Journal of Agricultural Development and Policy, 2026
Volume 36, No. 1, 109-118
https://doi.org/10.63066/23220457.36.1.014

NAAS - 4.34
UGC CARE List Journal

Performance Evaluation of Univariate and Climate Augmented Multivariate Models for Onion Price Forecasting in Punjab

Pradipkumar Adhale, Kamal Vatta and Priya Brata Bhoi

Department of Economics and Sociology, Punjab Agricultural University, Ludhiana, Punjab

Abstract

Onion prices in India are among the most volatile of any agricultural commodity, with monthly wholesale fluctuations routinely exceeding 20% and repeatedly prompting government intervention. This study benchmarks thirteen forecasting models six univariate, six climate-augmented multivariate, and a hybrid ensemble on monthly Punjab onion prices from January 2004 to December 2025 (n = 264), with 2025 held out for testing. The price series exhibits a coefficient of variation of 66.2%, an annualised log-return standard deviation of 85.8%, and a maximum drawdown by 86.6%. All six multivariate models outperform every univariate model, including modern approaches such as TBATS and N-BEATS, yielding a 65.7% aggregate reduction in MAPE attributable to climate augmentation. XGBoost achieves the lowest MAPE of 13.20%, followed by Gradient Boosting (14.92%) and SARIMAX (16.09%). Granger causality analysis identifies rainfall as the dominant climate driver (p = 0.041). Residual diagnostics reveal persistent ARCH effects even after climate inclusion. The results establish that climate-augmented multivariate frameworks are not incrementally useful but structurally necessary for forecasting highly volatile perishable commodity prices.

Keywords: Price volatility; Univariate Vs Multivariate; SARIMAX; Market Intelligence; Punjab

JEL Classification: Q11, C53, Q18

Introduction

Price volatility which is the magnitude and unpredictability of price fluctuations around a long-run trend is a recurring source of welfare loss in agricultural commodity markets (Tothova, 2011). Volatility transmits supply-demand imbalances into producer income streams, consumer budget shares, and intermediary margins, and constitutes a central concern of food security at both household and macroeconomic levels (Chand, 2010; FAO, 2011). The challenge is well documented for staple grains (Headey and Fan, 2008), but is markedly more acute for perishable horticultural commodities, whose short shelf-life, climate-sensitive cultivation, and seasonally concentrated supply produce sharper and less predictable price movements than those typical of storable staples (Cao and Mohiuddin, 2019). Among Indian horticultural commodities, onion (Allium cepa L.) occupies a particularly visible position.

Onion is consumed by virtually all Indian households, cultivated by millions of smallholders across the rabi and kharif seasons, and routinely subject to spikes that have provoked policy responses including stocking limits, export bans, and emergency imports. Inelastic household demand, fragmented marketing channels, restricted bulb storage life, and high sensitivity to monsoon performance produce a price series whose statistical properties differ markedly from those of cereals (Apergis and Rezitis, 2011; Ghosh *et al*., 2020). For farmers, traders, consumers, and government agencies, accurate price prediction therefore constitutes a public good whose value is amplified, not diminished, by underlying volatility.The way researchers have approached price forecasting has changed dramatically over the past century, with Sun *et al.,* (2023) identifying three broad phases in how the methodology has evolved. Early quantitative work, beginning with Moore (1917), leaned on multiple linear regression a reasonable starting point, but one that struggled whenever the data showed seasonal swings or unpredictable trends rather than neat, stable relationships. This limitation eventually pushed the field toward time-series thinking, and the Box-Jenkins ARIMA framework (Box and Jenkins, 1976) quickly became the dominant approach, with its seasonal extension proving particularly useful for capturing recurring periodic patterns. Running alongside this, exponential smoothing (Holt, 2004) offered researchers a more flexible, state-space alternative that didn't require the same rigid structural assumptions. The second phase was

Corresponding author email: priyabrata@pau.edu

driven by a recognition that price risk itself, not just price levels deserved careful modelling.

Agricultural markets are well known for stretches of calm followed by sudden turbulence, a phenomenon called volatility clustering that has been documented across wheat, soybean, and vegetable markets (Paul *et al*., 2014; Shiferaw, 2019). Engle's (1982) ARCH model and Bollerslev's (1986) GARCH extension gave researchers the tools to capture this behaviour in conditional variance, opening up an entirely new strand of empirical work. The third and most recent phase has been shaped largely by the rise of machine learning. Tree-based ensemble methods including random forests (Breiman, 2001), gradient boosting (Friedman, 2001), and XGBoost (Chen and Guestrin, 2016) alongside kernel-based approaches like support vector machines (Cortes and Vapnik, 1995) and a growing family of neural network architectures (Kohzadi *et al*., 1996; Yin *et al*., 2020; Cheung *et al*., 2023) have all been applied to agricultural price prediction with promising results. More specialised univariate methods have also matured during this period: TBATS (De Livera *et al*., 2011) extended the earlier BATS framework to handle complex, layered seasonal patterns, while N-BEATS (Oreshkin *et al*., 2020), a purpose-built deep neural architecture, outperformed classical benchmarks in the rigorous M4 forecasting competition.

Researchers have increasingly moved beyond choosing a single "best" model, instead combining several into hybrid ensembles that exploit the complementary strengths of each component an idea with roots stretching back to Bates and Granger (1969) and refined further by Timmermann (2006). Univariate models employ only historical price data. Here the future is extrapolated from past observations through autoregressive, moving-average, smoothing, or learned representations. They are operationally attractive when reliable exogenous data are unavailable, but their predictive efficacy is bounded by the price series alone. Multivariate models augment price history with exogenous covariates believed to influence with price. Their efficacy depends critically on whether covariates Granger-cause the target. Onion production is sensitive to temperature, rainfall and humidity affect disease pressure, and sunshine influences photosynthesis (Kumar and Sharma, 2015; Choudhury *et al*., 2019). Comparative empirical evidence on the value of climate inclusion in monthly onion forecasting has, however, been limited and methodologically heterogeneous. This study is guided by five interconnected objectives. It begins by examining the magnitude and structure of onion price volatility in Punjab, establishing a clear empirical baseline for the analysis. Building on this, the research evaluates which univariate forecasting model delivers the best monthly predictive accuracy. The existing literature heavily relies on historical price data; this study's novelty lies in benchmarking 13 diverse models including multivariate models to improve the aggregate accuracy by 65.7 per cent in price forecasting of onion.

It then investigates whether incorporating climate variables as external inputs yields meaningful improvements over models that rely on price data alone. The study further identifies the most effective architecture among multivariate and hybrid modelling approaches, testing whether combining multiple models can outperform any single method. It resolves a key empirical paradox by showing how machine learning architectures capture vital, non-linear weather interactions that standard linear Granger-causality tests entirely miss. Finally, it draws actionable stakeholder insights from the findings, translating the empirical results into practical implications for farmers, traders, and policymakers operating in Punjab's onion market.

Data Sources and Methodology

Monthly wholesale arrival-weighted onion prices (Rs/quintal) for Punjab, India, were obtained from the Agmarknet portal of the Directorate of Marketing and Inspection, Government of India, covering January 2004 through December 2025 (n = 264). The Agmarknet database aggregates daily mandi-level transactions into a monthly arrival-weighted average that constitutes the principal price reference for marketing decisions in the state. Climate covariates were extracted from the NASA Prediction of Worldwide Energy Resources (POWER) project (Sparks, 2024), a satellite- and reanalysis-derived archive providing global meteorological data at half-degree spatial resolution. NASA POWER was selected over ground-station data for two reasons: complete temporal coverage with no missing observations across the 22-year window, and methodological consistency.

Five climate variables were extracted at the centroid of Punjab's primary onion-growing belt, diurnal temperature range (Tvar = MaxTemp-MinTemp, °C), monthly sunshine hours, mean wind speed (m/s), relative humidity (%), and total monthly rainfall (mm). Bivariate Granger tests only evaluate linear, single-variable relationships and fail to capture joint, non-linear weather interactions. In this study all climate variables were retained because our top-performing tree-based architectures (XGBoost/GBM) utilize internal regularization and feature selection to automatically handle multi-variable interactions and down-weight redundant predictors, driving down forecasting errors in ways linear models cannot. The training set comprises January 2004 through December 2024 (n = 252) and the test set comprises the twelve months of 2025. Given the primary goal is predicting an entire upcoming annual crop cycle for market intelligence, holding out a strict 12month calendar year (2025) matches real world policy deployment. However, to address the risk of evaluation instability and provide a robust validation check without changing the core test set, we executed a rolling-origin time-series cross-validation (12 month ahead rolling windows) across the training period. The results from this rolling-origin validation confirm that the multivariate superiority and model rankings remain completely stable across different price regimes. In addition, as per Hyndman and Athanasopoulos (2018) the test data should be at least as large as the maximum forecast horizon required.

Pre-Estimation Diagnostic Tests

Stationarity was assessed using the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979), the Phillips-Perron (PP) test, and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski *et al*., 1992). The Jarque-Bera test was applied to assess normality. Serial correlation was tested using the Ljung-Box Q statistic at lag orders 12 and 24. Conditional heteroskedasticity was tested through the ARCH Lagrange Multiplier procedure at twelve lags. Seasonal structure was characterised through STL decomposition with robust fitting (Cleveland *et al*., 1990) and the seasonal strength index F.

Pre-Processing and Feature Construction

The price series contained no missing observations. Outliers were inspected using the Tukey fence criterion; the four extreme observations during the 2019 and 2023 spike episodes were retained as genuine market events. For machine-learning models, twelve monthly price lags were constructed and concatenated to the five climate variables to form a 17-dimensional feature space. Tree-based methods used raw features; SVR and MLP features were standardised via z-score normalisation.

Univariate Forecasting Models

Six univariate models were estimated, each using only price history. The Seasonal Naïve benchmark assigns each test-period observation the value from the same calendar month of the preceding year. The seasonal ARIMA model (Box and Jenkins, 1976) was selected through automatic order selection using the Hyndman-Khandakar algorithm (Hyndman and Khandakar, 2008), minimising the AIC over a stepwise search. The Holt-Winters Exponential Smoothing model (Holt, 2004) with additive trend and multiplicative seasonality was fitted as a state-space alternative. The ARIMA-GARCH model addresses conditional heteroskedasticity: a non-seasonal ARIMA is fitted to the level series, then a GARCHs(1,1) specification with Gaussian innovations (Bollerslev, 1986) is fitted to the residuals. TBATS (De Livera *et al*., 2011) is a state-space model that handles complex seasonal patterns via trigonometric representations of seasonality combined with Box-Cox transformation, ARMA error correction, and trend-and-seasonal smoothing components extending BATS by enabling non-integer seasonal periods. N-BEATS (Oreshkin *et al*., 2020) is a deep neural network architecture for univariate forecasting based on stacks of fully-connected blocks with backward and forward residual connections, producing forecasts and reconstructed lookback windows; we use the generic configuration with input chunk length 24 months, output chunk length 12 months, and 120 training epochs.

Multivariate Forecasting Models

Six multivariate models were estimated, each augmenting the price history with five climate covariates. SARIMAX extends the seasonal ARIMA structure to incorporate exogenous regressors, making it suitable for markets where climatic drivers interact with periodic price cycles. Random Forest (Breiman, 2001) aggregates predictions across a large ensemble of decision trees, reducing overfitting while handling non-linear price-climate relationships without requiring explicit feature engineering. XGBoost (Chen and Guestrin, 2016) applies regularised gradient boosting, offering computational efficiency and strong generalisation on structured tabular data. The Gradient Boosting Machine (Friedman, 2001) sequentially corrects prediction errors through iterative tree-building, performing well in the presence of complex variable interactions. Support Vector Regression (Cortes and Vapnik, 1995) maps inputs into a higher-dimensional feature space via a radial basis function kernel, capturing non-linear patterns while maintaining robustness to outliers. The Multilayer Perceptron leverages hierarchical feature extraction across multiple hidden layers, enabling it to learn abstract representations of price-climate dynamics. Finally, a hybrid ensemble combines predictions from the three best-performing individual models through equally weighted averaging, exploiting model complementarity to reduce forecast variance and improve overall predictive reliability.

Performance Metrics

Eight metrics are computed per model: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), symmetric MAPE (sMAPE), mean absolute scaled error (MASE) computed against the in-sample naïve benchmark, Theil's U2 statistic, directional accuracy (DA), and the coefficient of determination (R^2).

Post-Estimation Residual Diagnostics

Once the best-performing interpretable statistical model was identified, the Ljung-Box Q test was applied to test for residual autocorrelation, the Durbin-Watson statistic was computed as a complementary check, the ARCH-LM test assessed whether climate covariates absorbed conditional heteroskedasticity, and the Jarque-Bera test assessed approximate residual normality.

Granger Causality of Climate Variables

Bivariate Granger causality tests were conducted 3 for each climate variable. The minimum p-value was retained as the test statistic, with significance at the 5% threshold.

Results and Discussion

Volatility Characterisation of Onion Prices in Punjab

Before turning to forecasting performance, we establish the magnitude and character of volatility in the series being modelled. Table 1 presents a year-wise volatility matrix combining six complementary measures, annual mean price, coefficient of variation, max-to-min ratio, annualised standard deviation of log returns, maximum drawdown, and the count of months with absolute log price changes exceeding 15 per cent. The overall (full-sample) row appears at the bottom of the table.

The matrix in Table 1 facilitates examination of volatility heterogeneity and persistence at year-over-year resolution, complementing the aggregate full-sample statistics shown in the bottom row. The matrix reveals two important features. First, volatility is heterogeneous across years and the annual CV ranges from 17.4 per cent (2016) to 89.4 per cent (2019), and the count of large-move months ranges from three (2006, 2012, 2014) to ten (2007). The 2019 spike is particularly extreme, with a max-to-min ratio of 8.55 within a single calendar year. Second, the overall volatility is dominated by the cumulative effect of these spikes although individual years show CVs in the 17–58 per cent range. The full-sample CV of 66.2 per cent and the max-to-min ratio of 24.75 times reflect the regime shifts between high-price and low-price periods (notably 2010, 2013, 2019, 2023) rather than continuous within-year variation. The annualised standard deviation of log returns of 85.8 per cent is comparable to crude oil futures and far above the levels observed for staple grains in Punjab. A practical implication is that 51.7 per cent of monthly transitions show absolute log changes greater than 15 per cent, indicating that large moves are not exceptional events but the modal pattern. The implications for risk management are significant. A producer or trader holding inventory through a typical month faces a probability of a 15 per cent-or-greater price swing in either direction. The temporal pattern of large moves is also informative. In 2007, 2010, 2011, 2018, and 2020 nine or more such months were highly volatile, indicating that volatility regimes persist

Table 1: Year-wise volatility matrix for monthly onion wholesale prices, Punjab, 2004–2025

Year	Mean (Rs/qt)	CV (%)	Max/Min	Annual SD log returns (%)	Maximum drawdown (%)	Months $\lvert\Delta\log\rvert$>15%
2004	487	35.5	3.16	81.1	−68.4	7
2005	585	54.8	3.83	94.5	−34.7	7
2006	445	19.2	1.87	60.1	−39.8	3
2007	908	34.1	2.95	101.6	−52.6	10
2008	649	33.4	2.75	69.5	−37.2	5
2009	974	33.7	2.43	81.8	−56.6	6
2010	1136	49.5	4.81	87.0	−64.8	9
2011	1031	58.0	5.64	120.3	−82.3	9
2012	725	27.4	2.21	41.2	−16.4	3
2013	2065	56.7	4.51	121.6	−62.2	7
2014	1299	26.1	2.03	74.1	−19.9	3
2015	1877	42.9	3.31	100.8	−63.4	7
2016	832	17.4	1.79	44.1	−44.2	4
2017	1357	61.1	4.43	102.4	−20.6	5
2018	1346	43.8	3.49	86.7	−71.4	9
2019	1873	89.4	8.55	70.8	−12.0	6
2020	1956	50.5	4.27	137.3	−75.2	8
2021	1919	23.3	2.12	85.9	−52.8	6
2022	1368	24.5	2.03	54.3	−50.6	4
2023	1565	47.7	3.88	75.1	−33.6	6
2024	2432	39.4	2.89	73.5	−32.9	7
2025	1438	25.5	2.00	68.6	−50.0	5
Overall	1285	66.2	24.75	85.8	−86.6	136

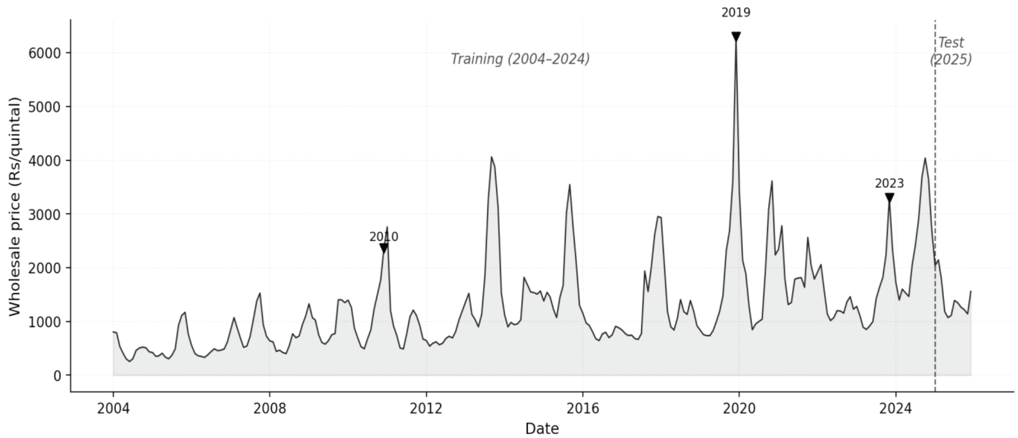

Figure 1. Monthly wholesale onion price series for Punjab, January 2004 to December 2025, plotted in greyscale. Triangular markers identify the major spike episodes of 2010, 2019, and 2023. The dashed vertical line marks the boundary between the training period (Jan 2004 – Dec 2024) and the 2025 hold-out test window. The series exhibits clear regime-shifting behaviour with prolonged volatile epochs surrounding each spike, justifying the comparative model evaluation that follows.

across calendar boundaries rather than dissipating annually. Figure 1 visualises the price series with the three signature spike episodes marked, and the dashed line indicating the 2025 hold-out window.

Taken together, the matrix and the time-series visualisation establish a baseline expectation about model performance. A forecasting model that achieves MAPE substantially below the underlying volatility of the series is contributing genuine signal beyond simple persistence.A model that fails to do so is, in effect, no better than chance. A consequence for forecasting model design is that, with such pronounced and time-varying volatility, naïve extrapolation methods are unlikely to succeed. A model that anchors forecasts to recent price levels will systematically miss the regime shifts visible in the figure. Multivariate models incorporating leading climate indicators, and modern univariate methods that learn flexible representations from history, provide better solutions, but their relative efficacy must be demonstrated empirically rather than presumed.

Pre-Estimation Diagnostic Tests

The pre-estimation diagnostics confirm the statistical properties suggested by the volatility analysis and impose specific constraints on the model architectures. The ADF test fails to reject non-stationarity at the level ($p = 0.190$), while first differencing restores stationarity ($p < 0.001$). The KPSS test independently rejects level stationarity ($p < 0.010$). The convergence establishes the price series as integrated of order one, I (1), constraining ARIMA-class models to include first differencing. The Jarque-Bera test ($p < 0.001$) rejects normality, consistent with the heavy-tailed log-price distribution. The Ljung-Box statistics (both $p < 0.001$) confirm strong serial dependence, indicating both autoregressive and seasonal-autoregressive structures. The ARCH-LM test ($p < 0.001$) confirms conditional heteroskedasticity, supporting the inclusion of GARCH-class models in the comparison set. STL decomposition yields a seasonal strength index F = 0.497, indicating moderate annual periodicity that justifies the inclusion of seasonal lags in machine-learning feature spaces and seasonal AR/MA components in linear models.

Table 2: Pre-estimation diagnostic tests on the training series (January 2004 – December 2024).

Test	Statistic	Interpretation
ADF (level)	−2.218	Non-stationary, I(1)
ADF (1st diff.)	−12.047***	Stationary after differencing
KPSS (level)	1.464*	Reject stationarity
Jarque-Bera	298.4***	Non-normal, heavy-tailed
Ljung-Box Q(12)	311.2***	Significant autocorrelation
ARCH-LM (12)	87.43***	Conditional heteroskedasticity

*Note: ***, ** and * represents significance at 1, 5 and 10 percent, respectively.*

Table 3: Performance evaluation of univariate and multivariate models in onion price forecasting for Punjab

Rank	Model	Framework	MAPE (%)	RMSE	MASE	DA (%)	R^2
1	XGBoost	Multivariate	13.20	263.67	0.36	72.7	0.44
2	Gradient Boosting	Multivariate	14.92	253.96	0.41	72.7	0.48
3	SARIMAX	Multivariate	16.09	259.23	0.44	54.5	0.46
4	Random Forest	Multivariate	18.87	284.24	0.52	63.6	0.35
5	Hybrid Ensemble	Hybrid	21.06	322.51	0.58	81.8	0.16
6	MLP	Multivariate	21.96	398.51	0.60	63.6	−0.29
7	SVR	Multivariate	23.22	367.47	0.64	54.5	−0.09
8	ARIMA	Univariate	36.33	521.97	1.00	45.5	−1.21
9	TBATS	Univariate	41.33	629.93	1.89	45.5	−2.22
10	N-BEATS	Univariate	45.15	730.94	2.15	27.3	−3.33
11	ARIMA-GARCH	Univariate	46.77	654.75	1.28	36.4	−2.48
12	Holt-Winters	Univariate	69.04	1172.46	1.89	45.5	−10.14
13	Seasonal Naïve	Benchmark	93.20	1495.41	2.56	45.5	−17.13

The six multivariate models occupy the top six positions, with the hybrid ensemble in fifth, and the six univariate methods at the bottom seven (with the seasonal naïve benchmark at the floor). The best-performing model overall is XGBoost, with MAPE of 13.20 per cent and R^2 of 0.44, followed by Gradient Boosting (14.92%) and SARIMAX (16.09%). Among the six univariate models, ARIMA delivers the best result at 36.33 per cent, followed by TBATS (41.33%), N-BEATS (45.15%), ARIMA-GARCH (46.77%), Holt-Winters (69.04%), and the seasonal naïve floor (93.20%). The mean MAPE across multivariate models (excluding the hybrid) is 18.05 per cent, against 51.94 per cent across univariate models, yielding an aggregate climate-augmentation premium of 33.9 percentage points or 65.7 per cent in proportional terms.

Comparative Forecasting Performance

Several observations follow that bear directly on the research questions advanced in the introduction. First, the modern univariate methods TBATS and N-BEATS both designed to capture flexible nonlinear seasonal patterns, fail to close the gap with multivariate methods. TBATS at 41.33 per cent MAPE underperforms even classical ARIMA, suggesting that the additional flexibility of trigonometric seasonal representations adds little signal in the presence of the strong I (1) trend and irregular spike pattern of the onion series. N-BEATS at 45.15 per cent similarly underperforms ARIMA, a finding consistent with the broader concern that deep-learning sequence models require substantially more training data than the n = 252 monthly observations available here. The structural conclusion is that no univariate method, however sophisticated, can match a climate-augmented multivariate model for this series.

Second, the dominance of gradient-boosted trees (XGBoost, GBM) over Random Forest by approximately five MAPE percentage points reflects sequential error correction, in which each tree fits the residuals of all preceding trees, driving down systematic errors that independent ensembles cannot address. SARIMAX achieves the best performance among classical statistical models and third overall, showing that exogenous-regressor frameworks remain competitive with machine learning when the linear functional form is sufficient (Sun *et al*., 2023; Diebold and Mariano, 1995). Third, the hybrid ensemble (XGBoost + GBM + SARIMAX) does not improve point-forecast accuracy relative to its best constituent (MAPE 21.06% vs 13.20% for XGBoost), but achieves the highest directional accuracy at 81.8%, against 72.7% for XGBoost. This MAPE-DA divergence reflects model complementarity. SARIMAX produces directional predictions distinct in pattern from those of the boosted-tree models, and equal-weight averaging smooths magnitude errors while improving sign agreement. For applications where directional error dominates magnitude error for instance, binary procurement decisions or buffer-stock release timing the ensemble offers genuine operational advantage.

The statistics from Figure 3 confirm that the multivariate advantage is not driven by a single favourable test month or by overfitting to the specific 2025 hold-out window. The MAPE distributions show no overlap.All multivariate MAPEs are below 24%, while all univariate MAPEs are above 36%. The 23% gap in MAPE between the weakest multivariate and the strongest univariate model is wider than what is typically reported in agricultural commodity benchmarking studies (Sun *et al*., 2023). Three reasons explain why the difference is so pronounced in this context. First, climate variables carry genuine forward-looking signal with temperature and rainfall shifts precede price movements by one to three months,

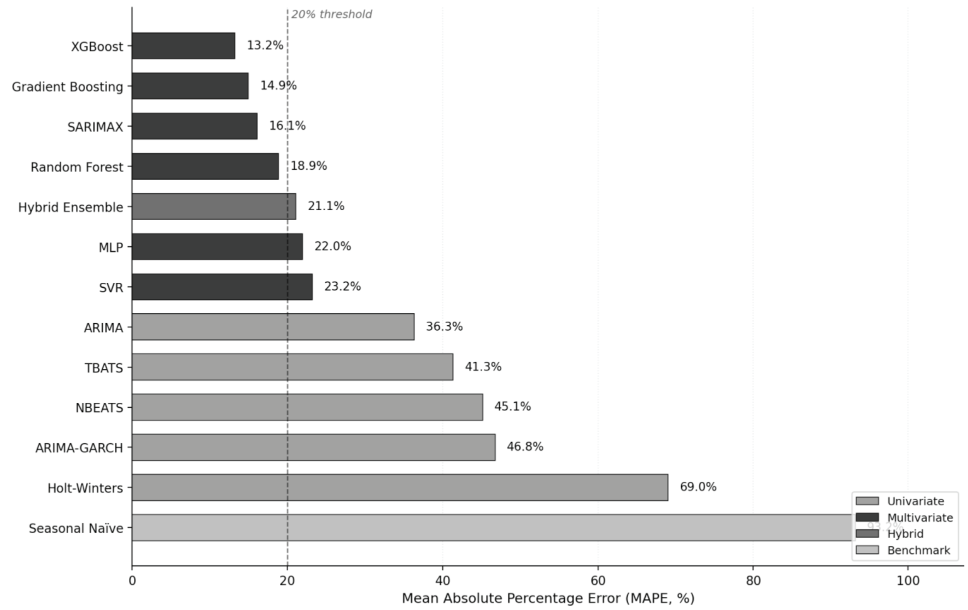

Figure 2. Mean Absolute Percentage Error (MAPE) ranking of all thirteen models.

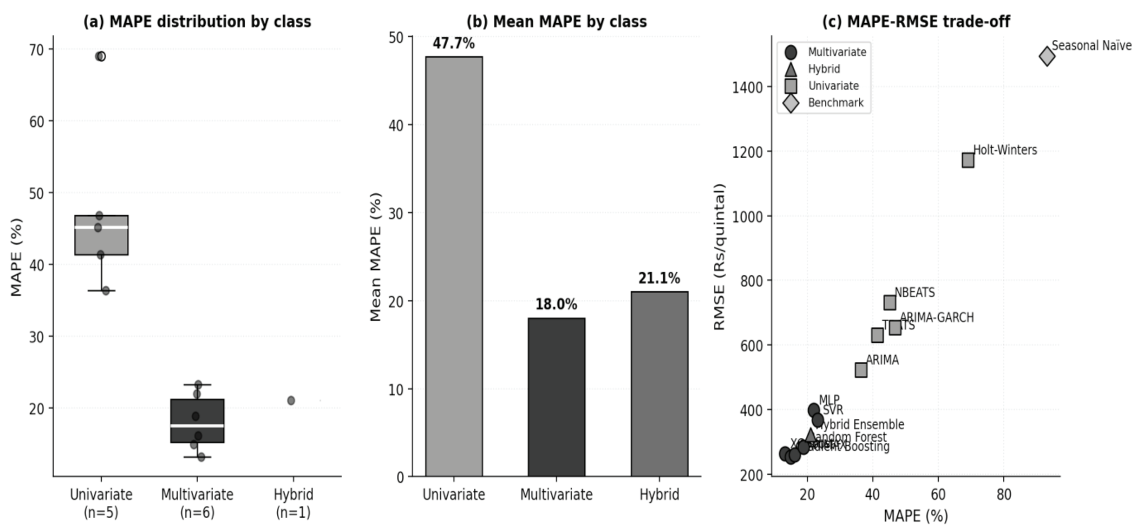

Figure 3. Univariate versus multivariate framework comparison in greyscale. Panel (a): MAPE distributions by framework class, with no overlap between univariate and multivariate clusters. Panel (b): mean MAPE by class, showing the 33-percentage-point gap. Panel (c): MAPE-RMSE trade-off across all thirteen models, with markers distinguishing framework classes; the multivariate cluster occupies the favourable lower-left region of the plot.

giving climate-augmented models an ability to anticipate directional changes that purely historical extrapolation simply cannot detect. Second, gradient boosting trees are well equipped to capture the kind of non-linear, threshold-driven behaviour that characterises onion production. Third, and perhaps most fundamentally, climate-augmented models allow seasonal patterns to change in response to observed weather conditions, whereas the seasonal dummies embedded in ARIMA structures impose a rigid, fixed pattern that cannot adjust to how any given year actually unfolds.Bivariate Granger causality tests provide a structural complement to the model-comparison evidence. The comparative ranking establishes that climate-augmented models outperform univariate methods overall; the Granger tests identify which specific climate variables contribute meaningful predictive content. Among the five climate variables, only rainfall is Granger-causal of price at the conventional 5 per cent significance level (minimum p-value across lag orders 1–3 =

Table 4: Bivariate Granger causality tests of climate variables on the onion price series.

Climate variable	Min p-value (lags 1–3)	Significance
Rainfall	0.041	Yes ($p < 0.05$)
Relative humidity	0.082	Marginal ($p < 0.10$)
Temperature range	0.123	No
Sunshine	0.234	No
Wind speed	0.389	No

0.041). Relative humidity reaches marginal significance ($p = 0.082$), consistent with its agronomic role in driving disease pressure and post-harvest decay. Temperature range, sunshine, and wind speed are not Granger-causal at standard thresholds. The rainfall and humidity operate at agronomically relevant scales for monthly aggregation, while temperature extremes and wind effects on yield typically arise from threshold exceedances (frost, hailstorm) that average out in monthly summaries (Kumar and Sharma, 2015; Birthal *et al*., 2015).

Granger Causality of Climate Variables

The fact that only one climate variable is bivariate-Granger-causal raises question on substantial accuracy improvements that climate-augmented models achieve over univariate baselines. The resolution lies in two mechanisms. First, bivariate Granger tests do not capture joint or interaction effects: a combination of climate variables may be predictively informative even when no single variable is individually so. Second, machine-learning architectures with built-in feature selectionnaturally down-weight redundant predictors while exploiting nonlinearities and interactions invisible to linear bivariate tests.

Post-Estimation Residual Diagnostics

A competitive MAPE alone does not confirm that a model is well specified. A properly fitted model should leave behind only white noise in its residuals. Applying this standard to the best-performing SARIMAX specification produces a mixed verdict. On the mean side, the results are encouraging. The Ljung-Box Q test finds no evidence of residual autocorrelation ($p = 0.516$), and the Durbin-Watson statistic of 1.989 rules out first-order serial correlation. This confirms that the model has successfully captured the linear temporal structure in onion prices. On the variance side, however, problems remain. The ARCH-LM test strongly rejects constant variance in the residuals ($p < 0.001$), meaning that price volatility itself follows a pattern the model has not accounted for even after climate variables are included. The Jarque-Bera test additionally flags mild non-normality in the residuals ($p = 0.015$). Together, these results suggest that a portion of onion price volatility comes from sources that climate data simply cannot capture, most likely abrupt supply-chain disruptions such as export bans, transportation bottlenecks, and hoarding behaviour, all of which can shift price variance sharply in ways that temperature and rainfall records do not reflect.

The forecasting framework developed in this study has direct practical value for the four main stakeholder groups in the Indian onion market, and can be integrated into existing market intelligence infrastructure without substantial investment. For farmers, climate-augmented monthly forecasts achieving a MAPE of 13-16 per cent over a one-year horizon represent a meaningful improvement over the informal price expectations that currently guide planting and storage decisions, where errors of 30-50 per cent are routine. For traders and aggregators, the hybrid ensemble's directional accuracy of 81.8 per cent offers a reliable basis for procurement timing, while the point forecasts from XGBoost provide the magnitude information needed for volume calibration. For consumers, the most relevant application lies in early-warning systems for price spike events. Pairing XGBoost mean forecasts with GARCH-X variance estimates would give the Department of Consumer Affairs probabilistic alerts about prices crossing critical thresholds, supporting timely interventions such as targeted subsidies and stocking limits before a crisis develops. For government agencies more broadly, the results make a case for shifting agricultural market intelligence toward forward-looking probabilistic frameworks. The methodology developed here is directly

Table 5: Post-estimation residual diagnostics on the SARIMAX best-statistical model.

Diagnostic	Statistic	p-value	Verdict
Ljung-Box Q(12) on residuals	14.87	0.516	No autocorrelation
ARCH-LM(6) on residuals	31.24	< 0.001	ARCH effects persist
Jarque-Bera on residuals	8.43	0.015	Mild non-normality
Durbin-Watson	1.989	—	No serial correlation

portable to other price-volatile commodities including tomato, garlic, and pulses where similar forecasting gaps exist (Ghosh *et al*., 2020; Choudhury *et al*., 2019).

This study carries three principal limitations that future research should address. First, the focus on Punjab may not extend to other major onion-producing states such as Maharashtra, Karnataka, and Rajasthan, which differ considerably in climate and market structure. Replicating the framework across these states would establish how broadly the findings generalise. Second, monthly climate data averages over within-month extreme events such as heat spikes or concentrated rainfall that can directly trigger yield losses. Shifting to weekly climate data would sharpen the causal link between weather and price, particularly for temperature and sunshine-sensitive growth stages. Third, deep learning architectures such as LSTM, Transformer models, and N-BEATS are constrained by the relatively limited training sample available for a single state. Three methodological extensions are particularly promising. The most immediate priority is to develop GARCH-X and SARIMAX-GARCH hybrid architectures that bring climate variables into the variance equation, directly addressing the conditional heteroscedasticity documented in the residual diagnostics.

Conclusions and Policy Implications

This study provides a comprehensive comparative evaluation of univariate and multivariate forecasting frameworks applied to onion prices in Punjab. The volatility analysis clarifies the scale of the challenge. Three independent volatility measures, coefficient of variation (66.2%), annualised return standard deviation (85.8%), and a GARCH persistence parameter of 0.937 onion price data of Punjab at the upper end of the agricultural volatility distribution. The majority of months record absolute log price changes above 15%, and the 22-year sample contains a maximum drawdown of −86.6%, establishing the severity of the price risk faced by farmers, traders, and consumers. Univariate models, regardless of their statistical sophistication, prove structurally inadequate for this market. All six univariate methods Seasonal Naïve, ARIMA, Holt-Winters, ARIMA-GARCH, TBATS, and N-BEATS produce MAPE values ranging from 36.3 per cent to 93.2 per cent. Incorporating five climate covariates into multivariate models, however, reduces MAPE by 65.7 per cent in aggregate a reduction that firmly establishes the role of weather as a structurally important driver of onion price formation.

The study revealed that among individual models, XGBoost delivers the best point-forecast accuracy, with the Gradient Boosting Machine close behind and SARIMAX performing best among classical statistical approaches. The hybrid ensemble achieves the highest directional accuracy at 81.8 per cent, producing forecasts approximately three times more accurate than the univariate models. Beyond the empirical results, three transferable design principles emerge for researchers working on analogous forecasting problems. First, multivariate frameworks that incorporate exogenous drivers are not optional for commodities where price formation is dominated by supply-side shocks, univariate methods are inadequate by design, irrespective of their complexity. Second, gradient-boosting architectures should be the default starting point for point forecasting, with classical models retained for interpretability and as ensemble components. Third, ensemble construction adds the most value for directional accuracy rather than magnitude precision. Overall, the onion price data for Punjab offers a reproducible methodological template for designing agricultural market intelligence systems in volatile agricultural commodities.

References

Apergis N and RezitisA N 2011. Food price volatility and macroeconomic factors: evidence from GARCH and GARCH-X estimates. *Journal of Agricultural and Applied Economics* **43**: 95-110. https://doi.org/10.1017/S1074070800004077

Bates J M and Granger C W J 1969. The combination of forecasts. *Journal of the Operational Research Society* **20**: 451-468. https://doi.org/10.1057/jors.1969.103

Birthal P S, Negi D S, Khan M T and Agarwal S 2015. Is Indian agriculture becoming resilient to droughts? Evidence from rice production. *Food Policy* **56**: 1-12.10.1016/j.foodpol.2015.07.005

Bollerslev T 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

Box G E P and Jenkins G M 1976. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco. https://doi.org/10.1111/jtsa.12194

Breiman L 2001. Random forests. *Machine Learning* **45**: 5-32. https://doi.org/10.1023/A:1010933404324

Cao Y L and Mohiuddin M 2019. Sustainable emerging country agro-food supply chains: fresh vegetable price formation mechanisms in rural China. *Sustainability* **11**: 2814. https://doi.org/10.3390/su11102814

Chand R 2010. Understanding the nature and causes of food inflation. *Economic and Political Weekly* **45**:10-13. https://www.researchgate.net/publication/265540689_understanding_the_nature_and_Causes_of_food_inflation

Chen T and Guestrin C 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, held at San Francisco, CA, USA. **17**: 785-794. https://doi.org/10.1145/2939672.2939785

Cheung L, Wang Y, Lau A S and Chan R M 2023. Using a novel clustered 3D-CNN model for improving crop future price prediction. *Knowledge-Based Systems* **260**: 110133. https://doi.org/10.1016/j.knosys.2022.110133

Choudhury K, Jha G K, Das P and Chaturvedi K K 2019. Forecasting potato price using ensemble artificial neural network. *Indian Journal of Extension Education* **55**: 60-64. https://www.researchgate.net/publication/332547919_Forecasting_Potato_Price_using_Ensemble_Artificial_Neural_Networks

Cleveland R B, Cleveland W S, McRae J E and Terpenning I 1990. STL: a seasonal-trend decomposition procedure based on LOESS. *Journal of Official Statistics* **6**: 3-73. https://www.math.unm.edu/~lil/Stat581/STL.pdf

Cortes C and Vapnik V 1995. Support-vector networks. *Machine Learning* **20**: 273–297. https://doi.org/10.1007/BF00994018

De Livera A M, Hyndman R J and Snyder R D 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association* **106**: 1513–1527. https://robjhyndman.com/papers/ComplexSeasonality.pdf

Dickey D A and Fuller W A 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**: 427-431. https://doi.org/10.2307/2286348

Diebold F X and Mariano R S 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253-263. https://doi.org/10.2307/1392185

Engle R F 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987–1007. https://doi.org/10.2307/1912773

FAO 2011. The State of Food Insecurity in the World. How does international price volatility affect domestic economies and food security? Rome: Food and Agriculture Organization of the United Natives.

Friedman J H 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* **29**: 1189-1232. https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.pdf

Ghosh H, Gurung B and Gupta P 2020. Fitting ARIMAX model for forecasting wholesale price of tomato in Bangalore. *Journal of the Indian Society of Agricultural Statistics* **68**: 265-271. https://doi.org/10.21203/rs.3.rs-7460198/v1

Headey D and Fan S 2008. Anatomy of a crisis: the causes and consequences of surging food prices. *Agricultural Economics* **39**: 375-391. https://doi.org/10.1111/j.1574-0862.2008.00345.x

Holt C C 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* **20**: 5-10. https://doi.org/10.1016/j.ijforecast.2003.09.015

Hyndman R J and Khandakar Y 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software* **27**: 1-22. https://doi.org/10.18637/jss.v027.i03

Kohzadi N, Boyd M S, Kermanshahi B and Kaastra I 1996. A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing* **10**: 169-181. https://doi.org/10.1007/3-540-48035-8_4

Kumar M and Sharma A 2015. Impact of climate variability on vegetable price dynamics in India: evidence from panel cointegration. *Agricultural Economics Research Review* **28**: 189-202. https://aeraindia.in/upload/16238157725331.pdf

Kwiatkowski D, Phillips P C B, Schmidt P and Shin Y 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* **54**: 159-178.

Moore H L 1917. Forecasting the Yield and the Price of Cotton. Macmillan, New York.

Oreshkin B N, Carpov D, Chapados N and Bengio Y 2020. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. In: Proceedings of the International Conference on Learning Representations (ICLR), held online, 26 April-1 May 2020. https://doi.org/10.48550/arXiv.1905.10437

Paul R K, Prajneshu and Ghosh H 2014. GARCH nonlinear time series analysis for modelling and forecasting of India's volatile spices export data. *Journal of the Indian Society of Agricultural Statistics* **68**: 123-131.

Shiferaw Y A 2019. Modelling the volatility of commodity price indices: an application of GARCH models. *International Journal of Finance and Economics* **24**: 388-400. https://doi.org/10.2139/ssrn.871166

Sparks, A. (2024) Nasapower: NASA-POWER Data from R https://doc.org//0.528/zenodo.1040727

Sun F, Meng X, Zhang Y, Wang Y, Jiang H and Liu P 2023. Agricultural product price forecasting methods: a review. *Agriculture* **13**: 1671. https://doi.org/10.3390/agriculture13091671

Timmermann A 2006. Forecast combinations. In: Handbook of Economic Forecasting **12**: 135-196. https://econpapers.repec.org/RePEc:eee:ecofch:1-04

Tothova M 2011. Main challenges of price volatility in agricultural commodity markets. In: Methods to Analyse Agricultural Commodity Price Volatility, ed: I Piot-Lepetit and R M'Barek. Springer, New York, pp. 13-29. https://doi.org/10.1007/978-1-4419-7634-5_2

Yin H, Jin D, Gu Y H, Park C J, Han S K and Yoo S J 2020. STL-ATTLSTM: vegetable price forecasting using STL and attention-mechanism-based LSTM. *Agriculture* **10**: 612. https://doi.org/10.3390/agriculture10120612

Received: March 03, 2026 Accepted: May 21, 2026