

# Forecast Alpha: Regime Dependence and Incremental Information Beyond Market Expectations

John E. Kaprich, CFA

August 12, 2026

## Working Paper V2.0.0

This working paper is distributed for review and discussion. Results and conclusions may be revised as additional feedback is incorporated.

### Abstract

This paper develops a real-time framework for evaluating whether inflation models contain incremental information beyond market expectations. The framework interprets model-market disagreement as a candidate active inflation signal and evaluates whether that disagreement contains information beyond the contemporaneous market benchmark. It combines real-time forecasting with a decomposition of market-relative inflation surprises into Forecast Alpha—the difference between the model forecast and the market benchmark—and Residual Surprise, then evaluates incremental information, recursive calibration, regime dependence, and subsequent market relevance.

The results indicate that the information content of inflation forecasts varies across forecast horizons and macroeconomic regimes. Forecast-encompassing regressions show that the Compact Inflation Model contributes incremental information conditional on the contemporaneous market benchmark at all evaluated horizons, while the magnitude of calibrated Forecast Alpha varies significantly across filtered macroeconomic states. Subsequent market-relevance evidence is more limited and horizon dependent, with the strongest evidence at the 12-month horizon, where Forecast Alpha exhibits statistically significant directional correspondence with subsequent breakeven repricing.

The framework provides a transparent methodology for evaluating the information content of inflation forecasts relative to market benchmarks and contributes to the literature on inflation forecasting, expectation formation, and market-based information aggregation.

## 1 Introduction

Markets are particularly sensitive to economic outcomes that depart from prior expectations. In his Nobel Prize Lecture, Milton Friedman argued that inflation’s economic effects depend not simply on

the level of inflation, but on deviations from what households, businesses, and investors previously expected. This insight has become a central principle of modern macroeconomics: economic outcomes are shaped in part by surprises relative to expectations rather than by expected outcomes alone. Because financial markets continuously incorporate new information into prices, market expectations provide a natural benchmark for evaluating whether a model contributes information beyond the prevailing market view. The central question is therefore not whether markets possess information, but whether a model contains incremental information conditional on what markets already reflect.

Inflation forecasts are produced using different information sets, economic theories, and statistical methodologies. A model may therefore disagree with contemporaneous market expectations even when both are formed from broadly available information. Such disagreement is not evidence of informational advantage by itself. It is a candidate signal whose value depends on whether it contains incremental information rather than noise.

This distinction is particularly important in financial markets. The valuation of nominal and inflation-sensitive securities depends on expectations regarding future inflation and the extent to which those expectations are already incorporated into prices. Consequently, the relevant question for an investment practitioner is often not simply whether inflation can be forecast, but whether a forecast adds information beyond the contemporaneous market benchmark. For an investment practitioner, the economic usefulness of a forecasting process ultimately depends on whether its information can improve subsequent economic or investment decisions.

Investor skepticism toward forecasting is longstanding. Writing specifically about short-term stock and bond price forecasts, Warren Buffett observed in Berkshire Hathaway’s 1980 Chairman’s Letter that “The forecasts may tell you a great deal about the forecaster; they tell you nothing about the future.”<sup>1</sup> Inflation forecasting literature provides a more formal basis for evaluation and commonly assesses models using forecast-error metrics such as root mean squared error (RMSE) and mean absolute error (MAE). These measures are essential for evaluating predictive accuracy, but they do not by themselves establish whether a model contributes information beyond what is already contained in a market benchmark. A model may improve forecast accuracy while largely reproducing the information embedded in market expectations. Forecast accuracy is therefore necessary for evaluating a model, but it is not sufficient for establishing incremental information relative to the market.

In this paper, market expectations serve as a passive inflation benchmark and model-based forecasts represent active inflation views. Raw Forecast Alpha is defined as the model’s deviation from the contemporaneous market benchmark. It does not presume that the model is superior to the market or that disagreement is informative. Rather, it identifies the candidate active signal that the empirical framework subsequently tests for incremental information, calibrates using prior realized outcomes, conditions on macroeconomic state, and evaluates for subsequent market relevance.

The information content of model–market disagreement may also vary with the prevailing

---

<sup>1</sup>Warren E. Buffett, *Berkshire Hathaway Inc. 1980 Chairman’s Letter to Shareholders*.

macroeconomic environment. Inflation persistence, expectations, labor-market conditions, policy settings, and financial conditions change over time, altering the environment in which forecasts are formed and realized. A given amount of model–market disagreement may therefore have different informational content across states. This motivates the use of regime identification as a conditioning device rather than as a separate forecasting engine.

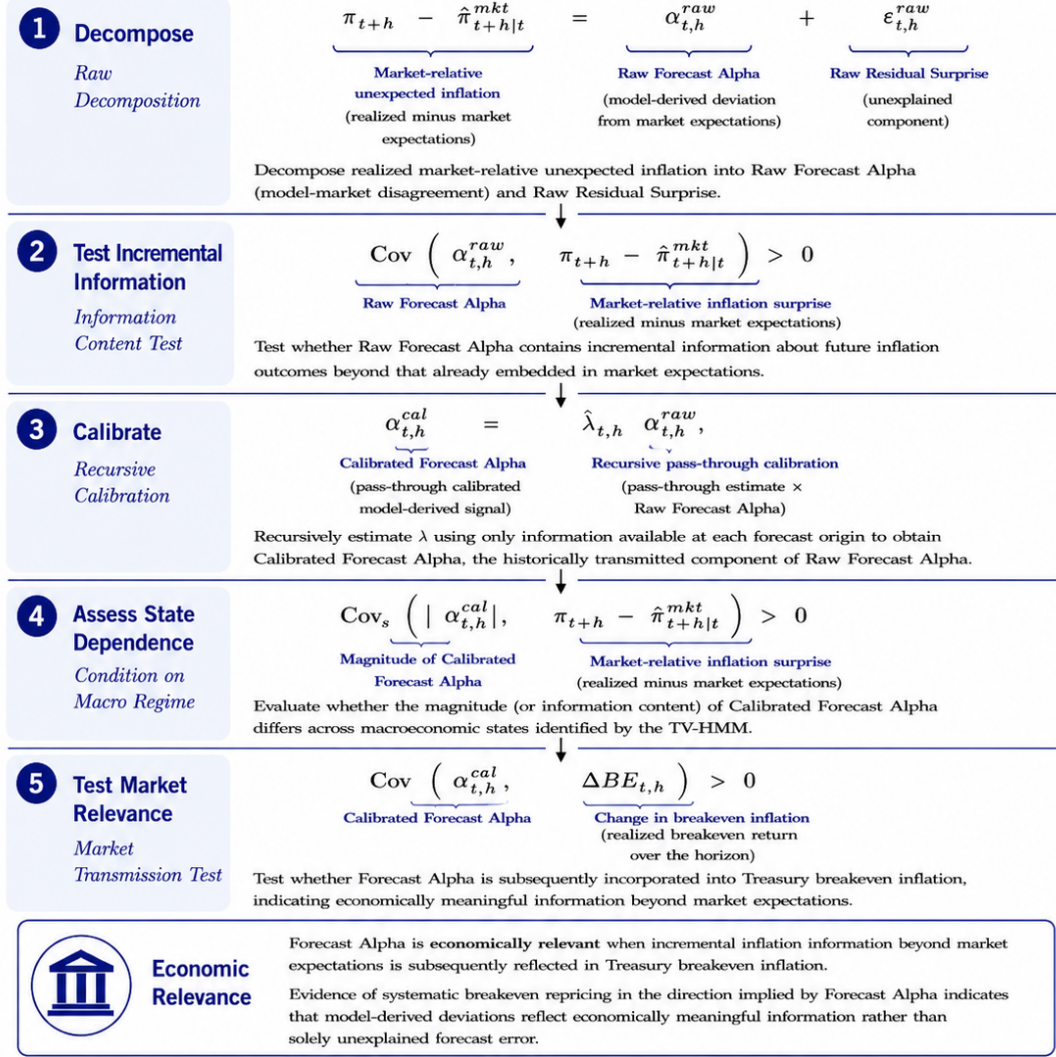
These observations motivate the central question addressed in this paper: under what conditions does model–market disagreement become informative? More specifically, this paper examines (1) whether model–market disagreement contains incremental information about subsequent inflation outcomes relative to the contemporaneous market benchmark, (2) whether the magnitude or information content of Forecast Alpha varies across macroeconomic regimes, and (3) whether Forecast Alpha is subsequently associated with market repricing.

To answer these questions, this paper develops the Forecast Alpha framework. Raw Forecast Alpha measures model–market disagreement at the forecast origin. After incremental information is evaluated, recursive calibration uses only prior realized outcomes to scale that disagreement, producing calibrated Forecast Alpha. Once the inflation outcome is realized, market-relative inflation surprise can be decomposed into calibrated Forecast Alpha and Residual Surprise. Residual Surprise is the ex post remainder after calibrated Forecast Alpha and is not itself an actionable forecasting signal. The decomposition provides an accounting framework for distinguishing the model-derived ex ante component from the portion of realized market-relative surprise that remains unexplained after calibration.

Figure 1 summarizes the conceptual structure of the Forecast Alpha framework. Raw Forecast Alpha first measures model–market disagreement observed at the forecast origin. The framework then tests whether the model contributes incremental information conditional on the contemporaneous market benchmark. Recursive calibration subsequently scales the disagreement using only prior realized outcomes. The magnitude of calibrated Forecast Alpha is then evaluated across filtered macroeconomic states, and its subsequent market relevance is assessed using breakeven repricing. After the corresponding inflation outcome is realized, Residual Surprise completes the market-relative surprise decomposition as the remaining ex post component. This sequence separates the existence of disagreement from evidence that the disagreement is informative and, separately, from evidence that it is relevant for subsequent market pricing.

## Figure 1. The Forecast Alpha Framework

A real-time framework for identifying, evaluating, calibrating, and assessing the economic relevance of model-derived incremental inflation information beyond market expectations.



*Notes:*  $\pi_{t+h}$  denotes realized inflation at horizon  $h$ ;  $\hat{\pi}_{t+h|t}^{mkt}$  denotes the market-based inflation benchmark observed at forecast origin  $t$ ;  $\alpha_{t,h}^{raw}$  denotes Raw Forecast Alpha;  $\alpha_{t,h}^{cal}$  denotes Calibrated Forecast Alpha obtained through recursive pass-through estimation;  $\varepsilon_{t,h}^{raw}$  denotes Raw Residual Surprise;  $BE$  denotes Treasury breakeven inflation; and  $\Delta BE_{t,h}$  denotes the change in Treasury breakeven inflation between the forecast origin and evaluation horizon. Calibrated Forecast Alpha is obtained through recursive pass-through estimation using only information available prior to each forecast origin.

Figure 1: Forecast Alpha Framework and Market-Relative Surprise Decomposition

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the data and market-based inflation benchmark. Section 4 presents the methodology, including the forecasting framework, regime classification, Forecast Alpha decomposition, and calibration procedure. Section 5 reports the empirical results. Section 6 evaluates the market relevance of Forecast Alpha. Section 7 concludes and discusses implications for economic decision-making and future research.

## 2 Literature Review

### 2.1 Expectations, Inflation Surprises, and Market-Based Expectations

Expectations are central to modern macroeconomics and inflation analysis. Milton Friedman argued that the economic consequences of inflation depend less on the level of inflation than on deviations from previous expectations [Friedman, 1977]. His insight motivated a deeper understanding of how unexpected inflation is measured, where expectations originate, and what information is embedded in market prices. These questions are central to understanding the influence of inflation expectations on asset prices, real returns, and the valuation of inflation-linked securities. Market-based measures, particularly Treasury breakeven inflation rates, aggregate the expectations of market participants and provide an observable benchmark for evaluating whether inflation forecasts contain information beyond contemporaneous market expectations [Ang et al., 2007, Haubrich et al., 2012, Chernov and Mueller, 2012].

Market prices are not pure measures of expected inflation. Treasury breakeven rates embed inflation risk premia, liquidity premia, and other market-specific distortions [D’Amico et al., 2018, Kajuth and Watzka, 2011, Andreasen et al., 2021, Bauer and Rudebusch, 2017]. In practice, differences between realized inflation and market-implied expectations may reflect information that was potentially forecastable *ex ante* as well as genuinely unpredictable shocks. These two sources of disagreement motivate the decomposition framework developed in this paper and raise the broader question of whether inflation forecasts contain incremental information beyond prevailing market expectations.

Despite these limitations, Treasury breakeven inflation rates remain among the most widely used market-based measures of inflation expectations because they reflect the collective inflation expectations of market participants as expressed through continuously updated market prices. Importantly, they directly influence the valuation of inflation-sensitive assets. Accordingly, breakeven inflation rates represent not only an estimate of future inflation but also the inflation outlook embedded in traded market prices. As such, they provide an observable benchmark for evaluating whether alternative inflation forecasts contain incremental information beyond prevailing market expectations.

## 2.2 Forecastable and Unforecastable Inflation Components

Early empirical approaches modeled inflation using statistical relationships, including autoregressive models and Phillips-curve frameworks [Phillips, 1958, Nelson and Schwert, 1977, Stock and Watson, 1999, Atkeson and Ohanian, 2001, Robinson et al., 2003, Gabrielyan, 2016, Martins and Gabriel, 2020, Jørgensen and Lansing, 2019]. Subsequent research showed that forecast accuracy varies substantially across periods and that no single forecasting methodology consistently dominates across economic environments [Stock and Watson, 2007, 2010, 2016, Faust and Wright, 2013]. These studies reflect the literature’s longstanding emphasis on forecast accuracy as the principal criterion for evaluating inflation forecasting models, motivating continued efforts to improve predictive accuracy through increasingly sophisticated forecasting methodologies.

Real-time forecasting studies further emphasize that forecast accuracy depends on the information set available at the forecast origin [Groen et al., 2012, Lieberman, 2012, Croushore, 2011]. Time-varying parameter models were developed in part to accommodate evolving relationships between inflation and its underlying economic determinants [Primiceri, 2005, Del Negro and Schorfheide, 2013]. These studies indicate that inflation forecastability is time varying, with the predictive content of economic information changing across periods and environments. As a result, forecasting models that assume stable relationships may fail to capture predictive information that emerges as economic conditions evolve. These findings reinforce the need for forecasting methodologies capable of adapting to evolving economic relationships.

These studies show that some portion of inflation is predictable, but they generally stop short of distinguishing between the forecastable and inherently unpredictable components of deviations between realized inflation and market expectations. As a result, relatively little attention has been devoted to evaluating whether improvements in forecast accuracy reflect genuinely incremental information or merely reduce forecast error, regardless of the forecasting methodology employed.

Recent advances in forecasting methodology have expanded the available toolkit through high-dimensional models, nonlinear methods, and regularization techniques capable of exploiting increasingly rich information sets. Regularized regression provides a particularly useful framework in data-rich settings by controlling model complexity while retaining predictive information across potentially correlated predictors. Zou and Hastie [2005] introduced the Elastic Net, which combines L1 and L2 penalties to perform shrinkage and variable selection while accommodating groups of correlated predictors. Subsequent applications of regularization and machine-learning methods have demonstrated their usefulness for macroeconomic and inflation forecasting [Kock and Medeiros, 2020, Medeiros et al., 2021, Goulet Coulombe et al., 2022, Huang et al., 2024, Naghi et al., 2024]. Yet regardless of the forecasting methodology employed, the literature continues to evaluate models primarily through forecast accuracy. Far less work has examined whether improved forecasts represent genuinely incremental information beyond market expectations.

## 2.3 Regime Dependence and State-Dependent Forecastability

The inflation surge of 2021–22 renewed interest in the possibility that inflation dynamics and forecastability may vary across macroeconomic environments, particularly during periods of rapid structural change. More broadly, a growing literature suggests that inflation is not governed by a single stable process. Inflation dynamics vary over time with changes in monetary policy, macroeconomic volatility, and broader structural conditions [Clarida et al., 2000, Sims and Zha, 2006, Cogley and Sargent, 2005, Bai and Perron, 2003]. This evidence motivates a shift toward state-dependent models of inflation.

A major step in this direction was the introduction of regime-switching models by Hamilton [1989]. By allowing the data-generating process to transition across latent states, these models capture nonlinear dynamics, structural breaks, and changes in persistence. In the inflation context, regime-switching frameworks have improved empirical fit and, in some cases, forecasting performance relative to constant-parameter alternatives [Ang et al., 2007, Bianchi, 2013]. More broadly, regime-dependent macro-finance models have helped explain time variation in inflation expectations, bond risk premia, and yield-curve behavior [Ang and Timmermann, 2012, D’Amico et al., 2018].

Subsequent research relaxed the assumption of constant transition probabilities by allowing regime dynamics to respond to observable economic conditions [Filardo, 1994, Diebold et al., 1994, Kim and Nelson, 1999]. These extensions, including time-varying Hidden Markov Models, provide a more flexible representation of macroeconomic state dynamics [Koop and Potter, 2007, Koop and Korobilis, 2013, Shi, 2016]. For inflation analysis, the implication is straightforward: forecastability itself may be state dependent, implying that both the relevance of predictors and the amount of forecastable inflation information may vary across macroeconomic environments. These studies establish that inflation dynamics and forecastability are state dependent. However, considerably less attention has been devoted to determining whether the forward-looking informational value of inflation forecasts also varies across macroeconomic regimes. Accordingly, evaluating the informational value of forecasts without accounting for regime dependence may obscure when model-derived information is most economically relevant.

## 2.4 Information Transmission and Breakeven Markets

If a forecast contains incremental information beyond prevailing market expectations, that information may subsequently be incorporated into market prices. Treasury breakeven inflation rates provide a natural setting in which to evaluate this proposition because they continuously incorporate inflation expectations and compensation for inflation risk into traded prices. The literature has extensively examined the informational content of breakeven inflation rates, their decomposition into expected inflation and risk-premium components, and their relationship to subsequent realized inflation. Related work on the anchoring of inflation expectations examines the extent to which changes in shorter-horizon expectations pass through to longer-horizon expectations. Yetman [2020], for example, evaluates pass-through from short-horizon to long-horizon inflation expectations as a measure of the degree to which longer-term expectations remain anchored. This literature establishes

an important cross-horizon transmission mechanism: changes in nearer-term inflation expectations need not remain confined to the horizon at which they originate.

The question considered here is distinct. Rather than asking whether movements in short-horizon expectations pass through to longer-horizon expectations, this study asks whether model-derived information that is not already contained in prevailing market expectations is subsequently incorporated into market pricing. Much less attention has been devoted to this question in the breakeven-inflation literature. This distinction is central to understanding whether forecast disagreement reflects genuinely incremental information or merely estimation error.

Demonstrating that a forecast contains incremental information does not by itself establish market relevance. A separate question is whether model–market disagreement is subsequently associated with changes in market pricing. Information that improves forecasts but bears no relationship to subsequent pricing may have more limited relevance for market-based investment decisions. Accordingly, the paper evaluates whether Forecast Alpha is associated with subsequent breakeven repricing as a distinct test of market relevance.

## 2.5 Present Study

The preceding literature establishes that market expectations provide an observable benchmark for evaluating inflation forecasts, that forecast accuracy alone does not establish informational value, that forecastability varies across macroeconomic regimes, and that economically relevant information may subsequently be incorporated into market prices. What remains missing is a unified framework for evaluating these questions and determining whether model-derived inflation forecasts contain forward-looking information beyond prevailing market expectations.

Existing inflation-forecasting studies primarily evaluate models by their ability to minimize forecast error relative to realized inflation. This paper instead asks whether model-based forecasts contribute information beyond market expectations. The distinction is important because a model may achieve low forecast error while largely replicating information already embedded in market prices, whereas a model with larger forecast errors may still provide decision-relevant information if its deviations from market expectations are systematically informative. Forecast Alpha formalizes this distinction by treating model–market disagreement as the object of analysis and evaluating when that disagreement contains incremental information beyond the contemporaneous market benchmark.

The contribution of this study is a framework for evaluating the forward-looking informational value of inflation forecasts by integrating market expectations, regime dependence, and market repricing into a unified approach. The analysis shifts the focus from forecast accuracy alone toward the economic relevance of model-derived inflation information. It does so by treating market expectations as the benchmark and model–market disagreement as the active signal to be evaluated.

## 3 Data

### 3.1 Inflation Target

The inflation target is year-over-year Core PCE inflation, constructed from the Core Personal Consumption Expenditures Price Index excluding food and energy (FRED series PCEPILFE; U.S. Bureau of Economic Analysis, 2026). Core PCE is used because it captures more persistent inflation dynamics than headline inflation and is commonly interpreted as a measure of the underlying inflation trend. It is also closely watched by the Federal Reserve in assessing inflation pressures and conducting monetary policy, making it relevant for evaluating model-based inflation information against market expectations and inflation-sensitive asset pricing [Federal Reserve Bank of Atlanta, 2026]. Forecasts are generated at the 1-, 3-, 6-, and 12-month horizons.

### 3.2 Predictor Variables

Predictor variables are organized around macro-financial channels through which inflation dynamics are commonly transmitted, including labor-market conditions, economic activity, monetary policy, financial conditions, commodity prices, and inflation expectations. The underlying series are obtained from the Federal Reserve Economic Data (FRED) database. Where appropriate, derived features are constructed from these underlying series to capture changes, spreads, lags, growth rates, and other inflation-relevant dynamics used in the forecasting specification.

The empirical forecasts reported in the main paper are generated using the retained Compact-Direct specification described in Section 4. Appendix Table 6 documents the broader development universe of predictor variables and derived features from which the Compact-Direct specification was established. Additional appendix tables report feature-stability and sign-interpretability diagnostics for the retained specification.

### 3.3 Market-Based Inflation Benchmark

With the inflation target and predictor information set defined, the framework next requires a market benchmark against which model forecasts can be evaluated. This paper uses the 5-Year Treasury Inflation Breakeven Rate (FRED series T5YIE) as the market-based inflation benchmark. Treasury breakeven inflation is derived from the difference between nominal Treasury yields and Treasury Inflation-Protected Securities (TIPS) yields and reflects inflation compensation embedded in financial markets.

Breakeven inflation is not a pure measure of expected inflation. It is linked to CPI-based securities rather than Core PCE inflation directly and may reflect inflation risk premia, liquidity effects, and other market-specific distortions. These limitations are important, but they do not eliminate its usefulness for this paper. Forecast Alpha is not designed to compare model forecasts against a perfectly measured expectation of future Core PCE inflation. It is designed to evaluate whether model-based inflation information differs from, and contains incremental information beyond, the

inflation outlook embedded in market pricing.

The use of Core PCE as the forecast target and Treasury breakevens as the market benchmark therefore creates an intentional market-relative comparison rather than a pure measurement exercise. The model forecasts Core PCE inflation, while the benchmark captures inflation compensation derived from market yields on nominal Treasury and TIPS securities. Differences arising from inflation definitions, forecast horizons, risk premia, liquidity conditions, or unanticipated inflation shocks may therefore contribute to Residual Surprise or affect the subsequent breakeven-repricing tests. The empirical framework does not assume that these components are mechanically separable *ex ante*.

Treasury breakevens are selected because they provide a transparent, observable, and market-based benchmark with a long historical record and direct relevance to inflation-sensitive fixed-income markets. Alternative measures, such as survey-based expectations or inflation swaps, provide useful information but differ in observability, investability, horizon availability, or historical coverage. The benchmark selection therefore reflects the paper's objective of evaluating model-derived inflation information against a market price rather than against a purely statistical or survey-based expectation measure.

Figure 2 illustrates the relationship between realized Core PCE inflation and the market-based inflation anchor used throughout the analysis.

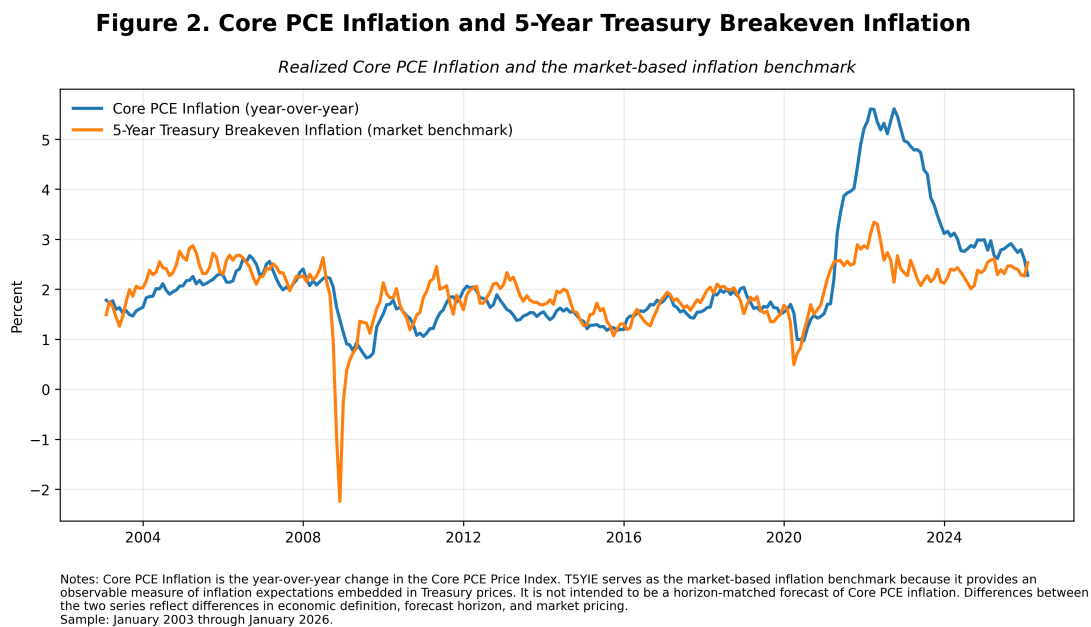


Figure 2: Core PCE inflation and the 5-year Treasury Inflation Breakeven Rate used as the market-based inflation anchor.

### 3.4 Data Frequency and Real-Time Alignment

The empirical implementation is monthly and uses month-end timestamps throughout. Variables are aligned to the forecast origin so that future observations are excluded and publication lags are respected in constructing forecasts for time  $t + h$ . This design approximates the information set available at the forecast origin while, as discussed below, not constituting a complete real-time data-vintage reconstruction. The same timing convention is applied to forecast generation, filtered regime classification, and model evaluation.

Where necessary, lower-frequency macroeconomic series are aligned to the monthly forecast calendar using the most recently available observation at the forecast origin. Publication-lag discipline is imposed where possible so that predictors enter the forecasting panel according to the specified publication-lag conventions. Because the implementation relies on FRED historical series rather than a full vintage database, the real-time design should be interpreted as forecast-origin aligned rather than as a complete real-time data-vintage reconstruction.

The unified monthly data panel begins in January 1978 and extends through December 31, 2025, the formal endpoint of the empirical study. The start date is selected so that the predictor panel is complete after the required transformations and lag construction, avoiding the need for imputation or unbalanced-panel adjustments. The effective evaluation sample varies across exercises depending on forecast horizon, market-benchmark availability, and calibration-window requirements. The full data period encompasses multiple inflation regimes, including the high-inflation environment of the late 1970s and early 1980s, the subsequent disinflation period, the low and stable inflation environment of the 2010s, and the inflation surge following 2020.

### 3.5 Data Integrity and Reproducibility

All series are obtained directly from FRED and cached locally to support reproducibility. When the data panel is refreshed, revised historical observations available through FRED are incorporated into the cached dataset. The data pipeline is designed to preserve forecast-origin information availability by preventing the inclusion of future-dated observations and supporting recursive model estimation. Model forecasts, market benchmarks, and filtered regime labels are aligned at the forecast origin, while realized inflation outcomes are joined only at the corresponding evaluation date. This timing convention preserves the ex-ante construction of Forecast Alpha while allowing Residual Surprise and subsequent market outcomes to be evaluated ex post.

Together, these data conventions define the forecast-origin panel from which the forecasting model, regime classification, Forecast Alpha decomposition, and breakeven repricing tests are implemented in the methodology that follows.

## 4 Methodology

Using information available at each forecast date, the empirical framework generates model-based inflation forecasts, compares those forecasts with a market-based inflation benchmark, evaluates

whether the model contains information beyond that benchmark, calibrates the resulting model–market disagreement in real time, conditions the calibrated signal on macroeconomic regimes, and tests whether the signal is subsequently reflected in breakeven repricing.

The methodology therefore proceeds through seven connected stages: construction of the real-time information set, recursive horizon-specific inflation forecasting, selection of the market benchmark, definition and validation of raw Forecast Alpha, recursive calibration of Forecast Alpha, regime-conditioned evaluation, and market-relevance testing. The forecasting model and regime-classification model are deliberately separated: the Compact-Direct model generates the inflation forecast, while the TV-HMM provides a real-time filtered state-conditioning layer used to evaluate when Forecast Alpha is most economically meaningful.

#### 4.1 Real-Time Information Set and Predictor Construction

The forecasting panel is constructed at each monthly forecast origin using only information available at that date. Forecast-origin features are drawn from a predefined set of economically motivated transformations applied to Core PCE inflation and selected macro-financial series. The Compact-Direct design combines an inflation-persistence backbone with a compact set of exogenous predictors intended to capture labor-market conditions, policy and rate dynamics, price pressures, and inflation expectations.

For each forecast origin, the predictor set is constructed from variables observable at, or lagged to, time  $t$ . The current Compact-Direct implementation uses the following feature blocks:

1. *Inflation persistence block*: lagged year-over-year Core PCE inflation, recent inflation changes, a short rolling inflation mean, and short rolling inflation volatility;
2. *Labor-market block*: unemployment level, an unemployment gap measure, and recent changes in unemployment;
3. *Policy and rates block*: the effective federal funds rate, the 2-year Treasury yield, changes in these rates, and a 2-year Treasury less federal funds rate spread;
4. *Price-pressure block*: producer-price and oil-price changes;
5. *Expectations block*: survey-based inflation-expectation information and recent changes in that measure.

All predictors are aligned to month-end timestamps. Publication-lag discipline is imposed where available so that the forecasting system approximates the information set observable at the forecast origin. The transformations depend only on contemporaneously available or lagged observations and include lags, differences, rolling moments, and selected spreads. The compact predictor set is fixed before recursive estimation rather than selected through a separate automated feature-selection procedure. Variable shrinkage occurs through the Elastic Net penalty during model estimation.

The resulting design is intentionally compact. The objective is not to approximate a large machine-learning feature universe or to perform regime-dependent feature selection, but to estimate a disciplined inflation forecast that complements persistence with a limited amount of additional macro-financial information. This structure preserves interpretability while maintaining the real-time information discipline required for the market-relative tests that follow.

## 4.2 Compact-Direct Inflation Forecast

Let  $\pi_t$  denote year-over-year Core PCE inflation at time  $t$ . In the empirical implementation,

$$\pi_t = 100 \left( \frac{P_t}{P_{t-12}} - 1 \right),$$

where  $P_t$  denotes the level of the Core PCE price index. For each forecast origin  $t$  and horizon  $h \in \{1, 3, 6, 12\}$  months, the objective is to form a direct forecast of the future level of year-over-year Core PCE inflation:

$$\hat{\pi}_{t+h|t}^{\text{model}} = E(\pi_{t+h} | \mathcal{I}_t),$$

where  $\mathcal{I}_t$  denotes the information set available at the forecast origin.

The empirical implementation uses a direct horizon-specific forecasting design rather than an iterated multi-step design. A separate forecast equation is estimated for each horizon. Direct forecasting avoids compounding specification error across intermediate steps and aligns naturally with the paper's market-relative framework, which evaluates the model's incremental information separately at each forecast horizon. This design is consistent with the broader real-time forecasting literature, which emphasizes that forecast construction and evaluation should respect the information set available at the forecast origin [Croushore, 2011, Faust and Wright, 2013, Groen et al., 2012].

Let  $X_{t,h}$  denote the horizon-specific predictor vector available at time  $t$ . The forecast equation is

$$\hat{\pi}_{t+h|t}^{\text{model}} = f_h(X_{t,h}),$$

where  $f_h(\cdot)$  denotes the horizon-specific forecasting rule. The canonical forecast engine is an Elastic Net regression estimated recursively using an expanding window. For each horizon  $h$ , the model solves

$$\hat{\beta}_{t,h} = \arg \min_{\beta_0, \beta} \left\{ \frac{1}{N_{t,h}} \sum_{\tau \in \mathcal{T}_{t,h}} (\pi_{\tau+h} - \beta_0 - X'_{\tau,h} \beta)^2 + \kappa_{t,h}^{EN} \left[ \frac{1 - \rho_{t,h}}{2} \|\beta\|_2^2 + \rho_{t,h} \|\beta\|_1 \right] \right\},$$

where  $\mathcal{T}_{t,h}$  denotes the expanding-window training sample available at forecast origin  $t$ , restricted to observations whose horizon-specific target values would have been observable by time  $t$ .  $N_{t,h}$  is the number of usable training observations,  $\kappa_{t,h}^{EN}$  is the Elastic Net penalty strength, and  $\rho_{t,h}$  is the mixing parameter between ridge and LASSO penalty components.

Elastic Net regularization is appropriate because the compact macro-financial predictor set

contains correlated variables, including inflation lags, rate variables, labor-market indicators, and price-pressure measures. The Elastic Net penalty combines the shrinkage properties of ridge regression with the sparse-selection properties associated with the LASSO penalty [Tibshirani, 1996, Zou and Hastie, 2005]. The implementation does not estimate a separate pure LASSO model; references to LASSO therefore refer to the  $\ell_1$  component of the Elastic Net penalty rather than to a distinct forecast engine.

Predictors are standardized within each expanding-window training sample, and the same transformation is applied to the predictor vector observed at the forecast origin. Elastic Net tuning parameters are selected within the available training sample using a recent validation block. The selected specification is then refit on the full available expanding-window sample before generating  $\hat{\pi}_{t+h|t}^{\text{model}}$ .

A key methodological distinction is that the forecast engine is not a weighted mixture of regime-specific forecasting models. The Compact-Direct model is the paper’s canonical forecast equation. The TV-HMM operates separately as a classification layer used to organize conditional analysis after the forecast and Forecast Alpha measures have been constructed.

### 4.3 Market-Based Inflation Benchmark

Let  $\hat{\pi}_{t+h|t}^{\text{mkt}}$  denote the market-based inflation benchmark associated with forecast origin  $t$  and horizon  $h$ . The empirical implementation uses the 5-year Treasury Inflation Breakeven Rate (T5YIE) as the market-based benchmark. T5YIE is derived from nominal Treasury and TIPS pricing and is therefore linked to CPI-based inflation compensation, whereas the forecasting model targets Core PCE inflation. The two series should therefore be interpreted as related but nonidentical inflation objects.

Because a clean horizon-matched market-implied Core PCE inflation series is not available for each horizon considered in the paper, T5YIE is not interpreted as a literal horizon-specific forecast of Core PCE inflation. Instead, it provides an observable market-pricing reference against which the model’s incremental information is evaluated. The accompanying Core PCE–T5YIE figure documents the broad historical relationship between the forecast target and the market benchmark without treating them as equivalent measures.

This interpretation is consistent with the literature showing that breakeven inflation rates contain market information about inflation expectations but may also embed inflation risk premia, liquidity premia, and other market-specific distortions [Ang et al., 2007, Haubrich et al., 2012, Chernov and Mueller, 2012, D’Amico et al., 2018, Kajuth and Watzka, 2011, Andreasen et al., 2021, Bauer and Rudebusch, 2017]. The paper therefore evaluates whether model forecasts contain information relative to market inflation pricing rather than claiming that T5YIE is a pure measure of expected Core PCE inflation at each forecast horizon.

The selection of Treasury breakeven inflation reflects a tradeoff among economic relevance, real-time observability, linkage to traded market prices, and historical coverage. Alternative measures of inflation expectations, including survey-based forecasts and inflation swap markets, provide useful

information but differ with respect to publication frequency, market-price linkage, and sample availability for recursive out-of-sample analysis.

Table 1: Benchmark Selection Considerations for Market-Relative Inflation Analysis

| Benchmark                          | Market-Based | Update Frequency | Linked to Traded Prices | Historical Coverage |
|------------------------------------|--------------|------------------|-------------------------|---------------------|
| Survey of Professional Forecasters | No           | Quarterly        | No                      | High                |
| Blue Chip Forecasts                | No           | Monthly          | No                      | Moderate            |
| Inflation Swaps                    | Yes          | Yes              | Yes                     | Moderate            |
| Treasury Breakevens                | Yes          | Yes              | Yes                     | High                |

Notes: The table summarizes broad benchmark characteristics relevant to recursive real-time inflation forecasting and market-relative evaluation. Historical coverage reflects practical sample availability for recursive out-of-sample analysis rather than absolute data existence.

This benchmark choice provides the passive market reference point used in the Forecast Alpha framework that follows.

#### 4.4 Raw Forecast Alpha and Market-Relative Surprise

Raw Forecast Alpha is defined as the model’s ex ante deviation from the market benchmark:

$$\alpha_{t,h}^{raw} \equiv \hat{\pi}_{t+h|t}^{\text{model}} - \hat{\pi}_{t+h|t}^{\text{mkt}}.$$

Positive raw Forecast Alpha indicates that the model anticipates higher inflation than implied by contemporaneous market pricing, while negative raw Forecast Alpha indicates lower anticipated inflation. At this stage, raw Forecast Alpha should be interpreted as model–market disagreement rather than as validated incremental information. Disagreement may reflect genuine inflation information, market-pricing wedges, model error, or some combination of these forces.

Market-relative unexpected inflation is defined as

$$s_{t,h} \equiv \pi_{t+h} - \hat{\pi}_{t+h|t}^{\text{mkt}}.$$

Raw Residual Surprise is

$$\varepsilon_{t,h}^{raw} \equiv \pi_{t+h} - \hat{\pi}_{t+h|t}^{\text{model}},$$

so that

$$s_{t,h} = \alpha_{t,h}^{raw} + \varepsilon_{t,h}^{raw}.$$

This identity separates model–market disagreement observed at the forecast origin from the portion of the subsequent market-relative inflation outcome not captured by the model. The identity alone does not establish that raw Forecast Alpha is informative, that it should be taken one-for-one, or that the components are orthogonal. Those propositions are evaluated empirically.

## 4.5 Incremental Information Beyond Market Expectations

The first empirical requirement for Forecast Alpha is that the model contain predictive information beyond that already embedded in the market benchmark. This is evaluated using horizon-specific forecast-encompassing regressions of realized inflation on both the market benchmark and the model forecast:

$$\pi_{t+h} = a_h + b_{m,h}\hat{\pi}_{t+h|t}^{\text{mkt}} + b_{f,h}\hat{\pi}_{t+h|t}^{\text{model}} + u_{t,h}.$$

The coefficient  $b_{f,h}$  measures whether the model forecast retains predictive information after conditioning on the market benchmark. A statistically informative model coefficient provides evidence that model–market disagreement is not merely arbitrary disagreement and establishes the empirical basis for treating raw Forecast Alpha as a candidate active signal.

The encompassing regression does not imply that the full magnitude of raw Forecast Alpha should be incorporated one-for-one into an active inflation view. It addresses a prior question: whether the model adds information conditional on the market. The amount of weight assigned to that disagreement is determined separately through recursive calibration.

## 4.6 Recursive Calibration of Forecast Alpha

Once model–market disagreement has an empirical information basis, the framework estimates how strongly that disagreement has historically mapped into subsequently realized market-relative inflation outcomes. For each horizon  $h$ , the calibration regression is estimated using only prior observations whose realized outcomes were available before the current forecast origin:

$$s_{\tau,h} = a_{t,h} + \lambda_{t,h}\alpha_{\tau,h}^{\text{raw}} + u_{\tau,h}, \quad \tau \in \mathcal{H}_{t,h},$$

where  $\mathcal{H}_{t,h}$  denotes the historical calibration sample available at forecast origin  $t$  for horizon  $h$ .

The coefficient  $\lambda_{t,h}$  is the horizon-specific recursive calibration weight. It is not interpreted as a structural inflation parameter, a fixed population coefficient, or a percentage of forecasts that are “correct.” Instead, it summarizes the historical mapping between raw model–market disagreement and subsequently realized market-relative inflation outcomes using only information available before the current forecast origin.

Calibrated Forecast Alpha is defined as

$$\alpha_{t,h}^{\text{cal}} \equiv \hat{\lambda}_{t,h}\alpha_{t,h}^{\text{raw}}.$$

A calibration weight of one leaves raw disagreement unchanged; a value below one attenuates the magnitude of the disagreement, while a value above one amplifies it. Because  $\hat{\lambda}_{t,h}$  is estimated recursively, the framework allows the historically supported magnitude of Forecast Alpha to evolve through time and across forecast horizons without using future realizations.

The calibration regression includes an intercept, but the intercept is not incorporated into calibrated Forecast Alpha. This convention keeps Forecast Alpha centered on directional model–

market disagreement rather than embedding a constant historical bias in the active signal. The intercept is retained as a calibration diagnostic.

The calibrated signal provides an active-passive interpretation of the forward inflation view:

$$\hat{\pi}_{t+h|t}^{FA} = \hat{\pi}_{t+h|t}^{\text{mkt}} + \alpha_{t,h}^{\text{cal}}.$$

Equivalently,

$$\hat{\pi}_{t+h|t}^{FA} = \hat{\pi}_{t+h|t}^{\text{mkt}} + \hat{\lambda}_{t,h} \left( \hat{\pi}_{t+h|t}^{\text{model}} - \hat{\pi}_{t+h|t}^{\text{mkt}} \right).$$

Market pricing therefore supplies the passive component, while calibrated Forecast Alpha supplies the active model-relative component.

Calibrated Residual Surprise is defined as

$$\eta_{t,h} \equiv s_{t,h} - \alpha_{t,h}^{\text{cal}},$$

which yields the calibrated decomposition

$$s_{t,h} = \alpha_{t,h}^{\text{cal}} + \eta_{t,h}.$$

Residual Surprise is interpreted as the remaining market-relative inflation outcome after the recursively calibrated Forecast Alpha component has been removed. Its behavior provides a diagnostic of the effectiveness of calibration, while Forecast Alpha remains the primary object of economic interest.

## 4.7 Regime Classification

The regime layer is used to evaluate whether the magnitude and relevance of calibrated Forecast Alpha depend on the prevailing macroeconomic environment. It is deliberately separate from the Compact-Direct forecast engine and does not generate the inflation forecast.

A preliminary analysis of the historical Core PCE inflation distribution is used to select the latent-state architecture and provide economic context for the state space. Candidate Gaussian and Student- $t$  finite-mixture specifications over alternative values of  $K$  are compared using the Bayesian Information Criterion. The minimum-BIC specification determines the governed number of TV-HMM states without reference to Forecast Alpha outcomes. The resulting inflation bands also provide an initialization candidate and a descriptive summary of nonlinear structure in historical inflation, but their estimated thresholds do not mechanically assign observations to latent states. State membership is instead determined recursively from filtered TV-HMM posterior probabilities. Detailed mixture specifications, boundary construction, and architecture-selection evidence are reported in the Appendix.

Macroeconomic regimes are identified using a time-varying Hidden Markov Model with  $K = 4$  latent states. The governed four-state architecture is the result of the preceding BIC-based state-count selection procedure. Let  $S_t \in \{1, \dots, 4\}$  denote the latent regime at time  $t$ . The TV-HMM is

estimated on a compact macro-financial state vector that includes inflation conditions, recent inflation changes, rolling inflation moments, selected labor-market indicators, short-rate and yield-curve variables, and commodity-price changes.

The latent state process follows a first-order Markov structure with time-varying transition probabilities:

$$P(S_t = j \mid S_{t-1} = i, Z_t) = p_{ij}(t),$$

where  $Z_t$  denotes the regime-detection information available at time  $t$ . For each prior state  $i$ , transition probabilities are modeled through a multinomial-logit mapping:

$$p_{ij}(t) = \frac{\exp(\gamma'_{ij} Z_t)}{\sum_{\ell=1}^K \exp(\gamma'_{i\ell} Z_t)}.$$

This specification follows the regime-switching tradition introduced by Hamilton [1989] and later time-varying transition-probability models in which transition dynamics respond to observable economic information [Filardo, 1994, Diebold et al., 1994, Kim and Nelson, 1999].

For each forecast origin, state inference uses filtered posterior probabilities

$$P(S_t = k \mid \mathcal{I}_t),$$

so that regime classification remains aligned with the information set available at time  $t$ . Smoothed probabilities are not used in the real-time panel. The dominant state is

$$\hat{S}_t = \arg \max_k P(S_t = k \mid \mathcal{I}_t).$$

Numerical state labels are aligned recursively through time rather than assigned by fixed inflation thresholds. The resulting state classifications are used for conditioning and interpretation, including formal tests of whether calibrated Forecast Alpha differs across macroeconomic states. Additional architecture-selection, multistart, convergence, label-alignment, and state-characterization details are reported in the Appendix.

## 4.8 Recursive Real-Time Implementation

The empirical production process is implemented as a sequence of forecast-origin calculations. For each forecast origin  $t$  and horizon  $h$ , the system:

1. constructs the month-end macro-financial information set using data available through, or lagged to, the forecast origin;
2. builds the predefined Compact-Direct predictor matrix;
3. estimates the horizon-specific Elastic Net model on an expanding window restricted to target observations observable at the forecast origin;

4. generates the direct inflation forecast  $\hat{\pi}_{t+h|t}^{\text{model}}$ ;
5. aligns the model forecast with the T5YIE-based market benchmark  $\hat{\pi}_{t+h|t}^{\text{mkt}}$ ;
6. computes raw Forecast Alpha and the associated market-relative decomposition;
7. estimates the horizon-specific recursive calibration weight using only prior observations whose realized outcomes were already available;
8. computes calibrated Forecast Alpha and Residual Surprise;
9. merges filtered TV-HMM state probabilities and dominant state labels for conditional evaluation; and
10. stores the out-of-sample forecast, Forecast Alpha measures, residual terms, regime classifications, calibration diagnostics, and market-relevance variables required for subsequent empirical tests.

This design preserves the real-time structure of both forecasting and calibration. For multi-month horizons, the calibration sample includes only historical observations whose target outcomes were already realized by the current forecast origin rather than all prior forecast-origin rows. This timing restriction prevents overlapping-horizon look-ahead bias and keeps the calibration exercise aligned with real-time forecast-evaluation practice [Croushore, 2011, Faust and Wright, 2013].

The forecast-encompassing regressions and regime-dependent inference are evaluation procedures applied to the recursively generated out-of-sample panel; they are not inputs to the live monthly forecast-generation loop. This distinction separates the operational production process from the empirical tests used to assess whether Forecast Alpha contains incremental information and whether its value varies across states.

## 4.9 Market-Relevance and Breakeven Repricing

The final methodological stage evaluates whether calibrated Forecast Alpha is subsequently reflected in market pricing. The objective is not to establish a complete trading strategy or portfolio backtest. It is to determine whether inflation information identified relative to the market benchmark is associated with subsequent adjustments in Treasury breakeven inflation.

Let

$$\Delta BE_{t,k} = BE_{t+k} - BE_t$$

denote the change in 5-year Treasury breakeven inflation over a post-signal evaluation window of  $k$  trading days. Market relevance is evaluated along complementary dimensions. First, linear specifications test whether the magnitude of calibrated Forecast Alpha is associated with the magnitude of subsequent breakeven repricing. Second, directional tests evaluate whether breakevens move in the direction implied by the Forecast Alpha signal. Third, directional-repricing measures summarize the average signed market movement following Forecast Alpha signals. Finally, event-time

analysis examines how directional repricing accumulates after the forecast origin in order to evaluate the timing of market incorporation.

These tests distinguish information content from subsequent market transmission. The incremental-information analysis asks whether the inflation model contains information beyond the contemporaneous market benchmark. The market-relevance analysis asks a separate downstream question: whether that information is subsequently incorporated into breakeven pricing. Evidence need not take the form of one-for-one linear repricing; directional and event-time tests allow economically meaningful incorporation to occur gradually and nonlinearly.

Inference is conducted using dependence-aware procedures appropriate for overlapping forecast horizons and serially correlated market data. The empirical results report both magnitude-based and directional evidence, while the event-time analysis provides additional information on the timing with which Forecast Alpha is reflected in market prices.

## 5 Results

This section presents the empirical evidence in the sequence implied by the Forecast Alpha framework. The analysis first tests whether the Compact Inflation Model contains incremental information beyond the contemporaneous market benchmark. It then evaluates how model–market disagreement is recursively calibrated into Forecast Alpha, whether the magnitude of calibrated Forecast Alpha varies systematically across filtered TV-HMM states, and whether the resulting signal is subsequently reflected in breakeven-market repricing. Distributional, calibration-path, and residual-surprise diagnostics are reported in the Appendix.

### 5.1 Incremental Information Beyond Market Expectations

The first empirical question is whether the model forecast contains information beyond the market-based benchmark. The encompassing specification is

$$\pi_{t+h} = a_h + b_h \hat{\pi}_{t+h|t}^{mkt} + c_h \hat{\pi}_{t+h|t}^{model} + u_{t,h},$$

where  $\pi_{t+h}$  is realized inflation,  $\hat{\pi}_{t+h|t}^{mkt}$  is the T5YIE market benchmark observed at the forecast origin, and  $\hat{\pi}_{t+h|t}^{model}$  is the Compact Inflation Model forecast formed at the same date. A statistically significant  $c_h$  indicates that the model forecast contributes incremental information conditional on the market benchmark.

Table 2 reports the forecast-encompassing regressions in their original units.

The Compact-model coefficient is positive and statistically significant at every forecast horizon. The corresponding incremental adjusted  $R^2$  from adding the model forecast to the market-only specification equals 0.505 at 1 month, 0.393 at 3 months, 0.307 at 6 months, and 0.095 at 12 months. The model’s incremental explanatory contribution therefore declines with the forecast horizon but remains statistically detectable through 12 months.

Table 2: Incremental Information Beyond Market Expectations

| Horizon  | Market Coef.       | Model Coef.         | Adj. $R^2$ | $N$ |
|----------|--------------------|---------------------|------------|-----|
| 1-month  | 0.141***<br>(2.99) | 0.922***<br>(31.55) | 0.955      | 214 |
| 3-month  | 0.330***<br>(3.42) | 0.826***<br>(14.01) | 0.873      | 210 |
| 6-month  | 0.456***<br>(3.63) | 0.725***<br>(9.15)  | 0.759      | 204 |
| 12-month | 0.586<br>(1.21)    | 0.545***<br>(2.90)  | 0.419      | 192 |

The table reports forecast-encompassing regressions of realized inflation on the T5YIE market benchmark and the Compact Inflation Model forecast. Regressions use raw model and market forecasts but are estimated on the horizon-specific sample for which timing-valid recursive Forecast Alpha calibration is available.

T-statistics based on heteroskedasticity- and autocorrelation-consistent standard errors are reported in parentheses. Newey–West lag length equals the forecast horizon.

Significance levels: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .

The market coefficient is statistically significant at the 1-, 3-, and 6-month horizons, indicating that the market benchmark retains conditional information beyond the model at shorter and intermediate horizons. At 12 months, the market coefficient is not statistically significant, while the model coefficient remains significant. This result is consistent with the Compact forecast containing incremental information at the longest horizon without establishing that the model generally encompasses the market benchmark.

Standardized encompassing coefficients and the corresponding conditional adjusted- $R^2$  contributions are reported in Appendix B. The level regressions in Table 2 provide the primary test of whether model forecasts contain incremental inflation information beyond contemporaneous market pricing.

**Forecast-engine robustness.** Detailed cross-engine decomposition, encompassing, and conventional forecast-accuracy comparisons are reported in Appendix B. The Compact Inflation Model is not selected because it minimizes RMSE at every horizon. The robustness exercise instead evaluates whether the Forecast Alpha framework and its incremental-information interpretation depend uniquely on the Compact forecasting architecture.

## 5.2 Recursive Calibration of Forecast Alpha

Having established that model forecasts contain incremental information conditional on the market benchmark, the next question is how the magnitude of model–market disagreement should be interpreted. Raw Forecast Alpha measures the disagreement observed at the forecast origin,

$$\alpha_{t,h}^{raw} = \hat{\pi}_{t+h|t}^{model} - \hat{\pi}_{t+h|t}^{mkt}.$$

Because the magnitude of that disagreement need not correspond one-for-one with subsequently

realized market-relative inflation surprise, the framework applies the recursively estimated calibration coefficient,

$$\alpha_{t,h}^{cal} = \hat{\lambda}_{t,h} \alpha_{t,h}^{raw}.$$

Table 3 summarizes the historical distribution of the recursive calibration coefficient at each forecast horizon.

Table 3: Recursive Calibration of Forecast Alpha

| Horizon | $N$ | Mean $\hat{\lambda}$ | Median $\hat{\lambda}$ | 5th–95th Percentile | % Outside $[0, 1]$ |
|---------|-----|----------------------|------------------------|---------------------|--------------------|
| 1M      | 214 | 0.864                | 0.855                  | [0.806, 0.939]      | 0.0%               |
| 3M      | 210 | 0.659                | 0.625                  | [0.544, 0.855]      | 0.0%               |
| 6M      | 204 | 0.510                | 0.468                  | [0.411, 0.758]      | 0.0%               |
| 12M     | 192 | 0.568                | 0.586                  | [0.324, 0.680]      | 1.6%               |

The calibration coefficient  $\hat{\lambda}_{t,h}$  is estimated recursively using only information available before each forecast origin and scales raw model–market disagreement into calibrated Forecast Alpha.

Mean and median summarize the typical recursive calibration weight, while the 5th–95th percentile range describes variation in the historical sequence without treating  $\hat{\lambda}_{t,h}$  as a fixed population parameter.

Values between zero and one attenuate raw model–market disagreement; values below zero imply an opposite-signed historical mapping, and values above one imply amplification. The final column reports the percentage of recursive estimates outside the interval  $[0, 1]$ . These ranges are descriptive and are not imposed as parameter restrictions.

The recursive estimates imply economically meaningful attenuation of raw model–market disagreement over most of the sample. Mean  $\hat{\lambda}$  equals 0.864 at 1 month, 0.659 at 3 months, 0.510 at 6 months, and 0.568 at 12 months. The corresponding medians are similar, although the 5th–95th percentile ranges show that the calibration weight varies through time rather than behaving as a fixed parameter.

At the 1-, 3-, and 6-month horizons, every recursive estimate lies within the interval  $[0, 1]$ . At 12 months, 98.4 percent of estimates lie within that interval and 1.6 percent are negative. The evidence therefore supports interpreting  $\hat{\lambda}_{t,h}$  as a time-varying historical calibration of model–market disagreement rather than as a fixed coefficient, a percentage of forecasts that are “correct,” or a mechanical market-transmission parameter. The full recursive paths and more detailed distributional diagnostics are reported in the Appendix.

### 5.3 State Dependence of Calibrated Forecast Alpha

Calibrated Forecast Alpha is next evaluated conditional on the dominant filtered TV-HMM state observed at the forecast origin. To make comparisons across forecast horizons independent of changes in sample composition, the state-conditioned statistics use the exact intersection of 192 monthly forecast origins available at the 1-, 3-, 6-, and 12-month horizons. The common-origin

sample extends from January 2009 through December 2024, with realized target dates ending no later than December 2025.

Figure 3 summarizes the mean absolute magnitude of calibrated Forecast Alpha across forecast horizons and the four filtered TV-HMM states.

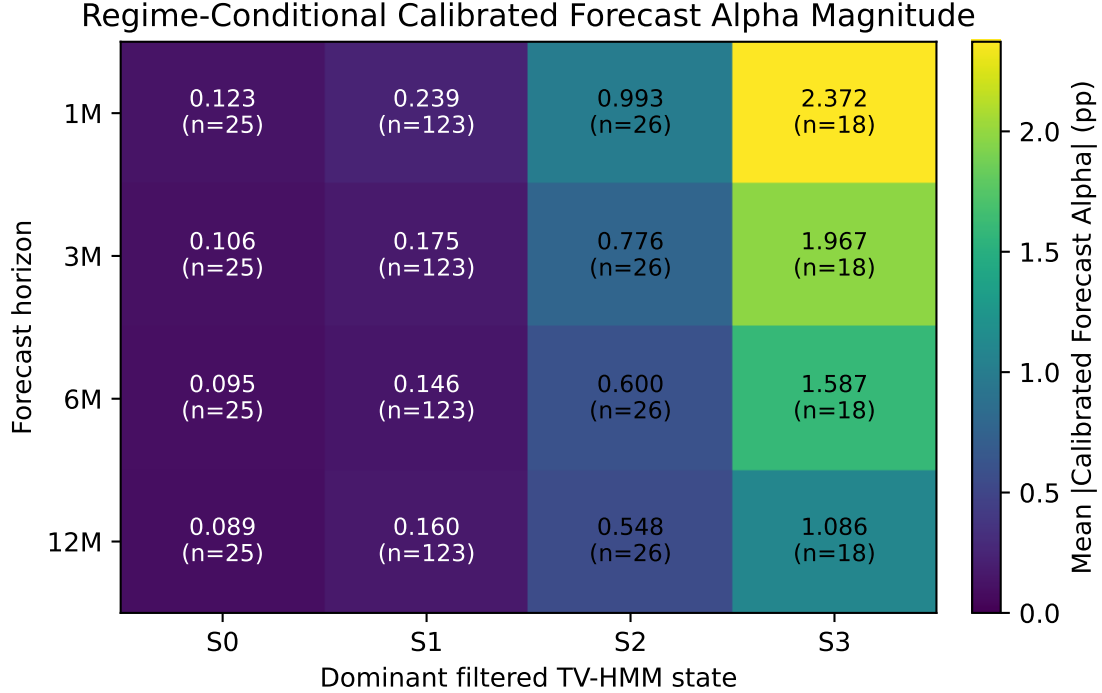


Figure 3: Regime-Conditional Calibrated Forecast Alpha Magnitude on a Common-Origin Sample

*Notes:* Each cell reports the mean absolute calibrated Forecast Alpha,  $|FA^{cal}|$ , in percentage points for the indicated forecast horizon and dominant filtered TV-HMM state. All cells use the same 192 forecast origins. State-level origin counts are 25 for State 0, 123 for State 1, 26 for State 2, and 18 for State 3. Shading is scaled linearly to the reported magnitude, with larger values indicating a larger calibrated model–market deviation in absolute terms.

Table 4 reports formal omnibus tests of state dependence in calibrated Forecast Alpha.

The results show statistically detectable state dependence in mean absolute calibrated Forecast Alpha at all four horizons on the common-origin sample. The finding is therefore not limited to a single forecast horizon and supports the interpretation that the economic magnitude of model–market disagreement depends on the prevailing latent macroeconomic state.

State 3 contains only 18 common-sample forecast origins and is concentrated in a limited historical episode. Fine distinctions involving that state should therefore be interpreted cautiously. The formal inference establishes state dependence of Forecast Alpha magnitude; it does not imply that every pairwise state comparison is estimated with equal precision.

Table 4: State Dependence of Calibrated Forecast Alpha

| Horizon | Mean $ FA^{cal} $<br>Statistic | Mean $ FA^{cal} $<br>$p$ |
|---------|--------------------------------|--------------------------|
| 1M      | 81.203                         | 0.0052                   |
| 3M      | 56.240                         | 0.0052                   |
| 6M      | 35.919                         | 0.0052                   |
| 12M     | 16.246                         | 0.0104                   |

Omnibus inference uses the governed circular-shift randomization test on the 192-origin common sample. Circular shifts preserve the observed persistence and exact counts of the K4 state sequence while breaking its alignment with Forecast Alpha outcomes.

Mean  $|FA^{cal}|$  measures the absolute magnitude of the calibrated active model-market deviation within each filtered TV-HMM state.

Each horizon uses 191 non-zero circular shifts. The tests evaluate state dependence of the observed historical alignment and should not be interpreted as iid ANOVA tests.

## 5.4 Market Relevance of Forecast Alpha

The final main-text empirical question is whether calibrated Forecast Alpha is subsequently reflected in breakeven-market repricing. Table 5 separates two distinct hypotheses: whether the magnitude of Forecast Alpha maps linearly into the magnitude of subsequent T5YIE changes, and whether subsequent repricing tends to occur in the direction implied by the Forecast Alpha signal.

Table 5: Market Relevance of Forecast Alpha

| Horizon | $N$ | Linear<br>$\beta$ | Linear<br>$p$ | Directional<br>Hit Rate | Hit<br>$p$ | Directional<br>Repricing (bp) | Repricing<br>$p$ |
|---------|-----|-------------------|---------------|-------------------------|------------|-------------------------------|------------------|
| 1M      | 214 | 0.008             | 0.8357        | 49.5%                   | 0.6671     | 0.08                          | 0.6275           |
| 3M      | 210 | 0.020             | 0.8077        | 49.0%                   | 0.6791     | -0.62                         | 0.7451           |
| 6M      | 204 | 0.016             | 0.9198        | 53.4%                   | 0.2985     | 4.59                          | 0.0924           |
| 12M     | 192 | -0.052            | 0.7726        | 62.5%                   | 0.0240     | 9.60                          | 0.0086           |

The linear transmission regression relates the horizon-matched change in T5YIE to calibrated Forecast Alpha and reports a two-sided HAC  $p$ -value with lag length equal to the forecast horizon.

Directional Hit Rate is the fraction of observations for which the sign of the horizon-matched T5YIE change equals the sign of calibrated Forecast Alpha. Directional Repricing is  $\text{sign}(FA^{cal})\Delta T5YIE$ , expressed in basis points.

Hit-rate and directional-repricing  $p$ -values are one-sided null-centered moving-block bootstrap tests with 5,000 replications, a 12-month block length, and fixed seed 20260806. The table distinguishes directional market relevance from linear magnitude transmission.

The linear transmission coefficients are not statistically significant at any reported horizon. The evidence therefore does not support a one-for-one or stable linear mapping from the magnitude of calibrated Forecast Alpha into the magnitude of subsequent breakeven-rate changes.

The directional evidence is different. At the 1- and 3-month horizons, hit rates are approximately 50 percent and directional repricing is economically small. At 6 months, the directional hit rate rises to 53.4 percent and mean directional repricing to 4.59 basis points, although neither result reaches conventional statistical significance. At 12 months, the directional hit rate is 62.5 percent with a one-sided bootstrap  $p$ -value of 0.0240, and mean directional repricing is 9.60 basis points with a one-sided bootstrap  $p$ -value of 0.0086. The strongest market-relevance evidence is therefore directional and concentrated at the longest forecast horizon rather than expressed as linear magnitude transmission.

Figure 4 reports the event-time path of subsequent directional T5YIE repricing for the H12 calibrated Forecast Alpha signal.

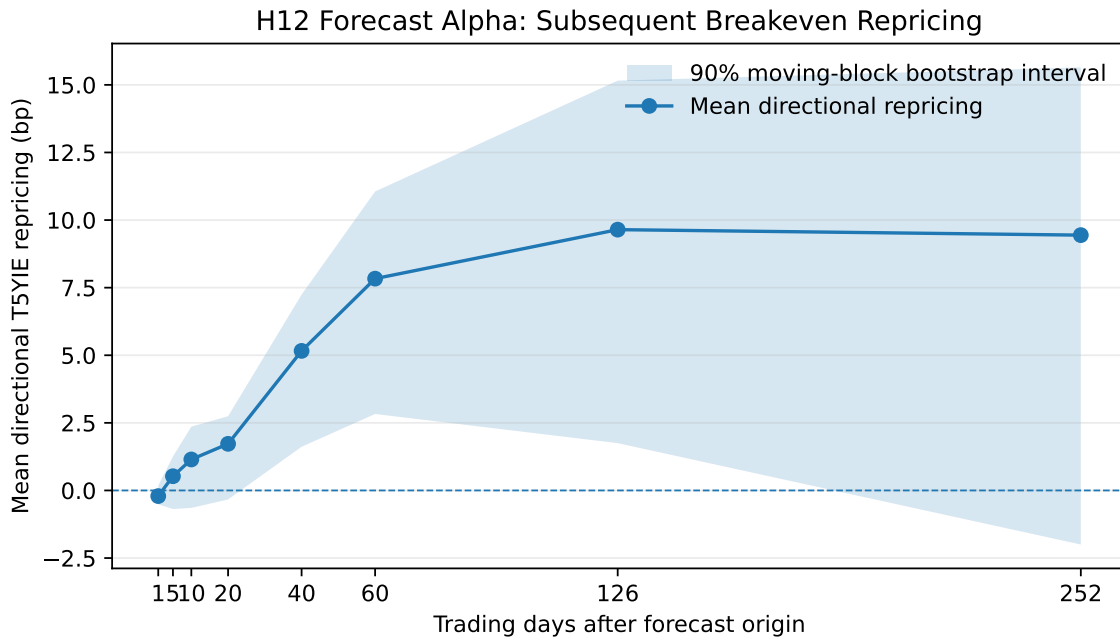


Figure 4: H12 Forecast Alpha and Subsequent Directional Breakeven Repricing

*Notes:* The figure reports mean directional T5YIE repricing following monthly H12 calibrated Forecast Alpha signals at 1, 5, 10, 20, 40, 60, 126, and 252 trading-day windows. Directional repricing is defined as  $\text{sign}(FA^{cal})\Delta T5YIE$  and is expressed in basis points. The shaded band reports the 90% moving-block bootstrap interval.

The event-time evidence complements Table 5 by showing when the H12 directional relationship emerges. Taken together, the results support a narrower claim than mechanical magnitude transmission: under the strongest documented configuration, calibrated Forecast Alpha contains directional information that is subsequently reflected in market pricing.

## 6 Evaluation

This section interprets the economic significance of the empirical results. Section 5 establishes four main findings: the Compact Inflation Model contains incremental information beyond the contemporaneous market benchmark; raw model–market disagreement is recursively calibrated

using only information available before each forecast origin; the magnitude of calibrated Forecast Alpha varies across filtered TV-HMM states; and the strongest market-relevance evidence appears as subsequent directional breakeven repricing at the 12-month horizon. The objective here is to interpret what those findings imply for investment decision-making without treating Forecast Alpha as a fully specified trading strategy.

## 6.1 What Forecast Alpha Adds Beyond the Market

The central empirical result is not simply that the model produces a different inflation forecast from the market benchmark. Model–market disagreement is economically relevant only if that disagreement contains information beyond what is already embedded in market pricing. The encompassing regressions in Table 2 provide that first test. The Compact-model coefficient remains positive and statistically significant across all four horizons, indicating that the model contributes information conditional on the contemporaneous T5YIE benchmark.

This distinction is important for interpretation. A raw forecast difference is not, by itself, an investable signal. The informational content of that difference must first be established relative to the market benchmark. Forecast Alpha therefore begins with model–market disagreement but acquires economic meaning only after the evidence shows that the disagreement contains incremental information and after its historical magnitude is calibrated using only prior outcomes.

The recursive coefficient summarized in Table 3 performs that second function. The calibration generally attenuates raw model–market disagreement, with mean and median values below one at every horizon. This behavior reinforces the interpretation of calibrated Forecast Alpha as a historically scaled active inflation view rather than a replacement for the market benchmark. The market remains the passive reference point; Forecast Alpha represents the model’s conditional deviation from that reference after recursive calibration.

A useful portfolio interpretation is therefore

$$\text{Forward Inflation View}_{t,h} = \text{Market Benchmark}_t + F A_{t,h}^{cal}.$$

More generally, when the active view is incorporated subject to portfolio-level risk controls, calibrated Forecast Alpha can be interpreted as the model-driven component of an active inflation view rather than as a stand-alone forecast target.

## 6.2 State and Horizon Dependence

The value of Forecast Alpha is not constant across macroeconomic environments or forecast horizons. Figure 3 and Table 4 show statistically detectable state dependence in the absolute magnitude of calibrated Forecast Alpha at all four horizons. This result implies that the degree of model–market disagreement judged historically informative is conditional on the prevailing latent macroeconomic state.

That finding matters for implementation because the same numerical Forecast Alpha signal

need not carry the same economic interpretation in every environment. The TV-HMM is used as a conditioning device rather than as a mechanism for mechanically switching the forecast model or determining a trade. Its role is to identify whether the historical relationship between model and market disagreement differs across macroeconomic states. In that sense, regime information helps characterize the context in which Forecast Alpha is observed.

Horizon dependence is equally important. Incremental information is statistically detectable through 12 months, but the character of the market-relevance evidence changes with horizon. At shorter horizons, the governed breakeven-repricing tests do not show economically compelling directional transmission. At 6 months, the evidence strengthens but remains below conventional significance thresholds. At 12 months, however, the directional hit rate reaches 62.5 percent and mean directional repricing is approximately 9.6 basis points, with both measures statistically significant under the governed moving-block bootstrap tests.

This pattern argues against treating Forecast Alpha as a horizon-invariant signal. The strongest evidence does not come from mechanically scaling the shortest-horizon model disagreement into a trading position. Instead, the empirical results suggest that the informational value of Forecast Alpha depends jointly on forecast horizon, macroeconomic state, and the subsequent market-pricing horizon being evaluated.

The horizon dependence of these results is consistent with a broader literature showing that inflation expectations and their adjustment dynamics differ across forecast horizons. Yetman [2020] examines pass-through from short-horizon to long-horizon inflation expectations as a measure of anchoring, while Moessner and Takáts [2020] shows that the relationship between short-term and long-term expectations, as well as persistent deviations of inflation from target, contains information about the degree of long-run anchoring. Gust et al. [2024] provides a complementary finite-horizon framework in which limited planning horizons generate systematic horizon-dependent adjustment in inflation forecasts. These studies provide useful context for the evidence here: Forecast Alpha contains incremental information across the evaluated horizons, but its downstream market relevance is weak at shorter horizons and strongest at the 12-month horizon. This correspondence does not establish that Forecast Alpha operates through the mechanisms identified in these studies; rather, it reinforces the importance of treating forecast horizon as an economically meaningful dimension of inflation-information transmission.

### 6.3 Implications for Investment Management

The market-relevance results support a deliberately narrow investment interpretation. Table 5 does not establish a stable linear mapping from the magnitude of calibrated Forecast Alpha to the magnitude of subsequent breakeven-rate changes. The linear coefficients are statistically insignificant at every horizon. The strongest evidence is instead directional and concentrated at the 12-month horizon.

For portfolio management, that distinction is consequential. Forecast Alpha is better interpreted as an input into the formation of an active inflation view than as a complete position-sizing

rule. A portfolio manager could begin with the market-implied inflation benchmark, incorporate calibrated Forecast Alpha as a model-based deviation from that benchmark, and then translate the resulting view into portfolio exposures subject to independent considerations such as valuation, carry, convexity, liquidity, transaction costs, risk budgets, and instrument selection.

The framework therefore separates three decisions that are often conflated. The first is whether the model contains information beyond the market. The second is how strongly that model-market disagreement should be weighted given its historical calibration and current macroeconomic state. The third is how any resulting inflation view should be implemented in an actual portfolio. The empirical analysis in this paper addresses the first two directly and evaluates evidence relevant to the third, but it does not claim to solve the complete portfolio-construction problem.

The strongest governed market result provides an example of this distinction. At the 12-month horizon, calibrated Forecast Alpha has statistically significant directional correspondence with subsequent T5YIE repricing, while the linear magnitude coefficient remains insignificant. This suggests that the signal may be useful for determining the direction of an inflation-sensitive portfolio tilt without implying that the size of the breakeven position should be mechanically proportional to the numerical magnitude of Forecast Alpha.

This interpretation is consistent with the broader objective of the framework. A forecast has economic value only to the extent that it changes a downstream decision. Forecast Alpha provides a disciplined mechanism for asking whether a model-based inflation view contains incremental information beyond the market, how that disagreement should be historically calibrated, and under what macroeconomic and horizon conditions the signal appears most economically relevant. Stress-period and supplementary robustness diagnostics are retained in the Appendix rather than treated as primary evidence for the framework.

## 7 Conclusion

This study develops a real-time framework for determining whether model-based inflation forecasts contain economically meaningful information beyond that already embedded in market pricing. The framework combines a direct horizon-specific inflation forecasting model, a filtered time-varying Hidden Markov Model (TV-HMM) used for regime classification, and a market-relative decomposition that distinguishes model–market disagreement from the remaining realized inflation surprise. Raw Forecast Alpha measures the model’s ex ante deviation from the market benchmark, while calibrated Forecast Alpha applies a recursively estimated historical calibration coefficient using only information available before each forecast origin.

The empirical evidence supports four principal conclusions.

First, the Compact Inflation Model contains incremental information beyond the contemporaneous market benchmark. In forecast-encompassing regressions, the model coefficient remains positive and statistically significant at the 1-, 3-, 6-, and 12-month horizons after controlling for T5YIE. This result is central to the interpretation of Forecast Alpha. Model–market disagreement is not assumed to be informative merely because the model differs from the market; its informational content is tested directly.

Second, the magnitude of that disagreement should not be taken at face value. Recursive calibration generally attenuates raw Forecast Alpha, with mean and median calibration coefficients below one across all horizons. Calibrated Forecast Alpha is therefore best interpreted as historically weighted model–market disagreement rather than as the portion of the model forecast that is mechanically “correct.” The calibration step converts raw disagreement into a disciplined active inflation view while preserving the market benchmark as the passive reference point.

Third, the magnitude of calibrated Forecast Alpha is state dependent. The filtered K=4 TV-HMM classifications show statistically detectable differences in absolute calibrated Forecast Alpha across macroeconomic states at every evaluated horizon. The regime layer does not mechanically determine the forecast or prescribe a trade. Instead, it provides economic context for interpreting when model–market disagreement has historically been larger or smaller. This finding reinforces the broader proposition that the value of an inflation forecast is conditional rather than invariant across macroeconomic environments.

Fourth, the strongest evidence linking Forecast Alpha to subsequent market pricing is directional rather than linear in magnitude. The linear relationship between calibrated Forecast Alpha and subsequent T5YIE repricing is not statistically significant at the evaluated horizons. At the 12-month horizon, however, the direction of Forecast Alpha is followed by same-direction breakeven repricing 62.5 percent of the time, and the corresponding mean directional repricing is approximately 9.6 basis points. Both results are statistically significant under the governed moving-block bootstrap procedure. The evidence therefore supports a narrower claim than mechanical price transmission: Forecast Alpha contains economically relevant directional information for subsequent breakeven repricing, with the strongest evidence concentrated at the 12-month horizon.

These findings have a direct investment-management interpretation. Forecast Alpha is not a

complete trading strategy or position-sizing rule. It is an input into the formation of an active inflation view. A practitioner can begin with the market-implied inflation benchmark, incorporate calibrated Forecast Alpha as the model-based deviation from that benchmark, and then determine how, or whether, to express that view through portfolio exposures subject to independent considerations including valuation, carry, liquidity, convexity, transaction costs, risk budgets, and instrument selection. In this sense,

$$\text{Forward Inflation View}_{t,h} = \text{Market Benchmark}_t + FA_{t,h}^{cal}.$$

The framework deliberately separates the informational question from the portfolio-construction question. The first asks whether a model contains information beyond the market. The second asks how that disagreement should be calibrated and interpreted given the prevailing macroeconomic state. Only after those questions are addressed does the practitioner confront implementation: which securities to use, how much risk to allocate, and how the signal interacts with other portfolio information. The empirical evidence in this study addresses the first two directly and provides evidence relevant to the third without claiming to solve the complete portfolio-allocation problem.

Several qualifications remain important. Forecast Alpha is defined relative to the forecasting system and market benchmark used in this study. Treasury breakeven inflation is linked to CPI-based inflation compensation, whereas the forecasting model targets Core PCE inflation, and breakeven rates may embed inflation risk premia, liquidity premia, and other market-specific distortions. Because a clean horizon-matched market-implied Core PCE expectations curve is not available for every horizon considered, T5YIE should be interpreted as an observable market-pricing anchor rather than as a literal horizon-specific forecast of Core PCE inflation. Forecast Alpha therefore measures model-based information relative to market inflation pricing, not disagreement with an unobservable measure of “true” expected inflation.

The results also caution against treating Forecast Alpha as horizon invariant. Incremental information is present across the evaluated horizons, but subsequent market relevance is heterogeneous. The strongest formal repricing evidence occurs at the 12-month horizon and is directional rather than proportional in magnitude. This distinction limits the claims that can be made from the present evidence while strengthening the interpretation that Forecast Alpha is most useful as a conditional decision variable whose relevance depends on horizon, state, and implementation context.

Taken together, the results support a qualified but economically meaningful conclusion. The contribution of the paper is not the claim that one inflation model uniformly dominates the market, nor that Forecast Alpha provides a finished trading strategy. Rather, it is a real-time methodology for determining whether model-based forecasts contain incremental information beyond market expectations, recursively calibrating that information, conditioning its interpretation on the prevailing macroeconomic state, and evaluating whether it subsequently appears in market pricing. The broader implication is that the value of a forecast should be judged not solely by forecast error, but by whether it contains information unavailable in the market benchmark and whether that

information can alter a downstream economic or investment decision.

## 7.1 Limitations and Future Research

The analysis has several limitations. The market benchmark is based on Treasury breakeven inflation rather than a clean horizon-specific Core PCE expectations curve. The market-relevance tests focus on subsequent breakeven repricing rather than a fully specified excess-return strategy, and the state-conditioned results should be evaluated further as additional real-time observations become available. The smallest TV-HMM states also contain relatively few common-sample observations, which limits the precision of state-specific comparisons.

Future research can extend the framework in several directions. One avenue is to evaluate alternative market-based inflation anchors, including measures that more closely match the forecast horizon or inflation concept. A second is to decompose breakeven inflation into expectations and risk-premium components to determine whether Forecast Alpha is associated more strongly with changes in expected inflation, inflation risk compensation, or both. A third is to develop portfolio-construction rules that translate calibrated Forecast Alpha into risk-controlled positions while preserving the real-time information discipline used in the present analysis.

Additional work can also examine whether the state dependence documented here survives across alternative inflation measures, asset classes, countries, and forecasting architectures. Because Forecast Alpha is defined relative to a model and market benchmark rather than to one specific forecasting algorithm, the framework naturally permits comparison across competing models. Future research could therefore examine whether combinations of independently generated Forecast Alpha signals improve the identification of incremental information without simply reproducing the consensus already embedded in market prices.

Although the framework is developed in the context of inflation, its underlying logic is more general. Whenever a market expectation can be observed or reasonably approximated, the same sequence can be applied: measure model–market disagreement, test whether that disagreement contains incremental information, calibrate it recursively, condition its interpretation on the relevant economic state, and evaluate whether it affects subsequent decisions or market outcomes. The broader contribution is therefore methodological: a framework for determining when a forecast contains economically meaningful information beyond market expectations and when that information matters.

## 8 Acknowledgements

Acknowledgments. The author used ChatGPT (OpenAI) for coding assistance. The author specified the intended functionality, modeling objectives, and analytical requirements, and ChatGPT was used to draft portions of the implementation. All code was reviewed, tested, revised, and validated by the author. The author was solely responsible for the study design, variable construction, model specification, backtesting protocol, empirical verification, interpretation of findings, and final manuscript content.

## 9 Disclosure Statement

The author reports no conflicts of interest.

## 10 Funding

No external funding was received for this research.

## References

- Martin M. Andreasen, Jens H. E. Christensen, and Simon Riddell. The TIPS liquidity premium. *Review of Finance*, 25(6):1639–1675, 2021. doi: 10.1093/rof/rfab025. Identifies and quantifies the liquidity premium embedded in TIPS yields using no-arbitrage term structure models.
- Andrew Ang and Allan Timmermann. Regime changes and financial markets. *Annual Review of Financial Economics*, 4:313–337, 2012. doi: 10.1146/annurev-financial-110311-101808.
- Andrew Ang, Geert Bekaert, and Min Wei. Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, 54(4):1163–1212, 2007. doi: 10.1016/j.jmoneco.2006.04.006.
- Andrew Atkeson and Lee E. Ohanian. Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1):2–11, 2001. Influential paper showing that naive random-walk models often outperform Phillips curve forecasts.
- Jushan Bai and Pierre Perron. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22, 2003. doi: 10.1002/jae.659. Provides methods for detecting and dating structural breaks, relevant to regime identification.
- Michael D. Bauer and Glenn D. Rudebusch. Resolving the spanning puzzle in macro-finance term structure models. *Review of Finance*, 21(2):511–553, 2017. doi: 10.1093/rof/rfw044. Examines the reliability of market-based inflation expectations and their informational content.
- Francesco Bianchi. Regime switches, agents’ beliefs, and post-world war ii u.s. macroeconomic dynamics. *The Review of Economic Studies*, 80(2):463–490, 2013. doi: 10.1093/restud/rds032.
- Mikhail Chernov and Philippe Mueller. The term structure of inflation expectations. *Journal of Financial Economics*, 106(2):367–394, 2012. doi: 10.1016/j.jfineco.2012.06.004.
- Richard Clarida, Jordi Galí, and Mark Gertler. Monetary policy rules and macroeconomic stability: Evidence and some theory. *The Quarterly Journal of Economics*, 115(1):147–180, 2000. doi: 10.1162/003355300554692.
- Timothy Cogley and Thomas J. Sargent. Drifts and volatilities: Monetary policies and outcomes in the post wwii u.s. *Review of Economic Dynamics*, 8(2):262–302, 2005. doi: 10.1016/j.red.2004.10.009.
- Dean Croushore. Forecasting with real-time data: What has changed since the early 1990s? *Annual Review of Economics*, 3:267–286, 2011. doi: 10.1146/annurev.economics.050708.142927. Reviews developments in real-time data forecasting over two decades.
- Stefania D’Amico, Don H. Kim, and Min Wei. Tips from TIPS: The informational content of treasury inflation-protected security prices. *Journal of Financial and Quantitative Analysis*, pages 1–35, 2018.

- Marco Del Negro and Frank Schorfheide. DSGE model-based forecasting. In Graham Elliott and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 2, pages 57–140. Elsevier, 2013.
- Francis X. Diebold, Joon-Haeng Lee, and Gretchen C. Weinbach. Regime switching with time-varying transition probabilities. In Colin P. Hargreaves, editor, *Nonstationary Time Series Analysis and Cointegration*, pages 283–302. Oxford University Press, Oxford, 1994.
- Jon Faust and Jonathan H. Wright. Forecasting inflation. In Graham Elliott and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 2, pages 2–56. Elsevier, 2013. doi: 10.1016/B978-0-444-53683-9.00001-3.
- Federal Reserve Bank of Atlanta. What is pce? explaining the fed’s preferred inflation measure. <https://www.atlantafed.org/what-we-study/inflation/2026/05/20/what-is-pce-explaining-the-feds-preferred-inflation-measure>, May 2026. Accessed July 7, 2026.
- Andrew J. Filardo. Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics*, 12(3):299–308, 1994.
- Milton Friedman. Nobel lecture: Inflation and unemployment. *Journal of Political Economy*, 85(3): 451–472, 1977. doi: 10.1086/260579. Lecture to the memory of Alfred Nobel, delivered December 13, 1976. Awarded the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, 1976.
- Lilit Gabrielyan. Forecasting inflation using Phillips curve. *Social Science Research Network*, 2016. Reviews Phillips curve specifications for inflation forecasting.
- Philippe Goulet Coulombe, Maxime Leroux, Dalibor Stevanovic, and Stéphane Surprenant. How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5): 920–964, 2022. doi: 10.1002/jae.2910.
- Jan J. J. Groen, Richard Paap, and Francesco Ravazzolo. Real-time inflation forecasting in a changing world. *Social Science Research Network*, 2012. doi: 10.2139/ssrn.1465985. Develops real-time inflation forecasting methods robust to structural change.
- Christopher Gust, Edward Herbst, and J. David López-Salido. Inflation expectations with finite horizon planning. Technical Report 2024-063, Board of Governors of the Federal Reserve System, August 2024. URL <https://www.federalreserve.gov/econres/feds/inflation-expectations-with-finite-horizon-planning.htm>. Shows that finite-horizon planning can generate initial underreaction and subsequent overreaction in inflation forecasts, with a four-quarter planning horizon matching key forecast-error evidence.
- James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.

- Joseph G. Haubrich, George Pennacchi, and Peter Ritchken. Inflation expectations, real rates, and risk premia: Evidence from inflation swaps. *The Review of Financial Studies*, 25(5):1588–1629, 2012. doi: 10.1093/rfs/hhs003.
- Yingying Huang, Paul Luk, and Jinsong Zhao. China’s inflation forecasting in a data-rich environment: Based on machine learning algorithms. *Applied Economics*, 2024. doi: 10.1080/00036846.2024.2322572. Evaluates machine learning algorithms for inflation forecasting in a data-rich environment.
- Peter Lund Jørgensen and Kevin J. Lansing. Anchored inflation expectations and the slope of the Phillips curve. *Federal Reserve Bank of San Francisco Working Paper*, 2019. doi: 10.24148/WP2019-27. Examines how anchored inflation expectations affect the slope of the Phillips curve.
- Florian Kajuth and Sebastian Watzka. Inflation expectations from index-linked bonds: Correcting for liquidity and inflation risk premia. *The Quarterly Review of Economics and Finance*, 51(3): 225–235, 2011. doi: 10.1016/j.qref.2011.03.004. Corrects breakeven inflation for liquidity and inflation risk premia using a term structure model.
- Chang-Jin Kim and Charles R. Nelson. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press, Cambridge, MA, 1999.
- Anders Bredahl Kock and Marcelo C. Medeiros. Penalized time series regression. In *Handbook of Statistics*. 2020. doi: 10.1007/978-3-030-31150-6\_7. Provides theoretical foundations for penalized (LASSO/Elastic Net) regression in time-series settings.
- Gary Koop and Dimitris Korobilis. Large time-varying parameter VARs. *Journal of Econometrics*, 177(2):185–198, 2013. doi: 10.1016/j.jeconom.2013.04.007.
- Gary Koop and Simon Potter. Estimation and forecasting in models with multiple breaks. *The Review of Economic Studies*, 74(3):763–789, 2007.
- Joelle Liebermann. Real-time forecasting in a data-rich environment. 2012. Examines the use of large datasets for real-time macroeconomic forecasting.
- Luis F. Martins and Vasco J. Gabriel. Forecasting inflation with the new Keynesian Phillips curve: Frequency matters. *Social Science Research Network*, 2020. doi: 10.2139/ssrn.3594397. Shows that data frequency affects the forecasting performance of New Keynesian Phillips curve models.
- Marcelo C. Medeiros, Gabriel F. R. Vasconcelos, Álvaro Veiga, and Eduardo Zilberman. Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39(1):98–119, 2021.
- Richhild Moessner and Előd Takáts. How well-anchored are long-term inflation expectations? Technical Report 869, Bank for International Settlements, June 2020. URL [https://www.bis.org/publ/other/moessner\\_takats\\_202006.pdf](https://www.bis.org/publ/other/moessner_takats_202006.pdf).

- [bis.org/publ/work869.htm](https://bis.org/publ/work869.htm). Examines how short-term inflation expectations and persistent deviations of inflation from target affect long-term inflation expectations.
- Andrea A. Naghi, Eoghan O'Neill, and Martina Danielova Zaharieva. The benefits of forecasting inflation with machine learning: New evidence. *Journal of Applied Econometrics*, 39(7):1321–1331, aug 2024. doi: 10.1002/jae.3088. URL <https://doi.org/10.1002/jae.3088>.
- Charles R. Nelson and G. William Schwert. Short-term interest rates as predictors of inflation: On testing the hypothesis that the real rate of interest is constant. *American Economic Review*, 67(3):478–486, 1977.
- A. W. Phillips. The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861–1957. *Economica*, 25(100):283–299, 1958. doi: 10.1111/j.1468-0335.1958.tb00003.x.
- Giorgio E. Primiceri. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852, 2005.
- Tim Robinson, Alan Stone, and Marileze van Zyl. The real-time forecasting performance of Phillips curves. 2003. Evaluates the real-time forecasting performance of various Phillips curve specifications.
- Shuping Shi. Bayesian model averaging under regime switching with application to cyclical macro variable forecasting. *Journal of Forecasting*, 35(8):739–750, 2016. doi: 10.1002/for.2375. Combines Bayesian model averaging with regime-switching for macroeconomic forecasting.
- Christopher A. Sims and Tao Zha. Were there regime switches in u.s. monetary policy? *American Economic Review*, 96(1):54–81, 2006. doi: 10.1257/000282806776157678.
- James H. Stock and Mark W. Watson. Forecasting inflation. *Journal of Monetary Economics*, 44(2):293–335, 1999.
- James H. Stock and Mark W. Watson. Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39(s1):3–33, 2007.
- James H. Stock and Mark W. Watson. Modeling inflation after the crisis. *Jackson Hole Economic Policy Symposium Proceedings*, pages 173–220, 2010.
- James H. Stock and Mark W. Watson. Core inflation and trend inflation. *The Review of Economics and Statistics*, 98(4):770–784, 2016. doi: 10.1162/REST\_a.00608.
- Robert Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. doi: 10.1111/j.2517-6161.1996.tb02080.x. Foundational LASSO paper; provides theoretical basis for the Elastic Net used in the paper.

U.S. Bureau of Economic Analysis. Personal consumption expenditures price index excluding food and energy. <https://www.bea.gov/data/personal-consumption-expenditures-price-index-excluding-food-and-energy>, 2026. Accessed July 7, 2026.

James Yetman. Pass-through from short-horizon to long-horizon inflation expectations, and the anchoring of inflation expectations. Technical Report 895, Bank for International Settlements, October 2020. URL <https://www.bis.org/publ/work895.htm>. Studies pass-through from short-horizon to long-horizon inflation forecasts as a measure of inflation-expectations anchoring.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005. doi: 10.1111/j.1467-9868.2005.00503.x. Original Elastic Net paper introducing the combined L1/L2 penalty for regularization and variable selection.

# Appendix

## A Forecast Model Validation and Robustness

This appendix reports validation results for the paper’s forecasting engine. The objective is to verify that the selected model is competitive relative to a persistence benchmark, stable across samples, economically interpretable, and robust across regimes.

### A.1 Candidate Predictor Universe

Table 6 reports the complete set of candidate variables available to the direct compact forecasting model. The specification combines inflation persistence measures with a limited set of labor-market, monetary-policy, interest-rate, producer-price, commodity-price, and survey-based indicators. Derived features used in estimation include lags, changes, rolling statistics, and selected spread variables constructed from the underlying series listed below.

Table 6: Variables Used in the Direct Compact Forecasting Model

| Category         | Variable                                  | FRED Series ID |
|------------------|-------------------------------------------|----------------|
| Inflation        | Core PCE Price Index                      | PCEPILFE       |
| Labor Market     | Unemployment Rate                         | UNRATE         |
| Monetary Policy  | Effective Federal Funds Rate              | DFF            |
| Interest Rates   | 2-Year Treasury Yield                     | DGS2           |
| Producer Prices  | Producer Price Index Commodities          | PPIACO         |
| Commodity Prices | WTI Crude Oil Spot Price                  | WTISPLC        |
| Survey-Based     | University of Michigan Consumer Sentiment | MICH           |

Notes: The forecasting model combines inflation persistence measures derived from Core PCE inflation with a compact set of exogenous macroeconomic and financial indicators. Engineered model inputs include lagged inflation terms, rolling statistics, changes in predictor variables, and selected spread measures constructed from the series listed above.

### A.2 Common-Sample Comparison

Tables 12A and 12B compare the direct compact model with the ARIMA benchmark on the exact overlapping forecast dates at each horizon.

Table 12A: Common-Sample Forecast Comparison with ARIMA: RMSE and MAE

| Horizon | $N$ | RMSE ARIMA | RMSE Direct | $\Delta$ RMSE | MAE ARIMA | MAE Direct | $\Delta$ MAE |
|---------|-----|------------|-------------|---------------|-----------|------------|--------------|
| 1       | 433 | 0.1539     | 0.1437      | -0.0102       | 0.1077    | 0.1043     | -0.0034      |
| 3       | 431 | 0.2937     | 0.2771      | -0.0166       | 0.2034    | 0.1951     | -0.0084      |
| 6       | 428 | 0.4543     | 0.4046      | -0.0497       | 0.3150    | 0.2853     | -0.0296      |
| 12      | 422 | 0.7667     | 0.6692      | -0.0974       | 0.5310    | 0.4616     | -0.0694      |

The table reports common-sample forecast performance for the ARIMA benchmark and the direct compact model.  $\Delta$  RMSE and  $\Delta$  MAE are computed as Direct minus ARIMA, so negative values indicate superior performance by the direct compact model.

The common-sample results confirm that the direct compact model’s gains are not driven by differences in forecast coverage. Relative to ARIMA, the direct compact model reduces RMSE and MAE at every horizon on the overlapping forecast sample.

Table 12B: Common-Sample Forecast Comparison with ARIMA: Bias

| Horizon | $N$ | Bias ARIMA | Bias Direct | $\Delta$ Bias |
|---------|-----|------------|-------------|---------------|
| 1       | 433 | 0.0094     | 0.0173      | 0.0079        |
| 3       | 431 | 0.0280     | 0.0453      | 0.0172        |
| 6       | 428 | 0.0586     | 0.0733      | 0.0147        |
| 12      | 422 | 0.1278     | 0.1348      | 0.0070        |

The table reports common-sample forecast bias for the ARIMA benchmark and the direct compact model on the exact overlapping forecast dates.  $\Delta$  Bias is computed as Direct minus ARIMA, so positive values indicate greater upward bias in the direct compact model relative to ARIMA.

Bias differences are modest relative to the error-level improvements, although the direct compact model exhibits somewhat higher positive bias than ARIMA at several horizons.

### A.3 Predictive-Accuracy Tests

Table 13 reports Diebold–Mariano tests of equal predictive accuracy between the ARIMA + Exog. specification and ARIMA.

Table 13: Diebold–Mariano Tests versus ARIMA

| Horizon | Loss | $N$ | Mean loss ARIMA | Mean loss ARIMA + Exog. | $\Delta$ loss | DM stat | $p$ -value |
|---------|------|-----|-----------------|-------------------------|---------------|---------|------------|
| 1       | AE   | 433 | 0.1077          | 0.1043                  | -0.0034       | -1.3134 | 0.1891     |
| 1       | SE   | 433 | 0.0237          | 0.0207                  | -0.0030       | -2.3677 | 0.0179     |
| 3       | AE   | 431 | 0.2034          | 0.1951                  | -0.0084       | -1.1767 | 0.2393     |
| 3       | SE   | 431 | 0.0863          | 0.0768                  | -0.0095       | -1.7765 | 0.0756     |
| 6       | AE   | 428 | 0.3150          | 0.2853                  | -0.0296       | -2.1386 | 0.0325     |
| 6       | SE   | 428 | 0.2064          | 0.1637                  | -0.0427       | -2.4849 | 0.0130     |
| 12      | AE   | 422 | 0.5310          | 0.4616                  | -0.0694       | -2.7218 | 0.0065     |
| 12      | SE   | 422 | 0.5878          | 0.4479                  | -0.1399       | -2.3184 | 0.0204     |

Diebold–Mariano tests of equal predictive accuracy are reported on the common sample. AE denotes absolute error and SE denotes squared error.  $\Delta$  loss is computed as ARIMA+Exog. minus ARIMA, so negative values indicate lower loss for the ARIMA+Exog. specification.

These tests assess whether the forecast-error improvements are statistically distinguishable from zero.

### A.4 Feature Stability and Interpretability

Table 14 summarizes the frequency with which predictors enter the model across recursive refits.

Table 14: Feature Stability Summary

| Horizon | Feature         | Nonzero freq. | Dominant sign | Mean $ \beta $ |
|---------|-----------------|---------------|---------------|----------------|
| 1       | infl_lag1       | 1.00          | +             | 0.5809         |
| 1       | infl_lag3       | 1.00          | +             | 0.5303         |
| 1       | infl_roll3_mean | 1.00          | +             | 0.4761         |
| 1       | infl_chg3       | 1.00          | +             | 0.1312         |
| 1       | mich            | 1.00          | +             | 0.0968         |
| 1       | infl_chg1       | 1.00          | +             | 0.0921         |
| 1       | infl_lag2       | 0.93          | +             | 0.2045         |
| 1       | oil_chg3        | 0.87          | +             | 0.0147         |
| 3       | infl_lag1       | 1.00          | +             | 0.5547         |
| 3       | infl_lag3       | 1.00          | +             | 0.4847         |
| 3       | infl_roll3_mean | 1.00          | +             | 0.4390         |
| 3       | mich            | 1.00          | +             | 0.1770         |
| 3       | infl_chg3       | 1.00          | +             | 0.1106         |
| 3       | infl_chg1       | 1.00          | +             | 0.0841         |
| 3       | infl_roll6_std  | 0.97          | +             | 0.0515         |
| 3       | infl_lag6       | 0.82          | -             | 0.1501         |
| 6       | infl_lag1       | 1.00          | +             | 0.5719         |
| 6       | infl_lag3       | 1.00          | +             | 0.4933         |
| 6       | infl_roll3_mean | 1.00          | +             | 0.4623         |
| 6       | mich            | 1.00          | +             | 0.2336         |
| 6       | infl_chg3       | 1.00          | +             | 0.1629         |
| 6       | infl_roll6_std  | 1.00          | +             | 0.1068         |
| 6       | infl_chg1       | 1.00          | +             | 0.0612         |
| 6       | infl_lag2       | 0.86          | +             | 0.2236         |
| 12      | infl_lag1       | 1.00          | +             | 0.4176         |
| 12      | infl_lag3       | 1.00          | +             | 0.3290         |
| 12      | infl_roll6_std  | 1.00          | +             | 0.2591         |
| 12      | infl_chg3       | 1.00          | +             | 0.1478         |
| 12      | mich            | 1.00          | +             | 0.2413         |
| 12      | infl_chg1       | 1.00          | +             | 0.0494         |
| 12      | infl_roll3_mean | 0.98          | +             | 0.3338         |
| 12      | infl_lag2       | 0.88          | +             | 0.2003         |

Top retained predictors by recursive nonzero frequency. The table reports the most stable features in the direct compact model by horizon.

Table 15 reports whether the dominant signs of the retained predictors align with standard economic intuition.

Table 15: Economic Interpretability Summary

| Horizon | Feature         | Expected sign | Dominant sign | Alignment  |
|---------|-----------------|---------------|---------------|------------|
| 1       | infl_lag1       | positive      | positive      | aligned    |
| 1       | infl_lag3       | positive      | positive      | aligned    |
| 1       | infl_roll3_mean | positive      | positive      | aligned    |
| 1       | infl_chg3       | positive      | positive      | aligned    |
| 1       | mich            | positive      | positive      | aligned    |
| 1       | infl_chg1       | positive      | positive      | aligned    |
| 1       | infl_lag2       | positive      | positive      | aligned    |
| 1       | oil_chg3        | positive      | positive      | aligned    |
| 3       | infl_lag1       | positive      | positive      | aligned    |
| 3       | infl_lag3       | positive      | positive      | aligned    |
| 3       | infl_roll3_mean | positive      | positive      | aligned    |
| 3       | mich            | positive      | positive      | aligned    |
| 3       | infl_chg3       | positive      | positive      | aligned    |
| 3       | infl_chg1       | positive      | positive      | aligned    |
| 3       | infl_roll6_std  | ambiguous     | positive      | not_scored |
| 3       | infl_lag6       | positive      | positive      | aligned    |
| 6       | infl_lag1       | positive      | positive      | aligned    |
| 6       | infl_lag3       | positive      | positive      | aligned    |
| 6       | infl_roll3_mean | positive      | positive      | aligned    |
| 6       | mich            | positive      | positive      | aligned    |
| 6       | infl_chg3       | positive      | positive      | aligned    |
| 6       | infl_roll6_std  | ambiguous     | positive      | not_scored |
| 6       | infl_chg1       | positive      | positive      | aligned    |
| 6       | infl_lag2       | positive      | positive      | aligned    |
| 12      | infl_lag1       | positive      | positive      | aligned    |
| 12      | infl_lag3       | positive      | positive      | aligned    |
| 12      | infl_roll6_std  | ambiguous     | positive      | not_scored |
| 12      | infl_chg3       | positive      | positive      | aligned    |
| 12      | mich            | positive      | positive      | aligned    |
| 12      | infl_chg1       | positive      | positive      | aligned    |
| 12      | infl_roll3_mean | positive      | positive      | aligned    |
| 12      | infl_lag2       | positive      | positive      | aligned    |

Economic interpretability audit for the most stable retained features. Alignment indicates whether the dominant estimated sign accords with standard economic intuition.

The feature-stability results indicate that the forecasting specification is not being driven by a rapidly changing or opaque predictor set. Across horizons, the most persistent inputs are inflation-persistence variables, including lagged inflation and rolling inflation summaries, while a smaller group of exogenous variables—most notably survey expectations and selected macro-financial controls—enters more selectively. The interpretability audit in Table 15 is broadly consistent with

that pattern: the dominant signs for the stable retained predictors generally accord with standard economic intuition, while the few variables with less clear directional priors are explicitly treated as ambiguous rather than forced into a stronger interpretation. Taken together, Tables 14 and 15 support the view that the selected specification combines a stable persistence block with a limited and economically interpretable set of exogenous predictors.

## A.5 Temporal Robustness

Table 16 reports subsample forecast performance.

Table 16: Subsample Forecast Robustness

| Horizon | Subsample | $N$ | RMSE ARIMA | RMSE Direct | $\Delta$ RMSE | MAE ARIMA | MAE Direct |
|---------|-----------|-----|------------|-------------|---------------|-----------|------------|
| 1       | Pre-2000  | 120 | 0.1332     | 0.1307      | -0.0025       | 0.1091    | 0.1077     |
| 1       | 2000–2019 | 240 | 0.1410     | 0.1243      | -0.0167       | 0.0936    | 0.0873     |
| 1       | Post-2020 | 73  | 0.2143     | 0.2089      | -0.0053       | 0.1517    | 0.1546     |
| 3       | Pre-2000  | 118 | 0.2074     | 0.2344      | 0.0270        | 0.1650    | 0.1915     |
| 3       | 2000–2019 | 240 | 0.2445     | 0.2091      | -0.0354       | 0.1777    | 0.1534     |
| 3       | Post-2020 | 73  | 0.4933     | 0.4699      | -0.0234       | 0.3502    | 0.3377     |
| 6       | Pre-2000  | 115 | 0.3368     | 0.3525      | 0.0157        | 0.2737    | 0.2879     |
| 6       | 2000–2019 | 240 | 0.3534     | 0.3169      | -0.0364       | 0.2576    | 0.2363     |
| 6       | Post-2020 | 73  | 0.7879     | 0.6586      | -0.1294       | 0.5685    | 0.4423     |
| 12      | Pre-2000  | 109 | 0.5774     | 0.5152      | -0.0622       | 0.5013    | 0.4646     |
| 12      | 2000–2019 | 240 | 0.4980     | 0.4748      | -0.0231       | 0.3864    | 0.3572     |
| 12      | Post-2020 | 73  | 1.4439     | 1.2048      | -0.2391       | 1.0507    | 0.8001     |

Subsample robustness for the common forecast sample.  $\Delta$  RMSE is computed as Direct minus ARIMA; negative values indicate superior performance by the direct compact model.

Figure 5 plots rolling RMSE differences between the selected specification and ARIMA. Negative values indicate windows in which the selected specification performs better.

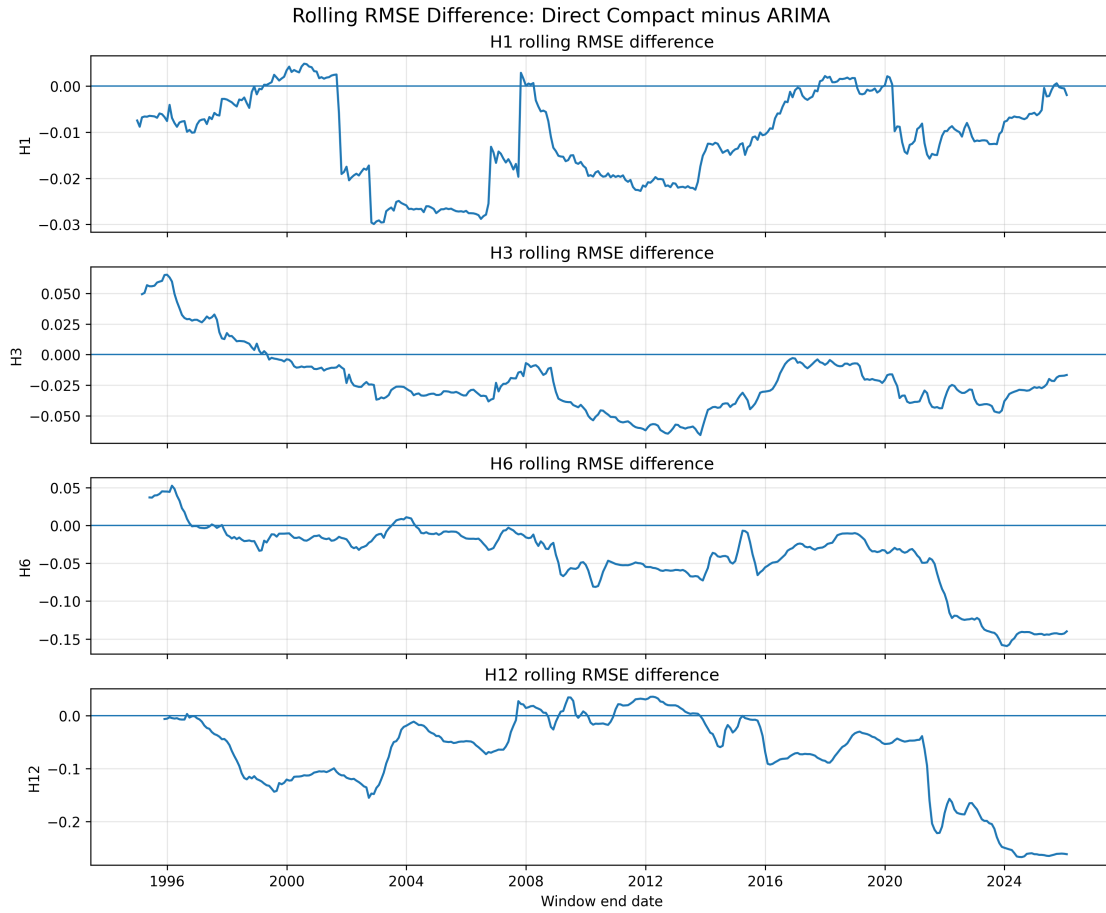


Figure 5: Rolling RMSE Difference: Direct Compact minus ARIMA  
 Negative values indicate rolling windows in which the direct compact model outperforms the ARIMA benchmark.

The temporal-robustness evidence is favorable but not uniform. Table 16 shows that the model’s gains relative to ARIMA are strongest in the 2000–2019 and post-2020 subsamples, particularly at intermediate and longer horizons, while the pre-2000 results are more mixed. Figure 5 reinforces that interpretation: rolling RMSE differences are not negative in every window, but they are predominantly negative over substantial portions of the evaluation sample, with the clearest and most sustained improvements appearing at the 6- and 12-month horizons. The results therefore indicate that the model’s advantage is not confined to a single historical episode, even though the magnitude of the gains varies over time.

## A.6 Regime Robustness

Table 17 reports forecast performance by horizon and regime.

Table 17: Regime-Specific Forecast Robustness

| Horizon | Regime | $N$ | RMSE ARIMA | RMSE Direct | $\Delta$ RMSE | MAE ARIMA | MAE Direct |
|---------|--------|-----|------------|-------------|---------------|-----------|------------|
| 1       | 0      | 126 | 0.1419     | 0.1300      | -0.0118       | 0.1029    | 0.0967     |
| 1       | 1      | 193 | 0.1541     | 0.1418      | -0.0122       | 0.1079    | 0.1064     |
| 1       | 2      | 113 | 0.1643     | 0.1587      | -0.0056       | 0.1109    | 0.1076     |
| 3       | 0      | 126 | 0.2405     | 0.2264      | -0.0141       | 0.1746    | 0.1645     |
| 3       | 1      | 192 | 0.2629     | 0.2410      | -0.0219       | 0.2001    | 0.1913     |
| 3       | 2      | 112 | 0.3833     | 0.3695      | -0.0138       | 0.2397    | 0.2336     |
| 6       | 0      | 126 | 0.3902     | 0.3022      | -0.0880       | 0.2745    | 0.2268     |
| 6       | 1      | 189 | 0.4045     | 0.3611      | -0.0434       | 0.3005    | 0.2827     |
| 6       | 2      | 112 | 0.5817     | 0.5471      | -0.0347       | 0.3826    | 0.3525     |
| 12      | 0      | 124 | 0.8075     | 0.6851      | -0.1224       | 0.5579    | 0.4473     |
| 12      | 1      | 189 | 0.6524     | 0.5805      | -0.0719       | 0.4918    | 0.4469     |
| 12      | 2      | 108 | 0.8956     | 0.7881      | -0.1075       | 0.5706    | 0.5067     |

Forecast robustness by TV-HMM regime on the common sample.  $\Delta$  RMSE is computed as Direct minus ARIMA; negative values indicate superior performance by the direct compact model.

The regime-specific results show that the selected specification remains competitive across the regime distribution rather than benefiting from a single narrow state. The RMSE differences are favorable across the reported regime-horizon cells, with the improvement generally becoming more pronounced at the 6- and 12-month horizons than at the 1- and 3-month horizons. This pattern is consistent with the broader validation evidence: the model’s gains are broad-based, but they are economically more meaningful at intermediate and longer horizons than at the very shortest horizon.

## A.7 Summary

Taken together, the validation results support the use of the selected forecasting specification in the main analysis. The common-sample comparisons show that the model’s gains relative to ARIMA are not an artifact of forecast coverage. The Diebold–Mariano tests indicate that those improvements

are statistically meaningful in several horizon-loss combinations, though not uniformly in every case. The feature-stability and interpretability tables show that the model relies on a stable and economically intelligible predictor set. The subsample, rolling-window, and regime-specific evidence further indicate that the model’s relative performance is reasonably robust across time and across states, with the strongest gains appearing at intermediate and longer horizons.

## **B Supplementary Forecast Alpha Exhibits**

The main text reports the exhibits most directly tied to the paper’s central claim-to-evidence chain. This section retains additional descriptive and robustness evidence that is useful for assessing distributional shape, regime decomposition, forecast-engine generality, and standardized incremental information without interrupting the main empirical sequence.

### **B.1 Forecast Alpha Distribution**

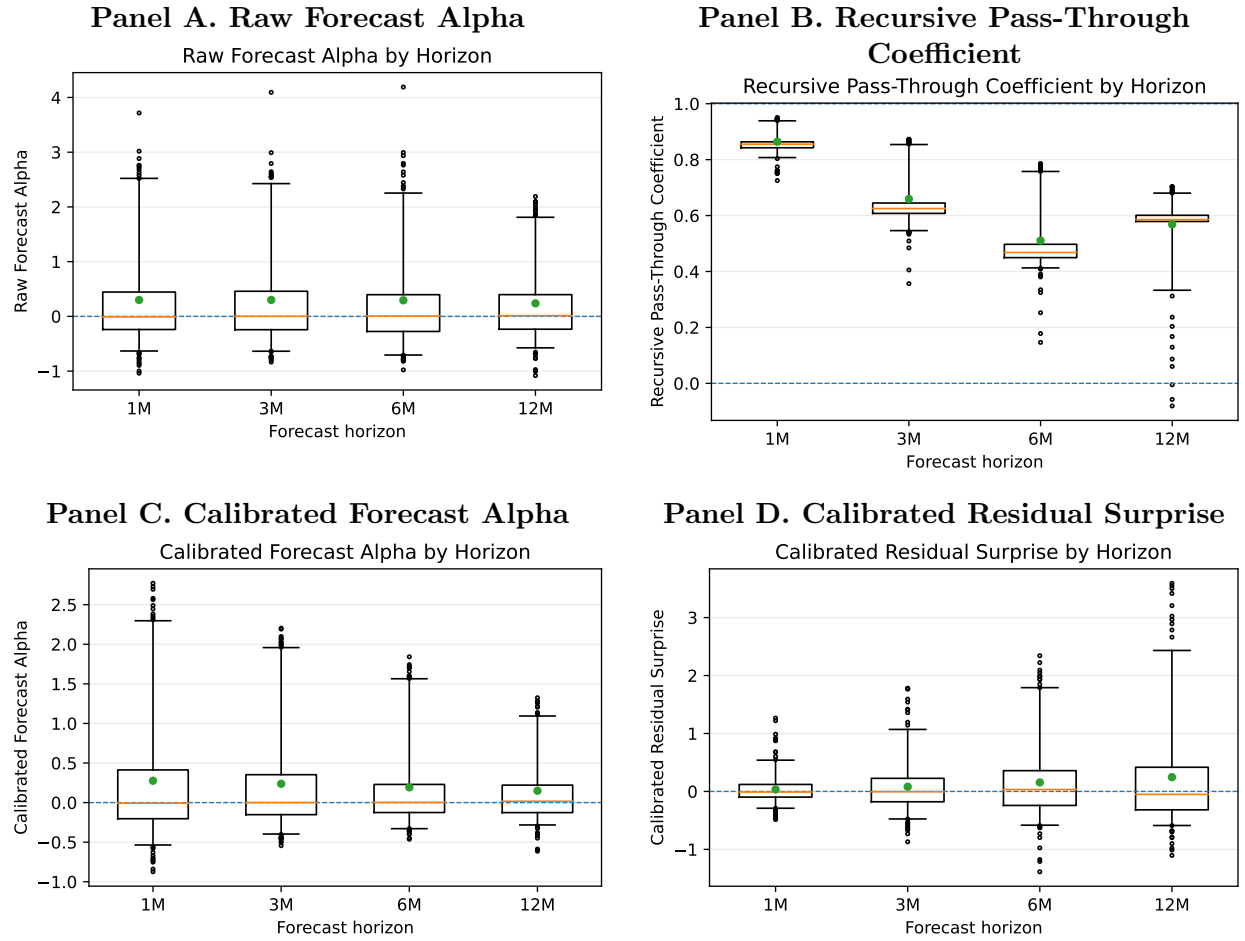


Figure 6: Distribution of Forecast Alpha Construction Components by Horizon

*Notes:* Boxes span the interquartile range, center lines report medians, markers report means, and whiskers span the 5th through 95th percentiles. Points beyond those percentiles are shown individually. Panels A, C, and D include a zero reference line. Panel B includes reference lines at zero and one. Samples are horizon-specific, and overlapping forecast horizons induce serial dependence.

## B.2 Detailed K4 Regime Decomposition

Table 18: Regime-Conditional Forecast Alpha Decomposition on a Common-Origin Sample

| Horizon | State | $N$ | Mean<br>$ \alpha^{raw} $ | Mean<br>$ \alpha^{cal} $ | Mean<br>$ \eta $ |
|---------|-------|-----|--------------------------|--------------------------|------------------|
| 1M      | S0    | 25  | 0.144                    | 0.123                    | 0.162            |
|         | S1    | 123 | 0.281                    | 0.239                    | 0.131            |
|         | S2    | 26  | 1.110                    | 0.993                    | 0.307            |
|         | S3    | 18  | 2.553                    | 2.372                    | 0.280            |
| 3M      | S0    | 25  | 0.186                    | 0.106                    | 0.209            |
|         | S1    | 123 | 0.282                    | 0.175                    | 0.242            |
|         | S2    | 26  | 1.053                    | 0.776                    | 0.606            |
|         | S3    | 18  | 2.437                    | 1.967                    | 0.425            |
| 6M      | S0    | 25  | 0.241                    | 0.095                    | 0.284            |
|         | S1    | 123 | 0.311                    | 0.146                    | 0.370            |
|         | S2    | 26  | 1.021                    | 0.600                    | 0.786            |
|         | S3    | 18  | 2.338                    | 1.587                    | 0.582            |
| 12M     | S0    | 25  | 0.241                    | 0.089                    | 0.486            |
|         | S1    | 123 | 0.300                    | 0.160                    | 0.560            |
|         | S2    | 26  | 0.855                    | 0.548                    | 0.845            |
|         | S3    | 18  | 1.848                    | 1.086                    | 0.567            |

The table reports the calibrated decomposition conditional on the dominant filtered TV-HMM state at the forecast origin. All horizons use the same 192 monthly forecast origins from January 2009 through December 2024; realized target dates do not extend beyond December 2025. State labels are numerical because the recursively aligned latent states are not ordered structurally by inflation. State-level origin counts are 25 for State 0, 123 for State 1, 26 for State 2, 18 for State 3.

## B.3 Robustness Across Forecast Engines

Table 19: Forecast Alpha Decomposition Across Forecast Engines

| Forecast Engine         | H  | N   | Mean $\hat{\lambda}$ | Mean $ \alpha^{raw} $ | Mean $ \alpha^{cal} $ | Mean $ \eta $ |
|-------------------------|----|-----|----------------------|-----------------------|-----------------------|---------------|
| Naive no-change         | 1  | 214 | 0.815                | 0.623                 | 0.529                 | 0.170         |
| Naive no-change         | 3  | 210 | 0.682                | 0.629                 | 0.462                 | 0.285         |
| Naive no-change         | 6  | 204 | 0.607                | 0.638                 | 0.424                 | 0.428         |
| Naive no-change         | 12 | 192 | 0.674                | 0.614                 | 0.418                 | 0.611         |
| AR(p)                   | 1  | 214 | 0.805                | 0.625                 | 0.521                 | 0.160         |
| AR(p)                   | 3  | 210 | 0.671                | 0.613                 | 0.439                 | 0.282         |
| AR(p)                   | 6  | 204 | 0.602                | 0.596                 | 0.388                 | 0.413         |
| AR(p)                   | 12 | 192 | 0.751                | 0.474                 | 0.362                 | 0.594         |
| Phillips curve          | 1  | 214 | 0.796                | 0.637                 | 0.529                 | 0.174         |
| Phillips curve          | 3  | 210 | 0.630                | 0.658                 | 0.450                 | 0.301         |
| Phillips curve          | 6  | 204 | 0.487                | 0.684                 | 0.375                 | 0.467         |
| Phillips curve          | 12 | 192 | 0.432                | 0.656                 | 0.286                 | 0.618         |
| Compact Inflation Model | 1  | 214 | 0.864                | 0.598                 | 0.528                 | 0.172         |
| Compact Inflation Model | 3  | 210 | 0.659                | 0.602                 | 0.424                 | 0.306         |
| Compact Inflation Model | 6  | 204 | 0.510                | 0.616                 | 0.342                 | 0.445         |
| Compact Inflation Model | 12 | 192 | 0.568                | 0.513                 | 0.290                 | 0.589         |

Notes: Statistics are evaluated on exact within-horizon common samples shared by all four forecast engines and the T5YIE market benchmark. Recursive calibration is estimated separately by engine and horizon using a minimum of 60 prior outcomes publication-available by each forecast origin. The calibration intercept is estimated but is not embedded in calibrated Forecast Alpha.

Table 20: Forecast-Encompassing Results Across Forecast Engines

| Forecast Engine         | H  | N   | Market Coef. | Model Coef. | Model $p$ | $\Delta$ Adj. $R^2$ |
|-------------------------|----|-----|--------------|-------------|-----------|---------------------|
| Naive no-change         | 1  | 275 | 0.208***     | 0.908***    | < 0.0001  | 0.567               |
| Naive no-change         | 3  | 273 | 0.330***     | 0.821***    | < 0.0001  | 0.464               |
| Naive no-change         | 6  | 270 | 0.379***     | 0.721***    | < 0.0001  | 0.357               |
| Naive no-change         | 12 | 264 | 0.234        | 0.526***    | < 0.0001  | 0.190               |
| AR(p)                   | 1  | 275 | 0.182***     | 0.920***    | < 0.0001  | 0.576               |
| AR(p)                   | 3  | 273 | 0.282***     | 0.865***    | < 0.0001  | 0.478               |
| AR(p)                   | 6  | 270 | 0.287***     | 0.800***    | < 0.0001  | 0.390               |
| AR(p)                   | 12 | 264 | 0.226        | 0.659***    | < 0.0001  | 0.179               |
| Phillips curve          | 1  | 275 | 0.194***     | 0.894***    | < 0.0001  | 0.562               |
| Phillips curve          | 3  | 273 | 0.318***     | 0.792***    | < 0.0001  | 0.447               |
| Phillips curve          | 6  | 270 | 0.385***     | 0.670***    | < 0.0001  | 0.321               |
| Phillips curve          | 12 | 264 | 0.272        | 0.477***    | 0.0001    | 0.144               |
| Compact Inflation Model | 1  | 275 | 0.110***     | 0.931***    | < 0.0001  | 0.563               |
| Compact Inflation Model | 3  | 273 | 0.255***     | 0.845***    | < 0.0001  | 0.449               |
| Compact Inflation Model | 6  | 270 | 0.328***     | 0.750***    | < 0.0001  | 0.355               |
| Compact Inflation Model | 12 | 264 | 0.288**      | 0.595***    | < 0.0001  | 0.169               |

Notes: Each row reports a forecast-encompassing regression of realized inflation on the T5YIE market benchmark and the indicated model forecast. All engines use the same origin dates within each horizon. Heteroskedasticity- and autocorrelation-consistent standard errors use a Newey–West lag length equal to the forecast horizon.  $\Delta$  Adj.  $R^2$  is the increase from adding the model forecast to a market-only regression. Significance levels: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .

Table 21: Out-of-Sample Forecast Accuracy on Exact Common Samples

| Horizon | Forecast                | $N$ | RMSE          | MAE    | Bias    |
|---------|-------------------------|-----|---------------|--------|---------|
| 1       | T5YIE market benchmark  | 275 | 0.8523        | 0.5168 | -0.1965 |
| 1       | Naive no-change         | 275 | 0.2491        | 0.1681 | -0.0082 |
| 1       | AR(p)                   | 275 | <b>0.2207</b> | 0.1577 | -0.0064 |
| 1       | Phillips curve          | 275 | 0.2614        | 0.1738 | -0.0240 |
| 1       | Compact Inflation Model | 275 | 0.2419        | 0.1607 | -0.0039 |
| 3       | T5YIE market benchmark  | 273 | 0.8421        | 0.5182 | -0.2024 |
| 3       | Naive no-change         | 273 | 0.4160        | 0.2768 | -0.0159 |
| 3       | AR(p)                   | 273 | <b>0.3804</b> | 0.2546 | -0.0173 |
| 3       | Phillips curve          | 273 | 0.4464        | 0.2960 | -0.0488 |
| 3       | Compact Inflation Model | 273 | 0.4193        | 0.2755 | -0.0055 |
| 6       | T5YIE market benchmark  | 270 | 0.8716        | 0.5547 | -0.2142 |
| 6       | Naive no-change         | 270 | 0.6094        | 0.3984 | -0.0302 |
| 6       | AR(p)                   | 270 | <b>0.5471</b> | 0.3735 | -0.0294 |
| 6       | Phillips curve          | 270 | 0.6654        | 0.4380 | -0.0861 |
| 6       | Compact Inflation Model | 270 | 0.5961        | 0.3933 | -0.0216 |
| 12      | T5YIE market benchmark  | 264 | 1.0160        | 0.6639 | -0.2395 |
| 12      | Naive no-change         | 264 | 0.9655        | 0.6132 | -0.0592 |
| 12      | AR(p)                   | 264 | <b>0.8988</b> | 0.5487 | -0.0909 |
| 12      | Phillips curve          | 264 | 1.0160        | 0.6669 | -0.1529 |
| 12      | Compact Inflation Model | 264 | 0.9208        | 0.5756 | -0.0438 |

*Notes:* All forecasts are evaluated on exact within-horizon common samples shared by the four forecast engines and T5YIE. Bias is forecast minus actual inflation. Forecast dates are aligned from the forecast origin to the realized  $t + h$  target, and all target dates end no later than December 2025. T5YIE is a market-based inflation-compensation proxy rather than a horizon-matched short-horizon inflation forecast. Bold indicates the lowest RMSE within each horizon.

The cross-engine evidence is interpreted as a robustness test of the Forecast Alpha framework rather than as a forecast-model horse race. In particular, the Compact Inflation Model does not minimize RMSE at every horizon, while the incremental-information results remain economically relevant to the framework’s market-relative objective.

## B.4 Standardized Incremental-Information Results

Table 22: Standardized Compact-Model Encompassing Results on the Calibrated Sample

| Horizon  | Std. Market<br>Coef. | Std. Model<br>Coef. | $\Delta$ Adj. $R^2$<br>(Model   Market) | $\Delta$ Adj. $R^2$<br>(Market   Model) | $N$ |
|----------|----------------------|---------------------|-----------------------------------------|-----------------------------------------|-----|
| 1-month  | 0.073***<br>(2.99)   | 0.929***<br>(31.55) | 0.504                                   | 0.003                                   | 214 |
| 3-month  | 0.171***<br>(3.42)   | 0.816***<br>(14.01) | 0.393                                   | 0.017                                   | 210 |
| 6-month  | 0.233***<br>(3.63)   | 0.708***<br>(9.15)  | 0.306                                   | 0.032                                   | 204 |
| 12-month | 0.239<br>(1.21)      | 0.456***<br>(2.90)  | 0.094                                   | 0.023                                   | 192 |

Realized inflation, the market forecast, and the model forecast are standardized within each horizon to mean zero and unit variance before estimation.

$\Delta$  Adj.  $R^2$  (Model | Market) is the increase from adding the model forecast to the market-only regression.  $\Delta$  Adj.  $R^2$  (Market | Model) is the increase from adding the market benchmark to the model-only regression.

T-statistics use Newey–West standard errors, with lag length equal to the forecast horizon.

Significance levels: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .

## C Formal Hypotheses

This appendix states the formal hypotheses corresponding to the principal empirical tests reported in the main text. The hypotheses follow the sequence of the Forecast Alpha framework: incremental information beyond the market benchmark, state dependence of calibrated Forecast Alpha, and subsequent market relevance. Recursive calibration and residual-surprise separation are treated as supporting diagnostics rather than as independent structural hypotheses.

### C.1 Incremental Information Beyond Market Expectations

The first empirical question is whether the model forecast contains information about future inflation beyond that already contained in the contemporaneous market benchmark. For each horizon  $h$ , the encompassing specification is

$$\pi_{t+h} = a_h + b_{m,h} \hat{\pi}_t^{market} + b_{f,h} \hat{\pi}_{t+h|t}^{model} + \varepsilon_{t+h}.$$

The formal hypothesis is

$$H_0 : b_{f,h} = 0 \quad \text{vs.} \quad H_1 : b_{f,h} \neq 0.$$

Rejection of the null indicates that the model contributes predictive information conditional on the contemporaneous market benchmark. This is the principal test establishing whether model–market disagreement can contain incremental information before recursive calibration is applied.

## C.2 State Dependence of Calibrated Forecast Alpha

The second empirical question is whether the magnitude of calibrated Forecast Alpha varies across filtered TV-HMM states. Let  $S_t$  denote the dominant filtered state at forecast origin  $t$ . The formal null is

$$H_0 : E\left[|FA_{t,h}^{cal}| \mid S_t = k\right] = E\left[|FA_{t,h}^{cal}| \mid S_t = j\right] \quad \forall k, j,$$

against the alternative

$$H_1 : E\left[|FA_{t,h}^{cal}| \mid S_t = k\right] \neq E\left[|FA_{t,h}^{cal}| \mid S_t = j\right] \quad \text{for some } k, j.$$

The empirical implementation uses persistence-preserving circular-shift inference rather than an i.i.d. ANOVA framework. Rejection of the null indicates that the magnitude of historically calibrated model–market disagreement is state dependent.

## C.3 Market Relevance of Calibrated Forecast Alpha

The third empirical question is whether calibrated Forecast Alpha is associated with subsequent breakeven repricing. The main text evaluates both linear magnitude transmission and directional correspondence.

For the linear specification,

$$\Delta BE_{t,h} = a_h + \beta_h FA_{t,h}^{cal} + u_{t,h},$$

the formal null is

$$H_0^{lin} : \beta_h = 0 \quad \text{vs.} \quad H_1^{lin} : \beta_h \neq 0.$$

The directional test evaluates whether the sign of calibrated Forecast Alpha is associated with the direction of subsequent breakeven repricing:

$$H_0^{dir} : P\left[\text{sign}(FA_{t,h}^{cal}) = \text{sign}(\Delta BE_{t,h})\right] = 0.5,$$

against the alternative

$$H_1^{dir} : P\left[\text{sign}(FA_{t,h}^{cal}) = \text{sign}(\Delta BE_{t,h})\right] > 0.5.$$

The governed implementation also evaluates the mean directional repricing differential. The market-relevance tests are interpreted as evidence about subsequent pricing response rather than as a fully specified trading-strategy backtest.

## C.4 Recursive Calibration and Residual-Separation Diagnostics

Recursive calibration is not treated as a fixed-parameter hypothesis test. The coefficient  $\hat{\lambda}_{t,h}$  is estimated recursively using only information available before each forecast origin and is allowed to vary through time. Its empirical distribution is therefore summarized descriptively rather than tested against a fixed population value such as one.

Residual-surprise diagnostics provide a supporting check on whether recursive calibration improves empirical separation between calibrated Forecast Alpha and the remaining market-relative inflation surprise. The relevant regressions are interpreted cautiously: failure to reject a zero relationship is not evidence of structural orthogonality. Instead, the diagnostics indicate whether the estimated association is statistically distinguishable from zero after calibration.

## C.5 Distributional and Calibration Diagnostics

This subsection characterizes the empirical distributions of raw Forecast Alpha, calibrated Forecast Alpha, calibrated Residual Surprise, and the recursively estimated calibration coefficient. The objective is to document location, dispersion, asymmetry, tail behavior, and serial dependence in the decomposition objects used throughout the paper.

Table 23 reports the principal distributional moments by forecast horizon.

Table 23: Distributional Moments of Forecast Alpha and Residual Surprise

| Variable                     | Horizon | <i>N</i> | Mean  | Median | SD    | Q1     | Q3    | IQR   | P5     | P95   | Skew. | Ex. Kurt. |
|------------------------------|---------|----------|-------|--------|-------|--------|-------|-------|--------|-------|-------|-----------|
| Raw Forecast Alpha           | 1M      | 214      | 0.300 | -0.010 | 0.929 | -0.240 | 0.445 | 0.685 | -0.644 | 2.522 | 1.581 | 1.777     |
|                              | 3M      | 210      | 0.300 | 0.001  | 0.922 | -0.244 | 0.459 | 0.702 | -0.637 | 2.486 | 1.595 | 1.967     |
|                              | 6M      | 204      | 0.294 | 0.005  | 0.936 | -0.276 | 0.396 | 0.672 | -0.709 | 2.318 | 1.563 | 1.937     |
|                              | 12M     | 192      | 0.238 | 0.013  | 0.733 | -0.235 | 0.397 | 0.632 | -0.610 | 1.825 | 1.090 | 0.429     |
| Calibrated Forecast Alpha    | 1M      | 214      | 0.276 | -0.008 | 0.828 | -0.204 | 0.413 | 0.616 | -0.543 | 2.301 | 1.599 | 1.691     |
|                              | 3M      | 210      | 0.237 | 0.001  | 0.672 | -0.152 | 0.353 | 0.505 | -0.400 | 1.961 | 1.674 | 1.827     |
|                              | 6M      | 204      | 0.193 | 0.002  | 0.546 | -0.125 | 0.230 | 0.355 | -0.331 | 1.567 | 1.674 | 1.802     |
|                              | 12M     | 192      | 0.150 | 0.017  | 0.426 | -0.126 | 0.221 | 0.347 | -0.288 | 1.101 | 1.267 | 0.851     |
| Calibrated Residual Surprise | 1M      | 214      | 0.035 | -0.010 | 0.264 | -0.097 | 0.121 | 0.218 | -0.291 | 0.543 | 1.811 | 5.541     |
|                              | 3M      | 210      | 0.082 | -0.005 | 0.465 | -0.178 | 0.226 | 0.404 | -0.475 | 1.111 | 1.665 | 3.605     |
|                              | 6M      | 204      | 0.155 | 0.033  | 0.647 | -0.242 | 0.359 | 0.600 | -0.590 | 1.791 | 1.299 | 2.293     |
|                              | 12M     | 192      | 0.247 | -0.051 | 0.936 | -0.317 | 0.417 | 0.734 | -0.588 | 2.536 | 1.939 | 3.557     |

*Notes:* SD is the sample standard deviation. Q1 and Q3 are the 25th and 75th percentiles; IQR is Q3 minus Q1. P5 and P95 are the 5th and 95th percentiles. Skew. is sample skewness, and Ex. Kurt. is excess kurtosis relative to the normal distribution. Statistics are descriptive; overlapping forecast horizons induce serial dependence.

Raw Forecast Alpha has a median close to zero at every horizon but a positive mean and positive skewness, indicating that the average model–market disagreement is influenced by relatively large positive observations. Recursive calibration reduces the dispersion and absolute magnitude of Forecast Alpha. In contrast, the dispersion of calibrated Residual Surprise increases with the

forecast horizon, consistent with a larger unresolved component of market-relative inflation variation at longer horizons.

Table 24 reports the distribution of the recursive calibration coefficient.

Table 24: Distribution of the Recursive Calibration Coefficient

| Horizon | $N$ | Mean  | Median | SD    | Q1    | Q3    | P5    | P95   | % < 0 | %[0, 1] | % > 1 |
|---------|-----|-------|--------|-------|-------|-------|-------|-------|-------|---------|-------|
| 1M      | 214 | 0.864 | 0.855  | 0.043 | 0.842 | 0.864 | 0.806 | 0.939 | 0.0%  | 100.0%  | 0.0%  |
| 3M      | 210 | 0.659 | 0.625  | 0.099 | 0.608 | 0.645 | 0.544 | 0.855 | 0.0%  | 100.0%  | 0.0%  |
| 6M      | 204 | 0.510 | 0.468  | 0.123 | 0.450 | 0.497 | 0.411 | 0.758 | 0.0%  | 100.0%  | 0.0%  |
| 12M     | 192 | 0.568 | 0.586  | 0.120 | 0.579 | 0.601 | 0.324 | 0.680 | 1.6%  | 98.4%   | 0.0%  |

*Notes:* The calibration coefficient  $\hat{\lambda}_{t,h}$  is estimated recursively from information available at each forecast origin. Values between zero and one imply attenuation of raw model–market disagreement; values below zero imply an opposite-signed historical mapping between model–market disagreement and subsequently realized market-relative inflation surprise; values above one imply amplification. These ranges are descriptive and are not imposed as parameter restrictions.

The calibration coefficient lies between zero and one for every calibrated observation at the 1-, 3-, and 6-month horizons and for 98.4 percent of observations at the 12-month horizon. This pattern indicates that recursive calibration generally attenuates raw model–market disagreement rather than amplifying it. A small share of 12-month estimates is negative, reflecting occasional opposite-signed historical mappings between model–market disagreement and subsequently realized market-relative inflation surprise at the longest forecast horizon.

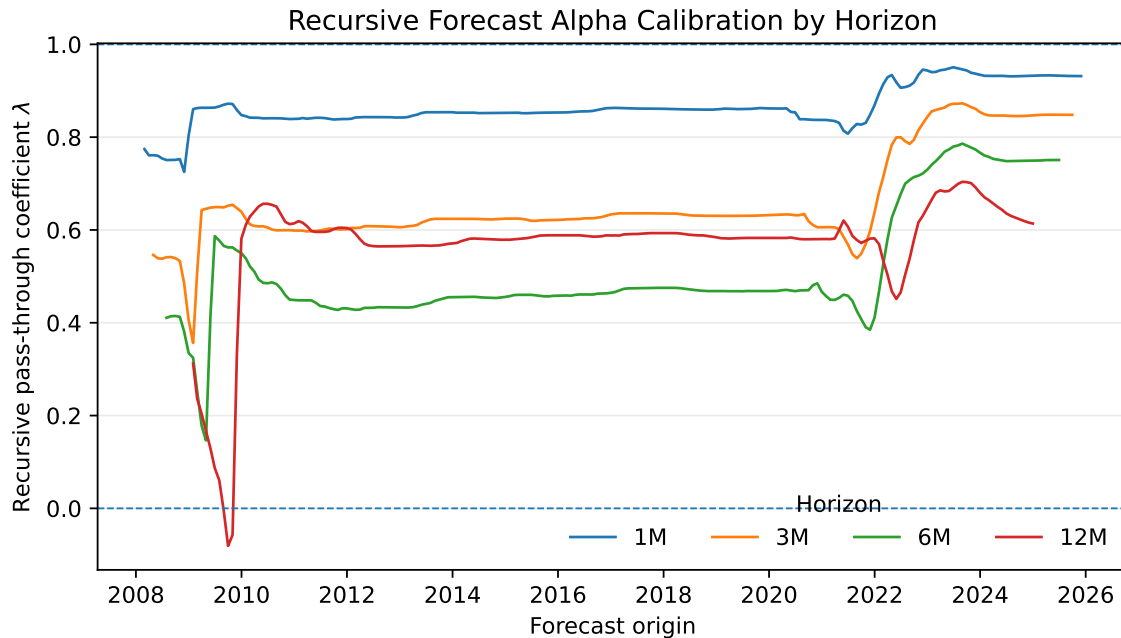


Figure 7: Recursive Forecast Alpha Calibration by Horizon

*Notes:* The figure plots the recursively estimated calibration coefficient,  $\hat{\lambda}_{t,h}$ , used to scale raw model–market disagreement into calibrated Forecast Alpha. Horizontal reference lines at zero and one are descriptive only; no bounds are imposed on the recursive coefficient.

Table 25 reports formal normality and serial-dependence diagnostics.

Table 25: Distributional and Serial-Dependence Diagnostics

| Variable                     | Horizon | $N$ | JB Statistic | JB $p$   | LB Lag | LB $p$   |
|------------------------------|---------|-----|--------------|----------|--------|----------|
| Raw Forecast Alpha           | 1M      | 214 | 113.90       | < 0.0001 | 1      | < 0.0001 |
|                              | 3M      | 210 | 119.05       | < 0.0001 | 3      | < 0.0001 |
|                              | 6M      | 204 | 111.27       | < 0.0001 | 6      | < 0.0001 |
|                              | 12M     | 192 | 38.65        | < 0.0001 | 12     | < 0.0001 |
| Calibrated Forecast Alpha    | 1M      | 214 | 113.40       | < 0.0001 | 1      | < 0.0001 |
|                              | 3M      | 210 | 123.68       | < 0.0001 | 3      | < 0.0001 |
|                              | 6M      | 204 | 119.35       | < 0.0001 | 6      | < 0.0001 |
|                              | 12M     | 192 | 55.70        | < 0.0001 | 12     | < 0.0001 |
| Calibrated Residual Surprise | 1M      | 214 | 373.80       | < 0.0001 | 1      | < 0.0001 |
|                              | 3M      | 210 | 202.34       | < 0.0001 | 3      | < 0.0001 |
|                              | 6M      | 204 | 97.97        | < 0.0001 | 6      | < 0.0001 |
|                              | 12M     | 192 | 212.82       | < 0.0001 | 12     | < 0.0001 |

*Notes:* JB is the Jarque–Bera normality test. LB is the Ljung–Box serial-correlation test using the reported horizon-matched lag. Rejection of normality or serial independence motivates heteroskedasticity- and autocorrelation-robust inference in subsequent tests.

Jarque–Bera tests reject normality, and horizon-matched Ljung–Box tests reject serial independence, for each reported variable and forecast horizon. These findings show that the decomposition objects are not well approximated by a Gaussian i.i.d. process and support the paper’s use of heteroskedasticity- and autocorrelation-consistent inference in subsequent empirical tests.

**Residual-surprise separation diagnostics.** As a supporting diagnostic, the raw and calibrated residual components are regressed on their corresponding Forecast Alpha measures. The exercise evaluates whether recursive calibration improves empirical separation between the model–market signal and the remaining realized market-relative inflation surprise; it is not interpreted as proof of structural orthogonality.

The raw decomposition rejects orthogonality at conventional significance levels at all four horizons. Following recursive calibration, the estimated relationship between calibrated Residual Surprise and calibrated Forecast Alpha is not statistically distinguishable from zero at conventional significance levels at the reported horizons. This pattern is consistent with improved signal–residual separation after calibration, while the diagnostic nature of the exercise warrants a more limited interpretation than structural independence.

Table 26: Residual-Surprise Orthogonality Diagnostics by Horizon

| Horizon | $N$ | Raw<br>$\beta$ | Raw<br>HAC $p$ | Cal.<br>$\beta$ | Cal.<br>HAC $p$ | Interpretation                   |
|---------|-----|----------------|----------------|-----------------|-----------------|----------------------------------|
| H1      | 214 | -0.071         | 0.020          | 0.040           | 0.289           | Calibration removes 5% rejection |
| H3      | 210 | -0.158         | 0.021          | 0.141           | 0.122           | Calibration removes 5% rejection |
| H6      | 204 | -0.260         | 0.006          | 0.216           | 0.219           | Calibration removes 5% rejection |
| H12     | 192 | -0.425         | 0.018          | -0.095          | 0.785           | Calibration removes 5% rejection |

The raw regression relates raw Residual Surprise to raw Forecast Alpha. The calibrated regression relates calibrated Residual Surprise to calibrated Forecast Alpha. Regressions include an intercept and use heteroskedasticity- and autocorrelation-consistent standard errors with lag length equal to the forecast horizon. Failure to reject in the calibrated regression indicates that the calibrated residual is not systematically related to the calibrated forecastable component at conventional significance levels.

## D Selected Stress-Episode Robustness

Table 27 examines Forecast Alpha behavior during three economically salient inflation stress episodes using the same governed, origin-aligned observations employed elsewhere in the paper. The exercise is descriptive and is not used to select the forecasting model, forecast horizon, TV-HMM architecture, or market-relevance specification.

The episodes illustrate why forecast information and subsequent market repricing should be treated as distinct empirical questions. During the 2021–22 inflation surge, the Compact Inflation Model records lower RMSE than the market benchmark at every evaluated horizon and calibrated Forecast Alpha is comparatively large, particularly at shorter horizons. Subsequent horizon-matched breakeven repricing, however, is not consistently aligned with the signal during that episode. The 2008–09 episode displays a different pattern, including positive H12 FA-aligned repricing, while the 2020 COVID episode provides weaker forecast-relative and market-repricing evidence.

These differences are consistent with the paper’s narrower interpretation of Forecast Alpha. A model can contain incremental information relative to a market benchmark without implying that the market must subsequently reprice in the same direction over every horizon and macroeconomic episode. Breakeven movements may also reflect changes in inflation risk compensation, liquidity conditions, policy expectations, and the basis between CPI-linked market pricing and the Core PCE forecast target. The stress-period evidence is therefore treated as robustness evidence rather than as a separate trading-performance test.

Table 27: Forecast Alpha Behavior Across Selected Inflation Stress Episodes

| <b>Panel A. Forecast Accuracy and Calibrated Forecast Alpha</b> |         |     |            |             |            |                   |  |
|-----------------------------------------------------------------|---------|-----|------------|-------------|------------|-------------------|--|
| Episode                                                         | Horizon | $N$ | Model RMSE | Market RMSE | RMSE Diff. | Mean $ FA^{cal} $ |  |
| 2008–09 GFC / commodity shock                                   | 1M      | 23  | 0.285      | 0.959       | -0.674     | 0.526             |  |
| 2008–09 GFC / commodity shock                                   | 3M      | 21  | 0.687      | 0.893       | -0.206     | 0.450             |  |
| 2008–09 GFC / commodity shock                                   | 6M      | 18  | 1.025      | 0.994       | 0.032      | 0.425             |  |
| 2008–09 GFC / commodity shock                                   | 12M     | 12  | 0.531      | 0.750       | -0.219     | 0.139             |  |
| 2020 COVID shock                                                | 1M      | 10  | 0.373      | 0.227       | 0.145      | 0.327             |  |
| 2020 COVID shock                                                | 3M      | 10  | 0.628      | 0.320       | 0.308      | 0.263             |  |
| 2020 COVID shock                                                | 6M      | 10  | 1.280      | 1.098       | 0.182      | 0.242             |  |
| 2020 COVID shock                                                | 12M     | 10  | 2.685      | 2.651       | 0.033      | 0.199             |  |
| 2021–22 inflation surge                                         | 1M      | 24  | 0.563      | 2.138       | -1.575     | 1.717             |  |
| 2021–22 inflation surge                                         | 3M      | 24  | 0.944      | 2.239       | -1.296     | 1.329             |  |
| 2021–22 inflation surge                                         | 6M      | 24  | 1.116      | 2.334       | -1.218     | 1.085             |  |
| 2021–22 inflation surge                                         | 12M     | 24  | 1.817      | 2.237       | -0.420     | 0.804             |  |

| <b>Panel B. Subsequent Breakeven Repricing</b> |         |     |                      |                                |  |
|------------------------------------------------|---------|-----|----------------------|--------------------------------|--|
| Episode                                        | Horizon | $N$ | Directional Hit Rate | Mean FA-Aligned Repricing (bp) |  |
| 2008–09 GFC / commodity shock                  | 1M      | 23  | 52.2%                | -8.74                          |  |
| 2008–09 GFC / commodity shock                  | 3M      | 21  | 42.9%                | -11.90                         |  |
| 2008–09 GFC / commodity shock                  | 6M      | 18  | 44.4%                | 12.72                          |  |
| 2008–09 GFC / commodity shock                  | 12M     | 12  | 66.7%                | 45.58                          |  |
| 2020 COVID shock                               | 1M      | 10  | 40.0%                | -5.60                          |  |
| 2020 COVID shock                               | 3M      | 10  | 40.0%                | -15.00                         |  |
| 2020 COVID shock                               | 6M      | 10  | 50.0%                | 7.40                           |  |
| 2020 COVID shock                               | 12M     | 10  | 50.0%                | 14.10                          |  |
| 2021–22 inflation surge                        | 1M      | 24  | 33.3%                | -2.75                          |  |
| 2021–22 inflation surge                        | 3M      | 24  | 50.0%                | -3.38                          |  |
| 2021–22 inflation surge                        | 6M      | 24  | 37.5%                | -10.46                         |  |
| 2021–22 inflation surge                        | 12M     | 24  | 20.8%                | -39.88                         |  |

The table reports descriptive robustness evidence for three fixed economically salient stress episodes. Episode windows are applied to forecast origins and are not selected using Forecast Alpha outcomes.

Panel A uses the governed Compact-model calibrated sample. RMSE Diff. is Model RMSE minus Market RMSE, so negative values indicate lower forecast error for the Compact Inflation Model. Mean  $|FA^{cal}|$  is the mean absolute magnitude of recursively calibrated model–market disagreement.

Panel B uses the governed Step 6 horizon-matched T5YIE response panel. Directional Hit Rate is the fraction of observations for which the sign of the subsequent breakeven move matches the sign of calibrated Forecast Alpha. Mean FA-Aligned Repricing is  $\text{sign}(FA^{cal})\Delta T5YIE$ , expressed in basis points; positive values indicate repricing in the direction of Forecast Alpha and negative values indicate repricing against it.

The two panels evaluate distinct questions. Lower inflation-forecast error or larger calibrated Forecast Alpha does not mechanically imply same-direction subsequent breakeven repricing. Because the episode samples are short, the results are interpreted descriptively rather than as separate within-episode inference tests.

## E Supplementary Regime and Forecast Diagnostics

### E.1 Inflation-Band Architecture Selection

The governed TV-HMM state count is selected using year-over-year core PCE inflation before the regime classifications are evaluated against Forecast Alpha outcomes. Candidate one-dimensional Gaussian and Student- $t$  mixture specifications are estimated over alternative values of  $K$ . The minimum-BIC mixture component count determines the governed TV-HMM state count. This architecture-selection step is distinct from latent-state assignment: the estimated inflation thresholds provide initialization information and economic context but do not mechanically classify observations into TV-HMM states.

The preferred mixture specification is selected using the Bayesian information criterion,

$$\text{BIC} = -2\ell(\hat{\theta}) + p \log(n),$$

where  $\ell(\hat{\theta})$  is the maximized log likelihood,  $p$  is the number of estimated parameters, and  $n$  is the number of observations. On the governed 792-observation architecture-selection sample, the minimum-BIC specification is a four-component Gaussian mixture. The resulting estimated decision boundaries are 2.412, 5.625, and 8.019 percent.

Table 28 reports the empirical inflation bands implied by those boundaries. The first two bands contain most observations, while the upper two bands represent progressively less common elevated-inflation environments. The highest-inflation band contains 35 observations, or 4.42 percent of the architecture-selection sample. That limited support is informative about the historical upper tail, but the inflation thresholds themselves do not mechanically assign observations to latent states.

The resulting governed regime layer is a four-state TV-HMM. State membership is determined recursively from filtered posterior probabilities rather than from the inflation thresholds in Table 28. The numerical state labels are recursively aligned through time and should therefore not be interpreted as mechanically ordered inflation bands. A separate timing-valid  $K=3$  exercise based on the initial 120-observation sample is retained as a historical sensitivity diagnostic and does not control the operational state count.

Table 28: Preliminary BIC-Guided Inflation-Band Diagnostic

| Preliminary Band | Inflation Range          | Mean Inflation | $N$ | Sample Share |
|------------------|--------------------------|----------------|-----|--------------|
| Band 0           | $\pi \leq 2.412$         | 1.651          | 397 | 50.13%       |
| Band 1           | $2.412 < \pi \leq 5.625$ | 3.826          | 286 | 36.11%       |
| Band 2           | $5.625 < \pi \leq 8.019$ | 6.637          | 74  | 9.34%        |
| Band 3           | $\pi > 8.019$            | 9.178          | 35  | 4.42%        |

Notes: Inflation bands are estimated from year-over-year core PCE inflation over the governed 792-observation architecture-selection sample. Candidate Gaussian and Student- $t$  mixture specifications with  $K = 2, \dots, 5$  are compared using BIC, and the minimum-BIC specification is a four-component Gaussian mixture. The selected component count determines the governed TV-HMM state count, while the estimated thresholds do not mechanically assign observations to the recursively estimated TV-HMM states. Inflation values are reported in percentage points.

## E.2 Economic Characterization of the Latent States

The four TV-HMM states are latent statistical classifications rather than prespecified macroeconomic regimes. State labels are recursively aligned through time, and neither their numerical ordering nor their economic interpretation is imposed by inflation thresholds. The characterization below is therefore descriptive and ex post: it summarizes the economic conditions historically associated with each recursively estimated state without treating those conditions as mechanical classification rules.

Figure 8 reports probability-weighted economic profiles over the governed recursive K4 history from June 1989 through December 2025. The standardized economic fingerprint condenses five dimensions—inflation pressure, labor-market tightness, activity, policy and financial restraint, and pricing pressure and expectations—onto a common scale centered on the recursive-sample average. The natural-unit measures beneath each fingerprint provide observable macroeconomic context, while the Forecast Alpha statistics show how the magnitude of calibrated model–market disagreement has historically differed across states.

## Economic Profiles of the Four Latent TV-HMM States

HOW TO READ | Fingerprint scores are standardized; 0 = recursive-sample average; red = above-average intensity and blue = below-average intensity. Typical conditions are probability-weighted historical state averages. Forecast Alpha is calibrated model-market disagreement; Alpha Share is its share of market-relative surprise. Endpoint arrows show recent movement within the current state—not the probability or forecast of another state.

### S0 Near-Target Inflation / High Utilization

#### Economic fingerprint

| Inflation | Labor | Activity | Policy | Pricing / Expect. |
|-----------|-------|----------|--------|-------------------|
| -0.16     | +0.02 | +0.30    | +0.12  | -0.11             |

#### Typical observable conditions

|              |       |                       |       |
|--------------|-------|-----------------------|-------|
| Core PCE     | 2.15% | Capacity utilization  | 80.2% |
| Unemployment | 5.56% | Fed funds             | 4.04% |
| CFNAI        | -0.05 | PPI 3M change         | +0.72 |
|              |       | Michigan expectations | 3.00% |

#### Forecast Alpha characteristics

|                   |                          |
|-------------------|--------------------------|
| Mean $ FA^{cal} $ | H1 0.12 pp → H12 0.09 pp |
| Alpha Share       | H1 43% → H12 15%         |

### S1 Low Inflation / Slack / Accommodation

#### Economic fingerprint

| Inflation | Labor | Activity | Policy | Pricing / Expect. |
|-----------|-------|----------|--------|-------------------|
| -0.38     | -0.34 | -0.50    | -0.61  | -0.18             |

#### Typical observable conditions

|              |       |                       |       |
|--------------|-------|-----------------------|-------|
| Core PCE     | 1.52% | Capacity utilization  | 75.8% |
| Unemployment | 6.58% | Fed funds             | 0.65% |
| CFNAI        | -0.11 | PPI 3M change         | +0.78 |
|              |       | Michigan expectations | 2.88% |

#### Forecast Alpha characteristics

|                   |                          |
|-------------------|--------------------------|
| Mean $ FA^{cal} $ | H1 0.24 pp → H12 0.16 pp |
| Alpha Share       | H1 65% → H12 22%         |

### S2 Elevated Inflation / Restrictive Policy / Soft Activity

Current state at Dec. 2025 | posterior probability 99.98%

#### Economic fingerprint

| Inflation | Labor | Activity | Policy | Pricing / Expect. |
|-----------|-------|----------|--------|-------------------|
| +1.16     | +0.50 | -0.24    | +0.97  | +0.73             |

#### Typical observable conditions

|              |       |                       |       |
|--------------|-------|-----------------------|-------|
| Core PCE     | 3.41% | Capacity utilization  | 77.4% |
| Unemployment | 4.45% | Fed funds             | 4.44% |
| CFNAI        | -0.14 | PPI 3M change         | +2.80 |
|              |       | Michigan expectations | 4.01% |

#### Forecast Alpha characteristics

|                   |                          |
|-------------------|--------------------------|
| Mean $ FA^{cal} $ | H1 0.99 pp → H12 0.55 pp |
| Alpha Share       | H1 76% → H12 39%         |

#### Direction within current state — prior 3 months

|                    |                          |            |
|--------------------|--------------------------|------------|
| Inflation ↓        | Labor ↓                  | Activity → |
| Policy restraint ↓ | Pricing / expectations ↓ |            |

### S3 Inflation Breakout / Very Tight Labor / Elevated Expectations

#### Economic fingerprint

| Inflation | Labor | Activity | Policy | Pricing / Expect. |
|-----------|-------|----------|--------|-------------------|
| +2.05     | +1.02 | -0.04    | +0.31  | +1.10             |

#### Typical observable conditions

|              |       |                       |       |
|--------------|-------|-----------------------|-------|
| Core PCE     | 5.19% | Capacity utilization  | 78.3% |
| Unemployment | 3.63% | Fed funds             | 2.86% |
| CFNAI        | -0.05 | PPI 3M change         | +2.65 |
|              |       | Michigan expectations | 4.76% |

#### Forecast Alpha characteristics

|                   |                          |
|-------------------|--------------------------|
| Mean $ FA^{cal} $ | H1 2.37 pp → H12 1.09 pp |
| Alpha Share       | H1 89% → H12 66%         |

Notes: Economic profiles are probability-weighted summaries over the governed recursive K4 history, June 1989–December 2025. The state descriptions are ex post economic characterizations of recursively aligned latent states; they do not mechanically define state membership. Forecast Alpha statistics use the governed 192-common-origin evaluation sample. Endpoint arrows describe recent within-state movement in standardized dimensions and are not next-state probabilities.

Figure 8: Economic profiles of the four latent TV-HMM states. Economic fingerprints and observable conditions are probability-weighted summaries of the governed recursive K4 history. Forecast Alpha statistics are calculated on the governed 192-common-origin evaluation sample. The state descriptions are interpretive labels applied after estimation and do not determine state membership. At the December 2025 paper endpoint, the filtered posterior assigns State 2 probability of approximately 99.98%. The directional indicators for State 2 summarize recent movement within the current state and should not be interpreted as forecasts or probabilities of transition to another state.

The four states have economically distinct profiles, but they do not form a simple low-to-high inflation ladder. State 0 is characterized by near-target inflation and relatively high utilization. State 1 combines lower inflation with greater economic slack and accommodative policy conditions. State 2 is associated with elevated inflation, restrictive policy and financial conditions, and comparatively soft activity. State 3 represents the most pronounced inflationary configuration, combining the highest inflation pressure with very tight labor conditions and elevated pricing pressure and expectations.

These distinctions also coincide with substantial variation in the magnitude of Forecast Alpha. Mean absolute calibrated Forecast Alpha rises monotonically from State 0 to State 3 at both the 1-month and 12-month horizons: from approximately 0.12 to 2.37 percentage points at H1 and from 0.09 to 1.09 percentage points at H12. These statistics are descriptive state-conditional results rather than evidence that the latent states were selected or optimized to maximize Forecast Alpha.

At the governed December 2025 endpoint, the model remains overwhelmingly classified as State 2. At the same time, the within-state indicators show moderation in inflation pressure, labor tightness, policy and financial restraint, and pricing pressure and expectations over the preceding three months, while activity is approximately stable. Thus, the endpoint can be summarized as remaining firmly within the elevated-inflation, restrictive-policy State 2 environment while several of its principal inflationary and policy-pressure dimensions are moderating. This within-state movement is distinct from, and does not imply, a forecast of the next latent state.