant-date benchmark, retaining surrounding vintages with truncated horizon sets when that benchmark remains identified. Because firms' names sometimes change in the survey, we occasionally need to use judgment to identify forecasters consistently.

Figure 1: *Optimal pooling estimates: Blue Chip versus the period-by-period benchmark.*



Notes: Black lines report point estimates from the period-by-period benchmark. In the intercept panel, respondent-specific benchmark intercepts are averaged across respondents within each survey date. Colored lines report model-averaged posterior means under the hyper-g/n prior, and shading shows pointwise 95 percent posterior credible bands.

3.2 Optimal Pooling in the Survey of Professional Forecasters

The SPF is the Federal Reserve Bank of Philadelphia's quarterly survey of U.S. macroeconomic forecasts. The survey asks professional forecasters for quarterly projections of key macroeconomic and financial variables, including output growth, unemployment, inflation, and interest rates.[13] We construct variables as follows. The policy rate is measured by the three-month Treasury bill rate, and inflation is measured as four-quarter CPI inflation less the 2 percent inflation target. Activity is measured using a constructed unemployment gap: we subtract forecasters' quarterly projections for the unemployment rate from their long-run unemployment-rate forecasts. The long-run unemployment-rate forecast is provided only in the third quarter of each year, beginning in 1996; we use the most recently available forecast to construct the series. The sample contains $T = 120$ quarterly survey dates from 1996:Q3 through 2026:Q2, $H = 5$ forecast horizons (beginning with the current-quarter nowcast), $J = 95$ forecasters, and 7,409 observations. A total of 160 restriction patterns satisfy our identification condition.

The posterior puts essentially all mass on

$$(S_\alpha, S_\gamma, S_\beta) = (\{h, t\}, \{h\}, \{t\}),$$

that is, horizon-date intercepts, horizon-specific activity responses, and date-specific inflation responses. Figure 2 reports the selected horizon-date intercept paths against the forecaster-date mean from the period-by-period benchmark, the selected horizon profile for activity, and the selected date path for the inflation response. The intercept paths (left panel) roughly track the forecaster-date mean from the period-by-period benchmark. These intercepts are by far the most important regressor—a fitted value decomposition implies the intercept term tracks the observed treasury bill well, and indeed, explains about 90 percent of the explained

[13] For background on the survey, see Croushore and Stark (2019).

variation in policy rule projections in both cases.

For γ, the posterior selects a horizon-specific coefficient. The estimated posterior means are small and increase over most of the forecast horizon:

$$\gamma = (0.002,\ 0.020,\ 0.056,\ 0.114,\ 0.194).$$

This stands in sharp contrast to the period-by-period benchmark, where the activity coefficient is assumed to vary across dates. That path is much more volatile, ranging from about -1.1 to 3.8. The model comparison therefore treats most of the movement in the period-by-period estimate of γ_t as overfit variation rather than as evidence of meaningful time variation in the perceived activity response. The selected inflation response also varies across dates, but its path is only weakly correlated with its period-by-period counterpart.

The Bayes factor against the period-by-period benchmark is large despite the benchmark's better in-sample fit. The selected model has a log marginal likelihood about 1,192 log points above the period-by-period benchmark, so prior odds greater than $10^{517} : 1$ in favor of the benchmark would be needed to reverse the ranking. The source of the comparison is transparent in the g-prior decomposition. The period-by-period benchmark explains more variation relative to the common-rule benchmark, with $R^2_{M\perp\emptyset} = 0.987$, compared with 0.978 for the selected model. But that gain comes from a much larger parameter space: the period-by-period benchmark uses $k_M = 1{,}730$ coefficients, while the selected model uses only $k_M = 725$ coefficients. The marginal likelihood therefore treats the benchmark's additional fit as too expensive. Moreover, the best in-sample fit in the model set is not the period-by-period benchmark, but

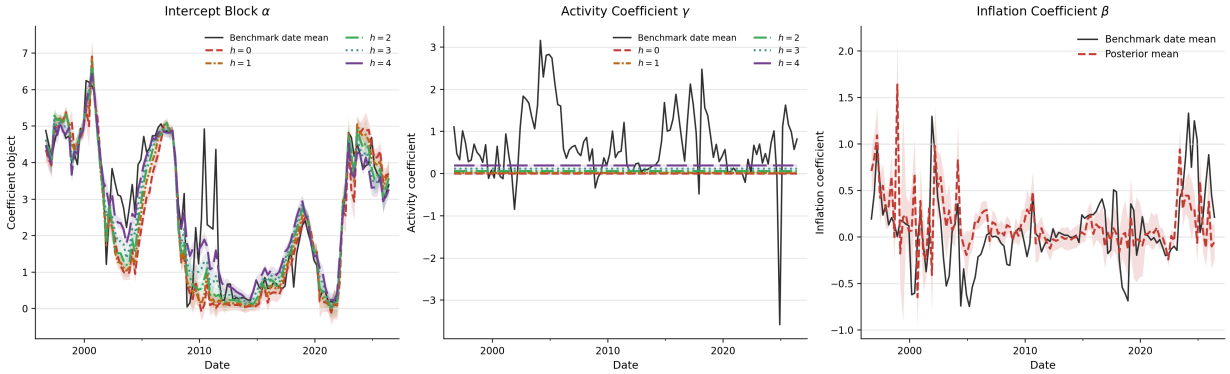
$$(S_\alpha, S_\gamma, S_\beta) = (\{j,t\}, \{h,t\}, \{j,t\}),$$

with $k_M = 3{,}580$ coefficients. Its $R^2_{M\perp\emptyset}$ is 0.995, only modestly above the period-by-period benchmark and the selected model, but the additional flexibility is heavily penalized. For the SPF, most of the available explanatory power is already captured by a time-varying intercept and low-dimensional variation in response coefficients; the remaining in-sample gains require thousands of additional coefficients.

3.3 Optimal Pooling in the SEP

The SEP reports FOMC participants' projections for real GDP growth, unemployment, total and core PCE inflation, and, since 2012, their "appropriate" federal funds rate. Anonymized ranges of these projections are released after every SEP (typically every other FOMC meeting), while the complete individual-level data are released with a lag of about five years. These projections include multiple calendar-year projections and "long-run" values for the unemployment rate, GDP growth, and core and headline PCE inflation. Starting in 2012, they also include values for the federal funds rate. We use the projected fourth-quarter federal funds rate for a given year as our policy rate. For activity, we subtract the projected

Figure 2: *Optimal pooling estimates: SPF versus the period-by-period benchmark.*



Notes: Black lines report point estimates from the period-by-period benchmark. In the intercept panel, respondent-specific benchmark intercepts are averaged across respondents within each survey date. Colored lines report model-averaged posterior means under the hyper-g/n prior, and shading shows pointwise 95 percent posterior credible bands.

fourth-quarter unemployment rate from the participant's long-run unemployment rate. For inflation, we use projected annual inflation for the forecast year.[14]

We use $H = 3$ years of forecasts. The panel contains $T = 36$ meetings from January 2012 through December 2020 and $J = 37$ distinct participants. Altogether, this dataset yields $N = 1{,}782$ observations. A total of 156 pooling patterns satisfy the identification requirement.

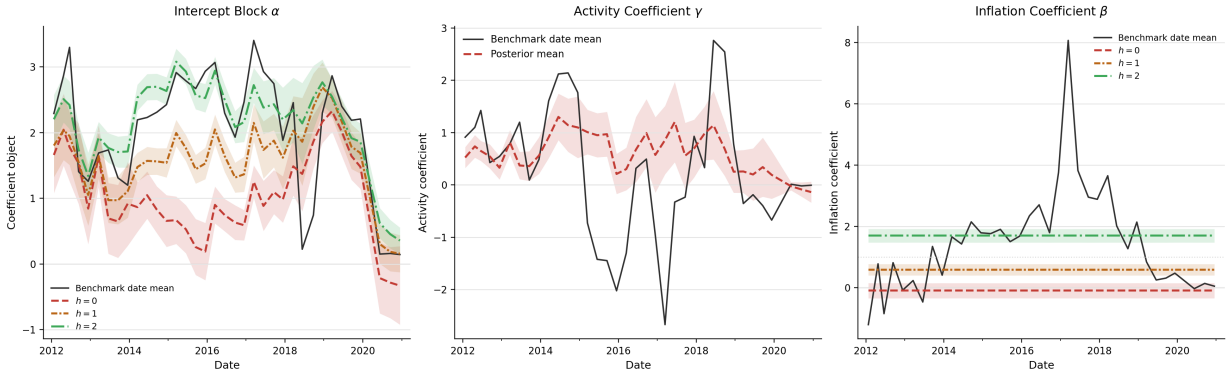
The posterior is degenerate, with only a single model receiving meaningful probability. The posterior is concentrated on the pooling pattern

$$(S_\alpha, S_\gamma, S_\beta) = (\{h, t\}, \{t\}, \{h\}).$$

The model features intercepts that vary across dates and horizons, an activity coefficient that varies across dates, and an inflation coefficient that varies only across horizons. Figure 3 displays the corresponding estimates alongside the period-by-period benchmark. The estimates of α_{ht} differ considerably across horizons during the middle of the last decade and generally increase with h. Viewed through the lens of the Taylor rule in (1), participants' projected policy-rate paths cannot be explained easily by their unemployment-gap and inflation projections, leaving considerable variation to be absorbed by α_{ht}. The estimated response to the unemployment gap varies over time, but the variation is modest and only weakly correlated with the period-by-period estimate. For most of the sample, the estimated coefficient is modestly positive, in line with benchmark values in the literature. Finally, the inflation coefficient varies only by horizon, with model-averaged posterior means $\beta_h \approx (-0.102, 0.583, 1.694)$. At the two-year horizon, the relationship between participants'

[14]The SEP also differs from the Blue Chip and SPF in that the forecasts are calendar-based, so the implicit horizon being forecast changes over time. It is straightforward to extend our framework so that coefficients can also depend on, for instance, the survey quarter. This extension does not meaningfully change the results reported here.

Figure 3: *Optimal pooling estimates: SEP versus the period-by-period benchmark.*



Notes: Black lines report point estimates from the period-by-period benchmark. In the intercept panel, respondent-specific benchmark intercepts are averaged across respondents within each survey date. Colored lines report model-averaged posterior means under the hyper-g/n prior, and shading shows pointwise 95 percent posterior credible bands.

inflation and federal funds rate forecasts is consistent with the original Taylor rule, whose inflation coefficient is 1.5. At shorter horizons, however, this relationship is weak at best. Overall, as with the other surveys, time variation in the response coefficients is limited in the posterior-dominant specifications.

4 Extensions

4.1 Inertial Rules

The framework can be extended to accommodate *inertial* rules by allowing the projected policy rate to respond not only to projected inflation and activity at horizon h, but also to the projected policy rate at the preceding horizon.[15] Define $R_{j,-1,t}$ as the observed policy rate one horizon earlier. The inertial specification is then

$$R_{jht} = \alpha_{jht} + \gamma_{jht} x_{jht} + \beta_{jht} \pi_{jht} + \rho_{jht} R_{j,h-1,t} + \varepsilon_{jht}, \qquad h = 0, \ldots, H-1. \tag{12}$$

(12) nests the non-inertial rule as the special case $\rho_{jht} = 0$. As in the baseline specification, (12) is not identified without restrictions. Within our Bayesian pooling framework, the inertial extension enlarges the model space in a straightforward way. We allow each of the four parameters $\{\alpha, \gamma, \beta, \rho\}$ to be pooled using any proper subset of $\{j, h, t\}$. There are therefore seven admissible pooling patterns per coefficient and

$$|\mathcal{M}| = 7^4 = 2401$$

[15] This kind of partial adjustment in the policy rate has been established empirically for the Federal Reserve by, for example, Clarida et al. (2000), and has often been found to be an important feature of optimal monetary policy; see Woodford (2005) and the references therein.

candidate models before imposing the rank condition. We proceed as before, maintaining the same prior hyperparameters of $a = 3$ and $\lambda = 0$.[16] As a benchmark, we use an inertial version of the participant-date fixed-effects specification. Its intercept varies by participant-date, while the activity, inflation, and inertia coefficients vary by date:

$$(S_\alpha, S_\gamma, S_\beta, S_\rho) = \big(\{j,t\}, \{t\}, \{t\}, \{t\}\big).$$

Blue Chip. The Blue Chip inertial panel contains 119,435 observations, 489 survey dates, 300 forecasters, and six forecast horizons. The posterior puts essentially all mass on

$$(S_\alpha, S_\gamma, S_\beta, S_\rho) = \big(\{j,h\}, \{h,t\}, \{h,t\}, \{t\}\big).$$

Thus the intercept varies jointly across forecasters and horizons, the activity and inflation responses vary jointly across dates and horizons, and inertia varies by date. The closest alternative retains participant-horizon intercepts but uses a common activity response together with date-specific inflation and inertia coefficients; it lies about 17,076 log marginal likelihood points below the selected model.

The modal model preserves substantial date and horizon variation in the direct response coefficients while pooling those responses across forecasters. At the same time, it organizes intercept heterogeneity by forecaster and horizon rather than by forecaster and date, and it pools the inertia coefficient across horizons within each date. Adding policy-rate persistence therefore changes where the model allocates heterogeneity, but it does not restore participant-specific activity or inflation responses.

SPF. For the SPF, four pooling patterns receive non-negligible posterior probability. The posterior modal pooling pattern is

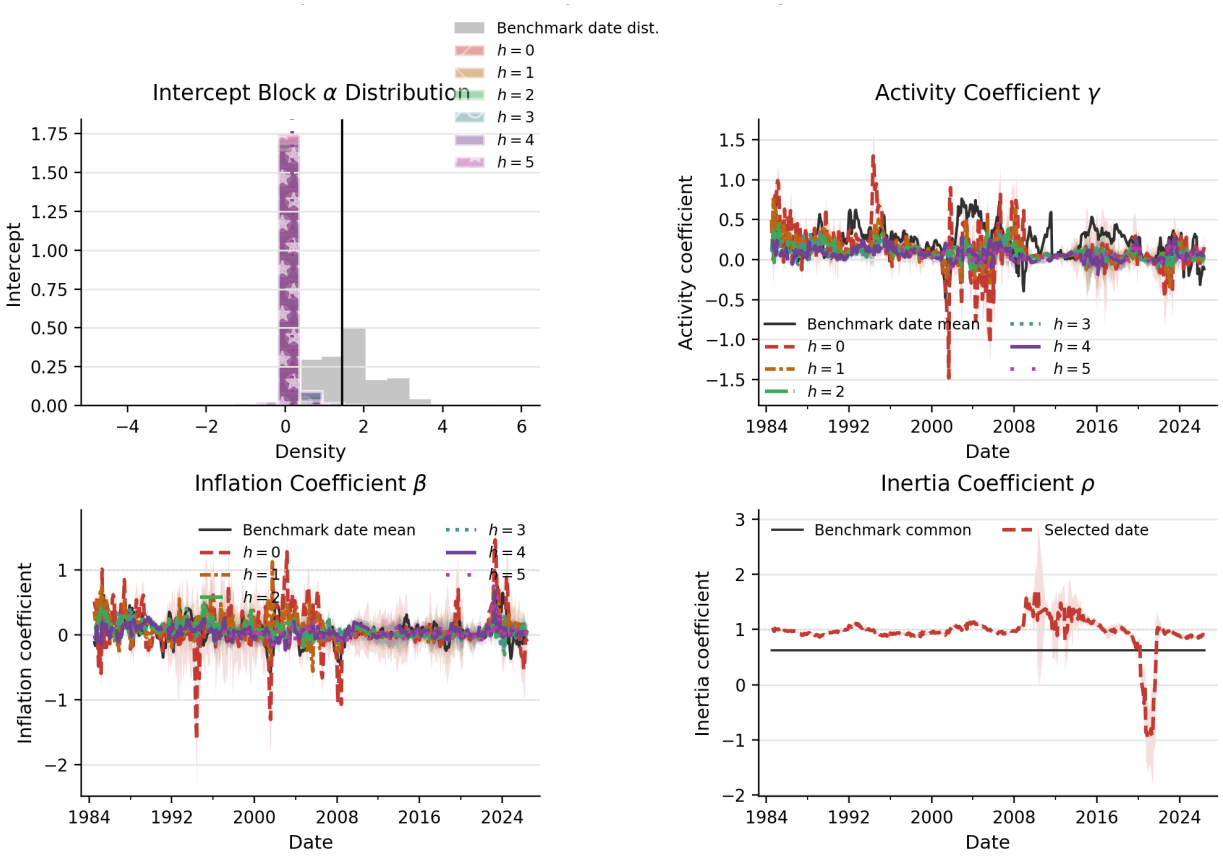
$$(S_\alpha, S_\gamma, S_\beta, S_\rho) = \big(\{j\}, \{h\}, \emptyset, \{h,t\}\big),$$

with posterior probability 0.670. The next pattern, $(\{j\}, \emptyset, \emptyset, \{h,t\})$, receives probability 0.236; the remaining probability is split between two patterns that allow a horizon-specific inflation response. All four patterns select forecaster-specific intercepts and horizon-date persistence, while differing in whether the direct activity and inflation responses vary by horizon. After averaging over these four models, the plotted activity coefficients are small:

$$\gamma = (0.007,\ 0.020,\ 0.023,\ 0.030,\ 0.027),$$

[16] Our treatment of the inertial parameter is identical to other Taylor rule coefficients, where we emphasize the role of heterogeneity by scoring models relative to a constant coefficient benchmark. We could instead adapt the framework so that the inclusion of this parameter and its heterogeneity are incorporated into the posterior probabilities. Because the lagged policy rate (or its forecast) explains much of the variation in policy rate forecasts, the results would be very similar. From the perspective of model fit, there is overwhelming evidence favoring the inertial specification over the static one.

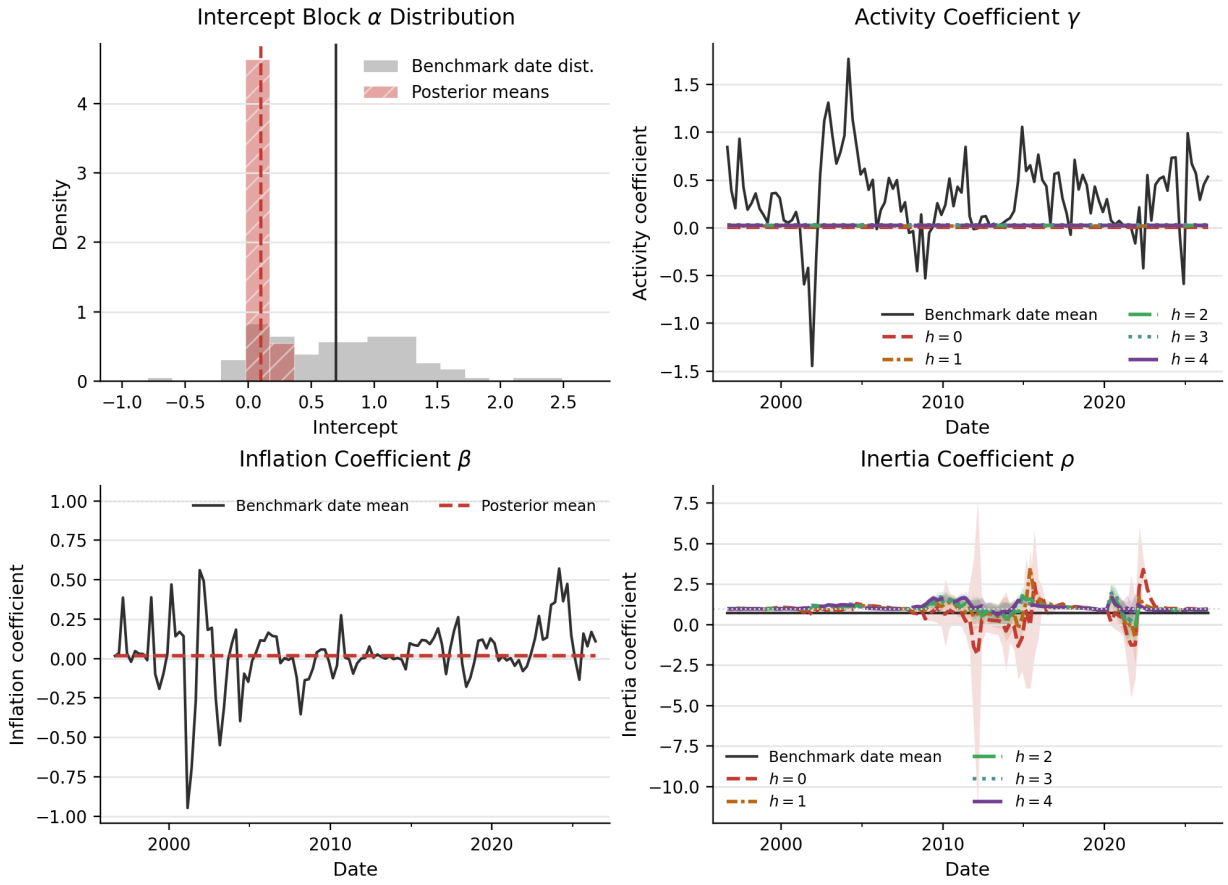
Figure 4: *Optimal pooling estimates: Blue Chip inertial rule versus the period-by-period benchmark.*



Notes: Black lines and gray distributions report point estimates from the inertial period-by-period benchmark. Colored lines and distributions report model-averaged posterior means under the hyper-g/n prior, and shading shows pointwise 95 percent posterior credible bands. The intercept panel summarizes the cross-sectional distribution of respondent-specific intercept estimates.

and the displayed model-averaged inflation object is 0.018. Most of the forecast-path heterogeneity is therefore assigned to persistence rather than to the direct response coefficients. Figure 5 displays these objects against the period-by-period reference.

Figure 5: *Optimal pooling estimates: SPF inertial rule versus the period-by-period benchmark.*



Notes: Black lines and gray distributions report point estimates from the inertial period-by-period benchmark. Colored lines and distributions report model-averaged posterior means under the hyper-g/n prior, and shading shows pointwise 95 percent posterior credible bands. The intercept panel summarizes the cross-sectional distribution of respondent-specific intercept estimates. For the nowcast, the lagged-rate regressor is the preceding-quarter Treasury bill rate; at later horizons, it is the respondent's Treasury bill forecast from the preceding horizon.

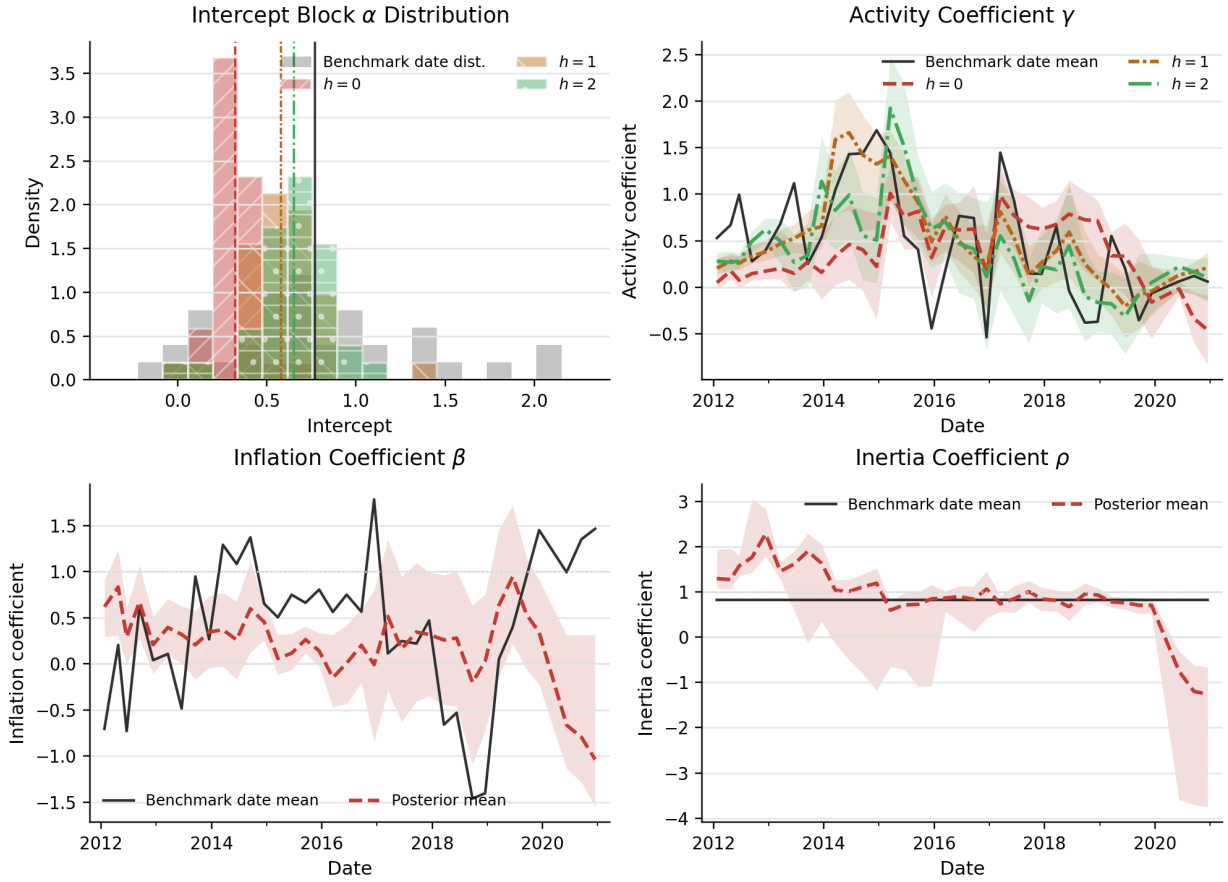
SEP. The SEP unemployment-gap inertial panel contains 1,782 observations across 36 survey dates, 37 participants, and three forecast horizons. For the first horizon, the lagged policy rate is the previous year-end federal funds rate; at later horizons, it is the participant's policy-rate projection at the preceding horizon. The posterior modal pooling pattern is

$$(S_\alpha, S_\gamma, S_\beta, S_\rho) = \big(\{j, h\}, \{h, t\}, \{t\}, \{t\}\big),$$

with posterior probability 0.951. The next model receives posterior probability 0.048 and replaces these objects with participant-specific intercepts, date-specific activity responses, horizon-specific inflation responses, and horizon-date inertia. Thus, even in the shorter SEP panel, the posterior sharply distinguishes between two economically different ways of organizing policy-rate persistence.

The modal model features intercepts that vary across participants and horizons, an activity coefficient that varies across time and horizon, and inflation and inertia coefficients that vary only across time. Figure 6 displays the corresponding coefficient objects. The horizon-date activity responses are largest around 2014–2015 and become smaller later in the sample. The direct inflation response is modest for most dates, while the inertia coefficient is near one through much of the sample before falling sharply at the end. Adding persistence therefore reallocates substantial variation across the intercept, activity, and inertia blocks, but it does not restore participant-date heterogeneity in response coefficients.

Figure 6: *Optimal pooling estimates: SEP unemployment-gap inertial rule.*



Notes: Black lines and gray distributions report point estimates from the inertial period-by-period benchmark. Colored lines and distributions report model-averaged posterior means under the hyper-g/n prior, and shading shows pointwise 95 percent posterior credible bands. The intercept panel summarizes the cross-sectional distribution of respondent-specific intercept estimates.

4.2 Intermediate Pooling Patterns for Survey Dates and Horizons

In the models considered above, along a given dimension, a parameter is either constant or fully heterogeneous. While this structure clarifies the source of heterogeneity and is consistent with previous work, it may be overly restrictive. Along ordered dimensions such as survey date and forecast horizon, adjacent observations may share a coefficient even when complete pooling is not favored by the data.[17] We therefore allow *partial pooling*, in which contiguous dates or horizons are assigned to groups that share coefficients. This intermediate structure can balance parsimony and fit while uncovering subtler forms of heterogeneity than the polar cases considered in Section 3.

We enlarge the model space by allowing the data to choose date and horizon pooling jointly. The date component is a contiguous partition of survey dates; each coefficient block can use that partition or retain unrestricted date-level variation. The horizon component is also estimated, allowing coefficients with horizon dependence to pool over contiguous groups of observed horizons. The search therefore asks whether variation in response coefficients is better described by every date and horizon separately, by a small number of date regimes, by grouped horizons, or by some combination of the two. Appendix B gives the notation and computational details.[18]

Blue Chip. For the Blue Chip, the joint search puts essentially all posterior mass on five date regimes and four horizon groups. The posterior modal specification is

$$(S_\alpha, S_\gamma, S_\beta) = (\{g, t\}, \{g, r\}, \{t\}),$$

with posterior probability 0.686, where r indexes the estimated date regimes and $g = \{\{0\}, \{1, 2\}, \{3, 4\}, \{5\}\}$. The selected regimes are July 1984–June 1988, July 1988–October 1998, November 1998–March 2020, April 2020–May 2024, and June 2024–June 2026. Figure 7 shows the resulting coefficient objects. The posterior probability that γ uses the date-regime partition is essentially one, while the intercept and inflation blocks remain date specific. Thus partial pooling compresses the activity response into a handful of persistent episodes without eliminating the higher-frequency movements absorbed by the other coefficients.

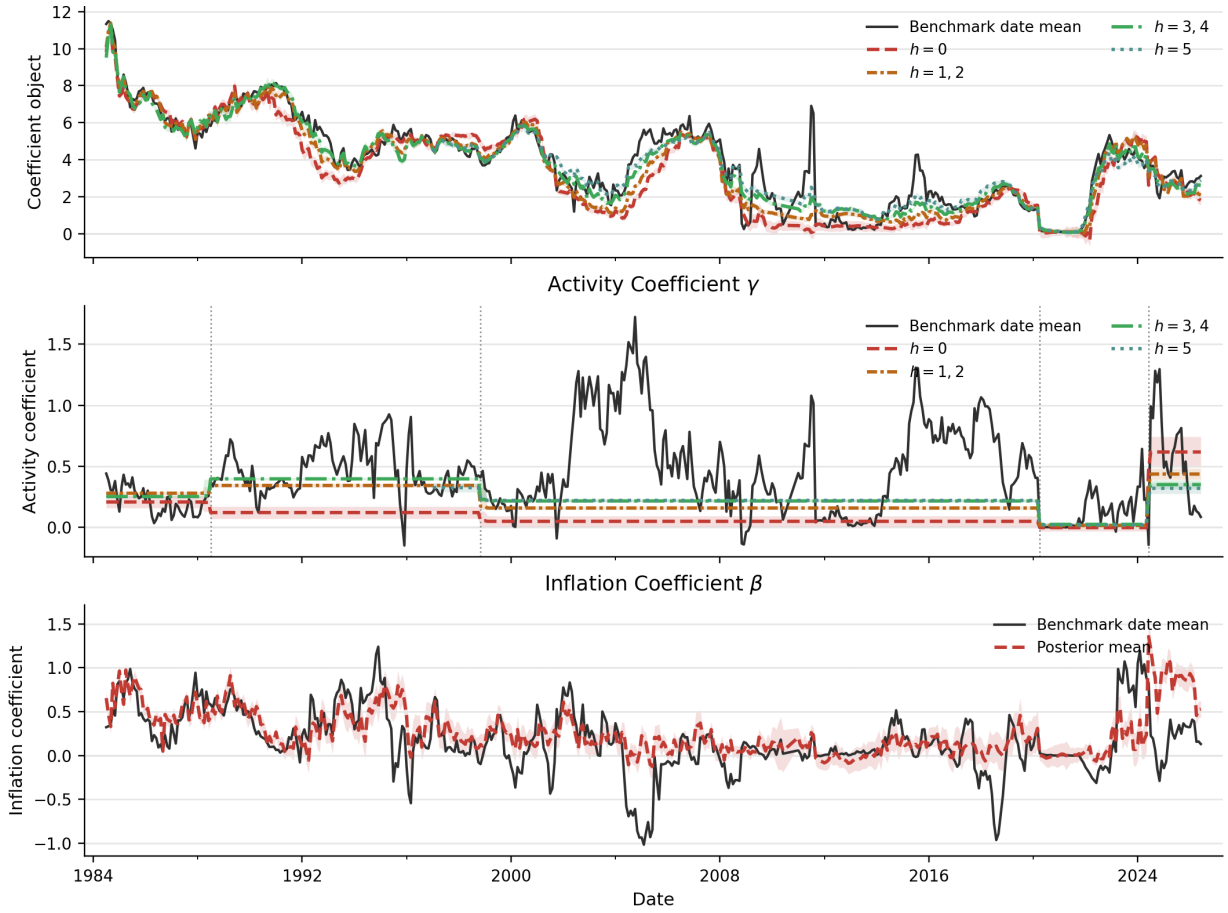
SPF. For the SPF, the joint search puts essentially all posterior mass on five date regimes and most posterior mass on three horizon groups. The top model is

$$(S_\alpha, S_\gamma, S_\beta) = (\{t\}, \{g, r\}, \{t\}),$$

[17] Groups of participants may also be of interest. This kind of clustering can be challenging, though not impossible, in our setting because participants lack a natural ordering. We leave participant clustering for future work.

[18] Each search allows at most five date regimes. The minimum regime lengths are 24 monthly survey dates for Blue Chip, eight quarterly survey dates for the SPF, and four projection rounds for the SEP. We retain the hyper-g/n prior with $a = 3$ and assign equal prior weight to each admissible restriction–partition model.

Figure 7: *Optimal pooling estimates: Blue Chip joint date/horizon pooling.*

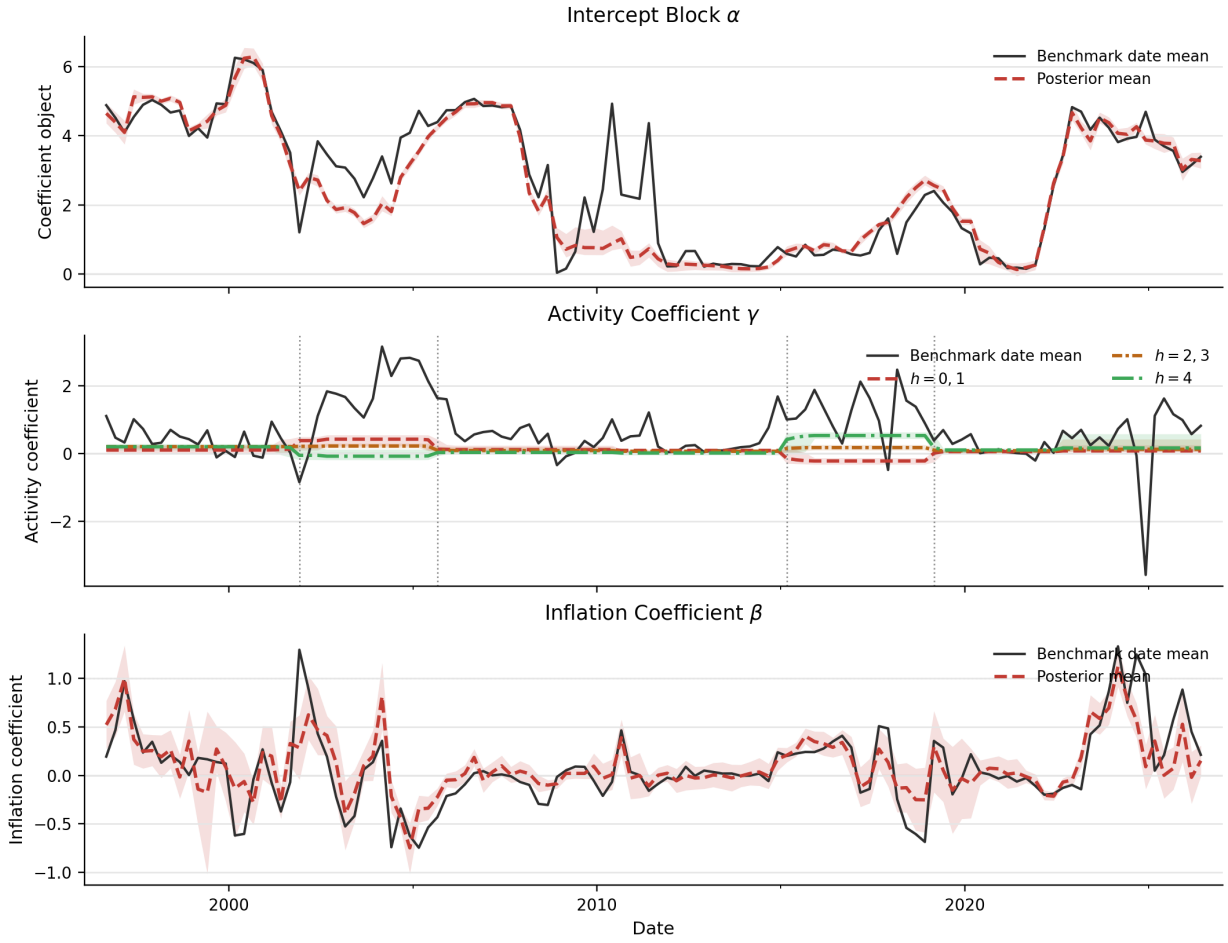


Notes: Black lines report point estimates from the period-by-period benchmark. In the intercept panel, respondent-specific benchmark intercepts are averaged across respondents within each survey date. Colored lines report model-averaged posterior means under the hyper-g/n prior, and shading shows pointwise 95 percent posterior credible bands. Dotted vertical lines mark the posterior modal model's regime boundaries.

where r indexes the estimated date regimes and $g = \{\{0,1\},\{2,3\},\{4\}\}$. Figure 8 shows that the date intercept and inflation response remain close to the period-by-period benchmark, while the activity response is summarized by three grouped-horizon regime paths. The selected date regimes are 1996:Q3–2001:Q3, 2001:Q4–2005:Q2, 2005:Q3–2014:Q4, 2015:Q1–2018:Q4, and 2019:Q1–2026:Q2. The posterior probability that γ uses the date-regime partition is essentially one, while the intercept and inflation blocks remain date specific. Thus the extra structure does not restore a fully local period-by-period rule; it compresses the time variation in the activity response into a small number of regimes and horizon groups.

SEP. The same joint exercise gives a different message in the SEP unemployment-gap panel. Date-regime pooling is useful for the response coefficients, but horizon pooling is not: posterior mass is essentially one on the singleton horizon partition $h = 0 \mid h = 1 \mid h = 2$.

Figure 8: *Optimal pooling estimates: SPF joint date/horizon pooling.*



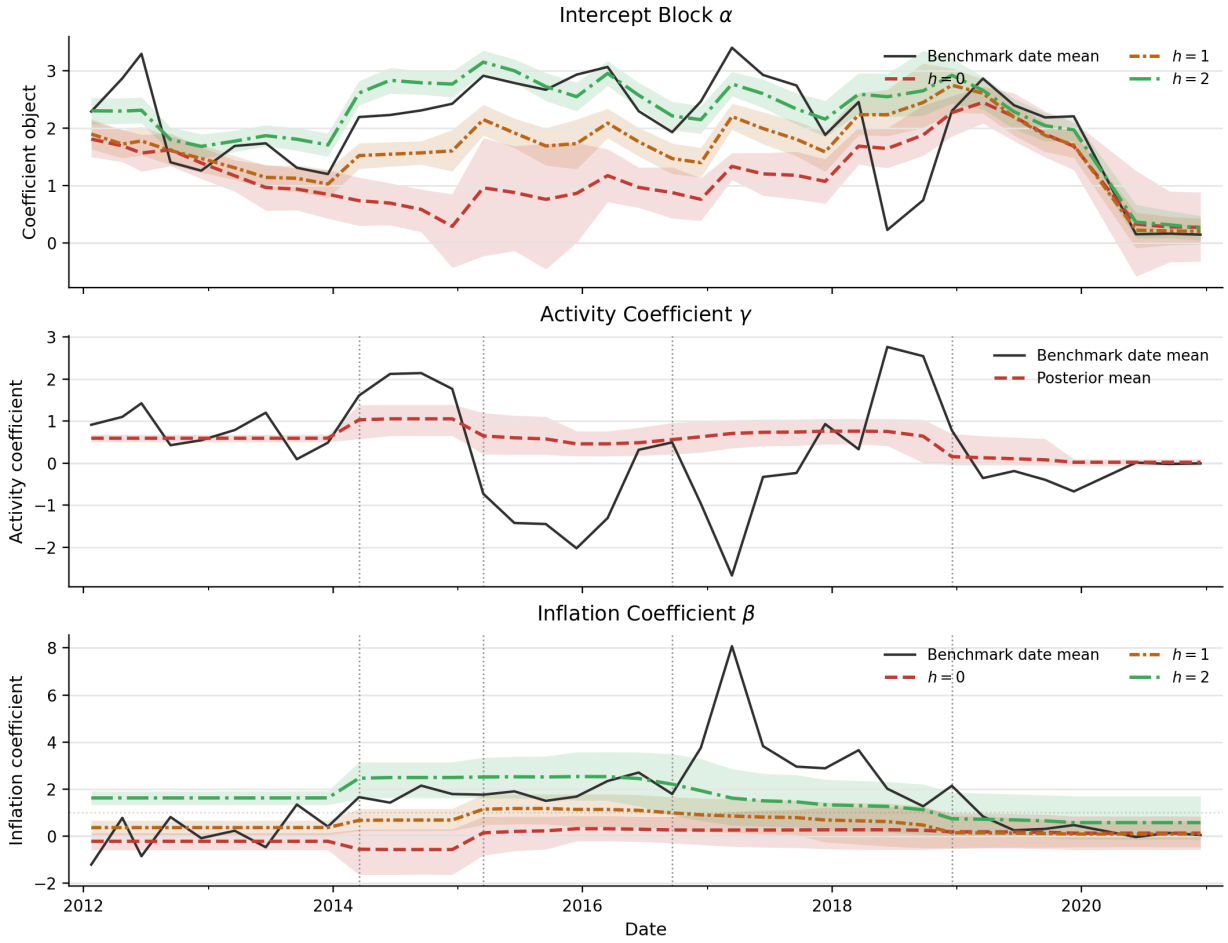
Notes: Black lines report point estimates from the period-by-period benchmark. In the intercept panel, respondent-specific benchmark intercepts are averaged across respondents within each survey date. Colored lines report model-averaged posterior means under the hyper-g/n prior, and shading shows pointwise 95 percent posterior credible bands. Dotted vertical lines mark the posterior modal model's regime boundaries.

The top model is

$$(S_\alpha, S_\gamma, S_\beta) = (\{h, t\}, \{r\}, \{h, r\}),$$

with five date regimes. Figure 9 shows the resulting objects. The intercept remains horizon-date specific, the activity coefficient is common across horizons within each date regime, and the inflation response varies by both horizon and date regime. The selected regimes are 2012:Q1–2013:Q4, 2014:Q1–2014:Q4, 2015:Q1–2016:Q2, 2016:Q3–2018:Q3, and 2018:Q4–2020:Q4. Posterior mass over the number of regimes is interior, with probability 0.811 on five regimes, 0.164 on four, and 0.025 on three. The result reinforces the main SEP interpretation: the data support low-dimensional variation in response coefficients over time, but they do not support pooling adjacent SEP horizons once date regimes are selected jointly.

Figure 9: *Optimal pooling estimates: SEP joint date/horizon pooling.*



Notes: Black lines report point estimates from the period-by-period benchmark. In the intercept panel, respondent-specific benchmark intercepts are averaged across respondents within each survey date. Colored lines report model-averaged posterior means under the hyper-g/n prior, and shading shows pointwise 95 percent posterior credible bands. Dotted vertical lines mark the posterior modal model's regime boundaries.

5 Conclusion

Across the Blue Chip, SPF, and SEP panels, the data strongly reject a common Taylor rule, but they do not favor the conventional specification with participant-date intercepts and date-specific response coefficients. The preferred specifications instead assign much of the systematic variation in policy-rate forecasts to intercepts that vary across forecast horizons and survey dates. The responses to activity and inflation are comparatively stable: when allowed to vary over time, they move far less than their period-by-period estimates. Partial pooling reaches the same conclusion, compressing this variation into a small number of date regimes and horizon groups, while inertial specifications assign more of the forecast path to persistence. The panels therefore contain clear evidence of heterogeneity, but much less support for a dominant role of time-variation in perceived policy response coefficients.

The results also suggest several extensions. The pooling patterns considered here are discrete: coefficients are either shared or allowed to differ across participants, dates, and horizons. Allowing coefficients to evolve smoothly, or tying changes in the pooling structure to macroeconomic conditions, could capture more gradual changes. More structure across forecast equations—for example, linking intercepts to beliefs about the inflation target, the neutral rate, or long-run outcomes—could also help distinguish changes in perceived policy behavior from horizon-specific forecasting conventions or omitted variables.

References

Arai, N. (2023): "The FOMC's New Individual Economic Projections and Macroeconomic Theories," *Journal of Banking & Finance*, 151, 106845.

Bauer, M. D., C. E. Pflueger, and A. Sunderam (2024): "Perceptions about Monetary Policy," *Quarterly Journal of Economics*, 139, 2227–2278.

Bayarri, M. J., J. O. Berger, A. Forte, and G. García-Donato (2012): "Criteria for Bayesian Model Choice with Application to Variable Selection," *Annals of Statistics*, 40, 1550–1577.

Bonhomme, S. and E. Manresa (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83, 1147–1184.

Bundick, B. (2015): "Estimating the Monetary Policy Rule Perceived by Forecasters," *Economic Review*, 100, 33–49, federal Reserve Bank of Kansas City.

Carlstrom, C. T. and M. M. Jacobson (2015): "Do Forecasters Agree on a Taylor Rule?" Economic Commentary 2015-10, Federal Reserve Bank of Cleveland.

Clarida, R., J. Gali, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115, 147–180.

Croushore, D. and T. Stark (2019): "Fifty Years of the Survey of Professional Forecasters," *Economic Insights*, 4, 1–11.

Engen, E. M., T. Laubach, and D. Reifschneider (2015): "The Macroeconomic Effects of the Federal Reserve's Unconventional Monetary Policies," Finance and Economics Discussion Series 2015-005, Board of Governors of the Federal Reserve System.

George, E. I. and R. E. McCulloch (1993): "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889.

González-Astudillo, M. and R. Tanvir (2023): "Hawkish or Dovish Fed? Estimating a Time-Varying Reaction Function of the Federal Open Market Committee's Median Participant," Finance and Economics Discussion Series 2023-070, Board of Governors of the Federal Reserve System.

Hamilton, J. D., S. Pruitt, and S. Borger (2011): "Estimating the Market-Perceived Monetary Policy Rule," *American Economic Journal: Macroeconomics*, 3, 1–28.

Kass, R. E. and A. E. Raftery (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

Knotek, E. S. (2019): "Changing Policy Rule Parameters Implied by the Median SEP Paths," Economic Commentary 2019-06, Federal Reserve Bank of Cleveland.

Leamer, E. E. (1978): *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, New York: Wiley.

——— (1983): "Let's Take the Con Out of Econometrics," *American Economic Review*, 73, 31–43.

Liang, F., R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger (2008): "Mixtures of g Priors for Bayesian Variable Selection," *Journal of the American Statistical Association*, 103, 410–423.

Madigan, D. and A. E. Raftery (1994): "Model Selection and Accounting for Model Uncertainty," *Journal of the American Statistical Association*, 89, 1535–1546.

Smith, S. C. (2023): "Structural Breaks in Grouped Heterogeneity," *Journal of Business & Economic Statistics*, 41, 752–764.

Su, L., Z. Shi, and P. C. B. Phillips (2016): "Identifying Latent Structures in Panel Data," *Econometrica*, 84, 2215–2264.

Taylor, J. B. (1993): "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.

Woodford, M. (2005): "Central-Bank Communication and Policy Effectiveness," *Proceedings - Economic Policy Symposium - Jackson Hole*, 399–474.

Zellner, A. (1986): "On Assessing Prior Distributions and Bayesian Regression Analysis With g-Prior Distributions," in *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, ed. by P. K. Goel and A. Zellner, Amsterdam: North-Holland/Elsevier, 233–243.

Appendix for
"Optimal Pooling in Taylor Rule Estimation with Multiple-Horizon Forecast Panels"

by Edward Herbst and Karen Page

A Unbalanced Panels

The notation in Section 2 uses a balanced panel to keep the construction of a pooling pattern transparent. The panels used in the empirical analysis are not balanced: forecasters enter and leave, some projections are missing, and the available horizon set can differ across survey dates. In that setting, the number of possible cells, the number of coefficient groups observed in the sample, and the number of identified coefficient directions are distinct objects. This section gives the corresponding sample-based definitions.

Observed support and coefficient groups. Let

$$\mathcal{O} = \big\{(j,h,t) : (R_{jht}, x_{jht}, \pi_{jht}) \text{ is included in the estimation sample}\big\}$$

denote the observed support, with $n = |\mathcal{O}|$. For a pooling pattern S_c, let $\rho_{S_c}(j,h,t) = \text{proj}_{S_c}(j,h,t)$ collect the labels along the dimensions in S_c, with $\rho_\emptyset(j,h,t) = 1$. The coefficient groups actually represented in the sample are

$$\mathcal{G}(S_c;\mathcal{O}) = \big\{\rho_{S_c}(j,h,t) : (j,h,t) \in \mathcal{O}\big\}, \qquad k_{\mathcal{O}}(S_c) = \big|\mathcal{G}(S_c;\mathcal{O})\big|.$$

For a balanced panel this reduces to the product formula in (2). In an unbalanced panel it does not: only combinations of labels that occur in $\mathcal{O}$ generate columns.

Accordingly, let $D^{\mathcal{O}}_{S_c}$ be the dummy matrix with rows indexed by $\mathcal{O}$ and one column for each element of $\mathcal{G}(S_c;\mathcal{O})$. The realized design matrix for the static model is

$$X^{\mathcal{O}}_M = \Big[D^{\mathcal{O}}_{S_\alpha}, \text{diag}(x)D^{\mathcal{O}}_{S_\gamma}, \text{diag}(\pi)D^{\mathcal{O}}_{S_\beta}\Big], \qquad k^{\mathcal{O}}_M = \sum_{c\in\{\alpha,\gamma,\beta\}} k_{\mathcal{O}}(S_c).$$

The inertial specification adds the corresponding block $\text{diag}(R_{-1})D^{\mathcal{O}}_{S_\rho}$.

Identification and realized rank. Removing empty cells is necessary, but it does not by itself establish identification. For the static rule let $X^{\mathcal{O}}_\emptyset = [\iota, x, \pi]$; for the inertial rule, augment this matrix with R_{-1}. We require the applicable common-rule block to have full

column rank, as it does in every application sample, and define its realized rank as

$$k_\emptyset = \operatorname{rank}(X_\emptyset^{\mathcal{O}}).$$

The associated annihilator matrix is

$$Q_\emptyset^{\mathcal{O}} = I - X_\emptyset^{\mathcal{O}} \left(X_\emptyset^{\mathcal{O}\prime} X_\emptyset^{\mathcal{O}} \right)^{-1} X_\emptyset^{\mathcal{O}\prime}.$$

Because every candidate pooling pattern nests the common rule, its number of identified heterogeneity directions is

$$r_M = \operatorname{rank}\big(Q_\emptyset^{\mathcal{O}} X_M^{\mathcal{O}}\big) = \operatorname{rank}(X_M^{\mathcal{O}}) - k_\emptyset. \tag{13}$$

We retain a candidate model only if $\operatorname{rank}(X_M^{\mathcal{O}}) = k_M^{\mathcal{O}}$. Conditional on that restriction, $r_M = k_M^{\mathcal{O}} - k_\emptyset$. Thus the familiar $k_M - 3$ expression is a balanced-panel, full-rank special case, not the general definition. For the static applications $k_\emptyset = 3$; for the inertial applications the common block also includes R_{-1}, so $k_\emptyset = 4$ when that block is full rank.

Marginal likelihoods. All sums of squares and projections are computed using the n observations in $\mathcal{O}$. Define

$$\nu = n - k_\emptyset, \qquad R^2_{M\perp\emptyset} = 1 - \frac{\mathrm{RSS}_M}{\mathrm{RSS}_\emptyset}.$$

Then the fixed-g marginal likelihood, up to terms common across models, is

$$\log p(Y \mid g, M) = \text{constant} - \frac{r_M}{2}\log(1+g) - \frac{\nu}{2}\log\left(1 - \frac{g}{1+g} R^2_{M\perp\emptyset}\right). \tag{14}$$

Integrating this expression under the hyper-g/n prior gives

$$p(Y \mid M) \propto \frac{a-2}{n(a+r_M-2)} F_1\left(1; \frac{a}{2}, \frac{\nu}{2}; \frac{a+r_M}{2}; 1 - \frac{1}{n}, R^2_{M\perp\emptyset}\right). \tag{15}$$

These are the formulas used in the empirical analysis. Unbalancedness changes the realized group counts, the rank penalty r_M, the residual degrees of freedom ν, and the sample sums of squares; it does not otherwise change the model-comparison argument.

A simple example. Suppose $j, t \in \{1, 2\}$ and, suppressing the horizon index, the observed support is

$$\mathcal{O} = \big\{(1,1), (1,2), (2,2)\big\}.$$

An intercept that varies jointly by participant and date has three observed groups, not the four implied by JT. After the empty $(2,1)$ cell is omitted, its dummy matrix is I_3. But if

a common activity slope is appended, the design is $[I_3\ x]$, which has four columns and rank only three because x lies in the span of the group dummies. The observed group count is therefore correct, but the slope is not separately identified.

Multiple horizons can break this collinearity. With a group-specific intercept α_{jt} and a common activity response γ, two horizons for an observed participant-date group imply

$$R_{jh_2t} - R_{jh_1t} = \gamma\big(x_{jh_2t} - x_{jh_1t}\big) + \big(\varepsilon_{jh_2t} - \varepsilon_{jh_1t}\big).$$

Whenever activity varies within at least one such group, the within-group difference contains information about γ. The empirical implementation follows this logic directly: it constructs groups from $\mathcal{O}$, omits empty dummy columns, forms the realized design, and retains only full-rank specifications. The marginal likelihood then penalizes the resulting identified rank r_M, rather than a notional Cartesian-product count.

B Partitions over Survey Dates and Forecast Horizons

The pooling-pattern framework in the main text allows a coefficient to be either constant or fully heterogeneous along a given dimension. We extend this framework to allow intermediate pooling through changepoint partitions of survey dates and forecast horizons. Date and horizon partitions may enter separately or jointly. The same date partition is shared by all coefficient blocks that use date regimes, and the same horizon partition is shared by all blocks that use horizon groups. We denote the corresponding grouping maps by $r_t(t)$ and $r_h(h)$.

Admissible partitions. A date partition with B_t breaks is characterized by breakpoints

$$0 = \tau_0 < \tau_1 < \cdots < \tau_{B_t} < \tau_{B_t+1} = T,$$

and the associated grouping map

$$r_t(t) = b \quad \Longleftrightarrow \quad \tau_{b-1} < t \leq \tau_b, \qquad b = 1, \ldots, B_t + 1.$$

We restrict $B_t \leq \overline{K}$ and impose a minimum regime length through

$$\tau_b - \tau_{b-1} \geq m.$$

In all three applications, $\overline{K} = 4$, allowing at most five date regimes. We set $m = 24$ for the monthly Blue Chip panel, $m = 8$ for the quarterly SPF panel, and $m = 4$ for the SEP.

A horizon partition is defined in the same way:

$$0 = \kappa_0 < \kappa_1 < \cdots < \kappa_{B_h} < \kappa_{B_h+1} = H,$$
$$r_h(h) = b \quad \Longleftrightarrow \quad \kappa_{b-1} \leq h < \kappa_b, \qquad b = 1, \ldots, B_h + 1.$$

We restrict $B_h \leq H-1$. A coefficient block with horizon dependence may use the shared map $r_h(h)$ or retain unrestricted horizon-level variation.

The one-group horizon partition corresponds to complete pooling across horizons, while the singleton partition corresponds to unrestricted horizon heterogeneity. Along the date dimension, a one-regime partition similarly gives complete pooling across dates for any block that uses it; unrestricted date heterogeneity remains available through the original date index t. Each distinct grouping map is counted only once.

Extended pooling patterns. The joint searches use the following dataset-specific baseline dimension sets:

$$\begin{aligned} \text{Blue Chip:} \quad & (S^0_\alpha, S^0_\gamma, S^0_\beta) &&= (\{h,t\}, \{h\}, \{t\}), \\ \text{SPF:} \quad & (S^0_\alpha, S^0_\gamma, S^0_\beta) &&= (\{t\}, \{h\}, \{t\}), \\ \text{SEP:} \quad & (S^0_\alpha, S^0_\gamma, S^0_\beta) &&= (\{h,t\}, \{t\}, \{h\}). \end{aligned}$$

For each coefficient block, the search may add date-regime dependence or replace unrestricted date dependence with the shared regime map. Thus a block with $t \in S^0_c$ may replace t by $r_t(t)$, while a block without date dependence may add $r_t(t)$. Independently, every occurrence of h in S^0_c is transformed by the shared horizon partition: it disappears under complete horizon pooling, becomes $r_h(h)$ under an intermediate partition, and remains h under the singleton partition. The joint extension does not introduce participant dependence that is absent from the corresponding baseline pattern.

All blocks that use date regimes share the same $r_t(t)$, but another block may remain fully date specific. For example,

$$R_{jht} = \alpha_t + \gamma_{h,r_t(t)} x_{jht} + \beta_{r_t(t)} \pi_{jht} + \varepsilon_{jht}$$

is admissible: the intercept remains date specific, while the activity and inflation coefficients use the same date partition.

Date and horizon grouping may enter the same coefficient block. For example,

$$\gamma_{r_h(h),r_t(t)}$$

allows the activity response to vary jointly across horizon groups and date regimes. A one-group horizon partition leaves only date-regime variation, while a one-regime date partition leaves only horizon-group variation.

Let $\widetilde{S}_c$ denote the dimensions for coefficient c after applying these date and horizon transformations, and let $D_{\widetilde{S}_c}$ be its dummy matrix. The design matrix for an extended pooling pattern M is

$$X_M = \left[D_{\widetilde{S}_\alpha}, \operatorname{diag}(x) D_{\widetilde{S}_\gamma}, \operatorname{diag}(\pi) D_{\widetilde{S}_\beta} \right].$$

The number of coefficients in block c is

$$k_c = \left|\left\{\text{proj}_{\widetilde{S}_c}(j,h,t) : (j,h,t) \text{ is observed}\right\}\right|,$$

and

$$k_M = k_\alpha + k_\gamma + k_\beta.$$

As in the main analysis, we restrict attention to models satisfying

$$\text{rank}(X_M) = k_M.$$

Marginal likelihoods. Every extended pooling model continues to contain the constant-coefficient Taylor rule

$$X_\emptyset = [\iota, x, \pi]$$

as a subspace. Here, $k_\emptyset = \text{rank}(X_\emptyset) = 3$. The common-rule decomposition and joint-contrast g-prior therefore apply without modification. In particular, the two model-specific quantities entering the marginal likelihood remain

$$r_M = \text{rank}(X_M) - k_\emptyset$$

and

$$R^2_{M\perp\emptyset} = 1 - \frac{\text{RSS}_M}{\text{RSS}_\emptyset}.$$

Conditional on a particular collection of date and horizon partitions, the fixed-g and hyper-g/n marginal likelihoods are consequently given by the expressions in the main text after replacing the original pooling pattern by the corresponding extended pattern M.

Prior over partition structures. Let $\mathcal{R}_t(B_t)$ denote the set of admissible date partitions containing B_t breaks, and let $\mathcal{R}_h(B_h)$ denote the analogous set of horizon partitions. The empirical searches assign equal prior weight to every unique admissible restriction–partition model. Equivalently, conditional on a break count, break locations are uniform,

$$p(r_t \mid B_t) = \frac{1}{|\mathcal{R}_t(B_t)|}, \qquad p(r_h \mid B_h) = \frac{1}{|\mathcal{R}_h(B_h)|},$$

while the induced marginal priors over break counts satisfy

$$p(B_t) \propto |\mathcal{R}_t(B_t)|, \qquad p(B_h) \propto |\mathcal{R}_h(B_h)|.$$

Thus the aggregate prior over the number of regimes or horizon groups is not uniform: counts admitting more partitions receive more prior mass. We set both the additional model-complexity penalty and the break-count penalty to zero. A shared partition is counted only once even when several coefficient blocks use it, and a partition that no block uses is not

part of the model. These conventions prevent observationally equivalent specifications from receiving prior probability multiple times.

Computation. The extension remains computationally tractable because dates and horizons are ordered. We first calculate the relevant cross-products at the most disaggregated date–horizon level. The cross-products for any contiguous group can then be obtained by summing these stored quantities, without reconstructing the observation-level design matrix. We enumerate all admissible horizon partitions and, conditional on each one, score every admissible date partition and coefficient-block restriction pattern. When the regression criterion is additive across regimes, candidate date partitions are evaluated in vectorized chunks using the stored regime-level cross-products. When some coefficients remain common across regimes, the same quantities are combined using a low-dimensional Schur complement. Each evaluation yields $\text{rank}(X_M)$ and RSS_M, the two model-specific inputs to the marginal likelihood. Posterior probabilities are normalized over the full enumerated model space, and probabilities reported for a regime count or horizon partition sum posterior mass over all models in the corresponding class.

C Robustness

The robustness exercises below isolate four choices that can affect the results: the prior scale for coefficient heterogeneity, the centering of inflation, the SEP activity measure, and the treatment of the SEP intercept. The first three comparisons change one design choice at a time and otherwise use the same variable construction and pooling-pattern search. The activity-measure comparison reports the small sample difference induced by the availability of longer-run unemployment projections. The final exercise instead fixes each participant-round intercept at a directly reported long-run value and repeats the pooling-pattern search over the remaining response coefficients.

C.1 Fixed $g = n$ Prior

The hyper-g/n prior used in the main analysis lets the scale of coefficient heterogeneity adapt to each candidate model. A conventional unit-information alternative fixes that scale at $g = n$. This is a demanding comparison in the present application because the posterior means of g under the hyper-g/n prior are far below the sample size. Fixing $g = n$ therefore places substantially more prior mass on large departures from the common rule and imposes a correspondingly larger marginal-likelihood penalty on high-dimensional pooling patterns.

Table 1 rescans the full admissible model set under $g = n$ for both the static and inertial rules. The static Blue Chip and SEP modal patterns are unchanged. For the SPF, fixing $g = n$ replaces the horizon-date intercept with a date intercept while leaving the response-coefficient patterns unchanged. The inertial selections move more: the fixed prior favors simpler direct-response and persistence blocks in all three panels. Even there, however, the alternative prior does not select participant-specific activity or inflation responses. The broad conclusion is therefore stable, although the exact allocation of heterogeneity across intercepts, response coefficients, and persistence is prior-sensitive.

Table 1: Sensitivity of the selected pooling pattern to fixing $g = n$

		Hyper-g/n			Fixed $g = n$	
Sample	n	Modal pattern	$p(M^\star \mid Y)$	$\mathrm{E}[g \mid M^\star, Y]$	Modal pattern	$p(M^\star \mid Y)$
Panel A. Static rule						
Blue Chip	119,435	$(\{h,t\},\{t\},\{t\})$	1.000	935	$(\{h,t\},\{t\},\{t\})$	1.000
SPF	7,409	$(\{h,t\},\{h\},\{t\})$	0.995	414	$(\{t\},\{h\},\{t\})$	1.000
SEP	1,782	$(\{h,t\},\{t\},\{h\})$	1.000	28.7	$(\{h,t\},\{t\},\{h\})$	0.938
Panel B. Inertial rule						
Blue Chip	119,435	$(\{j,h\},\{h,t\},\{h,t\},\{t\})$	1.000	26.0	$(\{j\},\{h\},\{h\},\{h,t\})$	1.000
SPF	7,409	$(\{j\},\{h\},\emptyset,\{h,t\})$	0.670	23.8	$(\{h,t\},\{h\},\emptyset,\{h\})$	0.549
SEP	1,782	$(\{j,h\},\{h,t\},\{t\},\{t\})$	0.951	19.3	$(\{t\},\{t\},\emptyset,\{t\})$	0.987

Notes: A modal pattern reports $(S_\alpha, S_\gamma, S_\beta)$ for the static rule and $(S_\alpha, S_\gamma, S_\beta, S_\rho)$ for the inertial rule. Each S_c lists the dimensions along which coefficient c varies: j, t, and h denote respondent, survey date, and forecast horizon, while $\emptyset$ denotes a common coefficient.

C.2 Inflation in Levels

The main specifications subtract two percentage points from inflation. Holding a pooling pattern fixed, this recentering can be absorbed by the intercept. Across pooling patterns, however, it need not be innocuous: if the intercept and inflation response vary along different dimensions, the transformation $\alpha + \beta(\pi - 2) = (\alpha - 2\beta) + \beta\pi$ can imply an intercept surface that is not available under the original restriction. We therefore repeat the complete model search with inflation in levels, leaving the sample, activity measure, and all other variable constructions unchanged.

Table 2 shows that the static Blue Chip selection is invariant to this change. The static SPF and SEP selections reallocate heterogeneity between the intercept and inflation blocks, but neither moves toward the participant-date benchmark. In contrast, the modal inertial pattern is unchanged in all three datasets. Inflation centering can therefore matter for the exact static decomposition, especially in the shorter panels, but it does not alter the paper's broader conclusion that the supported response surfaces are much lower-dimensional than the period-by-period specification.

Table 2: Sensitivity of the selected pooling pattern to inflation centering

		Centered inflation, $\pi - 2$		Inflation in levels, π	
Sample	n	Modal pattern	$p(M^\star \mid Y)$	Modal pattern	$p(M^\star \mid Y)$
Panel A. Static rule					
Blue Chip	119,435	$(\{h,t\},\{t\},\{t\})$	1.000	$(\{h,t\},\{t\},\{t\})$	1.000
SPF	7,409	$(\{h,t\},\{h\},\{t\})$	0.995	$(\{h,t\},\{h\},\{j\})$	1.000
SEP	1,782	$(\{h,t\},\{t\},\{h\})$	1.000	$(\{j,h\},\{t\},\{h,t\})$	1.000
Panel B. Inertial rule					
Blue Chip	119,435	$(\{j,h\},\{h,t\},\{h,t\},\{t\})$	1.000	$(\{j,h\},\{h,t\},\{h,t\},\{t\})$	1.000
SPF	7,409	$(\{j\},\{h\},\emptyset,\{h,t\})$	0.670	$(\{j\},\{h\},\emptyset,\{h,t\})$	0.670
SEP	1,782	$(\{j,h\},\{h,t\},\{t\},\{t\})$	0.951	$(\{j,h\},\{h,t\},\{t\},\{t\})$	0.951

Notes: A modal pattern reports $(S_\alpha, S_\gamma, S_\beta)$ for the static rule and $(S_\alpha, S_\gamma, S_\beta, S_\rho)$ for the inertial rule. Each S_c lists the dimensions along which coefficient c varies: j, t, and h denote respondent, survey date, and forecast horizon, while $\emptyset$ denotes a common coefficient.

C.3 SEP: Unemployment Gap versus Output Gap

The main SEP exercise uses the unemployment-rate gap, constructed from each participant's unemployment projection and longer-run unemployment projection. The alternative output-gap measure combines projected real GDP growth with a common real-time CBO potential-output path. This alternative facilitates comparison with the fixed-effects specification in Bauer et al. (2024), but introduces an external potential-output path that the unemployment-gap construction does not require.

Table 3 compares the two activity measures under the same inflation centering, forecast horizons, candidate pooling dimensions, and hyper-g/n prior. In the static rule, the intercept remains horizon-date specific and the inflation response remains horizon specific, while the activity response switches from date specific under the unemployment gap to participant specific under the output gap. The inertial selection changes more broadly: the unemployment-gap specification selects $(\{j,h\},\{h,t\},\{t\},\{t\})$, whereas the output-gap specification selects $(\{j\},\{t\},\{h\},\{h,t\})$. The allocation of heterogeneity across coefficient blocks is therefore sensitive to the activity measure, but neither specification restores participant-date heterogeneity in response coefficients of the period-by-period benchmark.

Table 3: Sensitivity of the selected SEP pooling pattern to the activity measure

	Unemployment gap			Output gap		
Rule	n	Modal pattern	$p(M^\star \mid Y)$	n	Modal pattern	$p(M^\star \mid Y)$
Static	1,782	$(\{h,t\},\{t\},\{h\})$	1.000	1,836	$(\{h,t\},\{j\},\{h\})$	1.000
Inertial	1,782	$(\{j,h\},\{h,t\},\{t\},\{t\})$	0.951	1,836	$(\{j\},\{t\},\{h\},\{h,t\})$	0.985

Notes: A modal pattern reports $(S_\alpha, S_\gamma, S_\beta)$ for the static rule and $(S_\alpha, S_\gamma, S_\beta, S_\rho)$ for the inertial rule. Each S_c lists the dimensions along which coefficient c varies: j, t, and h denote respondent, survey date, and forecast horizon, while $\emptyset$ denotes a common coefficient.

C.4 SEP: Participant-Reported Long-Run Real-Rate Intercepts

The flexible intercept objects selected in the main analysis absorb much of the variation in SEP policy-rate paths. As a final check, we replace the estimated intercept with an economically disciplined, participant-reported value. For participant j in projection round t, we fix the intercept at the participant's longer-run nominal federal funds rate less the participant's longer-run PCE inflation projection. This is the participant's reported longer-run real federal funds rate. We subtract this fixed offset from the projected policy rate and repeat the paper's pooling-pattern search over the unemployment-gap and inflation response coefficients, adding the persistence block for the inertial rule but never estimating an intercept block. Inflation enters in levels rather than as $\pi - 2$, and the output gap is not included.

Table 4 compares the selected patterns with the otherwise comparable inflation-in-levels searches that estimate the intercept. In the static rule, fixing the intercept changes the modal response-coefficient pattern from $(\{t\}, \{h, t\})$ to $(\{j\}, \{h, t\})$: the unemployment-gap response becomes participant specific, while the inflation response remains horizon-date specific. The fixed-intercept model receives essentially all posterior mass, but its model-averaged mean responses remain modest, at 0.464 for the unemployment gap and 0.074 for inflation. In the inertial rule, the modal remaining-coefficient pattern becomes $(\{h, t\}, \{h\}, \{h, t\})$, with posterior probability 0.787; a date-specific activity response and horizon-specific persistence pattern receives most of the remaining mass. The direct-response means again remain small.

Table 4: SEP pooling with participant-reported long-run real-rate intercepts

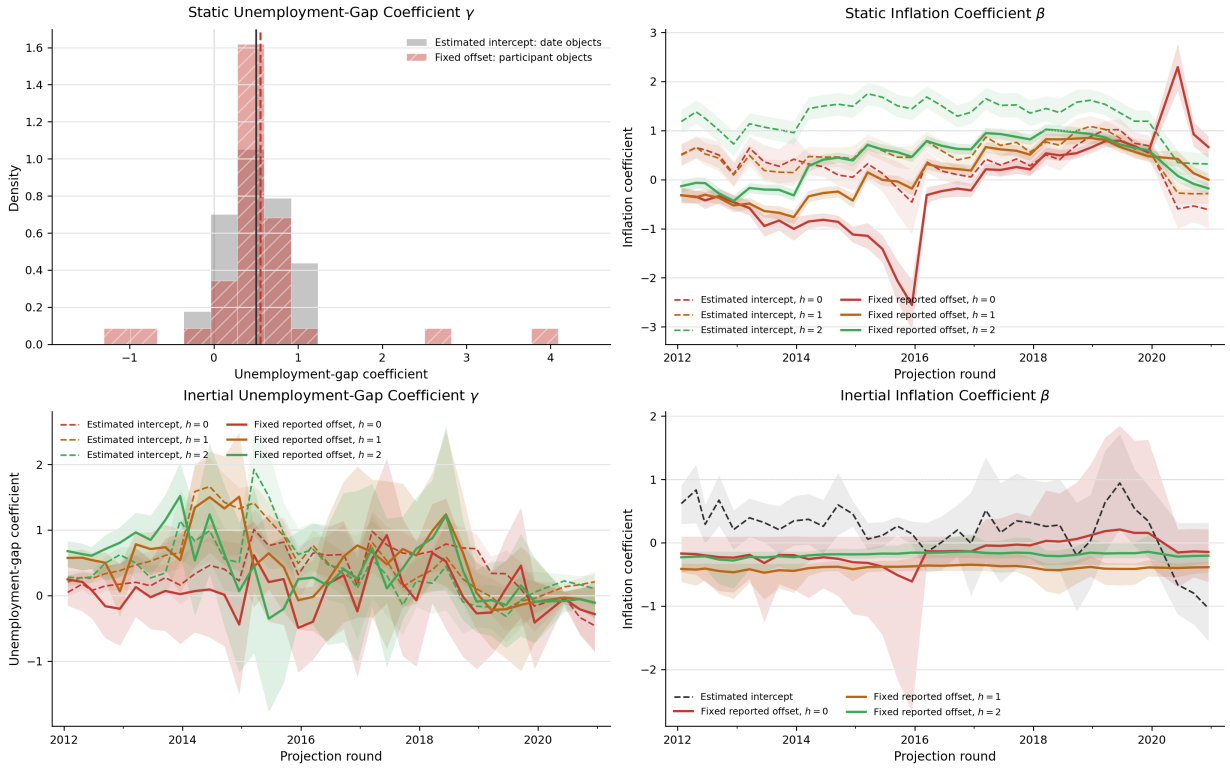
	Estimated intercept, raw inflation		Fixed long-run real-rate intercept	
Rule	Modal pattern	$p(M^\star \mid Y)$	Modal pattern	$p(M^\star \mid Y)$
Static	$(\{j, h\}, \{t\}, \{h, t\})$	1.000	$(\{j\}, \{h, t\})$	1.000
Inertial	$(\{j, h\}, \{h, t\}, \{t\}, \{t\})$	0.951	$(\{h, t\}, \{h\}, \{h, t\})$	0.787

Notes: The estimated-intercept columns report the otherwise comparable inflation-in-levels searches. Their modal patterns are $(S_\alpha, S_\gamma, S_\beta)$ for the static rule and $(S_\alpha, S_\gamma, S_\beta, S_\rho)$ for the inertial rule. Under the fixed-intercept specification, the intercept for participant j in projection round t is fixed at $\bar{R}_{jt} - \bar{\pi}_{jt}$, and the reported patterns omit S_α: (S_γ, S_β) and $(S_\gamma, S_\beta, S_\rho)$. Inflation enters in levels and the activity variable is the participant's unemployment gap. Each S_c lists the dimensions along which coefficient c varies: j, t, and h denote participant, projection round, and forecast horizon; $\emptyset$ denotes a common coefficient. Mean response coefficients average each model's coefficient objects over the estimation sample and then over models. The sample has 1,782 observations, 594 participant-round paths, and 37 participants.

Figure 10 compares the corresponding model-averaged response objects. Fixing the reported offset reveals substantial cross-participant dispersion in the static unemployment-gap coefficient and changes the horizon-specific inflation paths, but those inflation responses remain below one for most horizons and rounds. In the inertial rule, the offset restriction changes the timing and horizon allocation of both direct responses without producing a uniformly strong inflation response. Thus the restriction materially reallocates supported

heterogeneity, especially toward participant variation in the static unemployment-gap response, rather than simply shifting all response objects upward.

Figure 10: *Optimal pooling estimates: SEP response objects with estimated and reported long-run real-rate intercepts.*



Notes: Both specifications use the participant unemployment gap and inflation in levels; no output-gap object is included. Dashed lines and gray bars report the comparable SEP model in which the intercept remains estimated. Solid lines and hatched red bars report the model conditional on the participant-round offset $\bar{R}_{jt} - \bar{\pi}_{jt}$. Lines are model-averaged posterior means under the hyper-g/n prior, organized according to the modal pooling pattern; shading shows pointwise 95 percent posterior credible bands. The upper-left panel follows the paper's distribution convention because the fixed-offset static model selects participant-specific unemployment-gap coefficients; vertical lines denote distribution means. Dotted horizontal lines in the inflation panels mark a coefficient of one.