

Inflation Forecasting When the Phillips Curve Is Flat

Frank Schmid

This version: 31 October 2026

First version: 19 March 2026

Abstract. We develop a framework for forecasting Core Personal Consumption Expenditures (PCE) inflation from supply-side indicators that are never revised and are available to the forecaster at the time of the decision. During the U.S. inflation-targeting era the Phillips curve is flat: the output gap carries no incremental forecasting power for inflation, in sample or out of sample. Never-revised supply-side indicators—the New York Fed Global Supply Chain Pressure Index (GSCPI) and the prices-paid index of the Federal Reserve Bank of Philadelphia’s manufacturing survey—improve out-of-sample forecast accuracy relative to an autoregressive benchmark, and the margin widens with the forecast horizon under a protocol a forecaster can implement in real time. The real-time discipline is tested with first releases from the Philadelphia Fed’s real-time data set: the heavily revised unit-labor-cost series carries no resolved forecasting content on either vintage, while the lightly revised producer price index, used as first released, forecasts as its current vintage does and carries no resolved content next to the GSCPI. The sample does not separate the GSCPI alone from the model that adds the prices-paid index at any horizon. Granger causality tests attribute predictive power to the supply-side indicators and none to the federal funds rate. Richer parameterizations do not survive: asymmetric indicator terms change out-of-sample accuracy by a few hundredths of a point in either direction, and a two-regime Markov-switching model is not identified on the sample—a multistart search registered before it ran finds its likelihood maximized at a boundary and its optima and out-of-sample record dependent on the starting values—a parsimony result that bounds what the fourteen-year sample can support.

JEL classification: E31, E37, E52, C53.

This paper was developed with assistance from generative AI tools, including Claude, across the design, analysis, coding, and drafting stages. The author supervised all stages, verified the results, and takes full responsibility for the content.

Contents

Abbreviations and Acronyms	iii
Executive Summary	iv
1 Introduction	1
2 The Flat Phillips Curve: Motivation	2
2.1 Estimation	2
2.2 Why Is the Phillips Curve Flat?	2
3 Real-Time Data: Methodology and Classification	3
3.1 Variable Classification	3
3.2 Three Data Versions	3
3.3 Why Real-Time Data Is the Baseline	4
4 Supply-Side Leading Indicators	4
4.1 The Producer-Cost Pipeline	4
4.2 The Eight Indicators	4
4.3 The GSCPI as a Predictor	4
5 Forecasting Models and Estimation	5
5.1 General Framework	5
5.2 Model Specifications	5
5.3 Model Selection	5
6 In-Sample Results: Three Data Versions Compared	6
6.1 The Output Gap Does Not Improve Fit	6
6.2 Supply-Side Indicators Improve Fit	6
6.3 The Real-Time Discipline	7
7 Out-of-Sample Forecast Evaluation	7
7.1 Methodology	7
7.2 Results	8
7.3 Key Findings	8
8 Multi-Horizon Iterated Forecasts	9
8.1 Design	9
8.2 Findings	11
9 Four-Quarter-Ahead Direct Forecasts	11
10 The Pandemic and Its Aftermath: A Sub-Period Decomposition	12
10.1 Asymmetric GSCPI Specifications	13

11 2025 Genuine Out-of-Sample Validation	14
11.1 The 2025 Macroeconomic Context	14
11.2 Validation Results	15
11.3 Interpreting the Results	15
12 Regime-Switching Models	15
12.1 Specification and Estimation	15
12.2 A Registered Multistart Search	16
12.3 What the Evidence Supports	17
13 Does the Fed's Reaction Function Matter?	18
13.1 Granger Causality Tests	18
14 Implications for Monetary Policy	18
14.1 Better Forecasting Does Not Deliver Better Control	18
14.2 Where Better Forecasting Helps	18
15 Conclusions	19
A Data Sources and Provenance	20
Bibliography	22

Abbreviations and Acronyms

AIC	Akaike Information Criterion
AR(1)	First-order autoregressive model
BEA	Bureau of Economic Analysis
BLS	Bureau of Labor Statistics
CBO	Congressional Budget Office
CPS	Current Population Survey
COVID-19	Coronavirus disease 2019
EIA	Energy Information Administration
FAO	Food and Agriculture Organization
FFR	Federal funds rate
FOMC	Federal Open Market Committee
FRED	Federal Reserve Economic Data (St. Louis Fed)
GDP	Gross domestic product
GFC	Global Financial Crisis
GSCPI	Global Supply Chain Pressure Index (New York Fed)
MAE	Mean absolute error
MLE	Maximum likelihood estimation
NBER	National Bureau of Economic Research
OLS	Ordinary least squares
OOS	Out of sample
PCE	Personal Consumption Expenditures
PPI	Producer Price Index
RMSE	Root mean squared error
RTDSM	Real-Time Data Set for Macroeconomists (Philadelphia Fed)
ULC	Unit labor costs
WTI	West Texas Intermediate (crude oil)

Greek Letters

α	alpha (AL-fuh): regression intercept
β	beta (BAY-tuh): regression slope coefficient
ε	epsilon (EP-sih-lon): regression error term
π	pi (PIE): year-over-year Core PCE inflation
σ	sigma (SIG-muh): regime residual standard deviation

Executive Summary

This paper evaluates inflation forecasting models for the U.S. inflation-targeting era, 2012:Q1–2025:Q4, under a real-time data discipline: the preferred model uses only indicators that are never revised and are observable when the forecast is made. A reduced-form Phillips curve estimated on this sample yields an output-gap coefficient of 0.0098 ($t = 0.32$); adding the output gap to an autoregressive model raises in-sample R^2 by 0.001 and worsens the out-of-sample RMSE from 0.556 to 0.557. The demand side does not forecast inflation in this era.

The preferred specification, Model R, forecasts next-quarter Core PCE inflation from current inflation, the GSCPI, and the Philadelphia Fed prices-paid index. Over the 24-quarter evaluation window 2020:Q1–2025:Q4, Model R attains an out-of-sample RMSE of 0.506 against 0.556 for the AR(1) benchmark, an improvement of 9.1%. On a common target window at horizons of one through four quarters, with indicators held at their origin-date values so that the forecast is implementable in real time, the improvement rises from 13.3% at one quarter to 33.5% at four quarters; the GSCPI alone improves by 17.5% at one quarter and by a flat 23.7% to 27.0% across horizons, so in the point estimates the prices-paid index costs accuracy at one quarter and earns it at three and four. A Diebold–Mariano test of equal accuracy resolves none of the differences between the two models (p between 0.151 and 0.892 across the four horizons)¹; Model R is kept as the reference specification because the in-sample selection chose it before any out-of-sample comparison. A sub-period decomposition attributes the gains to 2022–2025, when supply-chain pressure abated and inflation descended and then held near its level; the model’s accuracy deteriorates in 2020 (demand collapse) and in the 2021 surge, when both models under-predicted and the AR(1) by less.

Three results discipline the interpretation. First, the real-time discipline is tested with first releases: the unit-labor-cost coefficient is not resolved with revised data ($t = -0.83$) or as first released ($t = -0.08$), and a measurement-error simulation with the observed revision moments leaves 3.1% of draws resolved; the producer price index as first released, by contrast, forecasts as its current vintage does (one-quarter-ahead RMSE 0.462 against 0.463), so first-release data, not only never-revised data, satisfy the discipline; next to the GSCPI its coefficient is not resolved ($t = 0.50$ as first released, 0.55 revised), so Model D’s one-quarter accuracy is the GSCPI’s. Second, the federal funds rate does not Granger-cause inflation ($F = 2.26$), while the GSCPI ($F = 9.15$) and the Philadelphia Fed prices-paid index ($F = 4.37$) do. Third, richer parameterizations add no accuracy the sample can attribute to them: the four asymmetric GSCPI specifications move the out-of-sample RMSE between -0.027 and $+0.012$ relative to Model R, and a two-regime Markov-switching model, whose parameters a registered multistart search shows the sample does not identify, records an out-of-sample RMSE between 0.500 and 1.111 across restart conventions, against the AR(1)’s 0.556. Its regime parameters are not interpreted.

Every series is read from its publisher’s file under the extraction rule recorded in Appendix A, which names the file, its digest, and its access date.

¹Francis X. Diebold and Roberto S. Mariano, “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics* 13, no. 3 (1995): 253–263; David Harvey, Stephen Leybourne, and Paul Newbold, “Testing the Equality of Prediction Mean Squared Errors,” *International Journal of Forecasting* 13, no. 2 (1997): 281–291. Squared-error loss; the variance of the loss differential by Newey–West with $h - 1$ lags at horizon h ; the small-sample correction of Harvey, Leybourne, and Newbold; two-sided p -values from the t distribution with $n - 1$ degrees of freedom.

1. Introduction

Inflation forecasting is central to the conduct of monetary policy: the Federal Reserve sets the federal funds rate on its assessment of where inflation is headed, not where it has been. For decades the standard instrument for that assessment has been some variant of the Phillips curve. The real-time data literature has demonstrated that analysis conducted on revised data can reach conclusions different from those available to the policymaker at the time.² This paper brings that discipline to inflation forecasting.

The paper documents that the Phillips curve is non-functional as a forecasting tool during the U.S. inflation-targeting era, 2012:Q1–2025:Q4. A reduced-form Phillips curve estimated on this sample yields an output-gap coefficient of 0.0098 with a t -statistic of 0.32; adding the output gap to an autoregressive inflation model raises R^2 by 0.001 and worsens the AIC. At the one-quarter horizon the output gap provides no forecasting value.

If the demand side of the economy cannot predict inflation, what can? We argue that the answer lies on the supply side. Supply-side pressures—energy costs, global supply-chain disruptions, import prices, producer cost pipelines—were the dominant drivers of inflation dynamics in this era, and they do not appear instantaneously in consumer prices: commodity prices move first, then producer prices, then import prices, then consumer prices. The lags in that sequence create a forecasting opportunity.

We compile eight supply-side leading indicators and evaluate their ability to forecast Core Personal Consumption Expenditures (PCE) inflation—the Federal Reserve’s preferred inflation measure—at one-quarter and four-quarter horizons, using three data versions: revised, first-vintage (first releases), and strictly real-time (a house term glossed at first use in Section 3: regressors restricted to series that are never revised). The paper is the second in a series of companion papers. The first, “Observational Equivalence in Taylor Rule Estimation,” demonstrates that the Federal Reserve’s reaction function cannot be identified from aggregate time series. The present paper extends the argument from the policy rule to the forecasting model: the reaction function is also unnecessary for forecasting inflation. A companion paper, “Recession Forecasting with Credit and Oil Price Channels,” applies the supply-side framework to recession forecasting with the Gilchrist–Zakrajšek excess bond premium as the financial-intermediation indicator.³

The route through the paper is as follows. Section 2 documents the flat Phillips curve. Section 3 sets out the real-time data methodology and classifies every variable by revision status. Section 4 describes the eight supply-side indicators. Section 5 specifies the forecasting models and the selection procedure. Sections 6 and 7 report in-sample and one-quarter-ahead out-of-sample results. Section 8, the principal section, reports the multi-horizon evaluation on a common target window; its central findings are that the real-time-feasible improvement over the AR(1) benchmark rises from 13.3% at one quarter to 33.5% at four quarters, and that conditioning on realized future indicator values—a diagnostic, not a forecast—raises the four-quarter figure to 43.0%. Section 9 corroborates the horizon profile with direct four-quarter-ahead estimation. Section 10 decomposes performance across five macroeconomic episodes and tests asymmetric indicator specifications. Section 11 conducts a genuine out-of-sample validation on 2025 data. Section 12 subjects a two-regime Markov-switching model to a registered multistart search and reports that the sample does

²Athanasios Orphanides, “Monetary Policy Rules Based on Real-Time Data,” *American Economic Review* 91, no. 4 (2001): 964–985; Dean Croushore and Tom Stark, “A Real-Time Data Set for Macroeconomists,” *Journal of Econometrics* 105, no. 1 (2001): 111–130.

³Simon Gilchrist and Egon Zakrajšek, “Credit Spreads and Business Cycle Fluctuations,” *American Economic Review* 102, no. 4 (2012): 1692–1720.

not identify its regimes. Section 13 tests whether the Fed’s policy stance adds forecasting power. Sections 14 and 15 draw out the implications and conclude.

2. The Flat Phillips Curve: Motivation

This section quantifies, for our sample, the flattening that motivates the supply-side approach. The flattening of the Phillips curve is among the most extensively documented phenomena in recent macroeconomics.⁴

2.1 Estimation

We estimate a backward-looking Phillips curve with two inflation lags, the lagged output gap, and a COVID-19 supply-shock dummy on quarterly data from 2012:Q1 to 2025:Q4 (54 effective observations after lags):

$$\pi_t = 0.123 + 1.165 \pi_{t-1} - 0.237 \pi_{t-2} + 0.0098 \text{gap}_{t-1} + 0.387 \text{COVID-19}_t + \varepsilon_t. \quad (1)$$

The R^2 is 0.949, but the fit is carried by inflation persistence: the lagged inflation coefficients sum to 0.928. The output-gap coefficient is 0.0098 with $t = 0.32$: a one-percentage-point output gap moves next-year inflation by 0.0098 percentage points, an amount indistinguishable from zero statistically and negligible economically. The COVID-19 supply-shock dummy (2020:Q2–2022:Q2) absorbs 0.387 percentage points per quarter ($t = 2.70$), consistent with the dominant inflationary episode in the sample being supply-driven.

2.2 Why Is the Phillips Curve Flat?

The literature identifies several contributing factors. First, anchored inflation expectations under credible inflation targeting mean that temporary changes in slack no longer feed into expected inflation and hence into price setting.⁵ Second, globalization has reduced the sensitivity of domestic prices to domestic slack.⁶ Third, there is an identification effect: a central bank that stabilizes inflation offsets the demand shocks that would render the Phillips curve visible in aggregate time series.⁷ Fourth, the sample contains unusual episodes—the slow recovery after the Global Financial Crisis (GFC), COVID-19—that further obscure the relationship. Whatever the cause, the practical implication is the same: the output gap does not forecast inflation in this sample, and a forecasting framework built on the Phillips curve starts from a position of irrelevance.

⁴Olivier Blanchard, “The Phillips Curve: Back to the ’60s?,” *American Economic Review: Papers & Proceedings* 106, no. 5 (2016): 31–34; Laurence Ball and Sandeep Mazumder, “Inflation Dynamics and the Great Recession,” *Brookings Papers on Economic Activity* (Spring 2011): 337–381; Marco Del Negro, Michele Lenza, Giorgio E. Primiceri, and Andrea Tambalotti, “What’s Up with the Phillips Curve?,” *Brookings Papers on Economic Activity* (Spring 2020): 301–373.

⁵Jonathon Hazell, Juan Herreño, Emi Nakamura, and Jón Steinsson, “The Slope of the Phillips Curve: Evidence from U.S. States,” *Quarterly Journal of Economics* 137, no. 3 (2022): 1299–1344.

⁶Kristin J. Forbes, “Has Globalization Changed the Inflation Process?,” BIS Working Papers No. 791 (Basel: Bank for International Settlements, 2019).

⁷Michael McLeay and Silvana Tenreyro, “Optimal Inflation and the Identification of the Phillips Curve,” *NBER Macroeconomics Annual* 34, no. 1 (2020): 199–255.

3. Real-Time Data: Methodology and Classification

This section classifies every variable by revision status and defines the three data versions used throughout. The motivation comes from the real-time literature: output-gap estimates as first reported can differ from final revised estimates by 2–5 percentage points.⁸ If revised data can distort the analysis of policy rules, it can distort the evaluation of forecasting models.

3.1 Variable Classification

Table 1 presents the classification. “Strictly real-time” designates, throughout this paper, a series that is never revised, so that the value observed at the time of the forecast is the value in the historical record.

Table 1: Real-time data classification. “Real-time” designates a series that is never revised; “near real-time,” revisions below 0.1 percentage points; “lightly revised,” 0.1–0.3; “heavily revised,” above 0.5. The typical-revision column reports year-over-year revision magnitudes as classified by the author from the sources named; the ranges are not recomputed.

Variable	Status	Typical revision	Source
Federal funds rate	Real-time	None	FRED FEDFUNDS
GSCPI	Real-time	None	New York Fed
Prices paid (Phila. Fed survey)	Real-time	None	Philadelphia Fed
WTI crude oil	Real-time	None	EIA market data
Trade-weighted USD	Real-time	None	Federal Reserve index
FAO food index	Real-time	None	FAO commodity data
Unemployment rate	Near real-time	$\pm 0.0\text{--}0.1$ pp	BLS CPS
Import price index	Lightly revised	$\pm 0.1\text{--}0.2$ pp	BLS
PPI finished goods	Lightly revised	$\pm 0.1\text{--}0.3$ pp	BLS
Core PCE Y/Y	Revised	$\pm 0.1\text{--}0.3$ pp	BEA
Unit labor costs	Heavily revised	$\pm 1.0\text{--}2.0$ pp	BLS
Output gap	Heavily revised	$\pm 2\text{--}5$ pp	CBO

3.2 Three Data Versions

We estimate models on three data versions to isolate the effect of revisions. *Revised data*: the most recent vintage of every variable, read from its publisher’s current file (Appendix A). *First-vintage data*: for the revised variables (Core PCE, ULC, PPI), the first release of each quarter as a forecaster saw it, read from the Philadelphia Fed Real-Time Data Set for Macroeconomists (RTDSM): the earliest vintage in which the quarter is published, with the year-over-year change computed within that vintage.⁹ The revisions from first release to current vintage have mean and standard deviation (mean $+0.109$, s.d. 0.168) for Core PCE, (mean -0.237 , s.d. 1.440) for ULC, and (mean -0.006 , s.d. 0.123) for PPI; Section 6.3 uses these moments in a measurement-error simulation beside the first releases themselves. *Strictly real-time data*: regressors restricted to never-revised series (GSCPI, the Philadelphia Fed prices-paid index, WTI oil, the trade-weighted dollar, the FAO food index).

⁸Orphanides, “Monetary Policy Rules,” 964–985.

⁹Dean Croushore and Tom Stark, “A Real-Time Data Set for Macroeconomists,” *Journal of Econometrics* 105, no. 1 (2001): 111–130; Dean Croushore, “Frontiers of Real-Time Data Analysis,” *Journal of Economic Literature* 49, no. 1 (2011): 72–100.

3.3 Why Real-Time Data Is the Baseline

A forecasting model can use only data that exists when the forecast is made; this is a practical necessity rather than a methodological preference. We therefore treat the strictly real-time specification as the baseline and present revised-data results as a comparison that measures whether revisions contain additional forecasting information. Sections 6–7 show what the comparison gives: the heavily revised unit-labor-cost series adds no resolved forecasting content on either vintage, while the lightly revised producer price index, as first released, forecasts better than the never-revised set at one quarter. The discipline is satisfied by data as first released, not only by data that are never revised.

4. Supply-Side Leading Indicators

This section names the eight indicators and the pipeline logic that makes them lead consumer prices. The common feature is that each measures upstream cost pressure that passes through the pricing chain with a lag of one to four quarters.

4.1 The Producer-Cost Pipeline

Prices propagate through the supply chain in a sequence: raw commodities, intermediate goods, finished producer goods, consumer prices. At each stage firms adjust pricing and pass costs forward with a delay. Crude oil, for example, reaches gasoline prices within weeks but takes two to three quarters to pass through transportation costs, manufacturing inputs, and core consumer goods.

4.2 The Eight Indicators

The dataset comprises: (1) the New York Fed GSCPI; (2) the prices-paid diffusion index of the Federal Reserve Bank of Philadelphia’s Manufacturing Business Outlook Survey, the share of firms reporting higher input prices less the share reporting lower, not seasonally adjusted; (3) WTI crude oil, year over year; (4) the BLS import price index, year over year; (5) the PPI for finished goods, year over year; (6) unit labor costs, year over year; (7) the Federal Reserve trade-weighted dollar index, year over year; and (8) the FAO food price index, year over year. Two of the eight (the GSCPI and the prices-paid index) are never revised and survey- or market-based; three (WTI, the dollar index, the FAO index) are market or administrative data that are likewise never revised; the remaining three (import prices, PPI, ULC) are subject to revision and enter only the revised-data and first-vintage specifications.

4.3 The GSCPI as a Predictor

The GSCPI integrates 27 variables from global transportation costs and manufacturing surveys across seven economies.¹⁰ In the data the GSCPI reaches 4.3 standard deviations above its mean in 2021:Q4, one quarter before Core PCE reaches its sample peak of 5.5% in 2022:Q1. That lead during the sample’s largest inflationary episode is the source of the forecasting value documented below.

¹⁰Gianluca Benigno, Julian di Giovanni, Jan J. J. Groen, and Adam I. Noble, “The GSCPI: A New Barometer of Global Supply Chain Pressures,” Federal Reserve Bank of New York Staff Reports No. 1017 (New York: Federal Reserve Bank of New York, 2022).

5. Forecasting Models and Estimation

This section specifies the candidate models, the estimation frame, and the selection procedure.

5.1 General Framework

All one-quarter-ahead models share the structure

$$\pi_{t+1} = \alpha + \beta_1 \pi_t + \beta_2' \mathbf{X}_t + \varepsilon_{t+1}, \quad (2)$$

where π_t is year-over-year Core PCE inflation and $\mathbf{X}_t$ is a vector of supply-side indicators observed at time t . All regressors are dated t or earlier, so the one-quarter-ahead equations carry no look-ahead by construction; the multi-horizon treatment of future indicator values is a separate protocol choice addressed explicitly in Section 8. Estimation is by ordinary least squares (OLS) with $t_{\text{start}} = 5$ (the first predicted quarter is 2013:Q3), giving $N = 50$ effective observations on the full sample.

5.2 Model Specifications

Table 2: Forecasting model specifications. k counts estimated parameters including the intercept. All models forecast π_{t+1} from regressors dated t ; estimation is OLS with $t_{\text{start}} = 5$, $N = 50$ on the full sample.

Model	Regressors	k	Rationale
A: AR(1)	π_t	2	Pure persistence benchmark
B: Phillips	π_t , output gap	3	Demand-side forecast
C: Energy	π_t , WTI Y/Y, import prices	4	Energy and import pipeline
D: Supply chain	π_t , GSCPI, PPI Y/Y	4	Bottleneck and producer pipeline
E: Full supply	π_t + all eight indicators	10	All indicators jointly
F: Revised-data comparison	π_t , GSCPI, ULC Y/Y	4	The heavily revised regressor (Section 6.3)
R: Real-time	π_t , GSCPI, prices paid	4	Reference; never-revised regressors
G: GSCPI only	π_t , GSCPI	3	The one-indicator real-time model (Section 7.3)
L: Policy stance	Model F + $(i_t - i_{t-4})$	5	Cumulative-tightening augmentation (Sections 11, 13)

5.3 Model Selection

An AIC forward stepwise procedure on the revised-data indicator set, starting from the AR(1) baseline (AIC = -98.7), selects the GSCPI first (AIC to -123.1) and the prices-paid index second (AIC to -125.8); no third variable improves the AIC by the pre-set threshold of 0.5. The selection on revised data therefore lands on the two never-revised indicators, that is, on Model R below. Model F, the revised-data comparison that carries the heavily revised regressor, is

$$\pi_{t+1} = 0.295 + 0.884 \pi_t + 0.217 \text{GSCPI}_t - 0.030 \text{ULC}_t. \quad (3)$$

The ULC coefficient is not resolved ($t = -0.83$), and Section 6.3 shows that it is not resolved as first released either. The preferred strictly real-time specification is

$$\pi_{t+1} = 0.188 + 0.841 \pi_t + 0.152 \text{GSCPI}_t + 0.0057 \text{PricesPaid}_t. \quad (4)$$

6. In-Sample Results: Three Data Versions Compared

This section compares in-sample fit across models and data versions; Table 3 carries the fit statistics and Table 4 the coefficient comparison.

Table 3: In-sample fit: revised, first-vintage, and strictly real-time specifications. $N = 50$ (2013:Q3–2025:Q4 predicted quarters). “1st vintage (first release)” rows use the first releases of Section 3.2 (Core PCE and ULC as first published, from the RTDSM); AIC comparable within the table. Source: author’s calculations.

Model	Data version	R^2	R^2_{adj}	RMSE	AIC
A: AR(1)	Revised	0.924	0.923	0.358	−98.7
B: Phillips	Revised	0.925	0.922	0.357	−97.1
C: Energy	Revised	0.941	0.937	0.316	−107.2
D: GSCPI+PPI	Revised	0.956	0.953	0.274	−121.5
E: Full supply	Revised	0.963	0.955	0.251	−118.4
F: GSCPI+ULC	Revised	0.956	0.953	0.273	−121.9
A: AR(1)	1st vintage (first release)	0.900	0.898	0.378	−93.3
F: GSCPI+ULC	1st vintage (first release)	0.931	0.926	0.315	−107.5
R: GSCPI+prices paid	Real-time	0.959	0.957	0.262	−125.8
G: GSCPI only	Real-time	0.955	0.953	0.275	−123.1

6.1 The Output Gap Does Not Improve Fit

Comparing Models A and B states the paper’s motivation in two rows. The output gap raises R^2 from 0.924 to 0.925—a gain of 0.001—while the AIC worsens from −98.7 to −97.1 because the additional parameter is not justified by the fit improvement. The gap coefficient itself is −0.022 ($t = -0.60$).

6.2 Supply-Side Indicators Improve Fit

The supply-chain model (D: GSCPI + PPI) attains $R^2 = 0.956$ and $\text{RMSE} = 0.274$, an error reduction of 23.4% against the AR(1) benchmark. The all-indicator Model E attains the highest raw R^2 (0.963) and a worse AIC (−118.4 against −125.8 for Model R): six of its eight indicators add parameters without commensurate fit, a first instance of the parsimony pattern that recurs out of sample in Sections 10 and 12.

6.3 The Real-Time Discipline

Table 4: Coefficient comparison across data versions. t -statistics in parentheses; *** $p < 0.01$, ** $p < 0.05$. The first-vintage column uses the first releases from the RTDSM. Monte Carlo distribution of the ULC t -statistic across 1,000 seeded draws of revision noise with the observed moments (seed 20260719): median -0.28 ; 10th–90th percentiles $[-1.37, 0.89]$; 3.1% of draws resolved at the 5% level (critical value 2.01). Source: author’s calculations.

Variable	Model R (real-time)	Model F (revised)	Model F (1st vintage)
Intercept	0.188 (2.03**)	0.295 (3.20***)	0.297 (2.81***)
π_t (persistence)	0.841 (23.67***)	0.884 (22.79***)	0.859 (17.67***)
GSCPI	0.152 (3.64***)	0.217 (5.52***)	0.186 (4.25***)
Prices paid (Phila. Fed)	0.0057 (2.13**)	—	—
Unit labor costs	—	-0.030 (-0.83)	-0.0025 (-0.08)

The GSCPI coefficient is resolved, that is, statistically distinguishable from zero at the 5% level, in every data version ($t = 3.64$ to 5.52). The ULC coefficient is not resolved with revised data ($t = -0.83$) and not resolved as first released ($t = -0.08$); the revisions to the series are large (standard deviation 1.44 percentage points), and neither vintage carries forecasting content for next-quarter inflation. A measurement-error simulation makes the same point from the other side: across 1,000 seeded draws of revision noise with the observed moments added to the revised series (seed 20260719), the median ULC t -statistic is -0.28 with a 10th–90th percentile range of $[-1.37, 0.89]$, and 3.1% of draws are resolved at the 5% level (critical value 2.01). Model R attains $R^2 = 0.959$ from never-revised regressors alone, and since it uses only data available at the time of the decision, it is the preferred specification.

7. Out-of-Sample Forecast Evaluation

This section evaluates one-quarter-ahead forecasting over the pandemic-era window.

7.1 Methodology

We conduct rolling pseudo-out-of-sample forecasts over the 24-quarter window 2020:Q1–2025:Q4. At each origin the model is re-estimated on an expanding window using data through the origin quarter, and the coefficients forecast the next quarter. Forecast origins run from 2019:Q4 through 2025:Q3.

7.2 Results

Table 5: Out-of-sample forecast results, one quarter ahead: 24 targets, 2020:Q1–2025:Q4, expanding estimation window with origins 2019:Q4–2025:Q3. Bias is the mean forecast error (forecast minus actual); positive improvement indicates a lower RMSE than the AR(1). The final two rows repeat Models D and F with the revised regressor replaced by its first release from the RTDSM. Source: author’s calculations.

Model	Data version	OOS RMSE	OOS MAE	Bias	Improv. vs. AR(1)
A: AR(1)	Revised	0.556	0.388	+0.095	—
B: Phillips	Revised	0.557	0.382	+0.115	−0.2%
D: GSCPI+PPI	Revised	0.463	0.279	−0.103	+16.8%
F: GSCPI+ULC	Revised	0.479	0.291	−0.092	+13.9%
R: GSCPI+prices paid	Real-time	0.506	0.313	−0.167	+9.1%
G: GSCPI only	Real-time	0.459	0.274	−0.077	+17.5%
D: GSCPI+PPI	PPI as first released	0.462	0.279	−0.101	+16.9%
F: GSCPI+ULC	ULC as first released	0.474	0.279	−0.066	+14.7%

7.3 Key Findings

Model R, from never-revised regressors alone, attains an OOS RMSE of 0.506, an improvement of 9.1% over the AR(1) benchmark. The revised-data models forecast better still (D at 0.463, F at 0.479), and the question the real-time discipline asks is whether a forecaster could have had them. Table 5 answers it with the revised regressor replaced by its first release from the RTDSM: Model D with the PPI as first released records 0.462, and Model F with ULC as first released 0.474. The PPI’s revisions are small (Section 3.2), so its first release carries the information its current vintage does, and the producer price index is real-time feasible. Model D’s advantage over Model R at one quarter is not the producer price index’s: next to the GSCPI its coefficient is not resolved ($t = 0.55$ revised, 0.50 as first released). The table carries one more row that the parsimony theme of this paper predicts: Model G, the AR(1) with the GSCPI alone, records 0.459, an improvement of 17.5%, below Model R’s 0.506, a difference that a Diebold–Mariano test does not resolve ($p = 0.466$ over the 24 targets)¹¹. The prices-paid index, which the AIC selects in sample (Section 5.3) and which Granger-causes inflation (Section 13), adds a parameter that the one-quarter forecast does not repay; the first-release producer-price index adds nothing beyond the GSCPI either (0.462 against 0.459). Model R remains the reference specification, the model the in-sample selection chose before any out-of-sample comparison, and the GSCPI-only result is carried through the multi-horizon evaluation and the 2025 validation below: Model R’s margin widens with the horizon while Model G’s does not (Section 8), and Model R is the more accurate of the two in 2025 (Section 11), differences the sample does not resolve at any horizon. The output gap worsens forecasts out of sample—Model B’s RMSE of 0.557 against 0.556 for the AR(1) is a 0.2% deterioration. During the 2021 surge the output gap was closing rapidly, which the Phillips specification read as demand pressure; the gap was in fact registering recovery from a supply-side shock.

¹¹Diebold and Mariano, “Comparing Predictive Accuracy,” 253–263; Harvey, Leybourne, and Newbold, “Testing the Equality of Prediction Mean Squared Errors,” 281–291.

8. Multi-Horizon Iterated Forecasts

This section, the principal empirical section, evaluates horizons of one through four quarters on a common target window and states the protocol governing future indicator values explicitly, because that choice determines what the numbers mean.

8.1 Design

At each forecast origin, Model R is estimated on all data through the origin; iterated forecasts are generated one through four quarters ahead, the model's own lagged forecast serving as the autoregressive input beyond one quarter. Two treatments of the indicator path are reported. Under the *real-time feasible* treatment, the GSCPI and prices-paid values are held at their origin-date values over the forecast horizon—an implementable rule, since both series are observed at the origin and the treatment assumes no knowledge of their future path. Under the *conditional* treatment, the realized future values of the GSCPI and the prices-paid index enter at each step; this is a conditional-forecast diagnostic that measures how much of the horizon profile the indicators' realized path explains, and it is not attainable in real time. The evaluation uses the common target window 2020:Q4–2025:Q4, giving $n = 21$ targets at every horizon so that rows are comparable; the corresponding one-quarter improvement (13.3%) differs from the 24-target figure of Section 7 (9.1%) only through the window composition. The AR(1) benchmark iterates on its own forecasts in both treatments.

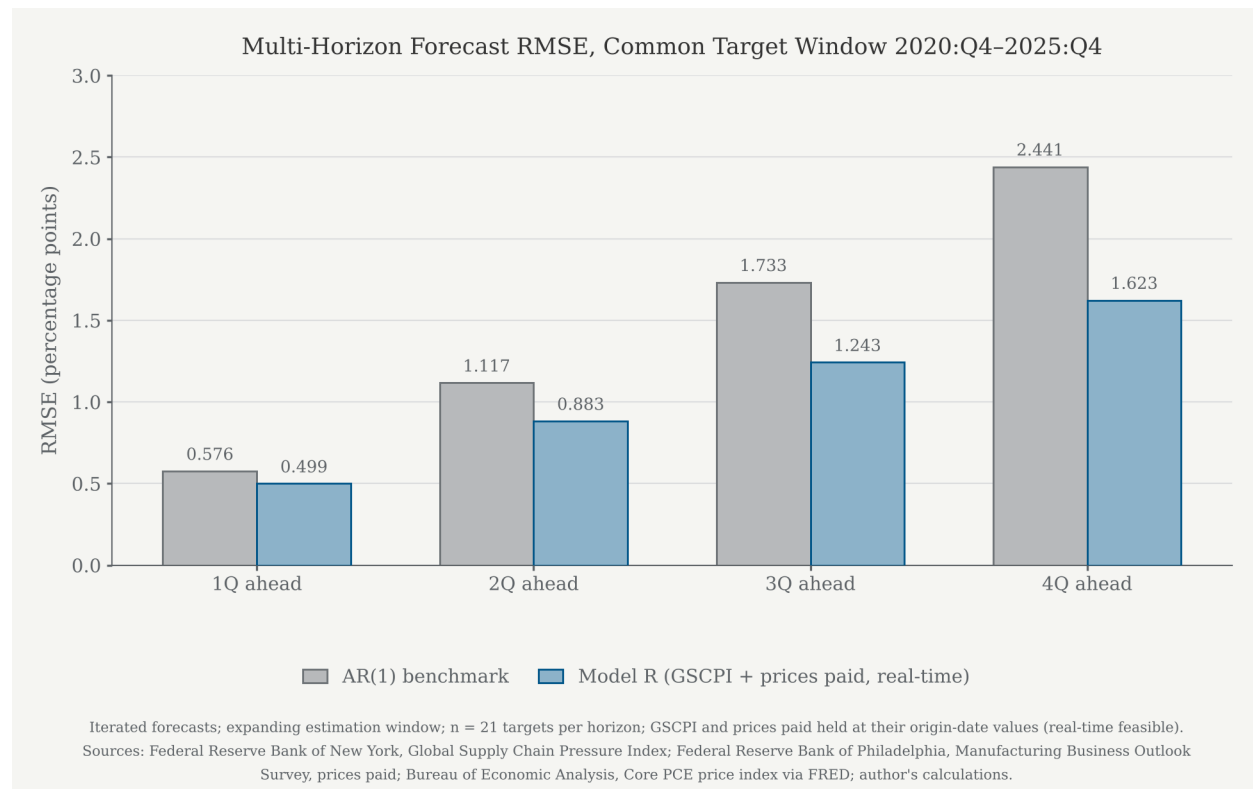


Figure 1: Multi-horizon forecast RMSE on the common target window 2020:Q4–2025:Q4. The horizontal axis groups forecast horizons of one through four quarters; the vertical axis is RMSE in percentage points of year-over-year Core PCE inflation, from a zero baseline. Gray bars (fill #B7B9BB, edge #6D7276): AR(1) benchmark. Blue bars (fill #8CB2C9, edge #005588): Model R (GSCPI + prices paid, strictly real-time) under the real-time feasible treatment, with the GSCPI and the prices-paid index held at origin-date values. Value labels above each bar report RMSE at three decimals. Each horizon uses the same $n = 21$ targets; estimation windows expand from each origin. Sources: Federal Reserve Bank of New York, Global Supply Chain Pressure Index; Federal Reserve Bank of Philadelphia, Manufacturing Business Outlook Survey, prices paid; Bureau of Economic Analysis, Core PCE price index via FRED; author's calculations.

Table 6: Rolling forecast RMSE by horizon on the common target window 2020:Q4–2025:Q4. Feasible: GSCPI and prices paid held at origin-date values. Conditional: realized future indicator values (diagnostic; not attainable in real time). Improvement = $1 - \text{RMSE}(\text{model})/\text{RMSE}(\text{AR1})$; the final two columns give Model G (GSCPI alone) under the feasible treatment. Bias (mean error), feasible treatment: Model R 1Q -0.197 , 2Q -0.392 , 3Q -0.521 , 4Q -0.634 ; AR(1) 1Q $+0.093$, 2Q $+0.198$, 3Q $+0.358$, 4Q $+0.551$. Source: author's calculations.

Horizon	n	AR(1) RMSE	R (feas.)	Improv.	R (cond.)	Improv.	G (feas.)	Improv.
1Q ahead	21	0.576	0.499	+13.3%	0.499	+13.3%	0.440	+23.7%
2Q ahead	21	1.117	0.883	+21.0%	0.848	+24.1%	0.825	+26.1%
3Q ahead	21	1.733	1.243	+28.3%	1.127	+35.0%	1.254	+27.6%
4Q ahead	21	2.441	1.623	+33.5%	1.390	+43.0%	1.782	+27.0%

8.2 Findings

Under the real-time feasible treatment the improvement rises from 13.3% at one quarter to 33.5% at four quarters (Figure 1). At short horizons persistence dominates: the AR(1) coefficient of 0.96 ties next quarter's forecast to this quarter's inflation. As the horizon lengthens, the AR(1) compounds its errors and drifts toward the sample mean regardless of where supply-side forces are pushing, while the origin-date GSCPI and prices-paid values carry directional information that remains useful several quarters out. Conditioning on the realized indicator path raises the four-quarter improvement to 43.0%; the 9.6-percentage-point gap between the two treatments measures the value of knowing the indicators' future path, and the feasible figure is the one attainable in real time. Model G, the GSCPI alone, improves on the AR(1) by 23.7% at one quarter and 27.0% at four under the same feasible treatment (Table 6): a profile that is flat across horizons, so the widening of Model R's margin is the prices-paid index's contribution in the point estimates. The index costs accuracy at one quarter and earns it at three and four. A Diebold–Mariano test of equal squared-error loss between Models G and R on the same targets resolves none of these differences ($p = 0.426, 0.719, 0.892, \text{ and } 0.151$ at one through four quarters)¹²: the sample does not separate the two models at any horizon.

The two models err in opposite directions. The AR(1)'s bias is positive and grows with horizon (+0.093 at 1Q to +0.551 at 4Q), reflecting over-prediction during the 2022–2024 disinflation. Model R's bias is negative (−0.197 to −0.634): the GSCPI normalizes faster than consumer prices, and the model forecasts the disinflation ahead of its arrival.

9. Four-Quarter-Ahead Direct Forecasts

This section corroborates the horizon profile with direct estimation, in which the dependent variable is π_{t+4} and no iteration occurs. The direct approach avoids compounding one-step errors at the cost of the intermediate forecasts.

Table 7: Four-quarter-ahead direct forecast results. Dependent variable π_{t+4} ; regressors dated t ; $N = 47$; in-sample statistics. Improvement is relative to the direct AR(1). Source: author's calculations.

Model	R^2	RMSE	RMSE improv. vs. AR(1)
A: AR(1)	0.416	1.008	—
B: Phillips	0.451	0.977	+3.1%
D: GSCPI+PPI	0.784	0.612	+39.2%
R: GSCPI+prices paid	0.740	0.672	+33.3%

At the four-quarter horizon the direct Model R reduces in-sample RMSE by 33.3% relative to the direct AR(1), in line with the 33.5% out-of-sample figure from the feasible iterated protocol. The Phillips specification improves to $R^2 = 0.451$ at this horizon—the demand channel has more time to operate—but remains behind the supply-side models. The agreement between direct and iterated results indicates that the horizon profile is a property of the indicators, not of the iteration mechanics.

¹²Diebold and Mariano, “Comparing Predictive Accuracy,” 253–263; Harvey, Leybourne, and Newbold, “Testing the Equality of Prediction Mean Squared Errors,” 281–291.

10. The Pandemic and Its Aftermath: A Sub-Period Decomposition

This section decomposes the 24-target one-quarter-ahead record of Section 7 into five macroeconomic episodes and then asks whether richer indicator parameterizations would have helped.

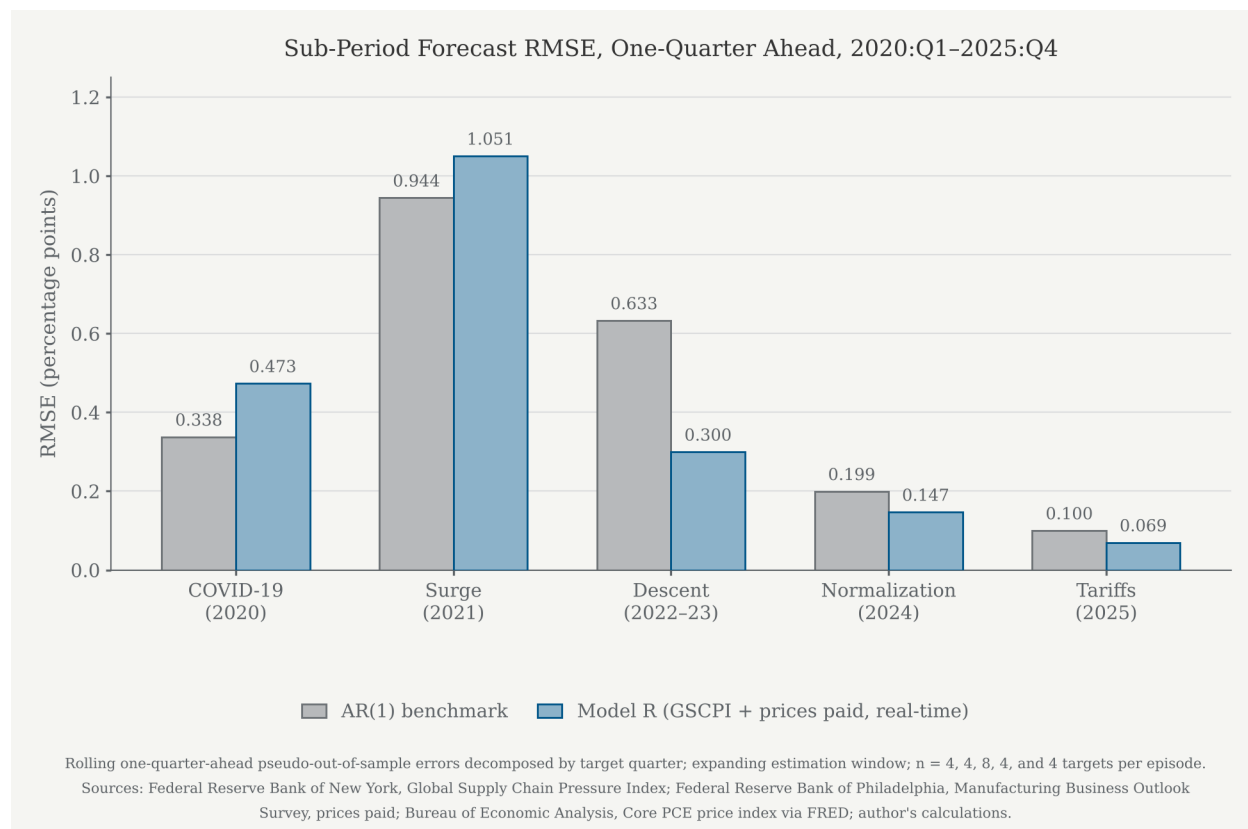


Figure 2: Sub-period forecast RMSE, one-quarter ahead, over the evaluation window 2020:Q1–2025:Q4. The horizontal axis groups the five episodes (COVID-19, 2020; surge, 2021; descent, 2022–23; normalization, 2024; tariff, 2025); the vertical axis is RMSE in percentage points of year-over-year Core PCE inflation, from a zero baseline. Gray bars (fill #B7B9BB, edge #6D7276): AR(1) benchmark. Blue bars (fill #8CB2C9, edge #005588): Model R (GSCPI + prices paid, strictly real-time). Value labels above each bar report RMSE at three decimals; episode sample sizes are $n = 4, 4, 8, 4, 4$ targets. Errors are those of the rolling expanding-window evaluation of Section 7. Sources: Federal Reserve Bank of New York, Global Supply Chain Pressure Index; Federal Reserve Bank of Philadelphia, Manufacturing Business Outlook Survey, prices paid; Bureau of Economic Analysis, Core PCE price index via FRED; author's calculations.

Table 8: Sub-period forecast RMSE and bias, one quarter ahead, decomposing the 24-target window of Table 5 by target quarter. Positive values in the “R vs. AR(1)” column indicate Model R attains the lower RMSE; bias is the mean forecast error. Source: author’s calculations.

Sub-period	Episode	n	R RMSE	AR(1) RMSE	R vs. AR(1)	R bias	AR(1) bias
2020:Q1– 2020:Q4	COVID-19 pandemic	4	0.473	0.338	−40.2%	+0.029	+0.067
2021:Q1– 2021:Q4	Inflation surge	4	1.051	0.944	−11.3%	−0.941	−0.765
2022:Q1– 2023:Q4	Peak and descent	8	0.300	0.633	+52.7%	+0.008	+0.543
2024:Q1– 2024:Q4	Normalization	4	0.147	0.199	+26.3%	−0.113	+0.137
2025:Q1– 2025:Q4	Tariff year	4	0.069	0.100	+30.4%	+0.007	+0.043

COVID-19 (2020). The AR(1) is more accurate (RMSE 0.338 against 0.473). The pandemic produced a simultaneous demand collapse and supply disruption: the GSCPI rose steeply in 2020:Q2, which Model R read as a signal of higher inflation, while Core PCE fell from 1.6% to 1.0% as the demand shock outweighed the supply disruption. The AR(1), carrying no supply-chain information, projected the prior quarter’s level and was closer to the outcome.

The inflation surge (2021). Both models under-predicted, and the AR(1) by less (RMSE 0.944 against 1.051 for Model R; biases −0.765 and −0.941). The GSCPI rose from 1.9 to 4.3 standard deviations through 2021 and pointed in the right direction, but inflation rose faster than a coefficient estimated on the pre-pandemic sample could carry it; the AR(1)’s largest miss came in 2021:Q2, a forecast of 1.72% against a realized 3.4%, and Model R’s miss in that quarter was larger still.

Peak and descent (2022–2023). Model R improves on the AR(1) by 52.7% (RMSE 0.300 against 0.633). As inflation peaked at 5.5% in 2022:Q1 and declined to 3.4% by 2023:Q4, the AR(1)’s persistence coefficient produced over-prediction quarter after quarter (bias +0.543), while the falling GSCPI signaled abating pressure (Model R bias +0.008).

Normalization (2024). Model R’s advantage is 26.3% (RMSE 0.147 against 0.199). With Core PCE in a 2.9–3.2% range and the GSCPI near zero, the model forecast stability while the AR(1) retained residual upward bias (+0.137).

The tariff year (2025). Model R’s advantage is 30.4% (RMSE 0.069 against 0.100). Core PCE held between 2.7% and 2.8% through the year of the tariff increases; the GSCPI stayed near zero and the prices-paid index rose and fell within the year, and the model’s forecasts tracked the level (bias +0.007) while the AR(1) placed it slightly above (bias +0.043).

The decomposition supports one sentence: Model R’s indicators add value once supply-chain pressure is the operative driver and inflation is descending or holding (2022–2025) and subtract value when inflation is driven by a demand collapse (2020) or rises faster than the estimated load-carrying (2021).

10.1 Asymmetric GSCPI Specifications

One response to the negative bias documented in Section 8 is to let the GSCPI enter asymmetrically. Table 9 evaluates four such specifications in the same rolling protocol.

Table 9: Asymmetric GSCPI specifications: rolling one-quarter-ahead OOS RMSE over the 24-target window, with in-sample adjusted R^2 and parameter count. Each specification replaces or augments the GSCPI term in Model R as described; the prices-paid term is retained. Source: author's calculations.

Code	Specification	OOS RMSE	vs. Model R	R^2_{adj}	k
R	GSCPI + prices paid (symmetric base-line)	0.506	—	0.957	4
G1	GSCPI trailing 4Q max + prices paid	0.506	+0.001	0.954	4
G3	GSCPI rising / falling split + prices paid	0.479	−0.027	0.961	5
G4	GSCPI positive part + prices paid	0.518	+0.012	0.958	4
G6	GSCPI positive trailing 4Q max + prices paid	0.505	−0.000	0.956	4

The four asymmetric specifications move the out-of-sample RMSE by between -0.027 and $+0.012$ relative to the symmetric Model R: two lower it by a few hundredths and two raise it. The rising/falling split (G3), with one parameter more, records the lowest RMSE of the group and the best in-sample adjusted R^2 ; the differences are of the order of one quarter's error, and with one major supply-chain cycle in fourteen years the data do not separate a symmetric loading from an asymmetric one. The symmetric loading is retained as the specification of record.

11. 2025 Genuine Out-of-Sample Validation

The rolling evaluations above re-estimate through the sample; a stronger test freezes the models earlier. This section estimates each model on data through 2024:Q4 only and forecasts 2025:Q1–2025:Q4 with no 2025 information in the coefficients, a year of tariff increases absent from the estimation sample.

11.1 The 2025 Macroeconomic Context

In the data, the federal funds rate declines from 4.33% to 3.90% (quarterly averages; Appendix A), while Core PCE inflation moves from 2.9% (2024:Q4) to 2.8% (2025:Q4), holding between 2.7% and 2.8% within the year. Import price inflation moves from 1.4% to 0.0% year over year. The GSCPI stays near zero and ends the year at 0.0 standard deviations; the prices-paid index rises from 45.3 in the first quarter to 58.0 at its peak and ends the year at 44.1.

11.2 Validation Results

Table 10: 2025 genuine out-of-sample validation. Models estimated on data through 2024:Q4 only (training pairs 2013:Q3–2024:Q4); one-quarter-ahead forecasts for 2025 use realized lagged inflation and origin-date regressors. Entries are forecast errors (forecast minus actual) in percentage points. Model L uses the documented FFR series of Appendix A. Source: author’s calculations.

Model	2025:Q1	2025:Q2	2025:Q3	2025:Q4	2025 RMSE
A: AR(1)	+0.171	−0.014	−0.058	+0.083	0.100
D: GSCPI+PPI	−0.042	−0.178	−0.204	−0.055	0.140
F: GSCPI+ULC	−0.056	−0.208	−0.186	−0.045	0.144
R: GSCPI+prices paid	−0.054	−0.047	+0.003	+0.114	0.068
G: GSCPI only	−0.035	−0.174	−0.188	−0.050	0.132
L: F + cumul. tightening	−0.015	−0.166	−0.145	−0.004	0.111

11.3 Interpreting the Results

Model R records the lowest 2025 RMSE (0.068 against 0.100 for the AR(1), 0.132 for the GSCPI-only Model G, 0.140 for Model D, and 0.144 for Model F). With inflation holding near its level, every model’s errors are small, and the biases sort the models: the AR(1) placed the year above the outcome (bias +0.046), the revised-data models D and F placed it below (−0.120 and −0.124), and Model R, whose prices-paid regressor rose through the year, came out unbiased (+0.004). Model L (Model F plus cumulative tightening, $i_t - i_{t-4}$) records 0.111 with bias −0.083: the Fed was cutting through 2025, and the policy stance pulled Model F’s forecasts part of the way back toward the outcome without changing the ordering. The year is a quiet one for the comparison; the Granger tests of Section 13 are the sharper test of the policy variable.

12. Regime-Switching Models

The model of this section was specified to separate a logistics channel, read through the GSCPI, from a trade and cost-push channel, read through import prices; in the one-regime frame of Section 5, import prices next to the GSCPI carry no resolved content ($t = 0.81$), and this section asks whether the sample identifies the two-regime model. The answer the sample gives is that the model’s parameters are not identified on its 50 predicted quarters: a search over starting values, registered before it ran, finds the likelihood maximized at a boundary, the leading interior optima disagreeing in their reading of the regimes, and the out-of-sample record depending on how the rolling re-estimation restarts. The section reports that evidence in place of parameter estimates.

12.1 Specification and Estimation

The Markov-switching model posits two latent regimes, each with its own regression:¹³

$$\text{Regime 1: } \pi_{t+1} = c_1 + a_1\pi_t + b_1 \text{GSCPI}_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_1^2), \quad (5)$$

$$\text{Regime 2: } \pi_{t+1} = c_2 + a_2\pi_t + b_2 \text{ImportPx}_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_2^2), \quad (6)$$

¹³James D. Hamilton, “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle,” *Econometrica* 57, no. 2 (1989): 357–384.

with first-order Markov transition probabilities p_{11} and p_{22} . Three estimation facts govern the reading of the results. First, the unconstrained likelihood of this model is unbounded: as either regime standard deviation tends to zero on a small set of exactly fitted points, the likelihood diverges, and an unconstrained optimizer finds that degenerate solution. Estimation therefore imposes a residual-standard-deviation floor of 0.05 on each regime. Second, the import price index is a revised series (Table 1), so the model departs from the strictly real-time discipline; the departure is accepted here for comparability with the specification's design and stated as such. Third, with the floor the likelihood is bounded but has many local maxima, and a single run of maximum-likelihood estimation (MLE) reports one of them. An earlier version of this paper reported the optimum reached from one set of starting values (log-likelihood 7.591) as the estimate; whether that optimum is the maximum, and whether the maximum is interior to the floor, is the question the search of the next subsection was registered to settle.

12.2 A Registered Multistart Search

The search was registered before it ran: the box from which starting values are drawn (intercepts in $[-2, 2]$, persistence coefficients in $[-0.5, 1.5]$, indicator coefficients in $[-1, 1]$, residual standard deviations in $[0.06, 2.05]$, transition probabilities with logits in $[-6, 6]$), 600 uniform random starts from catalogued streams together with two named starts (the previously reported optimum and the estimation stage's own starting vector), the optimizer and its settings, the rule by which two optima count as one basin, the rule by which an optimum counts as interior (both standard deviations at least 0.01 above the floor, both stay probabilities within $[0.01, 0.99]$), and the decision rule. Table 11 reports the outcome.

Table 11: Multistart search over the Markov-switching likelihood (predicted quarters 2013:Q3–2025:Q4), registered before it ran: box, catalogued random starts, basin and interior rules, and decision rule stated in advance; Nelder–Mead with a polish pass from every start under the residual-s.d. floor of 0.05. The share of starts is the share of the random starts whose optimum lies in the basin. The locations are reported as evidence on the likelihood surface, not as estimates: a parameter at the floor is not an estimate, and the sample does not identify the regimes. Source: author's calculations.

Quantity	Location	Share of starts
Starts: random from catalogued streams; named	600; 2	—
Starts not converging	73	—
Distinct basins; interior among them	342; 17	—
Maximum of the floored likelihood: log-likelihood 16.067	σ_2 at the floor (0.05), $b_2 = 0.4820$, $p_{22} = 0.0000$	5.3%
Best interior optimum: log-likelihood 13.776	$\sigma_1 = 0.1406$, $\sigma_2 = 0.1730$, $p_{11} = 0.9259$, $p_{22} = 0.4115$	2.3%
Previously reported optimum: log-likelihood 7.591	$\sigma_1 = 0.4351$, $\sigma_2 = 0.1216$, $p_{11} = 0.9070$, $p_{22} = 0.9682$	2.7%
Stay probabilities bounded below at 0.5 (200 starts): maximum	log-likelihood 14.768, a standard deviation at the floor	—

Three features of the likelihood surface stand out. First, the maximum of the floored likelihood (log-likelihood 16.067) stands on the floor: the import-price regime's residual standard deviation is the floor itself, its indicator coefficient 0.4820, and its stay probability 0.0000. That regime fits

a few outlying quarters almost exactly, which is the degeneracy the floor was imposed to exclude, relocated to the floor; a parameter at a bound is not an estimate, and the maximum is reached from 5.3% of the starts. Second, the best interior optimum (log-likelihood 13.776, reached from 2.3% of the starts) reverses the regime reading of the previously reported optimum: there the GSCPI regime is the persistent one with the smaller residual standard deviation ($\sigma_1 = 0.1406$ against $\sigma_2 = 0.1730$; $p_{11} = 0.9259$, $p_{22} = 0.4115$), whereas at the previously reported optimum (log-likelihood 7.591, reached from 2.7% of the starts) the GSCPI regime is the high-volatility one ($\sigma_1 = 0.4351$ against $\sigma_2 = 0.1216$). Third, the starts scatter over 342 distinct basins, 17 of them interior, and 73 starts do not converge; the optimizer's own convergence report does not distinguish one basin from another. Bounding both stay probabilities below at 0.5, a registered amendment run from 200 further starts, does not remove the degeneracy: the bounded maximum (log-likelihood 14.768) still stands on the floor, and the best interior optimum under the bound (log-likelihood 11.046) lies below it.¹⁴

12.3 What the Evidence Supports

Which optimum an optimizer reports, which regime carries the larger residual variance, and the stay probabilities all depend on the starting values. The regime parameters are therefore not interpreted here, and the labels the specification was designed around (logistics; trade and cost-push) are neither supported nor refuted by them: the sample does not identify the regimes. What the evidence does support is the forecasting comparison of Table 12, stated as a range.

Table 12: Static versus regime-switching comparison. OOS RMSE is the expanding-window one-quarter-ahead evaluation over the 24 targets of Table 5; the Markov-switching model is re-estimated at each origin with warm starts, and its row gives the range of its RMSE over the four restart conventions of Section 12; its in-sample fit is not reported, the parameters being unidentified. The 2025 column repeats the genuine out-of-sample RMSE of Table 10; the regime model was not run in that exercise. Source: author's calculations.

Model	In-sample RMSE	OOS RMSE	OOS improv. vs. AR(1)	2025 RMSE
A: AR(1), static	0.358	0.556	—	0.100
R: GSCPI+prices paid, static (real-time)	0.262	0.506	+9.1%	0.068
D: GSCPI+PPI, static	0.274	0.463	+16.8%	0.140
Markov-switching (s.d. floor 0.05)	not identified	0.500–1.111	10.2 to –99.7%	—

Under expanding-window re-estimation over the same 24 targets as Section 7, the model's one-quarter-ahead RMSE depends on the restart convention of the rolling refits (each origin warm-starts from the preceding origin's solution and restarts from a stated vector when the refit lands on the floor): 0.826 when the refits restart from the stage's starting vector, 1.111 when they restart from the previously reported optimum itself, and 0.590 or 0.500 for the best interior optimum under the same two conventions. Against the AR(1) benchmark's 0.556, the range runs from 0.500 to 1.111: the out-of-sample number is a property of the restart convention, not of the model. Ten parameters re-estimated on 40–50 observations produce unstable regime inference at the forecast origin, and the instability is what the range measures. The result extends the parsimony pattern of Section 10.1 from asymmetric terms to regime switching, with a qualification the search adds:

¹⁴The maximized likelihood over floor values of 0.05 (16.067), 0.075 (14.919), 0.1 (14.118), 0.125 (13.527), 0.15 (13.683), 0.2 (13.687), 0.25 (13.256) (log-likelihood in parentheses) does not fall monotonically with the floor. A floor is a device for bounding the likelihood, not a tuning parameter, and no value of it was found at which the maximum is interior.

on this sample the regime-switching model is not a competitor whose estimates can be compared with the static models', but a specification the sample does not identify.

13. Does the Fed's Reaction Function Matter?

The supply-side models omit the federal funds rate; this section tests whether the omission costs forecasting power, using the documented quarterly-average FFR series of Appendix A.

13.1 Granger Causality Tests

Table 13: Granger causality tests: candidate variable $\rightarrow$ Core PCE inflation. Bivariate regressions with two lags of inflation and two of the candidate; $n = 51$ common observations; H_0 : the candidate's lags add no predictive power beyond lagged inflation. The FFR series is the documented quarterly-average array of Appendix A. Source: author's calculations.

Candidate variable	F -statistic	5% critical value	p -value	Result
Federal funds rate	2.26	3.20	0.115	Fail to reject H_0
GSCPI	9.15	3.20	<0.001	Reject H_0
Prices paid (Phila. Fed)	4.37	3.20	0.018	Reject H_0

The federal funds rate does not Granger-cause Core PCE inflation ($F = 2.26$, $p = 0.115$).¹⁵ The GSCPI ($F = 9.15$, $p = < 0.001$) and the Philadelphia Fed prices-paid index ($F = 4.37$, $p = 0.018$) both do. Whatever forecasting power the funds rate appears to carry in isolation reflects reverse causality: the Fed reacts to inflation, so the funds rate proxies lagged inflation dynamics. The Model L result of Section 11 is consistent with it: conditioning on the policy stance changed the 2025 forecasts little. Under a flat Phillips curve this is the expected configuration—policy transmits through demand, and demand does not move inflation on this sample.

14. Implications for Monetary Policy

This section separates what the results say about forecasting from what they say about control.

14.1 Better Forecasting Does Not Deliver Better Control

Forecasting inflation and controlling it are different capabilities. The flat Phillips curve limits the Fed's leverage over inflation through the demand channel regardless of forecast quality; a policy-maker who sees supply-driven inflation coming holds no additional instrument for preventing it. The transmission lag literature reaches the adjacent conclusion from the policy side.¹⁶

14.2 Where Better Forecasting Helps

Three consequences follow from the results. If the Federal Open Market Committee (FOMC) monitors never-revised supply-side indicators alongside its structural models, it distinguishes supply

¹⁵Clive W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica* 37, no. 3 (1969): 424–438. Bivariate tests, two lags of inflation and two of the candidate; the null is that the candidate's lags add no predictive power beyond lagged inflation.

¹⁶Milton Friedman, "The Lag in Effect of Monetary Policy," *Journal of Political Economy* 69, no. 5 (1961): 447–466.

shocks from demand shocks in real time rather than after revisions arrive. If communications draw on that distinction, expectations formation confronts the supply-driven character of an episode while it is under way. If the forecasting suite's weight on Phillips-curve models falls during supply-driven periods, the suite's accuracy moves toward the profile documented in Sections 7–9.

15. Conclusions

This paper makes seven contributions.

First, the Phillips curve provides no forecasting power for Core PCE inflation during the U.S. inflation-targeting era: the output-gap coefficient is 0.0098 ($t = 0.32$) in the reduced form, and the gap worsens out-of-sample accuracy when added to an autoregressive model.

Second, never-revised supply-side indicators forecast. The strictly real-time Model R improves on the AR(1) benchmark by 9.1% in one-quarter-ahead RMSE over 2020:Q1–2025:Q4, and by 13.3% to 33.5% across horizons of one to four quarters on a common target window under a protocol implementable in real time; the GSCPI alone records a one-quarter improvement of 17.5%, and the producer price index adds nothing resolved to it; the second never-revised indicator costs accuracy at one quarter and widens the margin at three and four quarters and in the 2025 validation, and the sample does not resolve the differences between the one-indicator and the two-indicator model at any horizon.

Third, the real-time discipline is met by first releases, not only by never-revised series. Read from the Philadelphia Fed's real-time data set, the heavily revised unit-labor-cost series carries no resolved forecasting content on either vintage ($t = -0.83$ revised, -0.08 as first released; 3.1% of simulated revision draws resolved), while the lightly revised producer price index as first released forecasts as well as its current vintage.

Fourth, the Federal Reserve's reaction function is unnecessary for forecasting inflation: the funds rate does not Granger-cause inflation ($F = 2.26$), the GSCPI and the prices-paid index do ($F = 9.15, 4.37$), and conditioning on the policy stance changed the 2025 validation little.

Fifth, the forecasting advantage is episode-specific in an interpretable way: largest during the 2022–2023 descent, present through 2024 and the 2025 tariff year, absent during the 2020 demand collapse and the 2021 surge, when inflation rose faster than the estimated loading carried.

Sixth, structure beyond a single symmetric supply-side loading is not identified on this sample: the asymmetric GSCPI specifications move the out-of-sample RMSE by a few hundredths of a point in either direction, and the two-regime Markov-switching model is not identified—a registered multistart search finds its likelihood maximized at a boundary, its interior optima disagreeing in their reading of the regimes, and its out-of-sample RMSE running from 0.500 to 1.111 with the restart convention of the rolling refits.

Seventh, every series is read from its publisher's file under a stated extraction rule, with the file's digest and access date recorded (Appendix A); the data of record are reproducible from their sources.

The Granger configuration summarizes the policy dilemma of the supply-shock era: the variable the Fed can observe and forecast (supply-driven inflation) is not the variable it can control, and the variable it can control (demand, through the policy rate) does not predict the variable it targets.

A. Data Sources and Provenance

Every series of the data module is read from its publisher's file under a stated extraction rule; Table A.1 names, for each series, the publisher, the file as fetched, the rule, and the date of access; the file's SHA-256 digest is in the data record that travels with the computation. The quarterly value of a monthly or daily series is the arithmetic mean of the observations dated in the quarter; a year-over-year change is computed on the quarterly values; a quarter of a monthly series is taken when at least two of its three months are published, and the one quarter with a month the publisher did not release is named in the table. The first-vintage series are first releases from the RTDSM: for each quarter, the earliest vintage in which the quarter (or its third month) is published, with the year-over-year change computed within that vintage. Values are carried at six decimals; no statistic inherits a display precision. The data as entered in the first version of this paper approximated these sources at coarse precision and are superseded; the data record carries the series-by-series comparison.

Table A.1: Provenance of the data module. Each series is read from the file named, as fetched from the publisher on the date given, under the rule stated; file digests are in the data record. Y/Y: year-over-year change in percent on the quarterly values. FRED: Federal Reserve Economic Data, St. Louis Fed. RTDSM: the Philadelphia Fed real-time data set.

Series	Publisher	File	Rule	Accessed
Core PCE Y/Y	BEA (FRED)	PCEPILFE.csv	quarterly mean of the monthly index; Y/Y	2026-10-09
Core PCE Y/Y, first release	Philadelphia Fed RTDSM	rtdsm_pconxmvqd.xlsx	earliest vintage; Y/Y within it	2026-10-09
Output gap	BEA (FRED); CBO (FRED)	GDPC1.csv, GDPPOT.csv	100 (real GDP / potential – 1)	2026-10-09
WTI Y/Y	EIA (FRED)	DCOILWTIC0.csv	quarterly mean of daily spot; Y/Y	2026-10-09
Import prices Y/Y	BLS (FRED)	IR.csv	quarterly mean; Y/Y (October 2025 not released)	2026-10-09
GSCPI	New York Fed	gscpi_data.xlsx	quarterly mean	2026-10-09
PPI finished goods Y/Y	BLS (FRED)	WPSFD49207.csv	quarterly mean; Y/Y	2026-10-09
PPI Y/Y, first release	Philadelphia Fed RTDSM	rtdsm_pppimvmd.xlsx	earliest vintage with the third month; Y/Y within it	2026-10-09
Prices paid	Philadelphia Fed survey	bos_history.csv	quarterly mean; not seasonally adjusted	2026-10-09
Unit labor costs Y/Y	BLS (FRED)	ULCNFB.csv	Y/Y of the quarterly index	2026-10-09
ULC Y/Y, first release	Philadelphia Fed RTDSM	rtdsm_ulcqvd.xlsx	earliest vintage; Y/Y within it	2026-10-09
Dollar index Y/Y	Federal Reserve Board (FRED)	DTWEXBGS.csv	quarterly mean of daily; Y/Y	2026-10-09
FAO food index Y/Y	FAO	fao_food_price_index.csv	quarterly mean; Y/Y	2026-10-09
Federal funds rate	Federal Reserve Board (FRED)	FEDFUNDS.csv	quarterly mean of monthly	2026-10-09

The federal funds rate series used in Sections 11 and 13 is the quarterly mean of the monthly effective rate; the effective rate is not subject to revision. Table A.2 prints the series in full.

Table A.2: Effective federal funds rate, quarterly averages (percent), as entered from FRED series FEDFUNDS. The effective rate is not subject to revision; values are the author's quarterly averages of the monthly series.

Year	Q1	Q2	Q3	Q4
2012	0.10	0.15	0.14	0.16
2013	0.14	0.12	0.08	0.09
2014	0.07	0.09	0.09	0.10
2015	0.11	0.12	0.14	0.16
2016	0.36	0.37	0.40	0.45
2017	0.70	0.95	1.15	1.20
2018	1.45	1.74	1.92	2.22
2019	2.40	2.40	2.19	1.64
2020	1.26	0.06	0.09	0.09
2021	0.08	0.07	0.09	0.08
2022	0.12	0.77	2.19	3.65
2023	4.52	4.99	5.26	5.33
2024	5.33	5.33	5.26	4.65
2025	4.33	4.33	4.29	3.90

Bibliography

- Ball, Laurence, and Sandeep Mazumder. "Inflation Dynamics and the Great Recession." *Brookings Papers on Economic Activity* (Spring 2011): 337–381.
- Benigno, Gianluca, Julian di Giovanni, Jan J. J. Groen, and Adam I. Noble. "The GSCPI: A New Barometer of Global Supply Chain Pressures." Federal Reserve Bank of New York Staff Reports No. 1017. New York: Federal Reserve Bank of New York, 2022.
- Blanchard, Olivier. "The Phillips Curve: Back to the '60s?" *American Economic Review: Papers & Proceedings* 106, no. 5 (2016): 31–34.
- Croushore, Dean. "Frontiers of Real-Time Data Analysis." *Journal of Economic Literature* 49, no. 1 (2011): 72–100.
- Croushore, Dean, and Tom Stark. "A Real-Time Data Set for Macroeconomists." *Journal of Econometrics* 105, no. 1 (2001): 111–130.
- Del Negro, Marco, Michele Lenza, Giorgio E. Primiceri, and Andrea Tambalotti. "What's Up with the Phillips Curve?" *Brookings Papers on Economic Activity* (Spring 2020): 301–373.
- Diebold, Francis X., and Roberto S. Mariano. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13, no. 3 (1995): 253–263.
- Forbes, Kristin J. "Has Globalization Changed the Inflation Process?" BIS Working Papers No. 791. Basel: Bank for International Settlements, 2019.
- Friedman, Milton. "The Lag in Effect of Monetary Policy." *Journal of Political Economy* 69, no. 5 (1961): 447–466.
- Gilchrist, Simon, and Egon Zakrajšek. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102, no. 4 (2012): 1692–1720.
- Granger, Clive W. J. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37, no. 3 (1969): 424–438.
- Hamilton, James D. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57, no. 2 (1989): 357–384.
- Harvey, David, Stephen Leybourne, and Paul Newbold. "Testing the Equality of Prediction Mean Squared Errors." *International Journal of Forecasting* 13, no. 2 (1997): 281–291.
- Hazell, Jonathon, Juan Herreño, Emi Nakamura, and Jón Steinsson. "The Slope of the Phillips Curve: Evidence from U.S. States." *Quarterly Journal of Economics* 137, no. 3 (2022): 1299–1344.
- McLeay, Michael, and Silvana Tenreyro. "Optimal Inflation and the Identification of the Phillips Curve." *NBER Macroeconomics Annual* 34, no. 1 (2020): 199–255.
- Orphanides, Athanasios. "Monetary Policy Rules Based on Real-Time Data." *American Economic Review* 91, no. 4 (2001): 964–985.