(ISSN: 0973-7049(P))

The Bioscan AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES 21(4): S.I (4),1052-1065, 2026 www.thebioscan.com

Deconstructing India's Headline CPI: A Comparative Analysis of Climate, Macroeconomic, and Commodity Shocks

Aman Singh[1], Monil Patel[1], Sitanshu Nisar[1], Kshamta Mathur[1], Jaykrishna Joshi[1]

[1] Department of Data Science, SVKM's NMIMS Mukesh Patel School of Technology Management & Engineering, Mumbai, India

DOI: https://doi.org/10.63001/tbs.2026.v21.i04.S.I(4).pp1052-1065

Keywords:
Commodity prices, consumer price index (CPI), feature attribution, Indian economy, inflation forecasting, machine learning, SHAP, stacking ensemble, time-series econometrics.

Article Info

Received:14.10.2026

Accepted: 02.11.2026

Published: 12.11.2026

Abstract

Headline Consumer Price Index (CPI) inflation in India responds to concurrent domestic agricultural supply disruptions, global commodity price movements, and monetary conditions. This paper compares two forecasting-and-attribution paradigms on the same monthly data and the same chronological holdout. The machine learning (ML) paradigm comprises ElasticNet, XGBoost, Support Vector Regression (SVR), and a Stacking Regressor; the time-series (TS) econometric paradigm comprises a regularized ARDL(p, q), a block-structured Dynamic Factor Model, a VARX, and a Bates-Granger optimal forecast combination. Each model is estimated under four feature sets (Full, Climate-only, Macroeconomic-only, Commodity-only) with two CPI persistence controls. The Stacking Regressor under the Full configuration delivers the lowest holdout error (RMSE = 0.634, R^2 = 0.858), 21% below the best Full-configuration TS model, a gap that survives Benjamini-Hochberg-adjusted Diebold-Mariano testing. Grouped SHAP, Group Permutation Feature Importance (GPFI), and expanding-window stability diagnostics identify Commodity, in particular primary food prices, as the Core Driver of CPI in the ML paradigm and as the category with the largest permutation loss in the TS paradigm. Macroeconomic variables are a Redundant Driver once Commodity information is present, and Climate variables are Mixed/Inconclusive. The performance advantage of the ensemble over dynamic time-series specifications reflects sparse selection among collinear wholesale food indicators and the parameter variance costs associated with distributed lag expansion, though these mechanisms remain unisolated by controlled ablation.

INTRODUCTION AND RELATED LITERATURE

Headline Consumer Price Index (CPI) inflation in India exhibits substantial, recurring volatility driven largely by domestic food price shocks. The CPI basket is dominated by food and beverages, which carry a higher weight than other commodity groups and whose price changes explain a much higher share of the CPI's overall volatility [1]. Inflationary episodes are influenced by the cross-over shocks of insufficient monsoons, global crude and fertilizer price hikes, and domestic monetary expansion [2]. Isolating these climate, supply, and macroeconomic factors is a key empirical challenge for an economy that is exposed to imports and dependent on the monsoon [2], [3].

A. The Evolution of Inflation Forecasting

Vector-autoregressive (VAR) and structural time-series models have traditionally been used for empirical macroeconomics [4]. The autoregressive distributed lag (ARDL) model and the autoregressive integrated moving average (ARIMAX) model are still widely used in country-level inflation analysis, since neither imposes strong exogeneity conditions, allowing the establishment of short- and long-run relationships [5]. When applied to India, ARDL frameworks capture significant long-run effects of crude oil, temperature, rainfall, and money supply on food inflation [2]. Time-series architectures are well suited to representing linear dynamics and cointegrating structures, but they must be modified in complex ways to incorporate threshold effects or regime shifts [2]. Two extensions address the dimensionality problem that arises when many correlated indicators are available: dynamic factor models, which summarize large panels of predictors in a small number of common factors [6], and forecast combination, which weights competing models by their historical error variance [7]. In addition, traditional econometric coefficients are well interpretable in direction, but these magnitudes are not necessarily directly interpretable as out-of-sample predictive importance when features are severely collinear [8].

A complementary approach is machine learning (ML), which can estimate non-linear response surfaces [8] and use penalized regularization to suppress the impact of highly correlated features [9]. Kernel methods such as support vector regression offer a further route to non-linearity with controlled complexity [10], and stacked generalization combines heterogeneous learners through a meta-learner trained on held-out predictions [11]. In the context of Indian inflation forecasting, ML ensembles have shown promising results compared with traditional ARIMA models, especially in post-2020 samples [12]. Tree-based algorithms have been used to describe domestic agricultural procurement dynamics [13], and LASSO and random forest have proven to be strong nowcasters

[14]. Explainable ML extensions add non-linear expectation effects to traditional linear Phillips-curve models [15].

B. Drivers of Indian Inflation

Prior studies classify the drivers of Indian inflation into three broad categories. First, the relationship between monetary and macroeconomic conditions (broad money-supply growth) and domestic food inflation is similar in strength and direction to the magnitude of global crude-oil shocks [2] in the long run.

Second, disaggregated commodity studies show that domestic supply constraints are an important source of commodity volatility [1]. Because India's CPI is not extremely disaggregated, researchers instead use Wholesale Price Index (WPI) data [16]. Gradient-boosted ensembles suggest that these domestic procurement factors dominate global food prices and trade indicators [13], motivating the use of granular domestic commodity features.

Third, the climate-inflation literature has shifted from country-level aggregates toward more spatial metrics. High-latitude fixed-effects models show a long-lasting effect of temperature shocks on food inflation [3]. The long-term effect of temperature and rainfall anomalies is likewise confirmed by ARDL estimations for India [2]. Crop-price models using gridded NASA precipitation and temperature data show a pronounced meteorological influence on price volatility [17].

C. Research Gap

Notwithstanding recent developments in the use of machine learning for forecasting macroeconomic variables, current research suffers from a number of limitations. The first limitation is that the assessment of predictive ability tends to focus on aggregate forecast error without associating the predictive ability to any of the driver types. The second limitation is that the explanation of the drivers tends to be done through feature importance score assessments without structural categorization of such features. Finally, spatial climate indexes and disaggregated commodity prices tend not to be used for benchmarking with respect to mutually exclusive covariate blocks on the same holdout sample in order to determine transmission channels.

I. METHODOLOGY

The empirical framework bifurcates into parallel machine-learning and time-series estimation workflows following data ingestion, reconciliation, and stationarity screening, reuniting at the out-of-sample evaluation and diagnostic triangulation stages (Fig. 1). Summary is as follws:-

- Gather Data: Collects 24 years of monthly weather, inflation, and market numbers.
- Clean Data: Fixes missing information and removes misleading seasonal spikes.
- Group Factors: Sorts information into distinct categories like climate and economics.
- Split Timeline: Saves the most recent 20% of data for final testing.
- Train AI: Teaches advanced machine learning models to find hidden patterns.
- Train Statistics: Uses traditional time-series math to track historical trends.
- Test Accuracy: Competes 32 different models against the saved test data.
- Find Drivers: Calculates exactly which specific factors had the most influence.
- Cross-Check: Compares AI results with statistical results to ensure agreement.
- Deliver Insights: Groups economic forces into clear categories for final decision-making.



Fig. 1. Methodology flowchart.

A. Data Collection

Monthly headline CPI inflation (2001–2025) is compiled by the Ministry of Statistics and Programme Implementation (MoSPI) and cross-checked with the IMF STA database. Exogenous predictors include data from the India Meteorological Department (IMD) and NASA POWER (climate data), the Reserve Bank of India, MoSPI, and the IMF for US CPI (macroeconomic data), and the World Bank Pink Sheet and DPIIT for commodity data. This information is summarized in Table I.

TABLE I
DATA SOURCES AND REPRESENTATIVE VARIABLES BY CATEGORY

Category	Representative variables	Primary source(s)
Climate	Rainfall (total & intensity), mean temperature, relative humidity, soil-wetness index, wind speed, solar irradiance	IMD (2000–2024), NASA POWER (2025 onwards)
Macroeconomic	Repo rate (YoY diff), M_3 growth, IIP, US CPI (spillover control)	RBI DBIE, MOSPI, IMF
Commodity	Brent crude, palm oil, fertilizer index, WPI food/commodity sub-indices	World Bank Pink Sheet, DPIIT/WPI

B. Data Cleaning and Preprocessing

The raw series are base-year reconciled, and missing values are interpolated to provide uniform monthly frequencies. All variables are converted to percentage change (YoY) and seasonal anomalies (month-of-year z-scores), giving 45 substantive predictors (14 climate, 4 macroeconomic, 25 commodity, and 2 persistence variables). Predictors are standardized to zero mean and unit variance where necessary for algorithm-specific penalty functions and/or distances (e.g., ElasticNet, SVR) and for the factor extraction and ridge steps of the TS models:

$$z = \frac{x - \mu}{\sigma} \qquad (1)$$

Stationarity of every predictor and of the target was screened on the training sample with the Augmented Dickey-Fuller (ADF) test [18], whose null is a unit root, and the KPSS test [19], whose null is stationarity, both at the 5% level. The tests are diagnostic only: all models are estimated on the YoY and anomaly transformations described above rather than on further-differenced series, so that the ML and TS paradigms see identical inputs. Several series are not clearly stationary: broad money growth (M_3), the CPI target, and its lag-1 and lag-12 terms fail to reject a unit root in the ADF test (target ADF p = 0.58), and the KPSS test rejects stationarity for four of the seven climate anomaly series. This residual persistence is treated as a limitation of the TS estimates (Section IV-E).

C. Feature Grouping

The predictors are sharply segmented into three substantive categories that are entirely mutually exclusive: Climate (precipitation, temperature, soil wetness, humidity, wind, sun radiation), Macroeconomic (M_3, IIP, repo rate change, US CPI), and Commodities (granular domestic food and commodity index in WPI, Brent oil, palm oil, fertilizer). The CPI lag-1 and CPI lag-12 act as the persistent predictors for all models. The Lag-1 predicts the month-to-month persistence because the YoY inflation is a 12-month difference in which only 1 month of prices is changed each time from period to period. Lag-12 acts as a base-effect and seasonal control: it anchors each forecast to the inflation rate prevailing one year earlier, and hence to the base period that rolls out of the YoY window, and it carries the annual cycle in food prices. Rolling means of historical inflation are omitted to prevent rich autoregressive terms from masking exogenous shock transmission. Persistence controls are excluded from category-share calculations to ensure that attribution scores reflect only the exogenous Climate, Macroeconomic, and Commodity feature blocks. The same four configurations and the same predictor lists are used for every ML and TS model. How these two lags were chosen is discussed as a limitation in Section IV-E.

D. Time-Based Train–Test Split

An 80:20 split is applied to the 288 temporal observations to avoid temporal data leakage. The first 230 months are used for training and validation, and the last 58 months act as an untouched out-of-sample holdout test set. All forecasts are one-step-ahead, using observed persistence lags, and all eight model families are scored on exactly the same 58 months.

E. Machine Learning Models and Hyperparameter Tuning

This forecasting task relies on supervised learning, which predicts monthly YoY CPI inflation from two CPI persistence variables and variables from the Climate, Macroeconomic, and Commodity domains. Four feature configurations are used:

- Persistence + Climate
- Persistence + Macroeconomic
- Persistence + Commodity
- Persistence + Climate + Macroeconomic + Commodity (Full)

ElasticNet: a linear model combining L1 (Lasso) and L2 (Ridge) regularization for highly correlated predictors [9]; the L1 component performs embedded variable selection.

XGBoost: a gradient-boosted additive ensemble with a convex objective function equal to the sum of a convex loss (l) and a penalty term (Ω) for tree complexity [8].

Support Vector Regression (SVR): a kernel-based algorithm (radial basis function or linear kernel) that fits a regression hyper-tube within an epsilon-insensitive error margin [10].

Stacking Regressor: a stacked ensemble [11], built with chronological (time-series-aware) validation, whose base learners are an ElasticNet, an XGBoost model, and a LightGBM model [20]. The base learners are first fit on the earlier 75% of the training months and predict the later 25%; a non-negative Ridge meta-learner is fit on those held-out predictions, so that the combination weights are learned from genuinely out-of-sample, chronologically later data. The base learners are then refit on all 230 training months. The base-

learner hyperparameters are fixed shallow settings, so the stack carries no additional tuning burden.

Hyperparameters of ElasticNet, XGBoost, and SVR are optimized on the 230-month training set strictly using RandomizedSearchCV with chronological five-split TimeSeriesSplit and a root-mean-squared-error criterion. After hyperparameter optimization, each of the sixteen ML model/configuration combinations is refit on the full 230 observations to finalize the predictive architectures prior to out-of-sample evaluation.

F. Time-Series Econometric Models

The TS paradigm consists of four models; each estimated on the same four configurations (sixteen further model/configuration combinations) and the same 230-month training sample:

Regularized ARDL (p, q): an autoregressive distributed lag specification [5] in which the CPI persistence terms (p) are augmented with q = 2 monthly lags of every exogenous predictor, so that multi-month pass-through from commodity, macroeconomic, and climate shocks is represented explicitly. Because the lag expansion triples the number of regressors, the equation is estimated under ElasticNet shrinkage.

Block-Structured Dynamic Factor Model (Block-DFM): predictors are partitioned into the Macroeconomic, Commodity, and Climate blocks; two principal-component factors are extracted from each block [6] and the CPI equation is regressed on the CPI persistence terms and the block factors under ridge shrinkage. This is a factor-augmented regression in which the common factors, rather than individual predictors, carry the block information.

VARX: the CPI equation of a vector autoregression with exogenous covariates [4], containing the CPI persistence lags and the contemporaneous exogenous predictors, estimated with ridge shrinkage.

Bates-Granger optimal combination: the three models above are fit on the earlier 75% of the training months, and the combination weights are chosen to minimize the mean squared error of the combined forecast on the later 25% [7]:

$$\hat{y}_t = \sum_k w_k \hat{y}_{k,t}$$
$$w = \operatorname{argmin} \sum_t \left(y_t - \sum_k w_k \hat{y}_{k,t}\right)^2 \quad (2)$$
$$, \; w_k \geq 0, \; \sum_k w_k = 1$$

The component models are then refit on the full training sample.

G. Forecasting and Evaluation

Out-of-sample accuracy is assessed on the 58-month holdout test set using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2), computed in their standard form over the 58 forecasts.

Pairwise statistical significance between configurations, and between models, is evaluated using the Harvey-Leybourne-Newbold small-sample-corrected Diebold-Mariano statistic [21], [22] computed on squared forecast errors:

$$\mathrm{DM}^* = \mathrm{DM} \times \sqrt{\frac{T + 1 - 2h + \frac{h(h-1)}{T}}{T}} \quad (3)$$

To control the Type I error rate, *p*-values are adjusted using the Benjamini-Hochberg False Discovery Rate (FDR) procedure [23], with significance established at $Q = 0.05$. Within each paradigm the family consists of the 24 pairwise configuration comparisons (six per model). A third, cross-paradigm family compares the leading ML models with each TS model on the Full configuration, with the best TS model (Bates-Granger, Macro-only), and the leading ML models with one another (13 comparisons).

H. Attribution Analysis

A multi-diagnostic framework is used to quantify relative predictive importance on the holdout set. For the ML paradigm, native feature-level Shapley values are computed with algorithm-specific implementations: TreeSHAP for XGBoost, LinearSHAP for ElasticNet, and KernelSHAP for SVR and the Stacking Regressor [24], [25]. These absolute SHAP magnitudes are aggregated without persistence controls to give the normalized grouped SHAP share for each substantive category (G) [26]:

$$\mathrm{Share}^G_{\mathrm{SHAP}}(\%) = \frac{\mathrm{GroupedSHAP}^G}{\sum_{G'} \mathrm{GroupedSHAP}^{G'}} \times 100 \quad (4)$$

For the TS paradigm, native feature importance is calculated by summing the absolute standardized regression coefficients and distributed lag weights across predictors within each block for ARDL and VARX, and absolute block factor loadings for the Block-DFM, normalized across substantive categories to sum to 100% (Table VIII). Because linear econometric coefficients can be inflated by residual unit-root persistence in series such as M_3 and target inflation lags, native TS shares are paired with model-agnostic Group Permutation Feature Importance (GPFI). GPFI evaluates out-of-sample predictive loss following 100 joint permutations of each category's variables on the holdout set, with *p*-values adjusted via the Benjamini–Hochberg procedure across all within-paradigm tests.

I. Stability Diagnostics

The optimized ML estimators are refit within five expanding chronological training/validation windows to assess attribution robustness. Temporal sensitivity of the identified drivers is assessed via two measures: categorical rank consistency (Kendall's τ) and group-share dispersion (Coefficient of Variation, CV^G):

$$\mathrm{CV}^G = \frac{\sigma^G}{\mu^G} \quad (5)$$

J. Diagnostic Triangulation and Cross-Paradigm Synthesis

The results of native attribution dominance, GPFI error degradation, and category-only accuracy are combined within each paradigm. A category is classified as a Core Driver if it ranks first in native attribution, ranks in the top two on category-only accuracy, and has a positive mean GPFI loss that is significant in at least two models. A category that is accurate on

its own but whose GPFI evidence is weak (non-positive mean loss, or significance in fewer than two models) is classified as Redundant (the TS paradigm labels this Redundant / Linear Baseline). A category with a low native rank but strong GPFI evidence is a Non-Linear Synergy driver (ML); all remaining cases are Mixed/Inconclusive. Both paradigms apply identical diagnostic triangulation criteria across native attribution rank, category-only accuracy rank, and GPFI error degradation. In the TS paradigm, native ranks are derived from aggregate standardized elasticity shares. Finally, the two paradigms are compared on GPFI and category-only accuracy, and the accuracy results are used to discuss possible reasons for the difference between the paradigms.

I. Results

A. Machine Learning: Out-of-Sample Accuracy

Table II displays the holdout RMSE of all sixteen ML model/configuration combinations (upper half) on the 58-observation test set that was not used for training, together with MAE, MAPE, and R^2 for the Full configuration; the TS models in the lower half are discussed in Section III-B. The Stacking Regressor under the Full configuration yields the best overall forecast (RMSE = 0.634, MAE = 0.467, MAPE = 13.61%, R^2 = 0.858), 5.0% lower in RMSE than the best single-model alternative, ElasticNet Full (RMSE = 0.668, R^2 = 0.843). The Stacking Regressor with Commodity-only features is a close second (0.653). The optimal configuration is Full for the Stacking Regressor and ElasticNet, Commodity-only for XGBoost, and Macroeconomic-only for SVR. There is no single best configuration across every model. Climate-only is the weakest configuration for every model.

TABLE II
HOLDOUT RMSE BY MODEL AND FEATURE CONFIGURATION

Model	Full	Climate	Macro	Commodity	MAE	MAPE (%)	R^2
Machine learning							
Stacking Regressor	0.634*	0.806	0.799	0.653	0.467	13.61	0.858
ElasticNet	0.668*	1.085	0.801	0.673	0.537	15.42	0.843
XGBoost	0.816	1.012	0.927	0.786*	0.650	30.09	0.765
SVR	0.906	1.143	0.791*	0.929	0.720	18.70	0.710
Time series							
Bates-Granger	0.802	0.852	0.785*	0.820	0.636	22.54	0.773
Block-DFM	0.840	0.852	0.810*	0.820	0.661	24.71	0.751
Regularized ARDL	0.813*	0.993	0.865	0.880	0.646	21.93	0.767
VARX	0.922	1.179	0.792*	0.902	0.766	23.30	0.700

58-observation test set; * marks each model's lowest RMSE. MAE, MAPE, and R^2 refer to the Full configuration.

The Harvey-Leybourne-Newbold-corrected, Benjamini-Hochberg-adjusted Diebold-Mariano tests (24 pairwise comparisons) show that not every point-estimate gap in Table II is statistically distinguishable at the 58-observation holdout size. Only 4 of 24 comparisons survive FDR control. For ElasticNet, Climate-only is significantly worse than Full, Macro-only, and Commodity-only, whereas Full vs. Commodity-only, the two closest configurations, is not significant. For SVR, only Macro-only vs. Climate-only is significant. For the Stacking Regressor and XGBoost, no configuration difference survives: the Stacking Regressor's Full-versus-Climate-only and Full-versus-Macro-only comparisons have unadjusted p-values of 0.029 and 0.028 but BH-adjusted p-values of 0.086. Because the small-sample Harvey-Leybourne-Newbold correction enforces conservative critical thresholds at n = 58, several substantial point-estimate margins fail to achieve statistical significance under false-discovery-rate control.

(ISSN: 0973-7049(P))

The Bioscan AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES 21(4): S.I (4),1052-1065, 2026 www.thebioscan.com

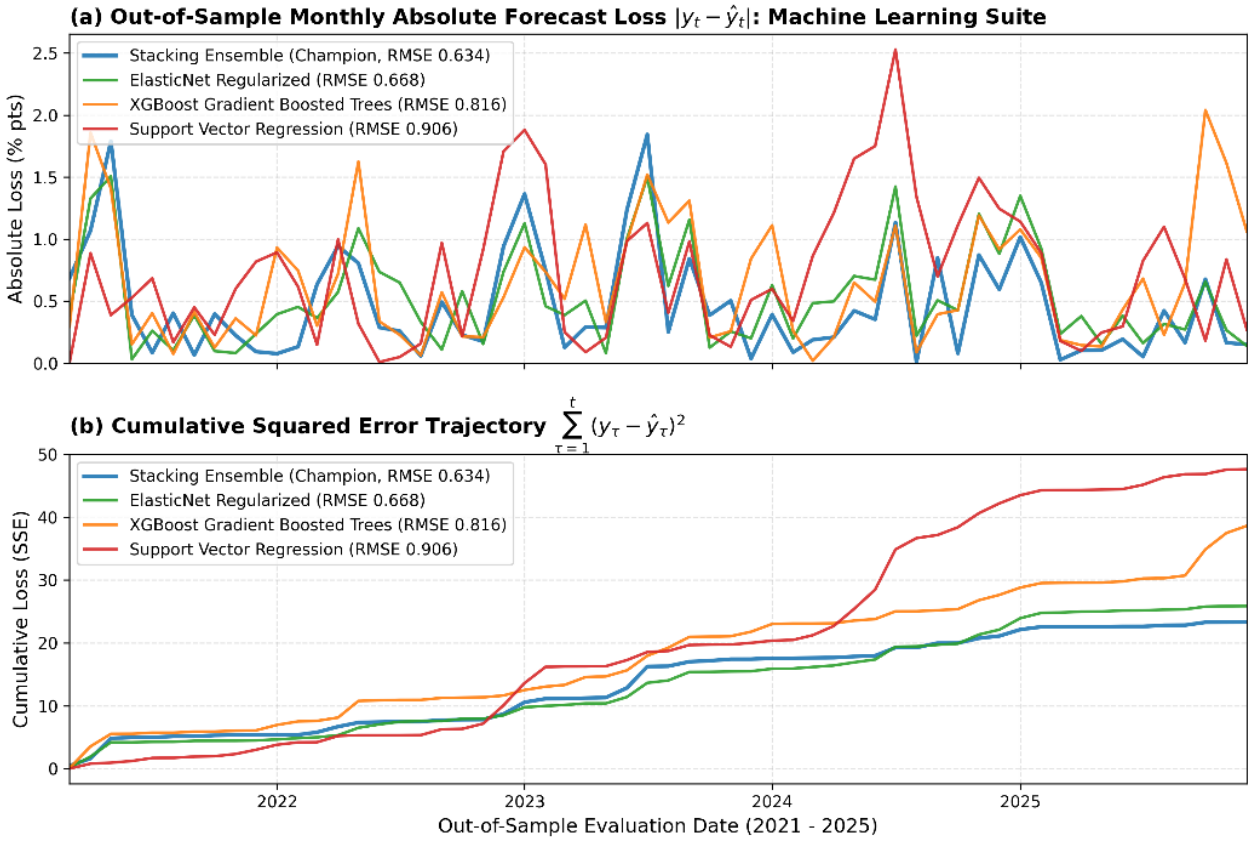


Fig. 2. (a) ML monthly absolute error in percentage points. (b) ML cumulative sum of squared errors (SSE) trajectory.

As shown in Fig. 2(a), monthly errors for the Stacking Regressor and ElasticNet stay under 1.0 percentage point across most of the holdout, spiking only briefly during the mid-2023 food price shock. SVR, by comparison, frequently misses by over 2.0 points. The cumulative squared error plot (Fig. 2(b)) highlights this gap: the Stack (SSE = 23.3) and ElasticNet (SSE = 25.8) maintain flat, steady trajectories, while SVR (47.6) and XGBoost (38.6) climb steeply after 2023.

B. Time Series: Out-of-Sample Accuracy

The lower half of Table II reports the sixteen TS model/configuration combinations. The Bates-Granger combination is the TS champion: it attains the lowest RMSE in the Macroeconomic-only configuration (0.785, $R^2 = 0.783$) and in the Full configuration (0.802), ahead of the regularized ARDL (0.813), the Block-DFM (0.840), and the VARX (0.922). For the regularized ARDL, Full is the best configuration; for the Block-DFM and the VARX, the Macroeconomic-only configuration is best (0.810 and 0.792). Adding the Climate and Commodity blocks to the Macroeconomic block increases the VARX holdout RMSE from 0.792 to 0.922, showing that expanding the exogenous feature set degraded out-of-sample accuracy in this specification. The Bates-Granger rows for the Climate-only and Commodity-only configurations coincide with the Block-DFM rows because the optimizer places its entire weight on the Block-DFM in those two configurations (weights of 0, 1, and 0 on the ARDL, DFM, and VARX); in the Full configuration the weights are 0.00, 0.91, and 0.09, and in the Macroeconomic-only configuration 0.04, 0.39, and 0.57.

None of the 24 within-model configuration comparisons is significant after FDR control; the smallest adjusted p-value is 0.079 (VARX, Climate-only vs. Macro-only). The TS models are therefore statistically indifferent to the choice of configuration at this sample size, even where point estimates differ by more than 0.3 RMSE units.

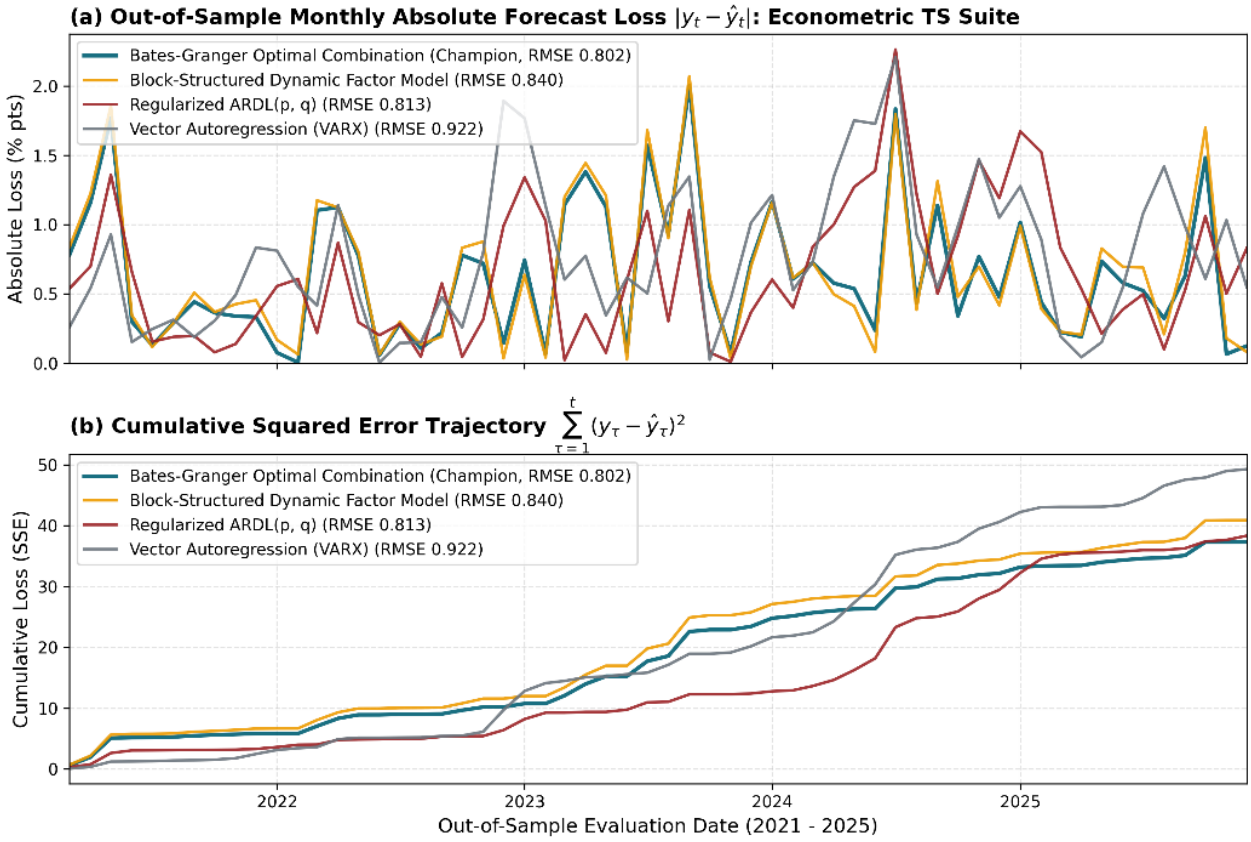


Fig. 3. (a) TS monthly absolute error in percentage points. (b) TS cumulative sum of squared errors (SSE) trajectory.

Figure 3 shows how error accumulates in econometric models. Bates-Granger results in the smallest cumulative error (SSE = 37.3) and follows the Block-DFM model closely due to the high weight placed on factor forecasts by the optimizer. The VARX model has trouble with abrupt spikes in monthly error greater than 1.5 points. ARDL is close for the first few years but experiences a large increase in error between 2023 and 2024 (Fig. 3(b)).

C. Cross-Paradigm Accuracy Comparison

Table III compares the paradigms. Averaged over all sixteen combinations the ML models are only modestly more accurate than the TS models (mean RMSE: ML 0.839 vs. TS 0.870), but the averages hide a clear pattern by configuration. With the Full and Commodity-only feature sets, which contain the 25 granular food and commodity price indicators, the ML paradigm is clearly ahead (ML 0.756 vs. TS 0.844 for Full; ML 0.760 vs. TS 0.855 for Commodity-only). With the Macroeconomic-only and Climate-only feature sets the TS models are, on average, at least as accurate (TS 0.813 vs. ML 0.829 for Macroeconomic-only; TS 0.969 vs. ML 1.012 for Climate-only), driven by weak individual ML models (ElasticNet and SVR on Climate-only, XGBoost on Macro-only). The point-estimate margin was largest between the top-performing ensemble on high-dimensional commodity inputs and the traditional models: the Stacking Regressor under the Full configuration achieved an RMSE of 0.634, compared to 0.785 for the best time-series model (Bates-Granger, Macroeconomic-only). However, this pairwise difference was borderline after false discovery rate control (p = 0.051), reflecting the conservative threshold of small-sample adjustment at n = 58.

TABLE III
MEAN AND BEST HOLDOUT RMSE BY PARADIGM AND FEATURE CONFIGURATION

Configuration	Mean RMSE: ML	Mean RMSE: TS	Best RMSE: ML	Best RMSE: TS
Full	0.756	0.844	0.634	0.802
Climate-only	1.012	0.969	0.806	0.852
Macro-only	0.829	0.813	0.791	0.785
Commodity-only	0.760	0.855	0.653	0.820
All configurations	**0.839**	**0.870**	**0.634**	**0.785**

(ISSN: 0973-7049(P))

The Bioscan AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES 21(4): S.I (4),1052-1065, 2026 www.thebioscan.com

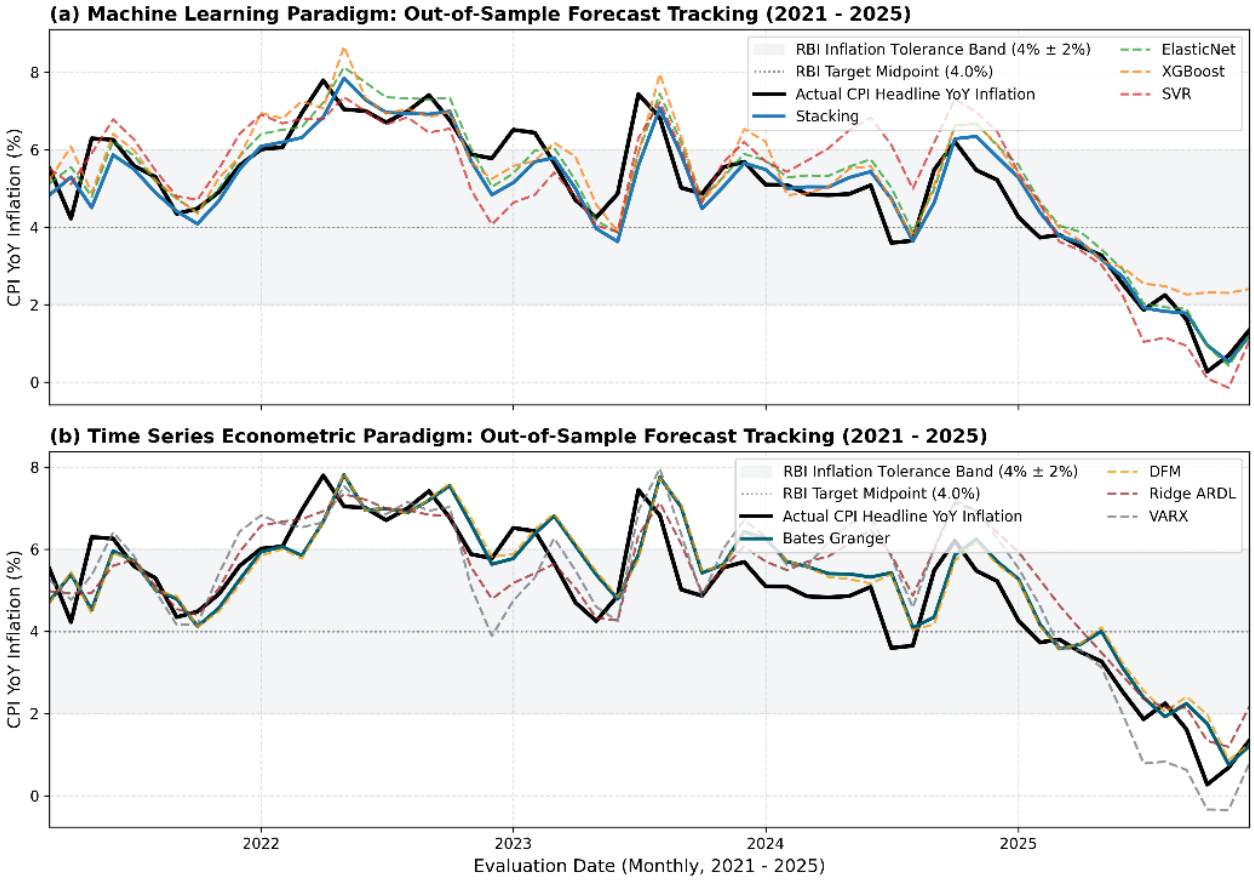


Fig. 4. Out-of-sample forecast tracking against actual Indian headline CPI (Full configuration). Shaded area shows the RBI tolerance band (4% ± 2%).

Both of the above paradigms represent the wide range of the Indian inflation cycle, which includes the period when it exceeds the 6% limit imposed by the RBI from 2022 to 2023 and the period of decline to 4% in late 2024 (Fig. 4). The key point is that the points of change are predicted differently: while the Stacking Regressor and ElasticNet identify the sudden changes with very small delay, the XGBoost and SVR consistently overestimate the peak values.

Table IV tests these gaps formally. The Stacking Regressor (Full) is significantly more accurate than every Full-configuration TS model after FDR control (adjusted p between 0.019 and 0.038), and so is ElasticNet (Full). Against the strongest TS benchmark, Bates-Granger Macroeconomic-only, the Stacking Regressor's advantage is borderline (unadjusted p = 0.043, adjusted p = 0.051), and ElasticNet's is not significant (adjusted p = 0.186). Within the ML paradigm, the Stacking Regressor beats XGBoost and SVR significantly but is statistically indistinguishable from ElasticNet (adjusted p = 0.382), consistent with the meta-learner placing most of its weight on ElasticNet (raw non-negative coefficients of 0.86, 0.12, and 0.00 on ElasticNet, XGBoost, and LightGBM in the Full configuration).

TABLE IV
CROSS-PARADIGM AND WITHIN-ML DIEBOLD–MARIANO TESTS ON THE HOLDOUT

Model A	Model B	RMSE A	RMSE B	Raw p	BH-FDR p	Sig.
Stacking Regressor Full	Bates-Granger Full	0.634	0.802	0.019	0.034	Yes
Stacking Regressor Full	Block-DFM Full	0.634	0.840	0.007	0.024	Yes
Stacking Regressor Full	Regularized ARDL Full	0.634	0.813	0.027	0.038	Yes
Stacking Regressor Full	VARX Full	0.634	0.922	0.003	0.019	Yes
Stacking Regressor Full	Bates-Granger Macro-only	0.634	0.785	0.043	0.051	No
ElasticNet Full	Bates-Granger Full	0.668	0.802	0.035	0.046	Yes
ElasticNet Full	Block-DFM Full	0.668	0.840	0.016	0.034	Yes

(ISSN: 0973-7049(P))

The Bioscan AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES 21(4): S.I (4),1052-1065, 2026 www.thebioscan.com

Model A	Model B	RMSE A	RMSE B	Raw *p*	BH-FDR *p*	Sig.
ElasticNet Full	Regularized ARDL Full	0.668	0.813	0.020	0.034	Yes
ElasticNet Full	VARX Full	0.668	0.922	0.002	0.019	Yes
ElasticNet Full	Bates-Granger Macro-only	0.668	0.785	0.172	0.186	No
Stacking Regressor Full	ElasticNet Full	0.634	0.668	0.382	0.382	No
Stacking Regressor Full	XGBoost Full	0.634	0.816	0.021	0.034	Yes
Stacking Regressor Full	SVR Full	0.634	0.906	0.007	0.024	Yes

13 comparisons; p-values adjusted with the Benjamini–Hochberg false discovery rate (BH-FDR) procedure.

D. ML Native Grouped-SHAP Attribution

Commodity dominates native SHAP attribution across all four ML models, ranging from 51.1% in SVR to 91.0% in ElasticNet (mean 68.4%). Macroeconomic and Climate shares average 24.3% and 7.2%, respectively (Table V). This ranking is unanimous across models, with Commodity ranked first in all four models. ElasticNet assigns Climate a share of exactly zero: its L1 penalty retains only 7 of the 45 Full-configuration predictors.

TABLE V
ML: GROUPED SHAP ATTRIBUTION SHARES (FULL CONFIGURATION)

Model	Climate (%)	Macroeconomic (%)	Commodity (%)
Stacking Regressor	6.00	20.62	73.38
ElasticNet	0.00	9.01	90.99
XGBoost	11.08	30.67	58.25
SVR	11.85	37.06	51.09
Mean	**7.23**	**24.34**	**68.43**

E. ML Group Permutation Feature Importance (GPFI)

GPFI (Table VI, Full configuration, 100 joint permutations per category) corroborates the SHAP ranking and sharpens it. Commodity shows a positive, statistically significant RMSE increase under permutation in all four models (from +0.418 for ElasticNet to +0.534 for SVR). Macroeconomic has a small positive mean loss overall (+0.025) but is significant in only one of four models (SVR); for the Stacking Regressor, ElasticNet, and XGBoost it is not significant after FDR control (XGBoost's adjusted *p*-value is 0.119). Climate produces non-significant zero or negative point estimates in all four models, confirming that permuting climate anomalies produces negligible degradation in out-of-sample predictive accuracy.

TABLE VI
ML: GPFI RESULTS (FULL CONFIGURATION, 100 PERMUTATIONS PER CATEGORY)

Model	Category	Mean ΔRMSE	BH-FDR *p*	Significant
Stacking Regressor	Climate	−0.001	1.000	No
Stacking Regressor	Macroeconomic	+0.005	0.178	No
Stacking Regressor	Commodity	+0.436	0.024	Yes
ElasticNet	Climate	+0.000	1.000	No
ElasticNet	Macroeconomic	+0.005	0.136	No
ElasticNet	Commodity	+0.418	0.024	Yes
XGBoost	Climate	−0.015	1.000	No
XGBoost	Macroeconomic	+0.015	0.119	No
XGBoost	Commodity	+0.492	0.024	Yes
SVR	Climate	−0.021	0.977	No
SVR	Macroeconomic	+0.075	0.024	Yes
SVR	Commodity	+0.534	0.024	Yes

F. ML Attribution Stability

Expanding-window stability (five chronological refits per model) is mixed. The coefficient of variation on category shares (CV^G) is between 0.067 (SVR, Climate) and 1.049 (ElasticNet, Climate). ElasticNet has high rank stability across chronological folds ($\tau = 0.963$), but the shares of the individual categories are quite variable across the windows, particularly for Climate ($CV^G = 1.049$). On the other hand, the Stacking

Regressor shows moderate variation in share sizes (CV^G 0.30–0.42) but the categorical rank ordering varies somewhat with increasing windows ($\tau = -0.067$).

G. Time-Series Attribution

Native econometric attribution for the TS models allocates the largest average share to Macroeconomic variables (49.33%, Rank 1), followed by Commodity (32.69%, Rank 2) and Climate (17.98%, Rank 3). This contrasts with the ML native attribution, which places Commodity decisively first (68.43%). However, this linear econometric weighting is scrutinized through model-agnostic GPFI on the holdout sample (Table VII).

Permuting Commodity variables raises holdout RMSE in all four TS models (mean +0.397), by far the largest loss of the three categories, and the raw permutation *p*-values are 0.020 or below in all four. After Benjamini-Hochberg adjustment across the twelve tests, however, the four Commodity adjusted *p*-values are 0.059 and none of the twelve tests is significant at 5%. The raw *p*-values sit at the resolution limit of 100 permutations (0.0099 or 0.0198), and the BH procedure scales them by the number of tests; a larger permutation count is needed to resolve this. The Commodity result is therefore best read as a consistently positive and largest loss that does not reach 5% significance under this design, not as evidence of no effect. Macroeconomic and Climate losses are close to zero (means +0.016 and +0.002).

TABLE VII
TS: GPFI RESULTS (FULL CONFIGURATION, 100 PERMUTATIONS PER CATEGORY)

Model	Category	Mean ΔRMSE	Raw *p*	BH-FDR *p*	Significant
Bates-Granger	Climate	+0.019	0.149	0.223	No
Bates-Granger	Macroeconomic	+0.006	0.188	0.251	No
Bates-Granger	Commodity	+0.147	0.020	0.059	No
Block-DFM	Climate	+0.020	0.099	0.170	No
Block-DFM	Macroeconomic	+0.006	0.248	0.297	No
Block-DFM	Commodity	+0.104	0.020	0.059	No
Regularized ARDL	Climate	+0.018	0.040	0.095	No
Regularized ARDL	Macroeconomic	−0.001	0.535	0.583	No
Regularized ARDL	Commodity	+0.767	0.010	0.059	No
VARX	Climate	−0.050	1.000	1.000	No
VARX	Macroeconomic	+0.054	0.059	0.119	No
VARX	Commodity	+0.570	0.010	0.059	No

H. Diagnostic Triangulation and Cross-Paradigm Convergence

The diagnostic triangulation process involves using native attribution rank, category-only accuracy, and GPFI evidence (Table VIII). Within the ML paradigm, Commodity meets all three criteria and is classified as a Core Driver. Macroeconomic variables rank second on native attribution, achieve only moderate category-only accuracy, and show significant GPFI loss in just one model (SVR); they therefore cannot be designated a Core Driver and are classified as a Redundant Driver under the triangulation protocol. Climate ranks last on attribution and category-only accuracy and is not GPFI-significant, resulting in a Mixed/Inconclusive classification.

In the TS paradigm, Macroeconomic ranks first in category-only accuracy (mean RMSE 0.813) and Commodity second (0.855); neither is GPFI-significant after adjustment, so both are labelled Redundant / Linear Baseline, while Climate is Mixed/Inconclusive. The Commodity label should be read with the caveat in Section III-G: it reflects the adjusted-*p*-value resolution, not an absence of GPFI loss (+0.397, versus +0.016 for Macroeconomic), and we do not interpret it as evidence that Commodity is redundant in the TS paradigm.

TABLE VIII
DIAGNOSTIC TRIANGULATION BY PARADIGM (FULL CONFIGURATION)

Paradigm	Category	Native rank (share)	Cat.-only RMSE (rank)	Mean GPFI ΔRMSE	GPFI-sig. models	Diagnosis
ML	Climate	3 (7.23%)	1.012 (3)	−0.009	0 / 4	Mixed / Inconclusive
ML	Macroeconomic	2 (24.34%)	0.829 (2)	+0.025	1 / 4	Redundant Driver
ML	Commodity	1 (68.43%)	0.760 (1)	+0.470	4 / 4	Core Driver
TS	Climate	3(17.98%)	0.969 (3)	+0.002	0 / 4	Mixed / Inconclusive
TS	Macroeconomic	1(49.33%)	0.813 (1)	+0.016	0 / 4	Redundant / Linear Baseline

Paradigm	Category	Native rank (share)	Cat.-only RMSE (rank)	Mean GPFI ΔRMSE	GPFI-sig. models	Diagnosis
TS	Commodity	2(32.69%)	0.855 (2)	+0.397	0 / 4	Redundant / Linear Baseline

Native rank is by grouped SHAP share (ML) and by aggregate standardized-coefficient share (TS). Category-only RMSE is the mean over models, with its rank in parentheses.

When comparing native rankings, there is a perceived gap between paradigms: Commodity is the top native driver identified by ML estimators (68.43% share, Rank 1) while TS estimators give the maximum native share to Macroeconomic indicators (49.33% share, Rank 1). Both paradigms agree that Climate is the least influential category (Rank 3 in both, averaging 7.23% in ML and 17.98% in TS).

Crucially, out-of-sample permutation diagnostics (GPFI) resolve this divergence. In both paradigms, scrambling the Commodity block causes the largest forecast error degradation (+0.470 in ML, +0.397 in TS), while permuting Macroeconomic variables produces minimal loss (+0.025 in ML, +0.016 in TS). This cross-paradigm convergence demonstrates that the high native weight assigned to Macroeconomic indicators in linear time-series equations reflects smooth autoregressive persistence and monetary trend alignment rather than indispensable out-of-sample predictive information once wholesale food price shocks are accounted for.

I. Variable-Level Importance Within Categories

We use the individual feature rankings from the category-specific configurations (Climate-only, Macroeconomic-only, and Commodity-only) of the ML models, ranked by native grouped SHAP attribution, to identify the indicators behind category-level performance. Wholesale primary food articles is the top-ranked variable in the Commodity-only configuration for all four ML models, and broad money-supply growth (M_3) is the top-ranked variable in the Macroeconomic-only configuration for all four. The Climate-only configurations show no such agreement: the top-ranked variable is a solar irradiance anomaly for the Stacking Regressor and XGBoost, a temperature anomaly for ElasticNet, and soil wetness for SVR. Tables IX and X list the three highest-ranked variables for every model and every configuration. In the Full configuration, wholesale primary food articles is the top-ranked variable for all four ML models and for three of the four TS models (the Block-DFM ranks the repo-rate change first). The TS rankings are by absolute implied coefficient and are descriptive only, given the residual-persistence caveat in Section IV-E.

TABLE IX
ML: TOP THREE VARIABLES BY MODEL AND CONFIGURATION, RANKED BY NATIVE GROUPED SHAP ATTRIBUTION

Model	Config	Rank 1	Rank 2	Rank 3
Stacking Regressor	Full	WPI Primary Food Articles	WPI Manufactured Food	WPI Brinjal
	Climate-only	Solar Irradiance Anomaly	Wind Speed Anomaly	Total Rainfall Anomaly
	Macro-only	M_3 Money Supply Growth	Industrial Production (IIP)	Repo Rate Change
	Commodity-only	WPI Primary Food Articles	WPI Manufactured Food	WPI Brinjal
ElasticNet	Full	WPI Primary Food Articles	WPI Manufactured Food	WPI Brinjal
	Climate-only	Mean Temperature Anomaly	Solar Irradiance Anomaly	Total Rainfall Anomaly
	Macro-only	M_3 Money Supply Growth	US CPI	Industrial Production (IIP)
	Commodity-only	WPI Primary Food Articles	WPI Manufactured Food	WPI Brinjal
XGBoost	Full	WPI Primary Food Articles	Brent Crude Oil	WPI Manufactured Food
	Climate-only	Solar Irradiance Anomaly	Wind Speed	Total Rainfall Anomaly
	Macro-only	M_3 Money Supply Growth	Industrial Production (IIP)	Repo Rate Change
	Commodity-only	WPI Primary Food Articles	WPI Mustard Oil	WPI Milk
SVR	Full	WPI Primary Food Articles	WPI Mustard Oil	WPI Potato
	Climate-only	Surface Soil Wetness Index	Relative Humidity	Mean Temperature
	Macro-only	M_3 Money Supply Growth	Industrial Production (IIP)	US CPI
	Commodity-only	WPI Primary Food Articles	WPI Manufactured Food	WPI Onion

WPI = Wholesale Price Index; growth rates are year-on-year.

TABLE X
TS: TOP THREE VARIABLES BY MODEL AND CONFIGURATION, RANKED BY ABSOLUTE IMPLIED COEFFICIENT (DESCRIPTIVE)

Model	Config	Rank 1	Rank 2	Rank 3
Bates-Granger	Full	WPI Primary Food Articles	Repo Rate Change	US CPI
	Climate-only	Relative Humidity Anomaly	Total Rainfall Anomaly	Rainfall Rate Anomaly
	Macro-only	M_3 Money Supply Growth	Repo Rate Change	US CPI
	Commodity-only	WPI Primary Food Articles	WPI Soybean Oil	WPI Mustard Oil
Block-DFM	Full	Repo Rate Change	US CPI	Rainfall Rate

(ISSN: 0973-7049(P))

The Bioscan AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES 21(4): S.I (4),1052-1065, 2026 www.thebioscan.com

Model	Config	Rank 1	Rank 2	Rank 3
	Climate-only	Relative Humidity Anomaly	Total Rainfall Anomaly	Rainfall Rate Anomaly
	Macro-only	M_3 Money Supply Growth	Industrial Production (IIP)	Repo Rate Change
	Commodity-only	WPI Primary Food Articles	WPI Soybean Oil	WPI Mustard Oil
Regularized ARDL	Full	WPI Primary Food Articles	WPI Milk	WPI Manufactured Food
	Climate-only	Relative Humidity Anomaly	Wind Speed Anomaly	Total Rainfall Anomaly
	Macro-only	M_3 Money Supply Growth	Industrial Production (IIP)	Repo Rate Change
	Commodity-only	WPI Primary Food Articles	WPI Milk	WPI Manufactured Food
VARX	Full	WPI Primary Food Articles	WPI Manufactured Food	M_3 Money Supply Growth
	Climate-only	Surface Soil Wetness Index	Solar Irradiance	Mean Temperature
	Macro-only	M_3 Money Supply Growth	Repo Rate Change	US CPI
	Commodity-only	WPI Primary Food Articles	WPI Arhar (Tur)	WPI Manufactured Food

II. Discussion

A. Drivers of Ensemble Predictive Superiority

Under the Full configuration, the Stacking Regressor achieved a point-estimate RMSE 21.0% lower than the best Full-configuration time-series benchmark (0.634 vs. 0.802). Because single-model non-linear architectures such as XGBoost (0.816) and SVR (0.906) did not outperform the Bates-Granger baseline (0.802), the performance difference cannot be ascribed to generalized machine learning superiority, and specific architectural explanations remain unisolated hypotheses.

Sparse selection and ensemble combination. The ML advantage appears only where the feature set contains the 25 granular commodity indicators (Full and Commodity only, Table III), which are highly collinear, and the sample has only 230 training months. ElasticNet's L1 penalty keeps just 7 of 45 predictors in the Full configuration and still reaches $R^2 = 0.843$. The ML models hold steady or improve when predictors are added (Stacking: 0.799 Macro-only to 0.634 Full), whereas the TS models do not (VARX: 0.792 to 0.922). The stack's meta-learner places 0.86 on the ElasticNet base learner, 0.12 on XGBoost, and 0.00 on LightGBM, so it behaves largely as a sparse linear model with a boosted component. Whether that component captures non-linear food-price effects was not tested.

The variance cost of added dynamic structure. The regularized ARDL uses the same penalty family as ElasticNet but adds two lags of each exogenous predictor, tripling regressor dimensionality over 230 training observations. The resulting Full-configuration RMSE (0.813) exceeds ElasticNet (0.668) by 21.7% ($p = 0.034$). While dynamic specifications permit multi-month pass-through, the additional lag parameters did not translate into lower out-of-sample forecast error in this sample. Bates-Granger weights are fitted on only about 58 observations and collapse to a corner (100% Block-DFM) in the Climate-only and Commodity-only configurations; the combination is the best TS model in two configurations, but it can only reweight three linear forecasts.

The advantage is conditional. With Macroeconomic-only inputs, a handful of slowly moving series dominated by money growth, the TS models match or beat the ML models (best RMSE 0.785 against 0.791 for SVR, 0.799 for the stack, and 0.801 for ElasticNet), and the Stacking Regressor's edge over the Bates-Granger Macroeconomic-only benchmark is borderline after FDR control (adjusted $p = 0.051$). Regularized machine-learning ensembles exploit high-dimensional, collinear commodity indices more effectively than dynamic linear benchmarks. Conversely, over low-dimensional, smooth macroeconomic aggregates, time-series specifications achieve parity with machine-learning models.

B. Commodity as the Consistent Driver

The study finds Commodity to be the dominant predictive category for Indian headline CPI. It ranks first in SHAP for all four ML models, produces significant GPFI loss in all four, and produces by far the largest GPFI loss in all four TS models, although none of the TS tests is significant after adjustment (Section III-G). Within the category, wholesale primary food articles is the top-ranked variable in every ML model (Section III-I). This is consistent with prior research highlighting the importance of wholesale commodity prices, particularly primary food articles, for Indian inflation [1], [13], and with a predominantly domestic supply-side channel. These conclusions characterize predictive association rather than structural macroeconomic causality.

C. Macroeconomic Variables: A Redundant Driver

Macroeconomic variables account for an average SHAP share of 24.34% (exceeding 30% in both XGBoost and SVR), and their category-restricted models show competitive standalone predictive accuracy. But in one of the four machine learning architectures (SVR), significant GPFI degradation appears. As a result, the diagnostic triangulation criteria categorize the macroeconomic indicators as a Redundant Driver and not a non-linear synergy. The empirical results do not support a strong non-linear macroeconomic mechanism as XGBoost does not suffer from a statistically significant loss of accuracy when using feature permutation after FDR control ($p = 0.119$).

In both paradigms, the broad money-supply growth rate (M_3) is always the most important macroeconomic indicator. In the time-series paradigm, macroeconomic indicators generate the lowest category-only error, but near-zero GPFI loss (+0.016) in the Full configuration, suggesting that macroeconomic features

offer little additional predictive gain when controlling for persistence and granular commodity indices.

D. Marginal Predictive Contribution of Climate Indicators

The predictive skill of the climate variables is generally poor for both paradigms. They have the smallest SHAP share in the ML paradigm and Climate-only models are the least accurate combination for each of the 8 models. The ML GPFI point estimates for Climate are zero or negative, and the TS estimates are close to zero and not significant. This does not contradict the literature that has already shown a relationship between climate and food prices and inflation in India, it suggests that, after the CPI persistence and commodity variables are in the feature set, climate variables do not have any measurable predictive value. Meteorological anomalies seem to have an indirect impact on prices, passing through domestic agricultural wholesale markets, but not directly as linear shocks on retail consumer price indices. Within the Climate-only configurations, the ML models do not agree on a leading variable (Section III-I).

E. Scope and Limitations

Several methodological boundaries frame these empirical findings. Methodologically, evaluations rely on one-step-ahead forecasts with observed persistence lags over a single 58-month holdout window without recursive multi-horizon projections or rolling-origin refits. Selection of autoregressive controls (lag 1 and lag 12) was made from holdout results from preliminary screens, and several series, both the target index and M_3 growth, display lack of rejection of the unit root hypothesis in ADF tests, implying possible inclusion of residuals' persistence in time-series parameters.

In terms of statistics, predictive hypotheses are rigorously confined to conservative inference: only 4 of 24 machine learning configurations contrast pass false discovery rate correction, and the difference between the Stacking Regressor and ElasticNet (p = 0.382) is statistically insignificant. Moreover, expanding windows stability tests reveal different allocation of categories and rank dynamics in estimation folds, which means that native attributions are to be viewed as one-time sample estimates and not structural population parameters.

Lastly, precision of attribution analysis is limited by 100-permutation resolution of GPFI within the time series framework, while benchmarks are based on single-equation regularized econometric models.

III. Conclusion

This study evaluated a dual-paradigm framework benchmarking four regularized machine-learning architectures against four dynamic econometric specifications across 25 years of Indian macroeconomic data. Out-of-sample results show that the Stacking Regressor generates superior predictive accuracy (RMSE = 0.634, R^2 = 0.858), lowering forecast error by 21.0% relative to the best Full-configuration time-series model, a gap that survives false-discovery-rate control (its margin over the best time-series model overall, Bates-Granger with Macroeconomic-only inputs, is borderline at an adjusted p of 0.051). The predictive gains of machine learning over time-series models are contingent and hold only when filtering dense and collinear agricultural price series and not in case of lower dimensional macroeconomic inputs where we observe near parity between models. For both types of models, diagnostic triangulation reveals domestic commodity shocks, specifically Wholesale Price Index primary food articles, as the main determinant of CPI movements. Macroeconomic aggregates become redundant once food prices and persistence are included in the equation, while spatial climate indicators do not have any measurable direct out-of-sample effect on retail headline indices. These results illustrate that predicting Indian consumer inflation calls for emphasizing high-frequency monitoring of domestic agriculture chain instead of monetary aggregates and weather indices.

IV. References

Data and Code Availability

All series are drawn from the public sources listed in Table I. The processed dataset and analysis code are available at https://github.com/AmannSinghh18/Capstone-project

References

[1] C. S. C. Sekhar, D. Roy, and Y. Bhatt, "Food inflation and volatility in India: Trends and determinants," *Indian Econ. Rev.*, vol. 53, pp. 65–91, 2018, doi: 10.1007/s41775-018-0017-z.

[2] P. Nayak and P. K. Jena, "The interplay of climate, subsidies and monetary factors in shaping India's food price inflation," *Indian Econ. J.*, online first, Feb. 26, 2026, doi: 10.1177/00194662261425167.

[3] M. Kotz, F. Kuik, E. Lis, and C. Nickel, "Global warming and heat extremes to enhance inflationary pressures," *Commun. Earth Environ.*, vol. 5, Art. no. 116, 2024, doi: 10.1038/s43247-023-01173-x.

[4] C. A. Sims, "Macroeconomics and reality," *Econometrica*, vol. 48, no. 1, pp. 1–48, 1980.

[5] M. H. Pesaran, Y. Shin, and R. J. Smith, "Bounds testing approaches to the analysis of level relationships," *J. Appl. Econometrics*, vol. 16, no. 3, pp. 289–326, 2001, doi: 10.1002/jae.616.

[6] J. H. Stock and M. W. Watson, "Forecasting using principal components from a large number of predictors," *J. Amer. Statist. Assoc.*, vol. 97, no. 460, pp. 1167–1179, 2002, doi: 10.1198/016214502388618960.

[7] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *Oper. Res. Quart.*, vol. 20, no. 4, pp. 451–468, 1969, doi: 10.1057/jors.1969.103.

[8] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[9] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Statist. Soc. B, Statist. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

[10] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 9, 1997, pp. 155–161.

[11] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992, doi: 10.1016/S0893-6080(05)80023-1.

[12] N. Singh and B. B. Bhoi, "Inflation forecasting in India: Are machine learning techniques useful?," *Reserve Bank India Occasional Papers*, vol. 43, no. 2, 2022.

[13] A. Malhotra and M. Maloo, "Understanding food inflation in India: A machine learning approach," 2017, arXiv:1701.08789.

[14] N. Singh and A. Rathi, "Nowcasting food inflation in India: Leveraging price and non-price signals through machine learning," *RBI Bull.*, vol. 78, no. 10, pp. 203–217, 2024.

[15] B. Pratap, A. Pawar, and S. Sengupta, "Non-linear Phillips curve for India: Evidence from explainable machine learning," *Comput. Econ.*, vol. 67, no. 3, pp. 2301–2344, 2026, doi: 10.1007/s10614-025-10942-z.

[16] P. Cashin and R. Anand, Eds., *Taming Indian Inflation*. Washington, DC, USA: Int. Monetary Fund, 2016, doi: 10.5089/9781513541259.071.

[17] A. Kumar *et al.*, "The impact of meteorological factors on crop price volatility in India: Case studies of soybean and brinjal," *EPJ Data Sci.*, vol. 14, Art. no. 74, 2025.

[18] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *J. Amer. Statist. Assoc.*, vol. 74, no. 366, pp. 427–431, 1979, doi: 10.1080/01621459.1979.10482531.

[19] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root," *J. Econometrics*, vol. 54, nos. 1–3, pp. 159–178, 1992, doi: 10.1016/0304-4076(92)90104-Y.

[20] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 3146–3154.

[21] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Statist.*, vol. 13, no. 3, pp. 253–263, 1995, doi: 10.1080/07350015.1995.10524599.

[22] D. Harvey, S. Leybourne, and P. Newbold, "Testing the equality of prediction mean squared errors," *Int. J. Forecasting*, vol. 13, no. 2, pp. 281–291, 1997, doi: 10.1016/S0169-2070(96)00719-4.

[23] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc. B, Methodol.*, vol. 57, no. 1, pp. 289–300, 1995, doi: 10.1111/j.2517-6161.1995.tb02031.x.

[24] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4765–4774.

[25] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.

[26] M. Jullum, A. Redelmeier, and K. Aas, "groupShapley: Efficient prediction explanation with Shapley values for feature groups," 2021, arXiv:2106.12228.