Australian National University
Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Forecasting Indian Core Inflation: Simple Made Simpler

CAMA Working Paper 83/2026
September 2026

Rishabh Choudhary
Indian Statistical Institute, Delhi

Chetan Dave
University of Alberta

Chetan Ghate
Indian Statistical Institute, Delhi
Centre for Applied Macroeconomic Analysis, ANU

Abstract

Managing rapid economic growth in India naturally brings into focus the role of inflation forecasts for monetary and fiscal policy making. We investigate whether any of a large set of forecasting models improves upon a univariate auto-regression in forecasting core consumer price inflation. Using a monthly panel of forty macroeconomic and financial indicators we estimate nine classes of models that range from an auto-regression to various factor models, quantile regressions and machine learning specifications. In doing so, we also account for inflation expectations and climate change variables. Our root mean square forecast error model comparison metric operates at horizons of one, three, six and twelve months. With respect to the conditional mean, no model produces a forecast error significantly below that of an auto-regression at any horizon over the full comparison sample. With respect to the conditional distribution, quantile regressions that include estimated factors reduce forecast loss relative to a specification without such factors at every quantile and horizon. Our approach is general enough to be applicable to forecasting core inflation in other emerging market economies.

Keywords

inflation forecasting in EMEs, core inflation, diffusion index models, quantile regression, machine learning, inflation targeting

JEL Classification

C22, C53, E31, E37, O23

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

Forecasting Indian Core Inflation: Simple Made Simpler.[1]

Rishabh Choudhary[2] Chetan Dave[3] Chetan Ghate[4]

September 18, 2026

Abstract

Managing rapid economic growth in India naturally brings into focus the role of inflation forecasts for monetary and fiscal policy making. We investigate whether any of a large set of forecasting models improves upon a univariate auto-regression in forecasting core consumer price inflation. Using a monthly panel of forty macroeconomic and financial indicators we estimate nine classes of models that range from an auto-regression to various factor models, quantile regressions and machine learning specifications. In doing so, we also account for inflation expectations and climate change variables. Our root mean square forecast error model comparison metric operates at horizons of one, three, six and twelve months. With respect to the conditional mean, no model produces a forecast error significantly below that of an auto-regression at any horizon over the full comparison sample. With respect to the conditional distribution, quantile regressions that include estimated factors reduce forecast loss relative to a specification without such factors at every quantile and horizon. Our approach is general enough to be applicable to forecasting core inflation in other emerging market economies.

Keywords: Inflation Forecasting in EMEs, Core Inflation, Diffusion Index Models, Quantile Regression, Machine Learning, Inflation Targeting.

JEL Code: C22, C53, E31, E37, O23

[1]We are grateful to seminar participants at the Indian Statistical Institute, Delhi, and the Institute for Economic Growth, Delhi, for comments. We thank Gaurav Kumar for excellent research assistance. An earlier version of this paper was entitled "Forecasting Core Inflation in India: A Demand Perspective." The views expressed are personal and do not necessarily reflect the views of the organizations the authors are affiliated to.

[2]Ph.D Candidate (Quantitative Economics), Indian Statistical Institute, Delhi. Email: rishabh23r@isid.ac.in

[3]Department of Economics, University of Alberta. Email: cdave@ualberta.ca

[4]Corresponding Author. Economics and Planning Unit, Indian Statistical Institute - Delhi and Centre for Applied Macroeconomic Analysis (CAMA), Australia. Email: cghate@isid.ac.in

It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.

— Albert Einstein, "On the Method of Theoretical Physics,"
Herbert Spencer Lecture, Oxford, 1933

1 Introduction

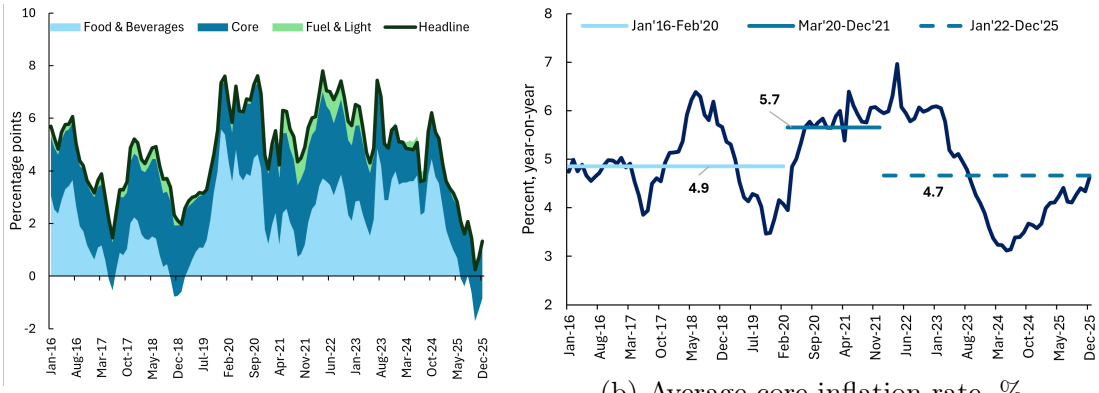
Inflation forecasting plays a pivotal role in guiding central bank policy decisions thereby promoting economic stability and shaping the decisions of businesses and households especially in fast growing emerging market economies. The ability to forecast using simple models then takes primacy so that policy can be informed quickly and efficiently. The adoption of a flexible inflation targeting (FIT) framework by the Reserve Bank of India (RBI) in 2016 constitutes an important step towards enhancing monetary policy effectiveness (Ghate and Ahmed (2023), Eichengreen, Gupta, and Choudhary (2020)). Under FIT, the RBI's policy decisions and communications rely heavily on accurate inflation projections. This paper asks a central question: which forecasting model should produce those projections? We compare a large set of models for core inflation in India, and find that a univariate auto-regression is very hard to improve upon.[5]

Why core inflation? First, core inflation, shown in Figure 1 below, constitutes a substantial portion (about 47.3%) of the (headline) consumer price index in India. Second, as Figure 1 shows, since 2016 core inflation has contributed an average of 2.3 percentage points to headline inflation, accounting for about 64% of the total on average.[6] Third, while the nominal anchor under FIT in India is headline consumer price inflation, it is widely recognized that core inflation proxies for demand side inflation in the economy (see RBI (2014)). Fourth, core inflation is markedly less volatile than headline inflation: between January 2016 to December 2025 the standard deviation of year on year core inflation was 0.9 percentage points, against 1.7 percentage points for headline inflation. From an average of 4.9 % during January 2016 to February 2020, core inflation rose to an average of 5.7 % between March 2020 and December 2021, before returning to an

[5]Throughout, the term headline inflation refers to CPI inflation; core inflation equals headline minus food, beverages, fuel and light inflation.

[6]These figures are computed using the January 2016 to December 2025 sample. The share of core inflation in the total has risen over this period because headline inflation moderated after 2022 while core inflation did not.

average of 4.7 % from January 2022 to December 2025. Fifth, there is evidence in the Indian context that headline inflation converges to core inflation (Blagrave and Lian (2020)), which suggests a significant role for core inflation in shaping headline inflation dynamics.



(a) Contribution to headline inflation

(b) Average core inflation rate, %

Figure 1: The Average Contribution of Core Inflation: Jan 2016 - Dec 2025
Source: RBI, MOSPI, Authors' Calculations.

A large literature finds that inflation in other countries is difficult to forecast with models richer than a univariate auto-regression. Atkeson and Ohanian (2001) show that for the United States a naive forecast is not improved upon by Phillips curve specifications, and Stock and Watson (2007) document that the predictable component of U.S. inflation declined over time. Duncan and Martínez-García (2019) find the same pattern across 14 emerging market economies, India among them, where a four-quarter moving average of past inflation attains lower prediction errors than autoregressions, factor models and Phillips curve specifications in most countries.

Another branch of the literature finds that richer information does improve forecasting. For instance, Stock and Watson (2002) propose diffusion index forecasts that summarize a large panel of predictors by a small number of estimated factors, and Medeiros et al. (2021) report that machine learning methods improve upon standard benchmarks for U.S. inflation in data rich environments. Beck and Wolf (2026) reweight the trees of a random forest to reduce U.S. and Swiss inflation forecast errors by three to nine percent, though against a machine learning benchmark and for point forecasts only. Our novelty lies in asking whether any model, machine learning or otherwise, improves on the simplest univariate benchmark for an emerging market with a far shorter indicator history, and we extend the comparison to the conditional distribution of inflation as well as its mean.

A large literature examines these questions for India specifically, and much of it makes use of high frequency indicators (HFI).[7] Jose et al. (2021) compare a range of models for Indian consumer price inflation and find that the ranking of the model depends on both the horizon and the measure of inflation being forecast: a seasonal autoregressive specification performs better at one quarter ahead, while Phillips curve variants for core inflation perform better at four quarters ahead. Sharma and Padhi (2020) construct a measure of economic slack from ten high frequency indicators using a dynamic factor model, and report that it forecasts core inflation better than the output gap or capacity utilization. Dua and Goel (2023) is closest to our exercise: they extract factors from 117 monthly series and find that a factor augmented vector autoregression forecasts Indian wholesale and consumer price inflation more accurately than vector error correction and univariate alternatives, a difference that a modified Diebold and Mariano test finds significant at almost all horizons.

Using a Phillips curve specification, Patra and Kapur (2012) and Behera, Wahi, and Kapur (2017) find a role for measures of slack in Indian inflation, as do Dua and Gaur (2010) for a panel of Asian economies. On aggregation, Chaudhuri and Bhadhuri (2019) find that forecasting inflation from its components outperforms forecasting the aggregate directly, and that combining component and aggregate forecasts improves accuracy further, in contrast to the Euro area evidence of Hubrich (2005). Kapur (2012) surveys the practical difficulties of forecasting inflation in India, and the Reserve Bank's own projections draw on a suite of models and a wide set of high frequency indicators (RBI (2025)). Forecasting using the high frequency indicators is appealing for a practical reason: quarterly national accounts, and therefore any measure of output gap constructed from them, are released about two months after the reference quarter. Indicators such as the purchasing managers' indices, motor vehicle sales and electricity generation are available within weeks.

There are three features of the above literature that leave open the question of whether richer models reliably improve upon a univariate autoregression for India. First, the evaluation samples end before or deliberately exclude the pandemic: Dua and Goel (2023) evaluate over July 2016 to January 2018, Sharma and Padhi (2020) forecast over thirteen recursive samples ending in 2019:Q4, and Jose et al. (2021) truncate their sample at 2019-20:Q4, noting data reporting difficulties during the pandemic lockdown. Second, the number of forecast comparisons is small, from six quarters to nineteen months. Third, differences in accuracy are

[7] Throughout, high frequency indicators are measured at the monthly frequency.

not always tested. Dua and Goel (2023) apply a modified Diebold and Mariano (DM) test, but Jose et al. (2021) apply the DM test only to compare direct with indirect aggregation,[8] while Sharma and Padhi (2020) report root mean squared errors alone. The reported improvements are often large: in Jose et al. (2021), a hybrid Phillips curve produces a four-quarters-ahead root mean squared error for core inflation of 0.74 against 1.54 for a random walk. Improvements of this size are informative, though with six overlapping forecast errors a longer evaluation sample would help establish how far they carry beyond the period examined.

Our contribution differs in three respects. First, we re-estimate every component of every model at each forecast origin, including the factors and the selection of indicators from which they are constructed. A constructed indicator improves upon a univariate benchmark at every horizon. Second, we compare a wider set of models, nine classes from a univariate autoregression through diffusion index, penalized and non-parametric estimators, under an identical protocol and against a common baseline. Third, we forecast the conditional distribution of core inflation as well as its conditional mean, which to our knowledge has not previously been done for India.

We conduct forecast comparisons for core inflation in India taking a univariate auto-regression as the baseline model. We then specify a sequence of models that adds information or flexibility to that baseline: an auto-regressive moving average model; auto-regressions augmented with one indicator at a time drawn from a panel of forty indicators; diffusion index models that summarize the panel by estimated factors; diffusion index models that use only those indicators whose past forecasts have been more accurate than the baseline; quantile regressions; specifications that add household inflation expectations; ridge, lasso, elastic net and random forest specifications; and specifications that add climate change related variables. Every model is estimated on the same data, forecasts the same variable at the same horizons, and is evaluated by the same criterion (root mean squared errors, RMSE). We ask whether *any* model produces a lower RMSE than the baseline, and whether *any* differences are statistically significant as per Diebold and Mariano (1995).

We report two main results. First, with respect to the conditional mean of core inflation, no model produces a RMSE significantly below that of our auto-regression (i.e., a relative to an autoregression RMSE ratio lower than one) over the full com-

[8]That is, between forecasts of headline inflation generated directly and those generated by modeling core, food and fuel inflation separately and combining them using their CPI weights.

parison sample. Six of the model and horizon combinations we consider produce a ratio of RMSEs below one, and in each case the difference is not significant at conventional levels.

We find that two sub-samples depart from this, the sample excluding March 2020 to December 2021, and the sample beginning January 2022. Once the pandemic months are excluded, a diffusion index model built from indicators screened on their real time forecast record reduces the error by about five percent at three and six months, and it wins again on the sample beginning January 2022; within the pandemic months the penalized regressions, which are the weakest models in every other sample, beat the auto-regression by a fifth or more at six months. Both samples were chosen after inspecting the forecast record, and the second contains twenty two forecast errors, so we treat neither as settled. Second, for the conditional distribution of core inflation, quantile regressions that include estimated factors reduce quantile forecast loss relative to a quantile auto-regression at all twenty combinations of quantile and horizon, ten of them significantly. The reduction is largest at the ninetieth percentile of inflation and at horizons of six and twelve months, and it remains when the pandemic months are excluded.

Two features of our forecasting efforts are worth making explicit. First, every model is re-estimated as the estimation sample grows, so that a forecast for a given month uses only data dated before that month. This is the scheme used by Stock and Watson (2002) in their diffusion index forecasts, by Marcellino, Stock, and Watson (2006) in comparing direct and iterated forecasts, and by Medeiros et al. (2021).[9] Section 3.1 describes the procedure in detail.

Second, every indicator enters the model with a one month lag, because most monthly indicators for India are released after the consumer price index for the same month. The consumer price index for month t is released by the Ministry of Statistics and Programme Implementation around the twelfth day of the next month $t+1$, whereas the Index of Industrial Production (IIP) for month t is released with a lag of about six weeks, that is, around the twelfth day of month $t+2$. HFIs like merchandise trade data are released in the middle of the following month, and scheduled commercial bank credit is available with a delay of about six weeks. A forecaster who constructs a projection on the day core inflation for month t is published therefore has industrial production and credit only through month $t-1$. The forecasts reported below are therefore constructed only from

[9] Giacomini and White (2006) discuss the properties of tests of predictive ability in such settings.

data a forecaster could actually have had in hand at the time, rather than from series that were published later. Section 2.3 states the convention and gives further examples.

Our analysis is structured as follows. Section 2 describes the data and the construction of the variable to be forecast. Section 3 states the estimation and forecast procedure and the evaluation criteria. Section 4 specifies the models. Section 5 reports the results. Section 6 reports subsample results and Section 7 concludes.

2 Data

2.1 Inflation Measurement

Let P_t denote the monthly consumer price index from January 2011 ($t = 1$) to December 2025 ($t = T$) for India excluding food, beverages, fuel and light. We adjust P_t for seasonality using the X-13ARIMA-SEATS procedure, and denote the adjusted index by the same symbol throughout.[10] We define month-on-month inflation as

$$\pi_t^{\text{MOM}} = 100 \left(\ln P_t - \ln P_{t-1} \right), \tag{1}$$

and year on year inflation as

$$\pi_t^{\text{YOY}} = 100 \left(\frac{P_t}{P_{t-12}} - 1 \right). \tag{2}$$

We estimate all models for π_t^{MOM}, and report all forecast errors for π_t^{YOY}. We estimate models for π_t^{MOM} because it is stationary: an augmented Dickey-Fuller test rejects the null hypothesis of a unit root.[11] We report forecast errors for π_t^{YOY} because year on year inflation is the rate in terms of which inflation outcomes in India are stated, and in terms of which the inflation target is defined. Section 3.3 states how a forecast of π_t^{MOM} is converted into a forecast of π_t^{YOY}.

[10] X-13ARIMA-SEATS is the seasonal adjustment program maintained by the United States Census Bureau. It fits a regression model with ARIMA errors to remove calendar and outlier effects before decomposing the series into seasonal, trend and irregular components.

[11] An ADF test rejects the null hypothesis of a unit root for π_t^{MOM} (Dickey-Fuller statistic 4.34, $p < 0.01$). The same test applied to π_t^{YOY} does not reject the null at conventional levels (Dickey-Fuller statistic 3.03, $p = 0.15$).

2.2 The Indicator Panel

We assemble $N = 40$ monthly indicators covering real activity, credit and money, external trade and prices, financial markets, and survey based measures, all drawn from the sources listed in Table 7. Denote the indicators by $x_t^{(i)}$ for $i = 1, \ldots, N$. Real activity is represented by five components of the index of industrial production, motor vehicle sales, electricity generation, coal production and reservoir levels; credit and money is represented by broad money and bank credit; external trade and prices is represented by imports, an import volume index, crude oil prices, the Food and Agriculture Organization food price index and eight non-ferrous and precious metal prices; financial markets are represented by the equity index, the ten year government security yield, the call money rate, the real effective exchange rate and four measures of foreign portfolio investment. The price block is represented by the headline consumer price index (CPI) and three wholesale price inflation (WPI) series; and survey based measures are represented by the manufacturing and services purchasing managers' indices, the Reserve Bank's consumer confidence index and the fiscal deficit. Table 7 in the Appendix lists each indicator, its transformation, and whether it was seasonally adjusted.

Each indicator is made stationary, as follows. Twenty seven indicators are positive level series, say denoted by x_t, and are transformed by $100(\ln x_t - \ln x_{t-1})$. The remaining thirteen are expressed as rates, yields, ratios, diffusion indices or year on year growth rates, and are transformed by first differencing.

Before being made stationary, each of the twenty seven level indicators is tested for seasonality using the combined test of Ollech and Webel (2020), and the thirteen for which the test rejects the null of no seasonality are seasonally adjusted by X-13ARIMA-SEATS. The thirteen first differenced indicators are not adjusted further. Two of them, the manufacturing and services purchasing managers' indices are already seasonally adjusted so that no further adjustment is required; the remainder are yields, policy rates, exchange rate indices, deficit ratios and year on year growth rates, in which seasonality is either absent by construction or removed by the year on year difference.

Household inflation expectations are taken from the RBI Inflation Expectations Survey of Households (IESH) (Reserve Bank of India 2026b). The survey is conducted in bi-monthly rounds, so a monthly series is constructed by carrying the most recently published round forward until the next round is released. A forecaster at month t therefore uses the latest reading actually available at t,

and the series does not use readings published later.[12] These series are excluded from the indicator panel throughout, and enter only in the models of Section 4.8. Climate variables are the Oceanic Niño Index and the all India monthly rainfall deviation from its long period average. These are also excluded from the indicator panel, and enter only in the models of Section 4.10.

2.3 Timing of the Information Set

Indicators for India are released on different schedules, and a forecast that uses all of them as though they were available at the same time would use information that no forecaster possess. We therefore adopt a single convention: every indicator enters the model with a lag of one month, so that a forecast constructed at month t uses π_s^{MOM} for $s \leq t$ and $x_s^{(i)}$ for $s \leq t-1$.

The convention is set by the release schedule of the slower indicators. The consumer price index for month t is published by the Ministry of Statistics and Programme Implementation around the twelfth day of month $t+1$. The index of industrial production is published on the same date but with a further month's delay, so that the reading for month t appears around the twelfth day of month $t+2$. Merchandise trade data are released in the middle of month $t+1$, motor vehicle sales in the second week of month $t+1$, petroleum consumption in the third week of month $t+1$, and scheduled commercial bank credit with a delay of about six weeks. The purchasing managers' indices are the fastest of the indicators used here, appearing in the first days of month $t+1$.

Consider a forecast constructed on the day core inflation for December is published, in the second week of January. On that date the forecaster observes core inflation through December, industrial production through November, trade and credit through November, and the purchasing managers' indices through December. The convention adopted here assigns all indicators the information set available through November. It is therefore conservative for the purchasing managers' indices, which are in fact available for December, and accurate for industrial production, trade and credit. No indicator is treated as available earlier than it is.

We do not use vintage data. The indicator series used here are the currently published series, and revisions that occurred after their initial release are embedded in them. A true real time database of the kind assembled for the United States is

[12] See Das, Lahiri, and Zhao (2019) on the properties of this survey, and Coibion and Gorodnichenko (2015) on the information rigidity that makes a step function of this kind a reasonable representation of the expectations available to a forecaster between rounds.

not available for this panel, and the convention above addresses the availability of an observation rather than the revision of its value.

3 Estimation and Forecast Procedure

3.1 The Estimation and Forecast Split

The sample runs from $t = 1$ to $t = T$. Let τ be a time period with $\tau < T$. The procedure for every model in this paper is as follows.[13]

1. Estimate the model using data from $t = 1$ to $t = \tau$.
2. Using those parameter estimates, construct forecasts of $\pi^{\text{MOM}}_{\tau+h}$ for $h = 1, \ldots, 12$.
3. Set $\tau' = \tau + 1$, re-estimate the model using data from $t = 1$ to τ', and construct forecasts of $\pi^{\text{MOM}}_{\tau'+h}$ for $h = 1, \ldots, 12$.
4. Repeat step 3, advancing τ' by one month at a time, until $\tau' = T - 1$.

The first value of τ is January 2017, so that the estimation sample at the first step covers January 2011 to January 2017 and the forecast comparison covers February 2017 to December 2025. The estimation sample grows by one month at each step.

The choice of January 2017 balances two requirements. The first estimation sample must be long enough to estimate the models of Section 4, the largest of which use forty predictors, and January 2011 to January 2017 provides seventy three monthly observations for that purpose. The comparison sample must in turn be long enough for the tests of Section 3.4 to have power, and February 2017 to December 2025 provides one hundred and seven forecast origins. Starting in 2017 also places the entire comparison within the flexible inflation targeting regime adopted in 2016, so that the comparison is not made across a change in the monetary policy framework.

3.2 Multi-step Forecasts

For each horizon h we estimate a separate equation in which the dependent variable is π^{MOM}_{t+h} and the regressors are dated t. For a model with regressor vector z_t,

$$\pi^{\text{MOM}}_{t+h} = \alpha_h + \beta_h' z_t + \varepsilon_{t+h}, \qquad h = 1, \ldots, 12. \tag{3}$$

[13] All estimation and forecasting was carried out in R 4.4.2. Replication code is available from the authors.

In estimating a separate equation for each h, rather than estimating one equation for $h = 1$ and iterating it forward, we follow Marcellino, Stock, and Watson (2006). The single exception is the auto-regressive moving average model of Section 4.3, which is iterated forward, and this is stated explicitly in that model description below.

3.3 Conversions

At τ, equation (3) delivers $\widehat{\pi}^{\text{MOM}}_{\tau+1}, \ldots, \widehat{\pi}^{\text{MOM}}_{\tau+12}$. We convert these into a forecast of the price index,

$$\widehat{P}_{\tau+h} = P_\tau \exp\left(\frac{1}{100}\sum_{s=1}^{h}\widehat{\pi}^{\text{MOM}}_{\tau+s}\right), \tag{4}$$

and then into a forecast of year on year inflation,

$$\widehat{\pi}^{\text{YOY}}_{\tau+h} = 100\left(\frac{\widehat{P}_{\tau+h}}{P_{\tau+h-12}} - 1\right). \tag{5}$$

For $h \leq 12$ the denominator $P_{\tau+h-12}$ is an index value dated at or before τ. It is therefore observed at the time the forecast is made, and is not itself a forecast. This is how base effects, which are large in India's inflation dynamics, enter the forecast.

Equation (5) has one implication for the comparison of models across horizons. At $h = 1$, eleven of the twelve monthly price changes that make up $\pi^{\text{YOY}}_{\tau+1}$ are already observed at τ, and models differ only through the single forecast month. At $h = 12$ all twelve monthly price changes are forecast. Differences between models are therefore compressed at short horizons. For this reason we also report RMSEs for π^{MOM}_{t+h} itself, into which no observed price changes enter. These are in Table 8 in the Appendix.

3.4 Forecast Evaluation

Let $\widehat{\pi}^{\text{YOY},m}_{\tau+h}$ denote the forecast of year on year inflation from model m made at τ for horizon h, and let $\mathcal{T}$ denote the set of values of τ over which forecasts are compared. The root mean squared forecast error of model m at horizon h is

$$\text{RMSE}^m_h = \sqrt{\frac{1}{|\mathcal{T}|}\sum_{\tau\in\mathcal{T}}\left(\pi^{\text{YOY}}_{\tau+h} - \widehat{\pi}^{\text{YOY},m}_{\tau+h}\right)^2}. \tag{6}$$

We report $\text{RMSE}^m_h/\text{RMSE}^{\text{T0}}_h$, i.e. the ratio of the RMSE of model m to that of the baseline auto-regression of Section 4.2. A ratio below one indicates that

model m produced a smaller forecast error than the baseline. Scaling accuracy by a parsimonious benchmark rather than reporting it in levels follows a convention dating to Theil's inequality coefficient U_2 (Theil 1966) and is standard in the forecast evaluation literature (Armstrong and Collopy 1992; Hyndman and Koehler 2006). The ratio is scale free and therefore comparable across horizons and price measures, whereas RMSE levels are not, and it isolates the question of interest, namely whether the information in x_t improves on what the history of inflation already contains. Following Stock and Watson (2003) and Rossi and Sekhposyan (2010), we use an auto-regression rather than the random walk of Theil's original U_2 as the denominator.

We test whether a difference in forecast accuracy is statistically significant using the test of Diebold and Mariano (1995), applied to the difference in squared forecast errors, with a heteroskedasticity and autocorrelation consistent variance estimator using $h-1$ lags and the small-sample correction of Harvey, Leybourne, and Newbold (1997). The alternative hypothesis is that model m is more accurate than the baseline. This is the test reported in every table below. Since most of the models below nest the baseline, the Diebold-Mariano statistic is not asymptotically normal under the null (Clark and McCracken 2001) and the test is conservative against the larger model, so a failure to reject should be read as an absence of evidence that the larger model forecasts more accurately, rather than as evidence that the additional regressors carry no predictive content.[14]

We report results for five samples: the full comparison sample, February 2017 to December 2025; the sample ending February 2020; the sample beginning January 2022; the full comparison sample excluding March 2020 to December 2021; and those pandemic months alone. The five are reported because the comparison sample contains the pandemic, during which both inflation and the indicators behaved unusually. The full sample gives the headline comparison; the samples ending before and beginning after the pandemic show whether any conclusion is specific to one regime; the sample that excludes the pandemic months shows whether a conclusion drawn on the full sample is an artifact of those months alone; and the pandemic months on their own show what drives that artifact where one is present.

[14]The distinction is that of Clark and West (2007), whose MSPE-adjusted statistic tests the second hypothesis rather than the first. Since our question is whether a forecaster gains by using the larger model, the Diebold and Mariano statistic is more relevant.

4 Models

4.1 Overview

Table 1 summarizes the nine classes of models. T0, a univariate autoregression, is the baseline against which all others are measured. Models T1 to T8 each add information or flexibility to T0, and are grouped as follows. T1 alters the univariate specification, replacing the direct projection with an iterated ARMA on the same information set, so that any difference reflects functional form rather than added information. T2 to T4 add the indicator panel, first one indicator at a time, then through factors extracted from the full panel, and then from a subset selected on its real-time forecast record. T5 replaces the conditional mean with conditional quantiles, asking whether the panel informs the spread of inflation when it does not inform its center. T6 adds survey expectations, held out of the panel elsewhere so that its contribution can be read separately. T7 replaces least squares with penalized and non-parametric estimators, allowing shrinkage and non-linearity over the same predictors. T8 adds climate variables. Throughout, we employ root mean squared errors as our model comparison metric, and test differences in accuracy using the Diebold and Mariano statistic described in Section 3.4.[15]

Four conventions apply throughout and are not repeated in each subsection. First, the regressor vector z_t in (3) contains only variables dated t or earlier, subject to the timing convention of Section 2.3. Second, every model is estimated and every forecast constructed by the procedure of Section 3.1. Third, every model that includes lags of π^{MOM}, other than T1, uses the three lags selected for T0. Fourth, models estimated by least squares drop any observation for which a regressor is missing. In T3 and T4 an indicator that is missing at t enters the standardized panel at its estimation sample mean, and in T7 missing indicator values are replaced by their estimation sample means before estimation, so that the penalized and non-parametric estimators use the full set of dates.

4.2 T0: Auto-Regression

The baseline model is

$$\pi^{\text{MOM}}_{t+h} = \alpha_h + \sum_{j=1}^{p} \phi_{h,j} \pi^{\text{MOM}}_{t-j+1} + \varepsilon_{t+h}. \tag{7}$$

[15] The exception is the quantile regressions, which are evaluated under quantile loss against a quantile auto-regression.

Model	Departure from the baseline	Versions	Estimator
T0	None (baseline)	1	Least squares
T1	ARMA in place of (7); forecast by iteration	1	Maximum likelihood
T2	Adds one indicator $x_t^{(i)}$ and its first lag	40	Least squares
T3	Adds factors $\widehat{F}_t$ from all 40 indicators	3	Principal components, least squares
T4	Adds factors from the subset $S(\tau)$ of indicators	3	Principal components, least squares
T5	Factors of T3 or T4; conditional quantiles in place of the mean	9	Quantile regression
T6	Adds survey expectations $E_t[\pi_{t+h}^{\text{MOM}}]$, alone or with factors	3	Least squares
T7	Adds all 40 indicators	4	Ridge, lasso, elastic net, random forest
T8	Adds ENSO phases and rainfall deviation	3	Least squares

Table 1: Summary of models

Note: "Versions" counts the specifications reported for each model at each horizon. T1 uses the same information set as the baseline and differs only in functional form and in generating the forecast path by iteration rather than by direct projection at each horizon. For T3 and T4 the three versions fix the number of factors r at one, at two, and at the value selected by the criterion of Bai and Ng (2002). For T5 the nine versions are three specifications of z_t at each of three quantiles. Every model is estimated at horizons $h = 1, 3, 6$ and 12 by the procedure of Section 3.1, and every model that includes lags of π^{MOM} uses the three lags selected for T0.

The lag order p is selected once by the Bayesian Information Criterion (BIC) over $p \in \{1, \ldots, 6\}$ on the full sample, and is held fixed at $p = 3$ across horizons and across values of τ. Fixing the order in this way, rather than re-selecting it at each τ, holds the benchmark constant across the comparison, so that differences in forecast accuracy are not attributable to changes in the specification of T0 itself. Every model below that includes lags of π^{MOM}, other than T1, uses these same three lags, so that differences in forecast accuracy across models are attributable to the additional regressors rather than to differences in auto-regressive order.

4.3 T1: Auto-Regressive Moving Average (ARMA) Model

T1 is an ARMA(p, q),

$$\pi_t^{\text{MOM}} = \alpha + \sum_{i=1}^{p} \phi_i \pi_{t-i}^{\text{MOM}} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \tag{8}$$

in which the orders p and q are selected at each τ by the corrected Akaike Information Criterion (AIC), using the algorithm of Hyndman and Khandakar (2008). T1 is the only model in the paper that is forecast by iteration rather than by direct projection. At τ we estimate (8) on data from $t = 1$ to $t = \tau$ and forecast $\pi_{\tau+1}^{\text{MOM}}$. To forecast $\pi_{\tau+2}^{\text{MOM}}$ we retain those parameter estimates and treat the forecast of $\pi_{\tau+1}^{\text{MOM}}$ as an observation, continuing in this way to $h = 12$. The comparison of T1 with T0 is therefore also a comparison of iterated with direct forecasts (see Marcellino, Stock, and Watson (2006)).

4.4 T2: Auto-Regression Augmented with One Indicator

For each indicator $i = 1, \ldots, N$ we estimate

$$\pi_{t+h}^{\text{MOM}} = \alpha_h + \sum_{j=1}^{p} \phi_{h,j} \pi_{t-j+1}^{\text{MOM}} + \beta_{h,0} x_t^{(i)} + \beta_{h,1} x_{t-1}^{(i)} + \varepsilon_{t+h}, \tag{9}$$

which yields $N = 40$ models at each horizon. Setting $\beta_{h,0} = \beta_{h,1} = 0$ returns (7), so T2 nests T0. T2 serves two purposes. It establishes whether any single indicator improves upon the baseline, which is the most conservative use that can be made of the panel. It also generates the sequence of forecast errors used to select the subset of indicators in T4.

4.5 T3: Diffusion Index using All Indicators

Let $X_t = (x_t^{(1)}, \ldots, x_t^{(N)})'$ and assume the factor structure

$$X_t = \Lambda F_t + e_t, \tag{10}$$

where F_t is an $r \times 1$ vector of factors, Λ is an $N \times r$ matrix of loadings, and e_t is an $N \times 1$ vector of idiosyncratic disturbances. Following Stock and Watson (2002) we estimate F_t by principal components applied to the standardized panel: each indicator is standardized to mean zero and unit variance using moments computed on the estimation sample $t = 1$ to $t = \tau$, so that indicators with larger variances do not dominate the estimated factors. Equation (10) is approximate in the sense of Chamberlain and Rothschild (1983) and Bai and Ng (2002): the covariance matrix $\Sigma_e = \mathbb{E}(e_t e_t')$ is not restricted to be diagonal, and its largest eigenvalue is bounded as N grows,

$$\lambda_{\max}(\Sigma_e) \leq C < \infty \quad \text{for all } N, \qquad \liminf_{N \to \infty} \frac{\lambda_r(\Lambda\Lambda')}{N} > 0, \tag{11}$$

so that the first r eigenvalues of the common component diverge with N while those of the idiosyncratic component do not. Weak cross-sectional and serial correlation in e_t is therefore permitted, and the two components remain separable in the limit.

The forecasting equation is

$$\pi_{t+h}^{\text{MOM}} = \alpha_h + \sum_{j=1}^{p} \phi_{h,j} \pi_{t-j+1}^{\text{MOM}} + \gamma_h' \widehat{F}_t + \varepsilon_{t+h}, \tag{12}$$

and setting $\gamma_h = 0$ returns (7). We report three versions of T3, with r fixed at

one, with r fixed at two, and with r selected at each τ by the information criterion of Bai and Ng (2002). The factors are re-estimated at every τ.

4.6 T4: Diffusion Index using a Subset of Indicators

T3 estimates factors from all N indicators, whether or not an indicator has been useful in forecasting. T4 estimates factors from a subset selected on past forecast accuracy. To proceed, we fix τ, and consider those values $s < \tau$ at which a one month ahead forecast was constructed and for which the outcome π_{s+1}^{MOM} is observed by τ. Let $\widehat{\pi}_{s+1}^{\text{MOM},(i)}$ denote the forecast from the T2 model that uses indicator i, and $\widehat{\pi}_{s+1}^{\text{MOM,T0}}$ the forecast from T0. Define

$$R_i(\tau) = \frac{\sum_s \left(\pi_{s+1}^{\text{MOM}} - \widehat{\pi}_{s+1}^{\text{MOM},(i)}\right)^2}{\sum_s \left(\pi_{s+1}^{\text{MOM}} - \widehat{\pi}_{s+1}^{\text{MOM,T0}}\right)^2}, \tag{13}$$

the ratio of the sum of squared one month ahead forecast errors of the model using indicator i to that of T0, computed over those forecasts whose outcomes are observed by τ. The subset is

$$S(\tau) = \{i : R_i(\tau) < 1\}. \tag{14}$$

Factors are then estimated from $\{x_t^{(i)} : i \in S(\tau)\}$ as in equation (10) and enter equation (12). As with T3, we report versions with r fixed at one, two, and selected by the criterion of Bai and Ng (2002).

Three features of (14) require explanation. First, the criterion compares sums of squared errors and applies no test of significance. An indicator enters $S(\tau)$ whenever its ratio falls below one on the record available at τ, however narrow the margin, so it may enter because it forecast well by chance rather than because it predicts inflation. This is deliberate: the subset reproduces what a forecaster could have selected in real time, not a claim about which indicators forecast inflation. Second, at early values of τ few forecasts have been constructed and equation (13) is uninformative. For the first twenty four values of τ we therefore select indicators whose coefficient in equation (9) at $h = 1$ has an absolute t statistic above 1.65 on the estimation sample, which is the targeted predictor criterion of Bai and Ng (2008). Third, when fewer than five indicators satisfy (14), the five with the lowest values of $R_i(\tau)$ are used instead, so that the factors are always estimated from at least five series. This occurs at 36 of the 107 origins, all of them from January 2022 onward, when few indicators improve upon T0 at any point in the record.

Because we recompute $S(\tau)$ at each τ, the indicators it contains change over the comparison sample.

4.7 T5: Quantile Regressions

Whereas T0 to T4 forecast the conditional mean of core inflation, T5 forecasts conditional quantiles. Define cumulative growth in the price index over h months,

$$y_{t,h} = 100\left(\ln P_{t+h} - \ln P_t\right). \tag{15}$$

Quantile forecasts are constructed for cumulative growth $y_{t,h}$ and then converted to year on year inflation. This works because the conversion is monotone. Given the index at the origin, P_t, and the base index P_{t+h-12}, which is realized at the origin for every horizon we consider,

$$\pi^{\text{YOY}}_{t+h} = 100\left(\frac{P_t \exp(y_{t,h}/100)}{P_{t+h-12}} - 1\right) \tag{16}$$

is increasing in $y_{t,h}$. Quantiles are preserved under increasing transformations, so the ςth quantile of $y_{t,h}$ maps into the ςth quantile of π^{YOY}_{t+h} exactly. The alternative, forecasting the quantiles of π^{MOM} at each month and summing them, is not valid: the ςth quantile of a sum is not in general the sum of the ςth quantiles.

Let $\varsigma \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ index the quantile. The estimator is

$$\widehat{\theta}_{h,\varsigma} = \arg\min_{\theta} \sum_{t} \rho_\varsigma\left(y_{t,h} - \theta' z_t\right), \qquad \rho_\varsigma(u) = u\left(\varsigma - \mathbf{1}\{u < 0\}\right), \tag{17}$$

which is the quantile regression estimator of Koenker and Bassett (1978). The check function ρ_ς weights positive and negative residuals differently: at $\varsigma = 0.90$ a negative residual receives weight 0.90 and a positive residual weight 0.10, so that the fitted value is the level below which 90% of observations fall. All observations enter the estimation for every ς; the estimation sample is not truncated.

The implied forecast of the ςth quantile of year on year inflation is

$$\widehat{Q}_\varsigma\left(\pi^{\text{YOY}}_{\tau+h}\right) = 100\left(\frac{P_\tau}{P_{\tau+h-12}} \exp\left(\frac{\widehat{Q}_\varsigma(y_{\tau,h})}{100}\right) - 1\right). \tag{18}$$

We estimate equation (17) for three specifications of z_t: the three lags of π^{MOM} used in (7), which we refer to as the quantile auto-regression; those lags together with the factors of T3; and those lags together with the factors of T4. The exer-

cise follows Adrian, Boyarchenko, and Giannone (2019), who forecast quantiles of output growth conditional on financial conditions.

Forecasts of quantiles are evaluated by mean quantile loss,

$$L^m_{h,\varsigma} = \frac{1}{|\mathcal{T}|} \sum_{\tau \in \mathcal{T}} \rho_\varsigma \left(\pi^{\text{YOY}}_{\tau+h} - \widehat{Q}^m_\varsigma \left(\pi^{\text{YOY}}_{\tau+h} \right) \right), \tag{19}$$

which is the function used in the estimation of equation (17). We report $L^m_{h,\varsigma}$ relative to the quantile auto-regression, and test the difference in ρ_ς evaluated at each τ.[16] Two points follow. The benchmark for T5 is the quantile auto-regression rather than T0, because the two must be compared under the same loss function. For the same reason the entries reported for T5 are not comparable with the root mean squared forecast errors reported for T0 to T4.

4.8 T6: Household Inflation Expectations

Let $E_t[\pi^{\text{MOM}}_{t+h}]$ denote the survey measure of household inflation expectations described in Section 2. T6 estimates

$$\pi^{\text{MOM}}_{t+h} = \alpha_h + \sum_{j=1}^{p} \phi_{h,j} \pi^{\text{MOM}}_{t-j+1} + \delta'_h E_t[\pi^{\text{MOM}}_{t+h}] + \varepsilon_{t+h}, \tag{20}$$

and the same specification with the factors of T3 and of T4 added, giving three versions. Because the survey measures are excluded from the indicator panel throughout, T6 isolates its contribution relative to the remaining indicators separately. Equation (20) is close to a hybrid Phillips curve as discussed in Galí and Gertler (1999), in which inflation depends on lagged inflation, expected inflation, and a measure of demand: the lag terms carry the backward looking component, the survey term the forward looking component, and, in the versions with factors added, the factors stand in for slack.[17]

[16] Mean forecasts are compared using the Diebold and Mariano statistic as implemented by Hyndman and Khandakar (2008), which applies the small sample correction of Harvey, Leybourne, and Newbold (1997). Quantile forecasts are compared using a heteroskedasticity and autocorrelation consistent t statistic on the difference in ρ_ς, with $h-1$ lags. For the quantile forecasts, the resulting t statistic is evaluated against the standard normal distribution. This process is more conservative in small samples than the Diebold–Mariano test, requiring stronger evidence to identify a statistically significant difference between forecasts.

[17] Galí and Gertler (1999) use a model consistent expectation one period ahead and a marginal cost term derived from the model, whereas here the expectation is an observed survey measure at horizon h and slack is proxied by estimated factors. Equation (20) shares the structure of the hybrid curve but does not estimate its parameters.

4.9 T7: Penalized and Non-Parametric Estimators

Let w_t collect the three lags of π^{MOM} used in equation (7) together with all N indicators. The penalized least squares estimator is

$$\widehat{\beta}_h = \arg\min_{\beta} \sum_{t} \left(\pi^{\text{MOM}}_{t+h} - \beta' w_t\right)^2 + \lambda \left[(1-a)\frac{1}{2}\|\beta\|_2^2 + a\|\beta\|_1\right], \tag{21}$$

which gives a ridge regression at $a = 0$, the lasso of Tibshirani (1996) at $a = 1$, and the elastic net of Zou and Hastie (2005) at $a = 0.5$. The penalty parameter λ is chosen at each τ by five fold cross validation, which splits the estimation sample into five parts, fits the model on four and measures its error on the fifth, and selects the λ with the lowest average error across the five rotations. The folds are drawn within the estimation sample available at τ, so no observation after the origin enters the choice of λ. Unlike T3 and T4, which impose a factor structure on the panel before forecasting, the penalized estimators use the indicators directly and restrict the coefficient vector instead, ridge shrinking all coefficients toward zero, lasso setting some exactly to zero, and the elastic net combining the two.

We also estimate a random forest (see Breiman (2001)), in which the forecast is the average of the forecasts of 500 regression trees, each estimated on a bootstrap re-sample of the estimation sample with $\lfloor\sqrt{d}\rfloor$ of the d elements of w_t drawn at random as candidates at each split, which is the default of the implementation of Wright and Ziegler (2017)[18]. This permits non-linear relationships and interactions between indicators, which none of the preceding models admits: the response of inflation to an indicator may differ across the range of that indicator and according to the values of the others, whereas a linear specification imposes a single coefficient in both respects. Medeiros et al. (2021) report that estimators of this kind improve upon standard benchmarks for United States inflation.

4.10 T8: Climate Variables

We also consider whether climate variables affect the forecast of core inflation. Cashin, Mohaddes, and Raissi (2017) find that El Niño episodes raise Indian consumer price inflation by about half a percentage point within a year, working through the large weight of food in the CPI basket. Core inflation excludes food, so climate variables cannot affect it directly. However, the indirect channel may exist: food price shocks raise household inflation expectations, which pass into

[18] The two randomizations reduce the correlation amongst the trees, so their average has lower variance than any single tree. Unlike the penalized regressions, the forecast is non-parametric and non-linear in w_t.

wage demands and into the services component of core. Anand, Ding, and Tulin (2014) estimate these second round effects for India and find them large, with the gap between headline and core inflation closing by about three quarters within a year. The channel operates with long and uncertain lags, so T8 is expected to help, if at all, at longer horizons rather than at one or three months.

Let ONI_t denote the Oceanic Niño Index and d_t the percentage deviation of all India rainfall from its long period average. Define $\text{EN}_t = \max(\text{ONI}_t, 0)$ and $\text{LN}_t = \max(-\text{ONI}_t, 0)$, which separate the warm and cool phases of the index and allow them to enter with different coefficients. T8 estimates

$$\pi_{t+h}^{\text{MOM}} = \alpha_h + \sum_{j=1}^{p} \phi_{h,j} \pi_{t-j+1}^{\text{MOM}} + \theta_1 \text{EN}_t + \theta_2 \text{LN}_t + \theta_3 \text{EN}_{t-3} + \theta_4 \text{LN}_{t-3} + \theta_5 d_t + \theta_6 \bar{d}_t + \varepsilon_{t+h}, \tag{22}$$

where $\bar{d}_t$ is the three month moving average of d_t. We report this specification, and the two specifications that include only the index and only the rainfall variables.

5 Results

The comparison sample runs from February 2017 to December 2025, and yields 107 forecast errors at $h = 1$, 105 at $h = 3$, 102 at $h = 6$ and 96 at $h = 12$. We report results for the conditional mean first, then for the conditional distribution, and then model by model.

5.1 The Conditional Mean

Over the full comparison sample no model forecasts the conditional mean of core inflation more accurately than the auto-regression. Six of the sixty-eight combinations of model and horizon in Figure 2 produce a ratio below one, and in every case the margin is too small to be informative: the largest is 1.5 %, for T4 estimated with one factor at $h = 3$, and the rest lie between 0.05 and 1.4%. None is significant, with p-values ranging from 0.349 to 0.490. The remaining models are worse than T0, in most cases substantially so.

A null of this kind invites the objection that the comparison is underpowered. It is not. Appendix D reports what happens when the indicator subset of T4 is chosen using the whole comparison sample rather than the record available at the time of the forecast. That version cannot be used to forecast, but it uses the same estimator, the same horizons, the same test and the same number of forecast errors, and it reduces the error by about 3% at $h = 3$ and $h = 6$, which the test

detects at conventional levels. The design can therefore identify a gain of that size over this sample. Section 6 shows that once the pandemic months are excluded, one of the models reported here delivers a comparable gain in real time.

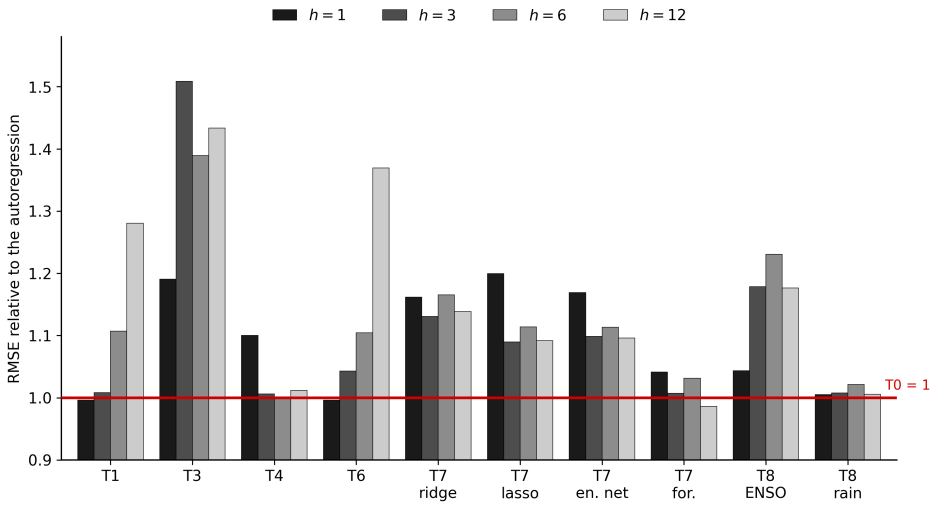


Figure 2: Root mean squared forecast error relative to T0

Note: Each bar is the root mean squared forecast error of the model divided by that of T0, the autoregression, for year on year inflation. The red line is drawn at one, so that bars below it indicate a smaller forecast error than T0. Shades denote horizons. For T3 and T4 the version estimated with a single factor is shown; Table 2 reports all three. Full comparison sample, February 2017 to December 2025. Source: Authors' estimates.

5.2 The Conditional Distribution

The same panel that fails to improve the central forecast does improve forecasts of the quantiles around it. Table 4 reports mean quantile loss for the quantile regression of T5 that includes the factors of T4, measured against a quantile autoregression estimated on the same three lags. Figure 3 shows the five quantiles.

All twenty combinations of quantiles and horizons produce a ratio below one.[19] Five are significant at the 5% level and five more at the 10% level. Where the gains are significant they run between 6 and 14%, and the largest single reduction, 0.859 at the ninetieth percentile at $h = 3$, is roughly ten times the best margin in Figure 2. The direction is uniform: not one of the twenty ratios exceeds one.

Two features of the pattern are worth drawing out. The gains are concentrated

[19]The median row of Table 4 also serves as a check under a second loss function. At $\varsigma = 0.50$ the check function in equation (19) is proportional to absolute error, so those entries compare the two models under mean absolute error rather than squared error: the factor specification reduces loss by 7.5% at $h = 3$ and 7.6% at $h = 6$.

at the longer horizons, with eight of the ten significant cells at $h = 6$ or $h = 12$ and none at $h = 1$, which is the reverse of the usual finding that predictability decays with the horizon. And they are largest away from the median: the ninetieth percentile is significant at three horizons and the tenth at two, while the median is significant at two and the seventy fifth at two. The panel appears to carry information about how widely core inflation is likely to be dispersed around its central path, and to carry it further ahead than it carries information about the path itself.

The contrast with the mean deserves emphasis, because the two exercises use the same estimated factors and differ only in the loss function under which they are fitted and evaluated. Under squared error loss the factors add nothing. Under quantile loss they add a good deal. The result does not come from the selection advantage that produces the bound in Appendix D: the same quantile regression built on the full-sample subset S^* is significant at one of the twenty combinations against five for the honest subset, so the favorably chosen version is the weaker of the two.

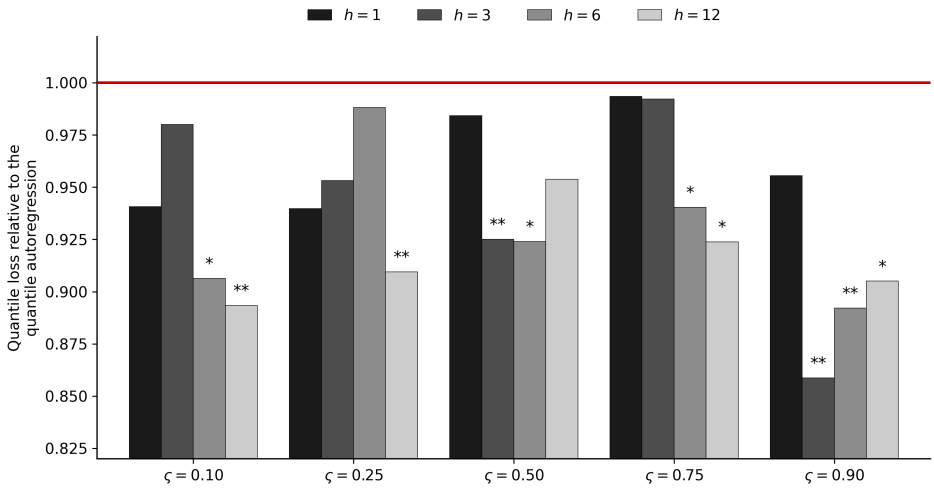


Figure 3: Mean quantile loss relative to the quantile autoregression

Note: Each bar is the mean quantile loss of the quantile regression that includes the factors of T4, divided by that of the quantile autoregression, at quantile ς and horizon h. The red line is drawn at one, so that bars below it indicate a smaller loss than the quantile autoregression. One asterisk denotes significance at the 10% level and two at the 5% level. Full comparison sample, February 2017 to December 2025. Source: Authors' estimates.

5.3 Results by Model

T2. Adding indicators one at a time is the most conservative use that can be made of the panel, and it produces very little by way of forecast improvements.

The count of indicators that beat T0 rises with the horizon, from three at $h = 1$ to ten at $h = 3$, nine at $h = 6$ and fourteen at $h = 12$, but the improvements are almost all too small to distinguish from sampling variation. Across forty indicators and four horizons, one indicator is significant at more than one horizon. The price of gold reduces the error by 2.9% at $h = 3$ and 3.0% at $h = 6$, with p-values of 0.016 and 0.041, and by 1.4 % at $h = 1$, with a p-value of 0.099. Silver reduces the error by 2.3% at $h = 6$, significant only at the 10% level. Nothing else survives at either level at more than one horizon, and the largest single reduction in the whole scan, 7.0% from passenger vehicle sales at $h = 12$, carries a p-value of 0.845. Table 3 reports the indicators significant at conventional levels.[20]

T3 and T4. Compressing the whole panel into factors produces the largest forecast errors in our analysis. The ratios for T3 run from 1.065 to 1.529 and are worst at three and six months, so the estimated factors are not neutral additions but active sources of error. One reading, consistent with the T2 results, is that with few indicators carrying usable signals, the leading principal components of the panel summarize variation that has little to do with core inflation, and the forecasting regression then projects onto it. Screening the panel on past forecast accuracy changes the picture. The ratios for T4 fall to between 0.985 and 1.100, and the version estimated with a single factor produces the lowest ratio of any model in the paper at $h = 3$. None of these differences is significant over the full sample, but Section 6 shows that they are once the pandemic is set aside. The gap between T3 and T4 is what screening the panel is worth.

T6. Survey expectations behave like a variable that is informative about the current level of inflation and progressively less informative about its future path. The ratio is 0.996 at $h = 1$, and rises monotonically to 1.369 at $h = 12$. Adding the survey to the factor models of T3 and T4 leaves every ratio above one, and in the case of T4 it undoes the gain that the screened factors deliver on their own. Nothing in these results suggests the survey should be part of a forecasting model for core inflation at horizons beyond one month.

T7. The penalized regressions are among the weakest models in our analysis, with ratios between 1.090 and 1.200 at every horizon, which is a poor showing for methods designed for exactly this setting of many predictors and a short sample. The random forest is the better of the two approaches and comes close to T0, with ratios between 0.986 and 1.042 and its best performance at the longest horizon,

[20] Results for all forty are available from the authors on request.

though the difference at $h = 12$ carries a p-value of 0.416. The penalized regressions are not uniformly poor, however: Section 6 shows they beat T0 decisively in the pandemic months, and only there.

T8. The climate variables divide into two groups. The rainfall deviation is essentially inert, with ratios between 1.005 and 1.021, so adding it neither helps nor hurts. The Oceanic Niño Index is costly, with ratios between 1.044 and 1.230, and adding both gives ratios between 1.054 and 1.269. There is no horizon at which either improves the forecast, including $h = 12$, where the channel set out in Section 4.10 would have the most time to operate.

5.4 Stability of the Comparison

A ratio computed over nine years can hide a model that forecasts well in one period and badly in another, and averaging over the pandemic is a particular concern here. Figure 4 therefore recomputes the ratio in equation (6) over rolling twenty four month windows at $h = 6$.

T3 is the clearest case. Its ratio rises above 2.7 during and after the pandemic and returns close to one from 2023, so almost all of its poor full sample performance comes from a single episode in which the factors moved sharply and core inflation did not. The elastic net is the mirror image, falling to 0.71 during the pandemic months and rising steadily afterward to end the sample above 1.3. T4 follows the auto-regression closely for most of the sample, between 0.92 and 1.30, rising above one only during the pandemic and settling slightly below one from 2023.

The remaining pattern is not visible in the full sample ratios at all. The iterated ARMA of T1 rises to 1.47 in 2021, crosses below one in late 2022 and continues to fall, reaching 0.82 by the end of the sample. Its full sample ratio of 1.107 therefore averages a reversal rather than describing a stable relationship. T1 and T4 are below one in most months from 2023 onward, as is T3 once its pandemic errors leave the window, and the second half of the comparison sample is the subject of the next section.

Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$
A. T0 and T1: univariate models				
T0: AR(p)	1.000	1.000	1.000	1.000
T1: ARMA(p, q)	0.996	1.008	1.107	1.281
B. T3: diffusion index, all indicators				
$r = 1$	1.065	1.529	1.520	1.394
$r = 2$	1.118	1.488	1.359	1.455
r by Bai and Ng	1.191	1.508	1.389	1.433
C. T4: diffusion index, subset of indicators				
$r = 1$	1.025	0.985	1.028	1.000
$r = 2$	1.079	0.994	1.001	1.002
r by Bai and Ng	1.100	1.006	1.000	1.012
D. T6: household inflation expectations				
T0 + expectations	0.996	1.043	1.104	1.369
T3 + expectations	1.109	1.283	1.217	1.430
T4 + expectations	1.132	1.144	1.196	1.371
E. T7: machine learning methods				
Ridge	1.162	1.131	1.165	1.138
Lasso	1.200	1.090	1.114	1.092
Elastic net	1.169	1.099	1.114	1.096
Random forest	1.042	1.007	1.031	0.986
F. T8: climate variables				
T0 + ENSO	1.044	1.179	1.230	1.176
T0 + rainfall	1.005	1.007	1.021	1.006
T0 + ENSO and rainfall	1.054	1.198	1.269	1.199

Table 2: Relative RMSE of h-month ahead forecasts of core inflation, February 2017 to December 2025

Note: Entries are the root mean squared forecast error of the model divided by that of T0, the autoregression, computed for year on year inflation at each horizon. Grey shading denotes the baseline, which equals one by construction; green shading denotes an entry below one. *, ** and *** denote rejection of the null of equal forecast accuracy in favour of the model, at the ten, five and one percent level, respectively, by the Diebold and Mariano (1995) test. T2 is reported separately in Table 3.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Indicators tested	40	40	40	40
Ratio below one	3	10	9	14
Ratio below one and significant at 5%	0	1	1	0
Indicators significant at the five percent level				
Gold	0.986^{*}	0.971^{**}	0.970^{**}	1.007

Table 3: T2: autoregression augmented with one indicator, summary

Note: Each of the forty indicators is added to T0 one at a time as in equation (9). The upper panel counts, at each horizon, the indicators whose ratio of root mean squared forecast errors to T0 is below one, and those for which that difference is significant at the five percent level. The lower panel reports every indicator that is significant at the five percent level at any horizon. Stars as in Table 2. Results for all forty indicators are available from the authors on request.

Quantile	$h = 1$	$h = 3$	$h = 6$	$h = 12$
$\varsigma = 0.10$	0.941	0.980	0.906^{*}	0.893^{**}
$\varsigma = 0.25$	0.940	0.953	0.988	0.910^{**}
$\varsigma = 0.50$	0.984	0.925^{**}	0.924^{*}	0.954
$\varsigma = 0.75$	0.993	0.992	0.940^{*}	0.924^{*}
$\varsigma = 0.90$	0.956	0.859^{**}	0.892^{***}	0.905^{*}

Table 4: T5: relative quantile loss, quantile regression with estimated factors

Note: Entries are the mean quantile loss of the quantile regression that includes the factors of T4, divided by the mean quantile loss of the quantile autoregression, at each quantile ς and horizon. Loss is the function ρ_{ς} defined in equation (17). Stars as in Table 2, with the test applied to the difference in ρ_{ς}. These entries are not comparable with those of Table 2, because the loss function differs.

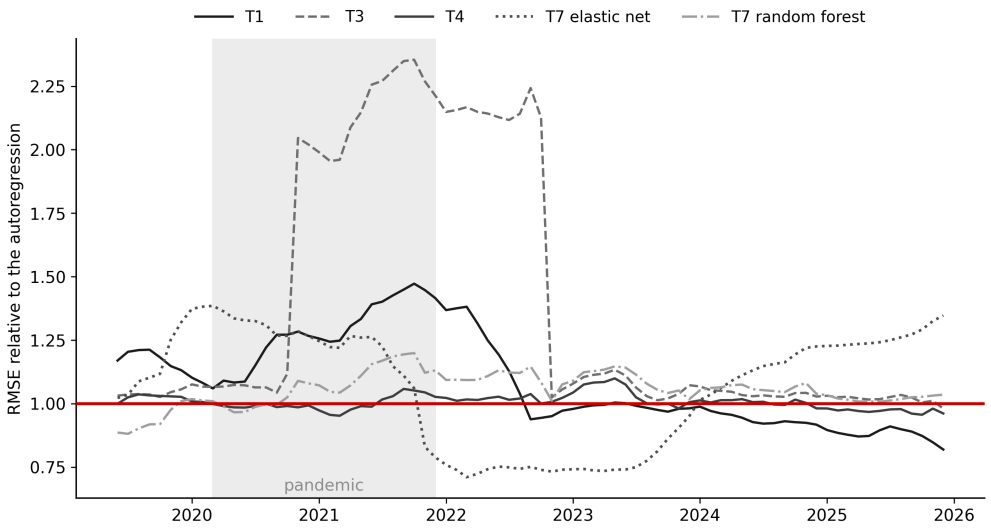


Figure 4: Rolling twenty four month RMSE relative to T0, $h = 6$

Note: Each line is the root mean squared forecast error of the model divided by that of T0, computed over the preceding twenty four months of forecast outcomes at a horizon of six months. The red line is drawn at one. The shaded region is March 2020 to December 2021. For T3 and T4 the version estimated with a single factor is shown. Source: Authors' estimates.

6 Sub-sample Results

6.1 The Conditional Mean

Table 5 reports the ratio in equation (6) at three and six months for four evaluation samples: the full comparison sample, the sample excluding March 2020 to December 2021, the sample beginning January 2022, and the pandemic months alone. Appendix Tables 9 and 10 report the sample ending February 2020 and the sample beginning January 2022.

The full sample result does not survive the exclusion of the pandemic. T4 estimated with a single factor has a ratio of 0.944 at $h = 3$ and 0.953 at $h = 6$, both significant at the 5% level, and 0.979 at $h = 12$, significant at the 10% level. Reductions of 5% at two horizons are larger than anything the model achieves over the full sample, and comparable to what the look-ahead specification of Appendix D achieves with the benefit of hindsight. The sample beginning January 2022 tells the same story on a shorter record: T4 with one factor has a ratio of 0.966 at $h = 6$, significant at the 5% level, and the iterated ARMA of T1 has a ratio of 0.914 at the same horizon, significant at the 10 % level.

What separates these samples from the full one is visible in the last column. In the pandemic months T4 has ratios of 1.076 at $h = 3$ and 1.253 at $h = 6$, and

the twenty two forecast errors in that window are large enough to dominate the full sample averages. The screened factors do poorly during the pandemic and are marginally better on either side of it, and squared error loss weights the first much more heavily than the second.

Three considerations require reading the result cautiously. It is strongest for one of the three versions of T4, the one estimated with a single factor; the versions with two factors and with the number chosen by the criterion of Bai and Ng are below one at the same horizons but not significant. It is absent at $h = 1$, where the ratio is 1.036. And the pandemic window was chosen by inspection rather than specified in advance, so the significance levels in the second and third columns are not those of a pre-specified test. The defensible statement is that the auto-regression is hard to beat over the full sample, that a screened factor model beats it outside the pandemic at horizons of three months and beyond, and that a longer sample will be needed to establish whether that is durable.

The penalized regressions of T7 move in the opposite direction and much more sharply. All three beat T0 at the 5 % level at $h = 6$ in the pandemic months, with ratios of 0.784 for ridge, 0.768 for the lasso and 0.749 for the elastic net, and ridge regression is significant at all four horizons. Excluding those months the same estimators have ratios above 1.10 at every horizon. This is the behavior these methods are built for: when one large shock moves most of the panel at once, many indicators carry a common signal, and shrinking across them extracts it more efficiently than a univariate model can. The cost is a poor showing in every other sample.

6.2 The Conditional Distribution

Table 6 splits the quantile results into the full sample, the sample excluding March 2020 to December 2021, and those months alone.

The distributional result is not a pandemic artifact. Excluding those months, four of the twenty combinations remain significant at the 5 % level and two more at the 10 % level, against five and five in the full sample, and the ratios at the significant cells are lower rather than higher: 0.872 at the tenth percentile at $h = 6$, 0.899 at the twenty fifth at $h = 12$, 0.905 at the median at $h = 6$, and 0.899 at the ninetieth at $h = 6$. Removing the most volatile episode in the sample leaves the gains intact and in some cases enlarges them.

Model	Full sample	Excluding pandemic	January 2022 onward	March 2020 to Dec. 2021
		A. Horizon $h = 3$		
T0: AR(p)	1.000	1.000	1.000	1.000
T1: ARMA(p, q)	1.008	0.965	0.923	1.106
T3: $r = 1$	1.529	0.993	0.988	2.369
T3: $r = 2$	1.488	1.022	1.003	2.246
T3: r by Bai and Ng	1.508	1.011	1.025	2.304
T4: $r = 1$	0.985	0.944**	0.971	1.076
T4: $r = 2$	0.994	0.982	0.963	1.022
T4: r by Bai and Ng	1.006	1.007	0.989	1.005
T6: T0 + expectations	1.043	1.022	1.014	1.093
T6: T3 + expectations	1.283	1.040	1.058	1.737
T6: T4 + expectations	1.144	1.085	1.024	1.276
T7: Ridge	1.131	1.235	1.136	0.825**
T7: Lasso	1.090	1.173	1.115	0.856**
T7: Elastic net	1.099	1.187	1.121	0.848**
T7: Random forest	1.007	1.049	1.100	0.898*
T8: T0 + ENSO	1.179	1.219	1.086	1.074
T8: T0 + rainfall	1.007	1.028	1.049	0.956
T8: T0 + ENSO and rainfall	1.198	1.248	1.158	1.067
		B. Horizon $h = 6$		
T0: AR(p)	1.000	1.000	1.000	1.000
T1: ARMA(p, q)	1.107	1.004	0.914*	1.408
T3: $r = 1$	1.520	1.009	0.997	2.609
T3: $r = 2$	1.359	1.021	1.013	2.156
T3: r by Bai and Ng	1.389	1.030	1.033	2.226
T4: $r = 1$	1.028	0.953**	0.966**	1.253
T4: $r = 2$	1.001	0.979	0.978	1.072
T4: r by Bai and Ng	1.000	0.995	0.986	1.014
T6: T0 + expectations	1.104	1.051	0.997	1.274
T6: T3 + expectations	1.217	1.079	1.039	1.609
T6: T4 + expectations	1.196	1.095	0.988	1.494
T7: Ridge	1.165	1.254	1.141	0.784***
T7: Lasso	1.114	1.195	1.143	0.768***
T7: Elastic net	1.114	1.198	1.140	0.749***
T7: Random forest	1.031	1.000	1.025	1.134
T8: T0 + ENSO	1.230	1.265	1.135	1.103
T8: T0 + rainfall	1.021	1.034	1.027	0.975
T8: T0 + ENSO and rainfall	1.269	1.302	1.194	1.147

Table 5: Relative RMSE by evaluation sample

Note: Entries are the root mean squared forecast error of the model divided by that of T0, computed on the stated evaluation sample. The pandemic sample is March 2020 to December 2021. Stars as in Table 2.

The horizon pattern noted in Section 5 is if anything sharper outside the pandemic. Every cell significant at the 5 % level in the second column lies at $h = 6$ or $h = 12$, and the entries at $h = 1$ are all within 2% of one, three of them above it. Whatever the factors are picking up about the dispersion of core inflation, they are not picking it up one month ahead.

The pandemic months tell a separate story, and a stronger one. Ratios there fall as low as 0.597 at the ninetieth percentile at $h = 12$ and 0.632 at $h = 3$, and six of the twenty combinations are significant at the 5 % level despite only twenty two forecast errors. The factors were informative about how far core inflation might run from its central path at precisely the time when it ran furthest. That is a separate finding from the one in the second column, and neither depends on the other.

7 Conclusion

We compare nine classes of monthly forecasting models for core consumer price inflation in India over February 2017 to December 2025. Our analysis is general enough to apply to inflation data in other emerging markets developing economies (EMDEs). Our analysis also supports monetary and fiscal policymaking in India, a large EMDE, given its rapid growth and transformation. Over the full comparison sample no model produces a RMSE significantly below that of a univariate auto-regression at any horizon. This is not because the comparison is too short to detect a difference. A version of the diffusion index model that picks its indicators using the whole sample, and so could not have been used to forecast, cuts the error by about 3% at three and six months, and the same test finds that difference significant.

Two sub-samples depart from this. Excluding the pandemic months of March 2020 to December 2021, a diffusion index model built from indicators screened on their real-time forecast record cuts the error by about 5% relative to the auto-regression at three and six months, and wins again on the sample beginning January 2022. The pandemic reverses the ranking: there the penalized regressions beat the auto-regression by a fifth or more at six months, having been the weakest models everywhere else. Both samples were chosen after looking at the forecast record, and the second contains twenty two forecast errors, so neither result should be treated as settled.

With respect to distributional results, quantile regressions that include esti-

Horizon	Full sample	Excluding pandemic	Pandemic only
		$\varsigma = 0.90$	
$h = 1$	0.956	1.014	0.858*
$h = 3$	0.859**	0.923	0.632***
$h = 6$	0.892***	0.899***	0.855
$h = 12$	0.905*	0.964	0.597***
		$\varsigma = 0.75$	
$h = 1$	0.993	1.004	0.972
$h = 3$	0.992	0.979	1.030
$h = 6$	0.940*	0.978	0.789*
$h = 12$	0.924*	0.938	0.832
		$\varsigma = 0.50$	
$h = 1$	0.984	0.999	0.950
$h = 3$	0.925**	0.937*	0.892*
$h = 6$	0.924*	0.905**	0.972
$h = 12$	0.954	0.976	0.832**
		$\varsigma = 0.25$	
$h = 1$	0.940	1.020	0.741***
$h = 3$	0.953	0.995	0.834*
$h = 6$	0.988	1.000	0.955
$h = 12$	0.910**	0.899**	0.948
		$\varsigma = 0.10$	
$h = 1$	0.941	0.983	0.825***
$h = 3$	0.980	1.008	0.910
$h = 6$	0.906*	0.872**	1.000
$h = 12$	0.893**	0.890*	0.905***

Table 6: Relative quantile loss by evaluation sample

Note: Entries are the mean quantile loss of the quantile regression that includes the factors of T4, divided by that of the quantile autoregression, computed on the stated evaluation sample. The pandemic sample is March 2020 to December 2021. Stars as in Table 2.

mated factors cut quantile loss relative to a quantile auto-regression at all twenty combinations of quantiles and horizons, ten of them significantly. The reductions are larger than any recorded for the mean. They remain significant when the pandemic months are excluded, so they do not depend on that episode and appear only at horizons of six and twelve months, never at one month.

Our analysis has a few implications for inflation forecasting by EMDE central banks, as well as for India. For the projections the Reserve Bank of India publishes under Flexible Inflation Targeting, richer monthly models add nothing at the one month forecasting horizon; we show that their contribution at longer horizons depends on which sample is used. While the indicator panel contains usable information, the panel data says more about how far core inflation might stray from its central path rather than about where that path lies, and a forecast evaluation based on squared error alone would miss that.

References

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone (2019). "Vulnerable Growth". In: *American Economic Review* 109.4, pp. 1263–1289.

Anand, Rahul, Ding Ding, and Volodymyr Tulin (Sept. 2014). *Food Inflation in India: The Role for Monetary Policy*. IMF Working Paper 14/178. International Monetary Fund.

Armstrong, J. Scott and Fred Collopy (1992). "Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons". In: *International Journal of Forecasting* 8.1, pp. 69–80. DOI: 10.1016/0169-2070(92)90008-W.

Atkeson, Andrew and Lee E. Ohanian (2001). "Are Phillips Curves Useful for Forecasting Inflation?" In: *Federal Reserve Bank of Minneapolis Quarterly Review* 25.1, pp. 2–11.

Bai, Jushan and Serena Ng (2002). "Determining the Number of Factors in Approximate Factor Models". In: *Econometrica* 70.1, pp. 191–221.

— (2008). "Forecasting Economic Time Series Using Targeted Predictors". In: *Journal of Econometrics* 146.2, pp. 304–317.

Beck, Elliot and Michael Wolf (2026). "Forecasting inflation with the hedged random forest". In: *Empirical Economics* 70.1, p. 23. DOI: 10.1007/s00181-025-02879-x.

Behera, H. K., G. Wahi, and M. Kapur (2017). "Phillips Curve Relationship in India: Evidence from State-Level Analysis". In: *RBI Working Paper Series* WPS (DEPR): 08/2017.

Blagrave, Patrick and Weicheng Lian (Nov. 2020). *India's Inflation Process Before and after Flexible Inflation Targeting*. IMF Working Paper WP/20/251. IMF.

Breiman, Leo (2001). "Random Forests". In: *Machine Learning* 45.1, pp. 5–32.

Cashin, Paul, Kamiar Mohaddes, and Mehdi Raissi (2017). "Fair weather or foul? The macroeconomic effects of El Niño". In: *Journal of International Economics* 106, pp. 37–54. DOI: 10.1016/j.jinteco.2017.01.010.

CEIC Data (2026). *India Premium Database*. Subscription database. URL: https://www.ceicdata.com (visited on 03/18/2026).

Chamberlain, Gary and Michael Rothschild (1983). "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets". In: *Econometrica* 51.5, pp. 1281–1304.

Chaudhuri, Kausik and Saumitra Bhadhuri (2019). "Inflation Forecast: Just use the Disaggregate or Combine it with the Aggregate". In: *Journal of Quantitative Economics* 17. DOI: 10.1007/s40953-019-00155-1.

Clark, Todd E. and Michael W. McCracken (2001). "Tests of Equal Forecast Accuracy and Encompassing for Nested Models". In: *Journal of Econometrics* 105.1, pp. 85–110. DOI: 10.1016/S0304-4076(01)00071-9.

Clark, Todd E. and Kenneth D. West (2007). "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models". In: *Journal of Econometrics* 138.1, pp. 291–311.

Coibion, Olivier and Yuriy Gorodnichenko (2015). "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts". In: *American Economic Review* 105.8, pp. 2644–2678.

Das, Abhiman, Kajal Lahiri, and Yongchen Zhao (2019). "Inflation Expectations in India: Learning from Household Tendency Surveys". In: *International Journal of Forecasting* 35.3, pp. 980–993.

Diebold, Francis X. and Roberto S. Mariano (1995). "Comparing Predictive Accuracy". In: *Journal of Business and Economic Statistics* 13.3, pp. 253–263.

Dua, Pami and Upasna Gaur (2010). "Determination of Inflation in an Open Economy Phillips Curve Framework: The Case of Developed and Developing Asian Countries". In: *Macroeconomics and Finance in Emerging Market Economies* 3.1, pp. 33–51.

Dua, Pami and Deepika Goel (2023). "Forecasting India's Inflation in a Data-Rich Environment: A FAVAR Study". In: Springer. Chap. Chapter 9, pp. 225–259.

Duncan, Roberto and Enrique Martínez-García (2019). "New perspectives on forecasting inflation in emerging market economies: An empirical assessment". In: *International Journal of Forecasting* 35.3, pp. 1008–1031. DOI: 10.1016/j.ijforecast.2019.04.004.

Eichengreen, Barry, Poonam Gupta, and Rishabh Choudhary (2020). *Inflation Targeting in India: An Interim Assessment*. Policy Research Working Paper 9422. Washington, DC: World Bank. URL: http://hdl.handle.net/10986/34593.

Galí, Jordi and Mark Gertler (1999). "Inflation dynamics: A structural econometric analysis". In: *Journal of Monetary Economics* 44.2, pp. 195–222. DOI: 10.1016/S0304-3932(99)00023-9.

Ghate, Chetan and Faisal Ahmed (Jan. 2023). "On Modernizing Monetary Policy Frameworks in South Asia". In: *South Asia's Path to Resilient Growth*. Ed. by Rahul Anand and Ranil M. Salgado. IMF Publication. Chap. 12, pp. 283–300.

Giacomini, Raffaella and Halbert White (2006). "Tests of Conditional Predictive Ability". In: *Econometrica* 74.6, pp. 1545–1578.

Harvey, David, Stephen Leybourne, and Paul Newbold (1997). "Testing the Equality of Prediction Mean Squared Errors". In: *International Journal of Forecasting* 13.2, pp. 281–291. DOI: 10.1016/S0169-2070(96)00719-4.

Hubrich, Kirstin (2005). "Forecasting Euro Area Inflation: Does Aggregating Forecasts by HICP Component Improve Forecast Accuracy?" In: *International Journal of Forecasting* 21.1, pp. 119–136.

Hyndman, Rob J. and Yeasmin Khandakar (2008). "Automatic Time Series Forecasting: The forecast Package for R". In: *Journal of Statistical Software* 27.3, pp. 1–22.

Hyndman, Rob J. and Anne B. Koehler (2006). "Another Look at Measures of Forecast Accuracy". In: *International Journal of Forecasting* 22.4, pp. 679–688. DOI: 10.1016/j.ijforecast.2006.03.001.

India Macro Indicators (2026). *India Services Purchasing Managers' Index*. URL: http://indiamacroindicators.co.in.

Jose, J. et al. (2021). "Alternative Inflation Forecasting Models for India–What Performs Better in Practice?" In: *Reserve Bank of India Occasional Papers* 42.1.

Kapur, M. (2012). *Inflation Forecasting: Issues and Challenges in India*. Working Paper 01. Reserve Bank of India.

Koenker, Roger and Gilbert Bassett (1978). "Regression Quantiles". In: *Econometrica* 46.1, pp. 33–50.

Marcellino, Massimiliano, James H. Stock, and Mark W. Watson (2006). "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series". In: *Journal of Econometrics* 135.1-2, pp. 499–526.

Medeiros, Marcelo C. et al. (2021). "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods". In: *Journal of Business and Economic Statistics* 39.1, pp. 98–119.

Moneycontrol (2026). *India Manufacturing Purchasing Managers' Index*. URL: https://www.moneycontrol.com.

National Securities Depository Limited (2026). *Foreign Portfolio Investment Monitor*. URL: https://www.fpi.nsdl.co.in.

Ollech, Daniel and Karsten Webel (2020). *A Random Forest-Based Approach to Identifying the Most Informative Seasonality Tests*. Tech. rep. Deutsche Bundesbank Discussion Paper 55/2020.

Patra, M. D. and M. Kapur (2012). "A Monetary Policy Model for India". In: *Macroeconomics and Finance in Emerging Market Economies* 5.1, pp. 18–41.

RBI (Jan. 2014). *Report of the Expert Committee to Revise and Strengthen the Monetary Policy Framework*. Tech. rep. Mumbai: Reserve Bank of India.

— (2025). *Monetary Policy Report*. Tech. rep. Reserve Bank of India.

Reserve Bank of India (2026a). *Database on Indian Economy (DBIE)*. URL: https://data.rbi.org.in.

— (2026b). *Inflation Expectations Survey of Households*. URL: https://www.rbi.org.in.

Rossi, Barbara and Tatevik Sekhposyan (2010). "Have Economic Models' Forecasting Performance for US Output Growth and Inflation Changed over Time, and When?" In: *International Journal of Forecasting* 26.4, pp. 808–835. DOI: 10.1016/j.ijforecast.2009.08.004.

Sharma, Saurabh and Ipsita Padhi (2020). "An Alternative Measure of Economic Slack to Forecast Core Inflation". In: *Reserve Bank of India Occasional Papers* 41.2.

Stock, James H. and Mark W. Watson (2002). "Macroeconomic Forecasting Using Diffusion Indexes". In: *Journal of Business and Economic Statistics* 20.2, pp. 147–162.

— (2003). "Forecasting Output and Inflation: The Role of Asset Prices". In: *Journal of Economic Literature* 41.3, pp. 788–829. DOI: 10.1257/002205103322436197.

— (2007). "Why Has U.S. Inflation Become Harder to Forecast?" In: *Journal of Money, Credit and Banking* 39.s1, pp. 3–33.

Theil, Henri (1966). *Applied Economic Forecasting*. Amsterdam: North-Holland.

Tibshirani, Robert (1996). "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society, Series B* 58.1, pp. 267–288.

World Bank (2026). *World Bank Commodity Price Data (The Pink Sheet)*. URL: https://www.worldbank.org/en/research/commodity-markets.

Wright, Marvin N. and Andreas Ziegler (2017). "ranger: A fast implementation of random forests for high dimensional data in C++ and R". In: *Journal of Statistical Software* 77.1, pp. 1–17.

Zou, Hui and Trevor Hastie (2005). "Regularization and Variable Selection via the Elastic Net". In: *Journal of the Royal Statistical Society, Series B* 67.2, pp. 301–320.

Appendix

A The Indicator Panel

Indicator	Transformation	Seasonally adjusted	Coverage
BSE Sensex	$100\Delta \ln x_t$	Yes	0.99
CPI Index	$100\Delta \ln x_t$	Yes	0.99
Coal and Lignite Production: Coal	$100\Delta \ln x_t$	Yes	0.99
Consumer Confidence Survey: RBI: Current Situation I	Δx_t	No	0.98
Copper Price	$100\Delta \ln x_t$	No	0.99
Crude Oil Prices (Avg)	$100\Delta \ln x_t$	No	0.99
Electricity Generation: Monthly: Utilities	$100\Delta \ln x_t$	Yes	0.99
FAO Food Price Index	$100\Delta \ln x_t$	No	0.99
FII	Δx_t	No	0.99
FPI Investment: AUC	$100\Delta \ln x_t$	No	0.77
Foreign Investment Inflow: Portfolio: FIIs	Δx_t	No	0.99
Full Reservoir Levels	$100\Delta \ln x_t$	No	0.69
Gold Price	$100\Delta \ln x_t$	No	0.99
Government Securities Yield: 10 Years	Δx_t	No	0.99
Govt. Fiscal Deficit	Δx_t	No	0.99

Continued on next page

Table 7: The Indicator Panel

Indicator	Transformation	Seasonally adjusted	Coverage
IIP Capital	$100\Delta \ln x_t$	Yes	0.91
IIP Consumer Durables	$100\Delta \ln x_t$	Yes	0.91
IIP Consumer Non-Durables	$100\Delta \ln x_t$	Yes	0.91
IIP Intermediary	$100\Delta \ln x_t$	Yes	0.91
IIP Primary	$100\Delta \ln x_t$	Yes	0.91
Import Volume Index	$100\Delta \ln x_t$	No	0.69
Imports	$100\Delta \ln x_t$	Yes	0.99
Lead Price	$100\Delta \ln x_t$	No	0.99
Manufacturing PMI: Headline: sa: India	Δx_t	No	0.83
Money Supply: M3	$100\Delta \ln x_t$	Yes	0.99
Monthly FPI Net Investments	Δx_t	No	0.99
Motor Vehicle Sales: Commercial Vehicle (CV)	$100\Delta \ln x_t$	No	0.94
Motor Vehicle Sales: DM: Passenger Vehicle (PV)	$100\Delta \ln x_t$	Yes	0.99
Motor Vehicle Sales: DM: Two Wheelers (TW)	$100\Delta \ln x_t$	Yes	0.99
Motor Vehicle Sales: Passenger Vehicle (PV)	$100\Delta \ln x_t$	No	0.74
Nickel Price	$100\Delta \ln x_t$	No	0.99
Real Effective Exchange Rate Index	Δx_t	No	0.99
Services PMI: Headline: sa: India	Δx_t	No	0.92
Silver Price	$100\Delta \ln x_t$	No	0.99

Continued on next page

Table 7: The Indicator Panel

Indicator	Transformation	Seasonally adjusted	Coverage
Tin Price	$100\Delta \ln x_t$	No	0.99
WPI: Fuel & Power: YoY	Δx_t	No	0.84
WPI: Manufactured Products (Mfg): YoY	Δx_t	No	0.84
WPI: Primary Articles: YoY	Δx_t	No	0.84
Weighted Average Call Money Rate	Δx_t	No	0.98
Zinc Price	$100\Delta \ln x_t$	No	0.99

Note: Transformations are applied as described in Section 2. Seasonal adjustment is applied where the combined test of Ollech and Webel (2020) rejects the null of no seasonality. Coverage is the share of months from January 2011 to December 2025 for which the indicator is observed. Sources: CEIC India Premium Database (CEIC Data 2026); World Bank Commodity Price Data (The Pink Sheet) (World Bank 2026); RBI Database on Indian Economy for the index of industrial production (Reserve Bank of India 2026a); NSDL for foreign portfolio investment (National Securities Depository Limited 2026); Moneycontrol for the manufacturing PMI (Moneycontrol 2026); India Macro Indicators for the services PMI (India Macro Indicators 2026).

Table 7: The Indicator Panel

B Forecast Errors for Month on Month Inflation

Table 8 reports the ratio in (6) computed for π_{t+h}^{MOM} rather than for π_{t+h}^{YOY}. As Section 3.3 notes, forecasts of year on year inflation at short horizons share observed price changes with the outcome, which compresses differences across models. No such compression arises here, and the spread across models is correspondingly wider, particularly at $h = 12$.

Nineteen of the sixty-eight entries fall below one, eight of them at $h = 12$, where the lowest ratio is 0.920. This does not overturn the conclusion of Section 5. The month on month forecasts are the inputs from which the year on year forecasts are constructed, so the two sets of errors are not independent, and a ratio below one on a single month's growth need not survive compounding over twelve months. The models that fall below one at $h = 12$ here are those whose twelfth month forecast happens to be least inaccurate, not models that track core inflation over the year. We report no test of equal forecast accuracy for this comparison and read the table as descriptive.

C Additional Samples

Table 9 reports the sample ending February 2020, which contains between 26 and 37 forecast errors depending on the horizon, and Table 10 the sample beginning January 2022, which contains 48 at every horizon. The second is the sample discussed in Section 6, reported here at all four horizons rather than at three and six.

Two results in the pre-pandemic sample are worth noting. T4 estimated with one factor has ratios of 0.918 at $h = 3$ and 0.937 at $h = 12$, both significant at the 10 % level, which is the same model and the same direction as the result reported for the sample excluding the pandemic. The random forest has a ratio of 0.821 at $h = 12$, significant at the 1 % level, and it is the only occurrence in the paper of a machine learning method beating the auto-regression outside the pandemic months. It rests on 26 forecast errors at a twelve month horizon, where the errors overlap heavily, and the same model has a ratio of 1.070 at $h = 12$ in the sample beginning January 2022, so we do not read it as evidence that the method forecasts core inflation well.

Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$
A. T0 and T1: univariate models				
T0: AR(p)	1.000	1.000	1.000	1.000
T1: ARMA(p, q)	0.997	0.989	1.062	1.112
B. T3: diffusion index, all indicators				
$r = 1$	1.065	1.475	1.107	0.920
$r = 2$	1.116	1.498	1.153	0.928
r by Bai and Ng	1.188	1.404	1.150	1.025
C. T4: diffusion index, subset of indicators				
$r = 1$	1.025	1.034	1.113	0.938
$r = 2$	1.079	0.999	1.121	0.949
r by Bai and Ng	1.099	1.088	1.126	1.162
D. T6: household inflation expectations				
T0 + expectations	0.996	1.063	1.131	1.185
T3 + expectations	1.108	1.250	1.194	1.380
T4 + expectations	1.132	1.147	1.209	1.370
E. T7: machine learning methods				
Ridge	1.163	1.001	1.053	0.988
Lasso	1.202	1.003	1.007	0.990
Elastic net	1.171	1.004	1.004	0.990
Random forest	1.043	1.020	1.063	0.969
F. T8: climate variables				
T0 + ENSO	1.043	1.138	1.022	1.036
T0 + rainfall	1.005	0.996	1.011	1.002
T0 + ENSO and rainfall	1.053	1.152	1.047	1.036

Table 8: Relative RMSE computed for month on month inflation

Note: Entries are the root mean squared forecast error of the model divided by that of T0, the autoregression, computed for month on month inflation at each horizon. Shaded grey denotes the baseline, which equals one by construction. Shaded green denotes a ratio below one. No test of equal forecast accuracy is reported for this comparison. T2 is reported separately in Table 3.

Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$
A. T0 and T1: univariate models				
T0: AR(p)	1.000	1.000	1.000	1.000
T1: ARMA(p, q)	0.990	1.003	1.121	1.184
B. T3: diffusion index, all indicators				
$r = 1$	1.006	0.998	1.025	1.014
$r = 2$	1.029	1.041	1.034	1.018
r by Bai and Ng	1.006	0.998	1.025	1.014
C. T4: diffusion index, subset of indicators				
$r = 1$	1.068	0.918*	0.936	0.937*
$r = 2$	1.174	1.000	0.982	0.959
r by Bai and Ng	1.254	1.023	1.008	0.958
D. T6: household inflation expectations				
T0 + expectations	0.983	1.030	1.123	1.324
T3 + expectations	0.991	1.022	1.133	1.351
T4 + expectations	1.312	1.139	1.233	1.328
E. T7: machine learning methods				
Ridge	1.522	1.322	1.400	1.174
Lasso	1.643	1.225	1.266	0.987
Elastic net	1.537	1.246	1.277	1.011
Random forest	1.089	0.997	0.962	0.821***
F. T8: climate variables				
T0 + ENSO	1.085	1.334	1.431	1.258
T0 + rainfall	1.032	1.008	1.045	0.988
T0 + ENSO and rainfall	1.129	1.328	1.443	1.235

Table 9: Relative RMSE of h-month ahead forecasts, sample ending February 2020

Note: Entries are the root mean squared forecast error of the model divided by that of T0, the autoregression, computed for year on year inflation at each horizon. Shaded grey denotes the baseline, which equals one by construction. Shaded green denotes a ratio below one. *, ** and *** denote rejection of the null of equal forecast accuracy in favour of the model, at the ten, five and one percent level, by the test of Diebold and Mariano (1995). T2 is reported separately in Table 3.

Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$
A. T0 and T1: univariate models				
T0: AR(p)	1.000	1.000	1.000	1.000
T1: ARMA(p, q)	0.973	0.923	0.914*	1.052
B. T3: diffusion index, all indicators				
$r = 1$	0.995	0.988	0.997	1.003
$r = 2$	0.998	1.003	1.013	1.009
r by Bai and Ng	1.028	1.025	1.033	1.049
C. T4: diffusion index, subset of indicators				
$r = 1$	0.998	0.971	0.966**	1.005
$r = 2$	1.000	0.963	0.978	1.009
r by Bai and Ng	1.033	0.989	0.986	1.025
D. T6: household inflation expectations				
T0 + expectations	0.995	1.014	0.997	1.171
T3 + expectations	1.029	1.058	1.039	1.227
T4 + expectations	0.999	1.024	0.988	1.183
E. T7: machine learning methods				
Ridge	1.085	1.136	1.141	1.166
Lasso	1.040	1.115	1.143	1.171
Elastic net	1.069	1.121	1.140	1.168
Random forest	1.089	1.100	1.025	1.070
F. T8: climate variables				
T0 + ENSO	1.009	1.086	1.135	1.130
T0 + rainfall	1.012	1.049	1.027	1.027
T0 + ENSO and rainfall	1.023	1.158	1.194	1.195

Table 10: Relative RMSE of h-month ahead forecasts, sample beginning January 2022

Note: Entries are the root mean squared forecast error of the model divided by that of T0, the autoregression, computed for year on year inflation at each horizon. Shaded grey denotes the baseline, which equals one by construction. Shaded green denotes a ratio below one. *, ** and *** denote rejection of the null of equal forecast accuracy in favour of the model, at the ten, five and one percent level, by the test of Diebold and Mariano (1995). T2 is reported separately in Table 3.

D An Upper Bound on the Value of Indicator Selection

T4 selects indicators using the record available at τ. A forecaster who knew in advance which indicators would prove useful would select differently. To bound what that knowledge is worth, we replace $S(\tau)$ with

$$S^* = \{i : R_i(T) < 1\}, \tag{23}$$

where T is the final origin, so that R_i is computed once over the whole comparison sample. Factors are estimated from S^* at every τ. The forecast made at τ therefore uses a set of indicators chosen with knowledge of how those indicators performed after τ, and is not a forecast that could have been made at the time. We report it for two reasons.

The first is that it bounds the gain available from better selection among these forty indicators. Table 11 reports ratios below one at nine of the twelve combinations of version and horizon, and three are significant at the 1 % level: 0.969 and 0.969 at $h = 3$ and 0.971 at $h = 6$. Two more are significant at the 10 % level. Perfect foresight about which indicators would work is therefore worth about 3% in root mean squared error at the intermediate horizons, and nothing at $h = 12$.

The second is that it establishes that the comparison can detect a gain of that size. The specification uses the same estimator, the same horizons, the same test and the same number of forecast errors as the models reported in Section 5. That those models do not improve upon T0 over the full sample is therefore not a consequence of too few forecast errors. It does not follow that no model could improve upon T0, only that better selection among these indicators is worth about 3% over this sample and that a forecaster at τ does not capture it.

Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$
$r = 1$	0.981*	0.969***	0.971***	0.999
$r = 2$	0.993	0.969***	0.981*	1.005
r by Bai and Ng	1.000	0.986	0.991	1.018

Table 11: Relative RMSE of T4 with the subset selected on the full sample

Note: Entries are the root mean squared forecast error of the model divided by that of T0, for year on year inflation, over the full comparison sample. Shaded green denotes a ratio below one. The subset S^* is selected once using the whole comparison sample, so these are not forecasts that could have been made at the time and the entries are favourable by construction. Stars as in Table 2.