

Article

A Weekly E-Commerce Seed Price Index and Short-Term Forecasting System: Evidence from Tomato, Pepper, and Lettuce

Qianchuan Li ^{1,2}, Xiaodong Wang ^{3,*}, Xiaojing Qin ^{1,2,*}, Feng Yu ^{1,2}, Rupeng Luan ^{1,2}, Yang Ping ^{1,2}, Weijia Yang ³ and Lin Zhang ³

¹ Institute of Data Science and Agricultural Economics, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; liqianchuan@baafs.net.cn (Q.L.)

² Key Laboratory of Intelligent Seedling Technology Innovation, Ministry of Agriculture and Rural Affairs, P. R. China, Beijing 100097, China

³ Beijing Digital Agriculture and Rural Affairs Promotion Center, Beijing Municipal Bureau of Agriculture and Rural Affairs, Beijing 101117, China

* Correspondence: wangxd@nyncj.beijing.gov.cn (X.W.); qinxiaojing@baafs.net.cn (X.Q.)

Abstract

Measuring prices in agricultural e-commerce markets is complicated by promotional spikes, rapid product turnover, and specification heterogeneity. The aim of this study is to develop and validate a reproducible framework that couples promotion-robust weekly price measurement with calibrated short-term forecasting for agricultural e-commerce input markets. Drawing on 452,489 de-identified raw transaction records (448,308 retained after quality control) of tomato, pepper, and lettuce seeds from major Chinese e-commerce platforms (July 2020–October 2025), we construct a weekly Paasche-robust price index with a Top-10 SKU basket and winsorized prices, benchmarked against Laspeyres, Fisher, Törnqvist, and GEKS-TPD formulations. A time-dummy hedonic regression is applied for quality adjustment. Short-term forecasts are generated by a two-level ensemble combining SARIMAX, Temporal Fusion Transformer, and LightGBM via inverse-MAPE weighting, with adaptive conformal intervals for uncertainty quantification. The Paasche-robust index attenuates promotional noise while co-moving closely with superlative formulas (MA-12 correlations > 0.80). Machine-learning models reduce the one-week-ahead MAPE by 42–58% relative to SARIMAX, and the weighted ensemble achieves the best four-week-ahead accuracy for two of three crops. Conformal intervals deliver near-nominal coverage at short horizons. The framework offers a reproducible, deployable tool for real-time monitoring of agricultural e-commerce input markets and supports evidence-based decisions by farmers and policymakers.

Keywords: seed price index; e-commerce; ensemble forecasting; machine learning; conformal prediction; Paasche-robust; vegetable seeds



Academic Editors: Pascal Ghazalian and Deborah Bentivoglio

Received: 30 April 2026

Revised: 26 June 2026

Accepted: 13 July 2026

Published: 29 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Seeds are intermediate agricultural inputs whose prices transmit upstream cost and scarcity signals to downstream planting decisions [1,2]. In China, the rapid expansion of e-commerce platforms for agricultural supplies has generated high-frequency transaction data that offer an unprecedented opportunity for real-time price monitoring [3–5]. However, e-commerce seed markets exhibit three features that complicate price measurement: (i) pronounced promotional extremes that create artificial spikes, (ii) frequent product substitution as stock-keeping units (SKUs) enter and exit the market, and (iii) substantial

specification heterogeneity in pack size, breeding type, and germination rate [6,7]. These factors introduce chain drift, substitution bias, and promotion-induced volatility into conventional index formulas, while simultaneously challenging short-term forecasting through nonstationarity and episodic variance shifts [8,9].

The price index literature has developed several remedies. Multilateral methods such as GEKS and time-product-dummy (TPD) variants reduce chain drift and address product substitution, while time-dummy hedonic (TDH) regressions control for measurable quality differences [10–13]. A related scanner-data study demonstrates how multilateral approaches can eliminate chain drift arising from high-frequency chaining [14]. On the forecasting side, forecast combination methods that integrate complementary learners have demonstrated horizon-wise gains across multiple domains [15–18], and conformal prediction provides distribution-free coverage guarantees without parametric assumptions [19–21]. Recent work on vegetable price forecasting using optimized neural-network time-series models has shown promising results with daily and weekly data from wholesale markets [22,23]. Yet the literature lacks a fully reproducible, weekly, e-commerce seed price index integrated with operational forecasting and calibrated uncertainty quantification.

Within agricultural economics, short-term price forecasting has progressed from classical time-series models toward machine-learning and deep-learning approaches. For vegetables specifically, recent Chinese studies report substantial accuracy gains from non-linear learners: optimized neural-network time-series models outperform conventional baselines on weekly vegetable series [22], long short-term memory (LSTM) networks trained on Beijing wholesale-market data achieve coefficients of determination above 0.95 for short-term vegetable price prediction [23], comprehensive reviews document a clear shift from ARIMA-type and econometric specifications toward tree-based ensembles and neural architectures that better accommodate the nonlinear, non-stationary behavior of fresh-produce prices [24], and mixture-of-experts architectures further improve robustness across heterogeneous price regimes [25]. These contributions establish the value of non-linear learners but share three limitations relevant to the present study: they rely almost exclusively on offline wholesale-market prices rather than e-commerce transactions; they target point forecasts without calibrated uncertainty quantification; and they are decoupled from the upstream measurement problem of constructing a consistent, promotion-robust index from heterogeneous micro-data. Forecast combination theory, by contrast, consistently shows that pooling complementary learners reduces error and improves robustness relative to selecting any single model [15–18], a property that is especially valuable when the dominant learner changes with the forecast horizon. The present study addresses these gaps jointly within a single e-commerce setting.

This work makes four contributions. First, we construct a weekly seed price index centered on a Paasche-robust formulation (report-period weights with winsorized prices) and benchmark it against Laspeyres, Fisher, Törnqvist, and GEKS-TPD for long-run consistency. Second, we implement a lightweight TDH quality adjustment to control for specification heterogeneity. Third, we build a multi-model forecasting system (TFT, SARI-MAX, LightGBM → inverse-MAPE weighted ensemble) evaluated with Diebold–Mariano tests and equipped with conformal prediction intervals. Fourth, we release a minimal reproducibility package (weekly indices and scripts) with controlled audit access to micro-records. The framework is demonstrated on tomato, pepper, and lettuce seeds—three crops that collectively represent a substantial share of China’s vegetable seed market—using 452,489 e-commerce transactions spanning five years.

The overarching aim of this study is to develop and validate an integrated, reproducible framework that couples high-frequency price measurement with operational short-term forecasting for agricultural e-commerce input markets. To this end, we address four

research questions. RQ1: Can a Paasche-robust formulation with a Top-10 basket and winsorized prices attenuate promotion-induced noise while remaining consistent with superlative index formulas? RQ2: Do machine-learning learners (LightGBM, Temporal Fusion Transformer) deliver materially lower short-horizon forecast error than a seasonal econometric benchmark (SARIMAX)? RQ3: Does an inverse-error weighted ensemble provide more robust accuracy across forecast horizons than its constituent base learners? RQ4: Do adaptive conformal prediction intervals attain near-nominal coverage for these series? Correspondingly, we test three hypotheses: H1, the robust index co-moves with superlative formulas (MA-12 correlation > 0.8) while exhibiting lower return dispersion than the Laspeyres index; H2, machine-learning models reduce one-week-ahead error relative to SARIMAX; and H3, the weighted ensemble is no worse than, and at longer horizons superior to, the best individual learner.

The remainder of this paper is organized as follows. Section 2 describes the dataset, quality control procedures, price index construction methods, and ensemble forecasting architecture. Section 3 presents empirical results on transaction volumes, weekly price indices, index formula comparisons, forecast accuracy, and sensitivity analyses. Section 4 discusses the methodological implications, policy relevance, and limitations of the proposed framework. Section 5 concludes the paper.

2. Materials and Methods

2.1. Methodological Overview

The methodology comprises four stages (Figure 1): (1) data collection and quality control, (2) price index construction, (3) ensemble forecasting, and (4) evaluation and deployment. Each stage is described below.

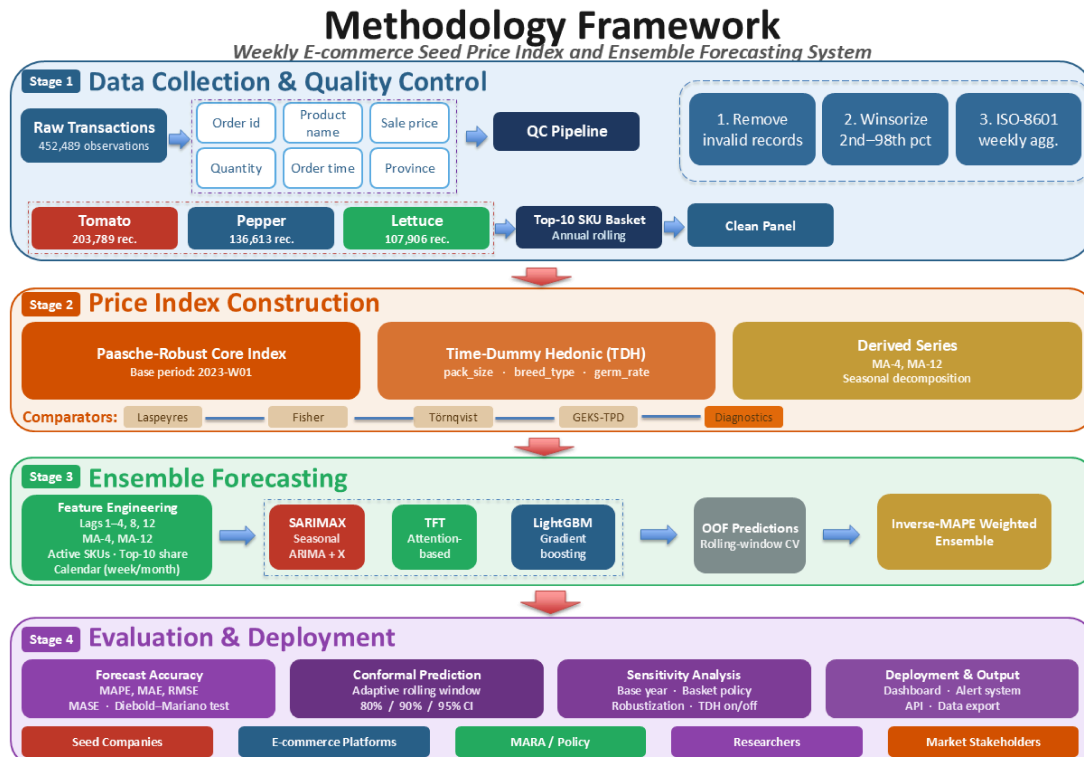


Figure 1. Methodological framework. The pipeline proceeds through four stages: data collection and quality control (blue), price index construction with Paasche-robust and comparator formulas (orange), ensemble forecasting with three complementary base learners (green), and evaluation with conformal prediction intervals (purple). Arrows indicate data flow between stages.

2.2. Data Source and Quality Control

We compiled de-identified product-level transaction records from major Chinese e-commerce platforms covering three vegetable seed categories: tomato (*Solanum lycopersicum*), pepper (*Capsicum* spp.), and lettuce (*Lactuca sativa*). The dataset spans July 2020 to October 2025 and contains 452,489 raw observations. Each record includes a unique order identifier, product name, unit sale price (CNY), quantity sold, order timestamp, and buyer geographic information (province and city, covering 32–33 provinces). All records were de-identified prior to analysis in compliance with platform terms of service. This study analyzed de-identified commercial transaction records and did not involve human subjects; institutional ethics review was therefore not required per the guidelines of the Beijing Academy of Agriculture and Forestry Sciences.

For reproducibility, the data provenance is specified as follows. Records were obtained from the principal agricultural input storefronts operating on China’s leading e-commerce platforms (Taobao/Tmall and JD.com). Prices are recorded in Chinese yuan (CNY) per order and quantities as integer units sold; the buyer-location field resolves to the provincial and municipal levels. The data cover 31 provincial-level administrative divisions of mainland China; Hong Kong, Macao, and Taiwan, and—in a small number of weeks—Tibet and Qinghai, were not consistently represented and are therefore omitted from geographic summaries.

Quality control (QC) proceeded in three steps. First, we removed records with missing or non-positive prices and quantities. Second, we filtered extreme prices (sale price > 500 CNY) that likely represent bundled or mis-coded entries. Third, we aggregated transactions to ISO 8601-1 weekly frequency (Monday-start weeks) [26] to construct product-week panels. After QC, the working dataset comprised 203,789 tomato records (270 complete weeks), 136,613 pepper records (163 weeks), and 107,906 lettuce records (184 weeks). The difference in weekly coverage across crops reflects platform data availability for the Top-10 basket products in each category. One anomaly was identified in the pepper index: 2024-W11 recorded a zero value attributable to a single active product with a price anomaly; this observation was retained but flagged for sensitivity analysis.

Because the three crops entered the e-commerce channel at different times and with different listing densities, the resulting index series differ in length and contain occasional internal gaps (Table 1). A product-week was admitted to the index only when at least four Top-10 basket products recorded valid transactions, so that each weekly index rests on a sufficient cross-section; weeks failing this minimum-coverage rule were treated as missing rather than imputed with potentially spurious values. The tomato series satisfies this criterion almost continuously from 2020-W29, whereas the pepper and lettuce series become continuously populated only from late 2021 (2021-W52), reflecting the later online expansion of these categories. For the forecasting experiments, each series was restricted to its longest continuously covered span, and short internal gaps (two weeks or fewer) were linearly interpolated on the log scale; no extrapolation was applied across longer gaps. This procedure accounts for the differing weekly coverage across crops (tomato 270 weeks, pepper 163, and lettuce 184) reported in Table 1 and reported in Section 3.2.

Table 1. Dataset summary after quality control.

	Tomato	Pepper	Lettuce
Raw transactions	206,478	137,696	108,315
After QC	203,789	136,613	107,906
Complete weeks	270	163	184
Raw data range	2020-07 to 2025-10	2020-07 to 2025-10	2020-07 to 2025-10

Table 1. Cont.

	Tomato	Pepper	Lettuce
Index series range	2020-W29 to 2025-W43	2021-W52 to 2025-W43	2021-W52 to 2025-W43
Price range (CNY)	0.01–500.00	0.02–500.00	0.01–500.00
Median weekly products	24 [19–22,24–32]	8 [4,6–10]	6 [4,6–11]
Winsor bounds (2–98%)	[1.98, 25.80]	[1.41, 20.03]	[1.11, 23.23]
Top-10 basket share	47.2%	52.8%	58.1%

Values in brackets denote interquartile ranges. The index base period is 2023-W01. Raw data range refers to the time span of original transaction records; index series range refers to weeks with sufficient product coverage to compute the Paasche-robust index. Top-10 basket share is the cumulative sales share of the ten highest-volume SKUs during 2022–2024. Winsor bounds refer to the 2nd and 98th percentile unit price cutoffs used for robust index construction. Pepper index at 2024-W11 recorded zero due to a single product-week anomaly.

2.3. Price Index Construction

Figure 2 summarizes the index construction pipeline. We targeted a rolling Top-10 SKU basket ranked by cumulative annual sales volume. The basket is fixed within each calendar year and re-selected annually to absorb structural market changes without sacrificing cross-period comparability. The base period is the first full week of 2023 (2023-W01), chosen to avoid first-week missingness and to align with a period of stable market conditions.

Index Construction Pipeline

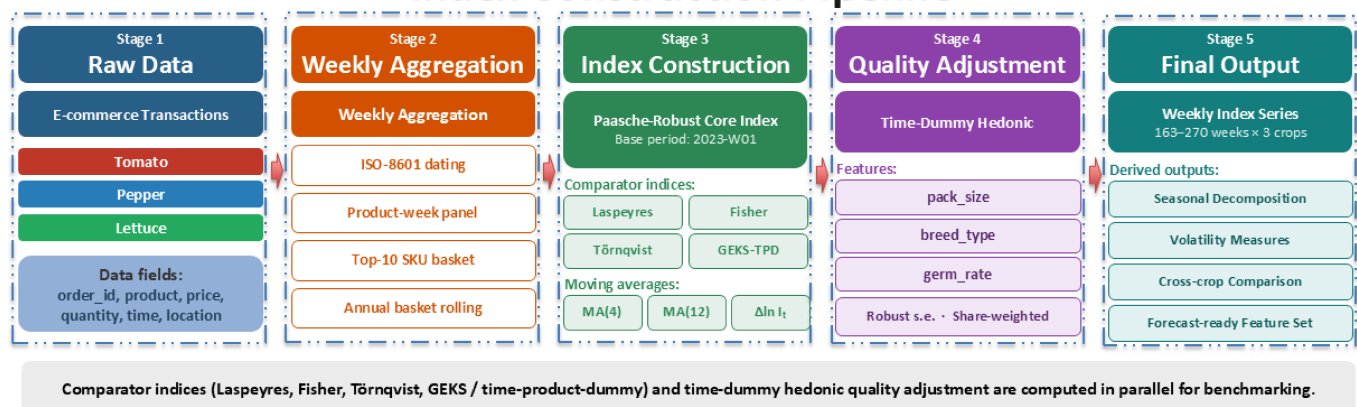


Figure 2. Index construction pipeline. Raw transactions undergo weekly aggregation with ISO-8601 dating, followed by Paasche-robust index computation with Top-10 basket weighting, quality adjustment via winsorization (2nd–98th percentile), and final index series output with moving averages (MA-4 and MA-12).

The primary index is the Paasche-robust formulation:

$$P_t = 100 \times \frac{\sum_{i \in S_t} \tilde{p}_{t,i} \cdot q_{t,i}}{\sum_{i \in S_t} \tilde{p}_{0,i} \cdot q_{t,i}} \quad (1)$$

where $\tilde{p}$ denotes winsorized prices (2nd–98th percentile), $q_{t,i}$ is the report-period quantity for product i , and subscript 0 denotes the 2023-W01 base. The Paasche formulation uses current-period weights, making it responsive to contemporary market structure while the winsorization suppresses promotional outliers [27,28]. For benchmarking, we computed Laspeyres (base-period weights), Fisher (geometric mean of Laspeyres and Paasche), Törnqvist (logarithmic mean with share weights), and GEKS-TPD (multilateral transitive) indices on rolling windows [29–31].

The 4-week (MA-4) and 12-week (MA-12) moving averages were chosen to summarize monthly and quarterly dynamics, respectively: MA-4 suppresses isolated single-week promotional spikes while preserving sub-seasonal movement, whereas MA-12 isolates the seasonal-to-trend component used for cross-crop comparison. Both window lengths align with the natural monthly and quarterly reporting cadences of agricultural price-monitoring systems.

To control for specification heterogeneity, we estimated a time-dummy hedonic (TDH) regression on product-week prices:

$$\ln p_{it} = \beta_0 + \sum_k \beta_k x_{k,it} + \gamma_t + \varepsilon_{it} \quad (2)$$

where x includes pack size, breeding type (hybrid/open-pollinated), germination rate, and platform fixed effects. The quality-adjusted index is $I_t = \exp(\hat{\gamma}_t)$. We used share-weighted least squares with robust standard errors [32].

2.4. Ensemble Forecasting

We forecast weekly price indices using a two-level ensemble forecasting framework (Figure 3). At Level 1, three complementary base learners generate out-of-fold (OOF) predictions via rolling-window cross-validation: (i) SARIMAX, which captures seasonal and autoregressive structure with exogenous covariates [33]; (ii) LightGBM, a gradient-boosted decision tree that models nonlinearities and feature interactions [34,35]; and (iii) Temporal Fusion Transformer (TFT), an attention-based architecture designed for multi-horizon time-series forecasting with interpretable variable selection [36]. At Level 2, an inverse-MAPE weighted averaging scheme combines the base-learner forecasts to produce final forecasts at horizons $h = 1, 2, 3$, and 4 weeks ahead. Weights are computed as $w_k = \frac{(\frac{1}{MAPE_k})}{\sum_j (\frac{1}{MAPE_j})}$, where $MAPE_k$ is the mean absolute percentage error of base learner k on a calibration window comprising the first half of the rolling evaluation period. This non-parametric approach, rooted in the forecast combination literature [37–39], avoids the overfitting inherent in regression-based meta-learning with limited calibration samples, while ensuring that more accurate base learners receive proportionally higher weight. The mean absolute percentage error (MAPE) used here for weighting is formally defined in Section 2.5.

The feature set comprised lagged index values (lags 1–4, 8, 12), 4-week and 12-week moving averages (MA-4 and MA-12), weekly log returns ($\Delta \ln$ index), the number of active SKUs, Top-10 basket concentration share, and calendar indicators (week-of-year, month). Hyperparameters were tuned via rolling-window cross-validation with a minimum training window of 52 weeks and a step size of 4 weeks. SARIMAX seasonal orders were selected by AICc; the LightGBM learning rate and tree depth were grid-searched; and TFT used early stopping with dropout regularization.

2.5. Evaluation Metrics and Uncertainty Quantification

Forecast accuracy was assessed using mean absolute percentage error (MAPE), mean absolute error (MAE), root mean squared error (RMSE), and mean absolute scaled error (MASE) [40]. Diebold–Mariano tests [41] were used to evaluate whether the ensemble significantly outperformed each base learner at each forecast horizon, with Newey–West corrections for serial correlation in multi-step forecast errors.

For completeness, the accuracy metrics are defined as follows. Let y_t denote the realized index and $\hat{y}_t$ the forecast over an evaluation set of size n . Then, $MAPE = \frac{100}{n} \sum_t \frac{|y_t - \hat{y}_t|}{|y_t|}$, $MAE = \frac{1}{n} \sum_t |y_t - \hat{y}_t|$, $RMSE = \sqrt{\frac{1}{n} \sum_t (y_t - \hat{y}_t)^2}$, and $MASE = \frac{MAE}{\frac{1}{n-m} \sum_t |y_t - y_{t-m}|}$, where the

denominator is the in-sample mean absolute error of a seasonal naive forecast with seasonal period m ($m = 52$ for weekly data). MAPE and MAE are interpreted on the percentage and index scales, RMSE penalizes large errors more heavily, and MASE is scale-free and therefore comparable across crops. Although all four metrics were computed, MAPE is reported in the main text as the primary percentage metric, with the remaining metrics available in the reproducibility package.

$$DM = \frac{d}{\sqrt{\frac{\hat{\sigma}_d^2}{n}}} \quad (3)$$

For uncertainty quantification, we implemented adaptive conformal prediction with a rolling calibration window of 26 weeks. At each forecast origin, prediction intervals at nominal coverage levels of 80%, 90%, and 95% were constructed from the empirical distribution of recent absolute residuals. We report realized coverage rates and average interval widths to assess calibration quality.

$$C_t(\alpha) = [\hat{y}_t - \hat{q}_{1-\alpha}, \hat{y}_t + \hat{q}_{1-\alpha}] \quad (4)$$

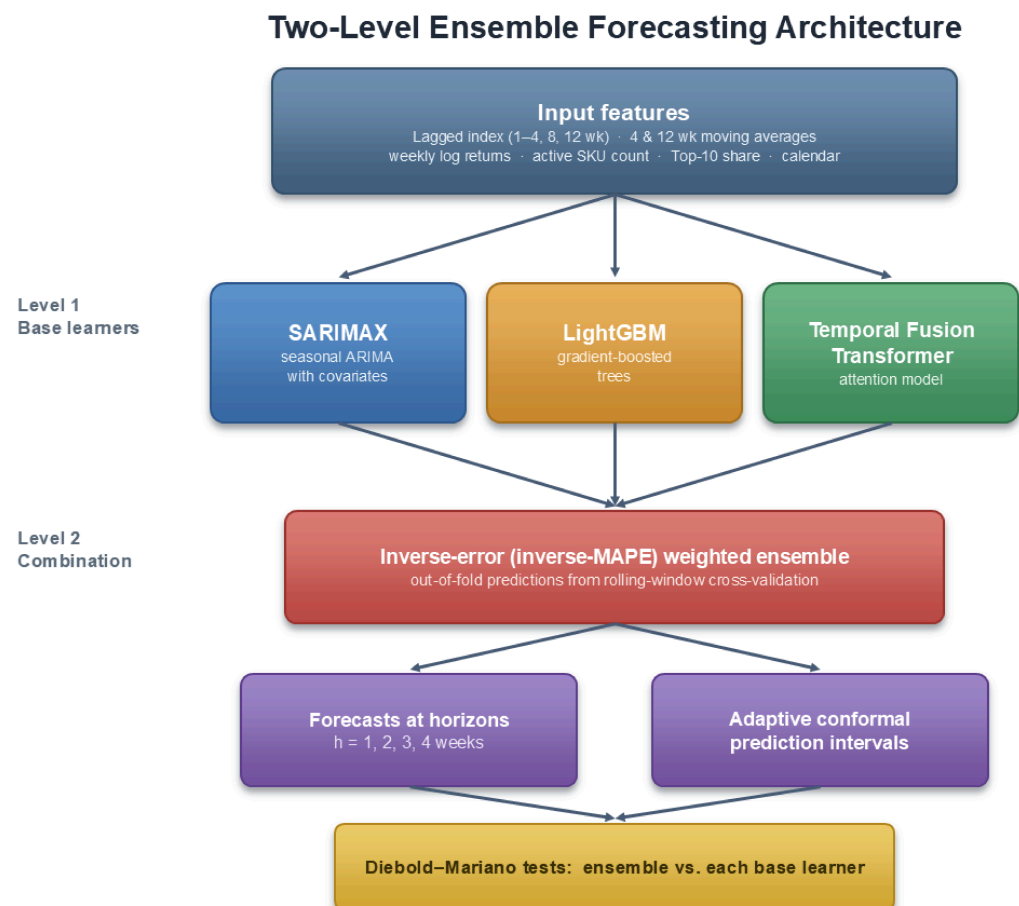


Figure 3. Ensemble forecasting architecture. Input features (lagged index values, moving averages, active SKU counts, Top-10 share, and calendar variables) feed three Level-1 base learners (SARIMAX, TFT, LightGBM). Out-of-fold predictions from rolling-window cross-validation are combined by an inverse-MAPE weighted averaging scheme at Level 2. Diebold–Mariano tests compare the ensemble output against each base learner.

2.6. Software and Reproducibility

The original analyses were implemented in Python 3.11. To support reproducibility, the supplied code is specified for the following environment: NumPy 1.26.4, pandas 2.2.2, statsmodels 0.14.2, LightGBM 4.5.0, scikit-learn 1.5.1, PyTorch 2.4.0, and PyTorch Forecasting 1.2.0. These versions define the environment for rerunning the supplied code and should not be interpreted as a reconstruction of the original historical environment, which was not retained. The core functions for index construction, the comparator formulas, the inverse-MAPE ensemble, and the conformal prediction routine are provided in the Supplementary Materials (Code S1); the weekly index series and quality-control statistics are provided in Tables S1 and S2. Random seeds were fixed for all stochastic components to ensure run-to-run reproducibility.

3. Results

3.1. Transaction Volume and Seasonal Patterns

Figure 4 presents the data overview. Weekly transaction volumes (Figure 4a) reveal a structural shift across the study period: moderate and stable volumes during 2022–2024 (dominated by seed packets in standard e-commerce packaging) gave way to a sharp surge in early 2025, driven primarily by seedling products. Tomato transactions increased from approximately 24,000 in calendar year 2023 to over 156,000 in the first seven months of 2025, reflecting the growing integration of seedling sales into e-commerce channels.

Seasonal decomposition (Figure 4b–d) revealed crop-specific seasonal patterns rather than a uniform planting-season dynamic. Tomato prices peaked in August (mean index 143.4) and December (140.7), reflecting summer transplanting demand and year-end promotional cycles, with a trough in March (94.2). Pepper exhibited an autumn-dominated pattern, peaking in September (103.7) with its nadir in December (74.4). Lettuce showed a September peak (117.1) and December trough (84.6). These differences underscore the importance of individualized forecasting models. Tomato displayed the highest volatility (coefficient of variation = 0.62), followed by pepper (0.28) and lettuce (0.20).

3.2. Weekly Price Index Dynamics

Figure 5 displays the weekly Paasche-robust price indices (2023-W01 = 100) for all three crops. Each panel shows the raw weekly index alongside 4-week (MA-4) and 12-week (MA-12) moving averages. Several features emerge. First, all three crops exhibit seasonal patterns with crop-specific timing, consistent with the monthly analysis reported above. Second, tomato displayed the widest index range (13.8–505.8, mean = 114.6) and highest volatility (CV = 0.62), while lettuce was the most stable (range 32.5–175.2, mean = 103.5, CV = 0.20). Pepper occupied an intermediate position (range 0.0–120.7, mean = 86.5, CV = 0.28). Third, a pronounced upward shift is visible in early 2025 across all crops, reflecting the structural entry of seedling products into the e-commerce marketplace.

The shorter pepper and lettuce series (beginning 2021-W52, versus 2020-W29 for tomato; Figure 5b,c) follow directly from the later online entry of these categories and the minimum-coverage rule described in Section 2.2. The common, fully populated window across all three crops spans 2022-W01 to 2025-W43 and is used for all cross-crop comparisons; crop-specific spans are retained only for the single-crop index plots in Figure 5.

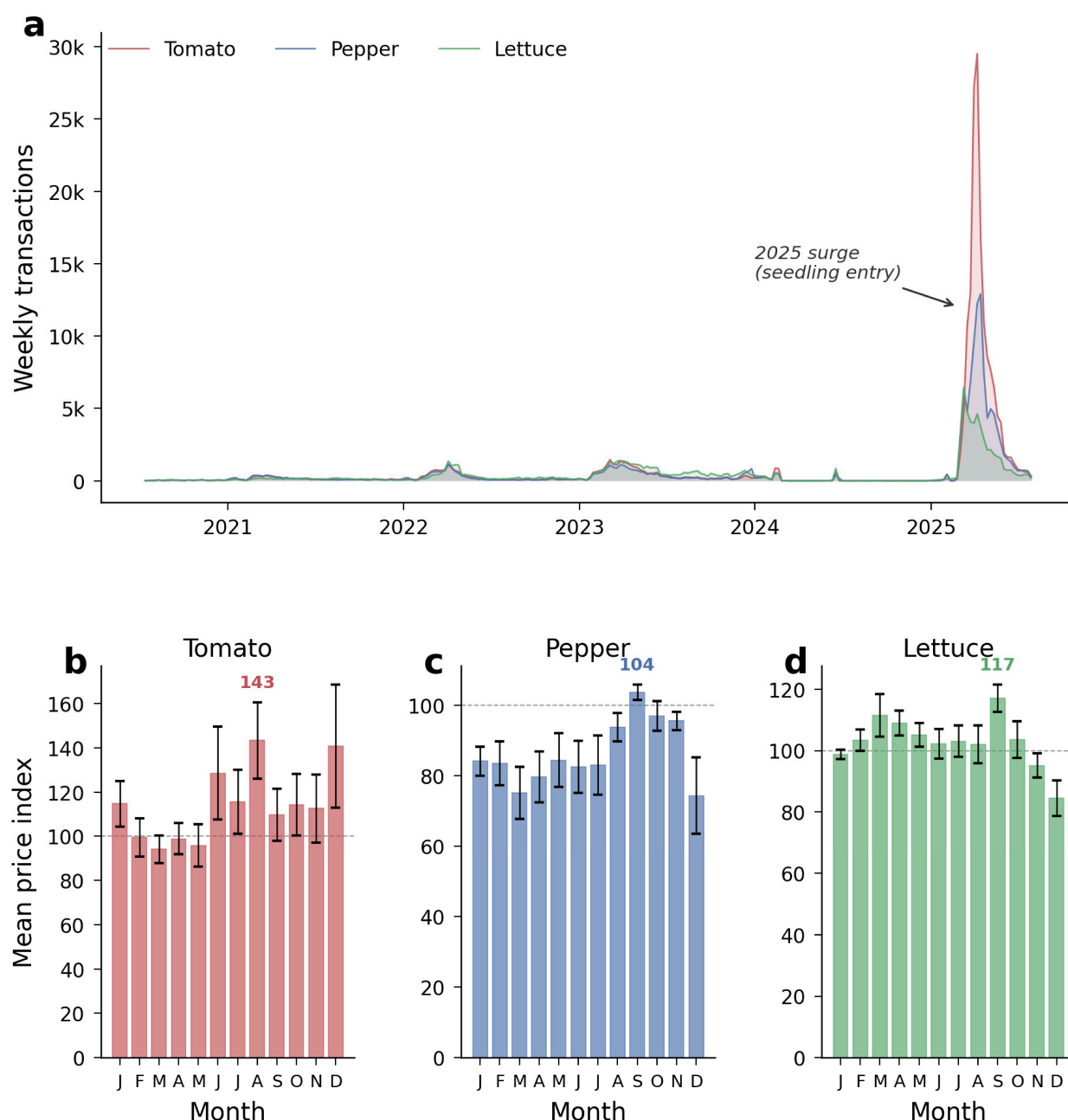


Figure 4. Data overview and seasonality. (a) Weekly transaction volume time series for three crops (filled areas). (b–d) Monthly mean transaction volumes with standard error bars for tomato, pepper, and lettuce, respectively. Tomato prices peak in August and December, pepper in September, and lettuce in September, with crop-specific seasonal dynamics.

The three-crop overlay (Figure 6) facilitates direct comparison of long-run trends using the MA-12 smoother. While all three indices co-move at the seasonal frequency, their amplitudes and timing differ. Pairwise MA-12 correlations ranged from 0.30 (tomato–pepper) to 0.80 (pepper–lettuce), indicating moderate seasonal co-movement with substantial crop-specific variation. Tomato leads the autumn ramp-up relative to pepper and lettuce, likely reflecting earlier transplanting schedules. The convergence of all three indices near the base level (100) during mid-2023 validates the base-period selection.

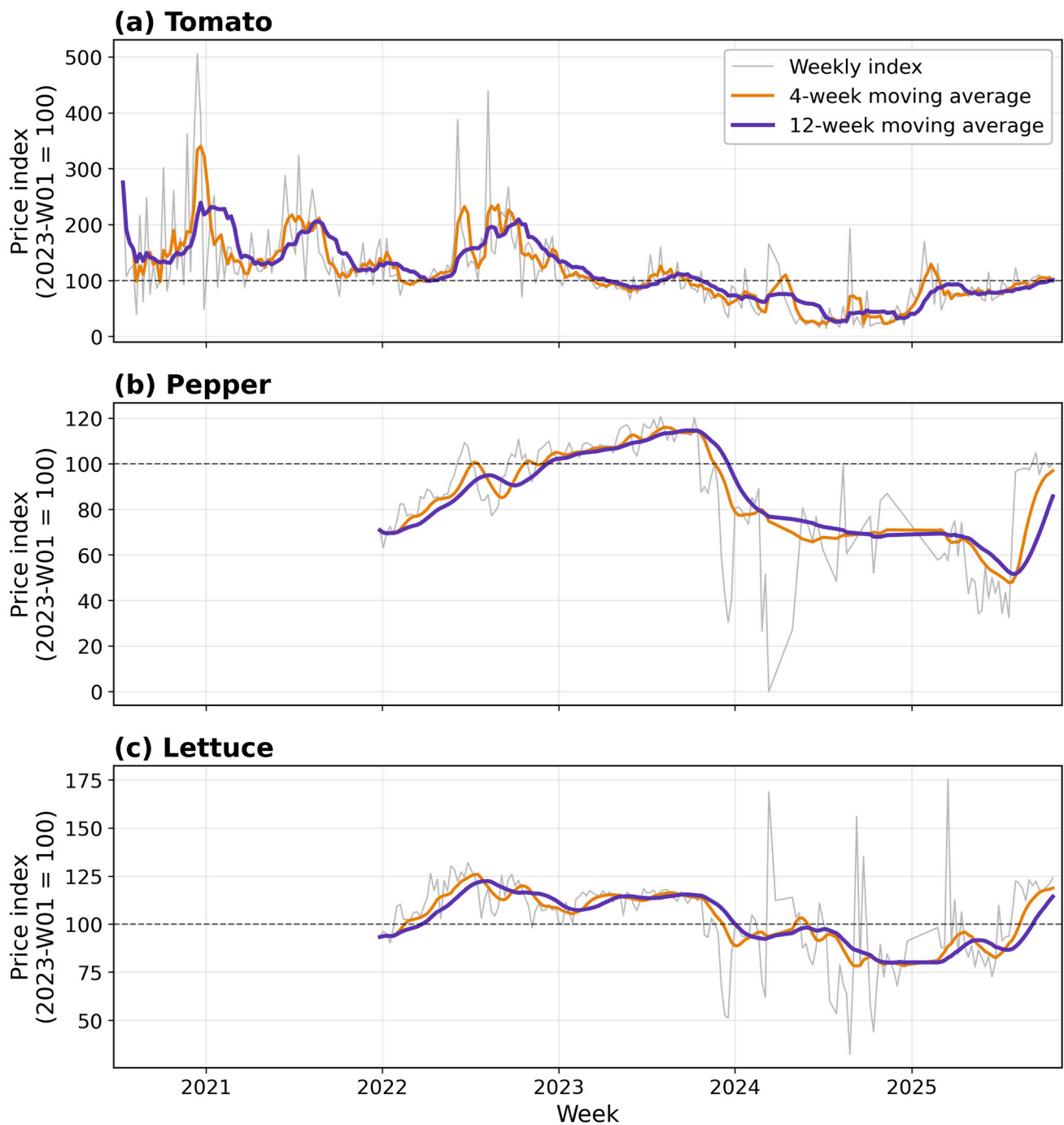


Figure 5. Weekly Paasche-robust price indices (2023-W01 = 100) for (a) tomato, (b) pepper, and (c) lettuce. Thin lines show raw weekly values; orange lines show the 4-week moving average (MA-4); and purple lines show the 12-week moving average (MA-12). Shaded areas highlight the index envelope. The horizontal dashed line marks the base level (100).

3.3. Index Formula Comparison

Figure 7 compares the Paasche-robust index with four alternative formulations for the tomato exemplar. The Paasche-robust index (thick red line) tracks closely with the Fisher and Törnqvist superlative indices, with pairwise MA-12 correlations exceeding 0.80. The Laspeyres index (blue dashed) exhibits systematically higher volatility and is more sensitive to promotional episodes, as indicated by the annotated spikes. The GEKS-TPD multilateral index provides a drift-resistant long-run anchor but sacrifices some short-run responsiveness.

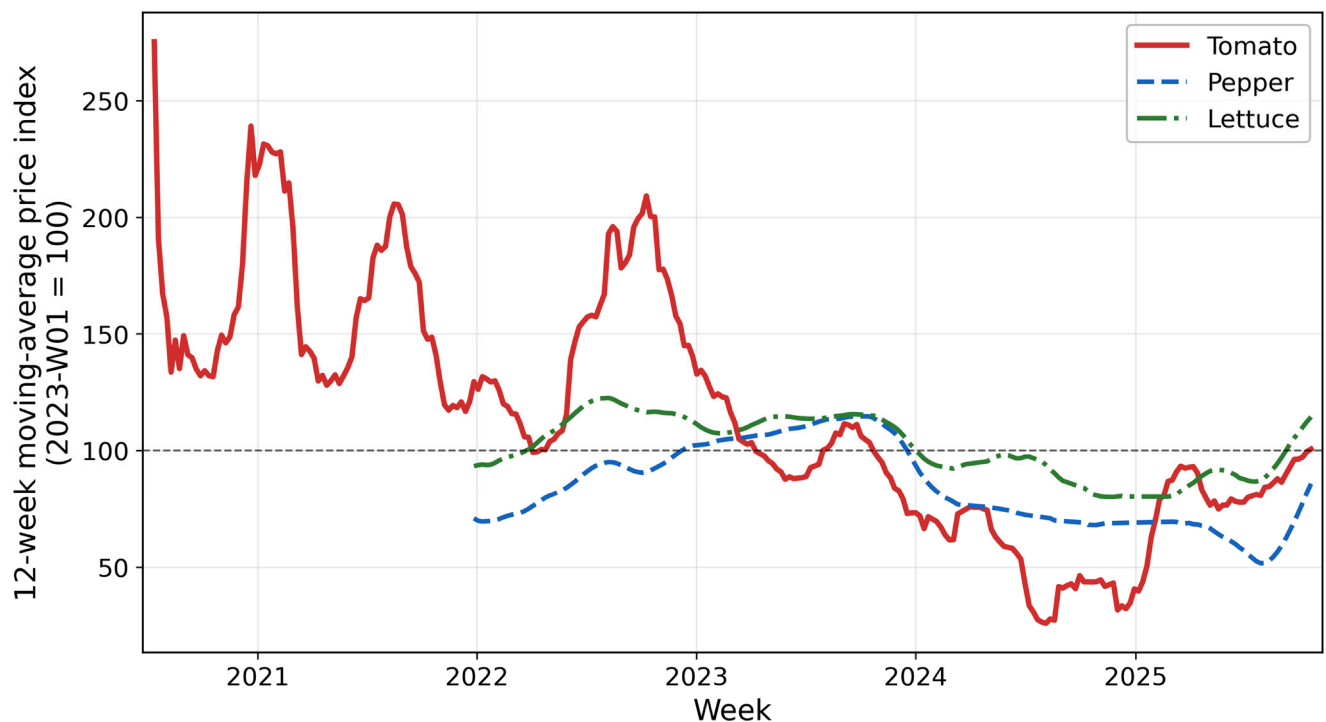


Figure 6. Three-crop MA-12 overlay comparison. The 12-week moving averages for tomato (red), pepper (blue dashed), and lettuce (green dash-dot) are plotted on a common axis. The horizontal dashed line marks the base level (100). Pairwise MA-12 correlations: tomato–pepper $\rho = 0.30$, tomato–lettuce $\rho = 0.65$, pepper–lettuce $\rho = 0.80$.

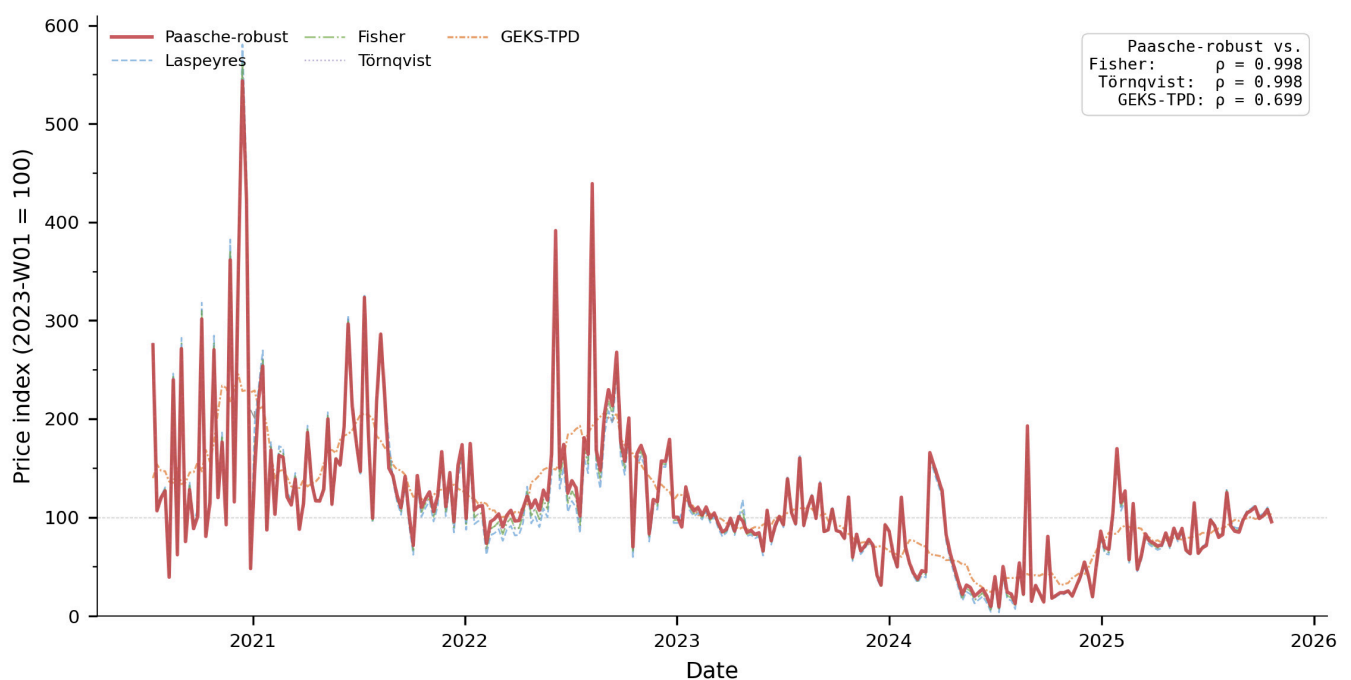


Figure 7. Index formula comparison for tomato (2023-W01 = 100). Five index families are shown: Paasche-robust (thick red), Laspeyres (blue dashed), Fisher (green dash-dot), Törnqvist (purple dotted), and GEKS-TPD (orange). Arrows annotate episodes where the Laspeyres index diverges due to promotional sensitivity. The five index families are almost perfectly correlated over the common period (Fisher and Törnqvist $\rho \sim 0.998$; GEKS-TPD $\rho \sim 0.70$), so their trajectories largely coincide on a single axis; this near-coincidence is itself the central finding, indicating that the winsorization-based robustization preserves the underlying price signal rather than distorting it.

Figure 7 uses tomato as a representative exemplar because it exhibits the widest price range and the highest volatility, making formula divergences most visible. Forecast accuracy for all three crops is summarized in Table 2 and discussed in Section 3.4.

Table 2. Forecast accuracy (MAPE, %) and Diebold–Mariano test results.

Crop	Horizon	SARIMAX	TFT	LightGBM	Ensemble	DM <i>p</i> -Value
Tomato	h = 1	13.0	7.5	7.7	7.5	0.343
	h = 2	15.0	11.0	11.1	11.5	0.386
	h = 3	16.3	13.0	16.5	13.8	0.220
	h = 4	18.1	17.5	17.0	16.8	0.440
Pepper	h = 1	16.1	7.3	6.8	10.9	<0.001
	h = 2	18.5	12.9	12.6	15.4	0.004
	h = 3	21.4	18.3	19.4	20.2	0.018
	h = 4	23.8	26.3	28.1	25.1	0.898
Lettuce	h = 1	7.3	3.6	3.5	3.9	0.149
	h = 2	8.5	5.0	5.8	5.8	0.050
	h = 3	9.5	7.8	8.0	8.2	0.605
	h = 4	10.5	10.6	10.6	10.1	0.140

Values are MAPE (%) on the rolling evaluation window. Bold indicates best performance at each horizon. DM *p*-values test the null hypothesis of equal predictive accuracy between the ensemble and the best individual machine-learning model at each horizon, with Newey–West standard errors. Test-period sample sizes: tomato *n* = 40, pepper *n* = 24, lettuce *n* = 27.

Figure 8 quantifies the distributional differences through violin plots of weekly log returns ($\Delta \ln$ index). For all three crops, the Paasche-robust return distribution is narrower (lower interquartile range) and exhibits fewer extreme outliers compared with the raw (unwinsorized) counterpart. The reduction in upper-tail outliers is particularly notable for lettuce, where promotional activity is most frequent. These results confirm that the winsorization embedded in the Paasche-robust formulation effectively attenuates promotion-driven noise without distorting the central tendency.

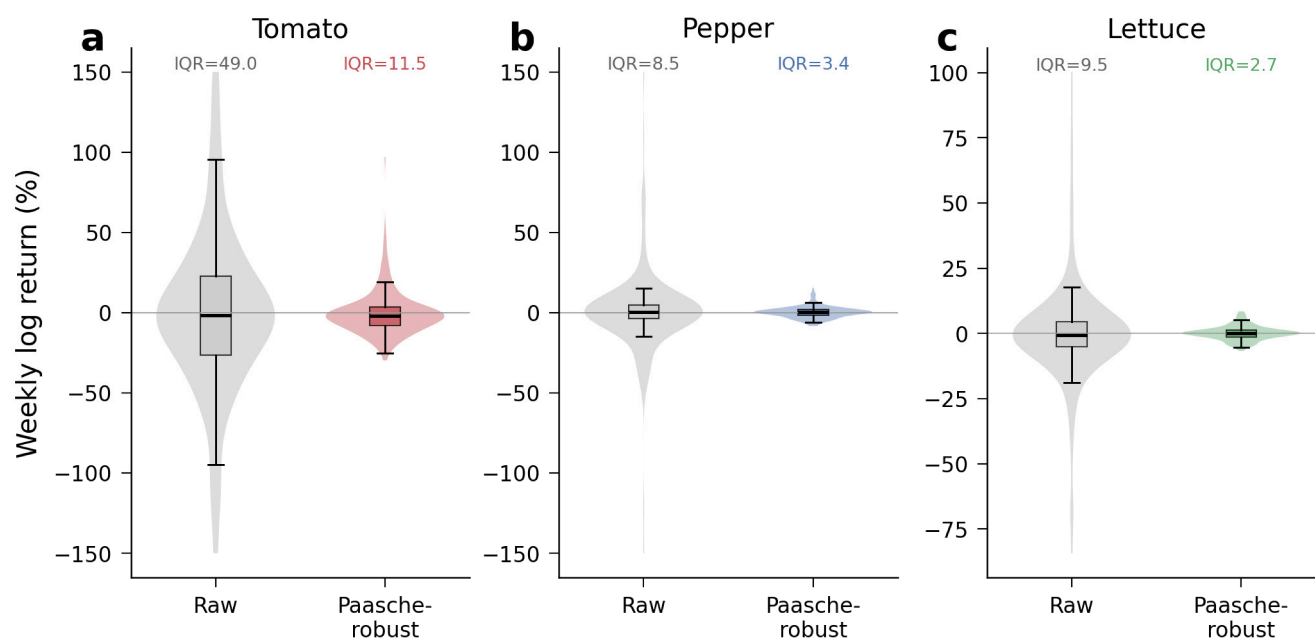


Figure 8. Weekly return distributions ($\Delta \ln$ index) for (a) tomato, (b) pepper, and (c) lettuce. In each panel, violin plots (density contours) overlaid with box plots (median and interquartile range) compare the raw (unwinsorized) and Paasche-robust formulations. The paired distributions show that Paasche-robust returns have narrower interquartile ranges and fewer extreme outliers for all three crops.

3.4. Forecast Accuracy and Ensemble Improvement

Figure 9 presents rolling one- to four-week-ahead forecasts with conformal prediction intervals for the tomato index. At the shortest horizon ($h = 1$), the best individual model (TFT) achieves an MAPE of 7.5% and the ensemble 7.7%, both closely tracking realized values. As the horizon extends, forecast uncertainty increases: the best-model MAPE rises to 11.0% ($h = 2$), 13.0% ($h = 3$), and 16.8% ($h = 4$). The conformal intervals widen commensurately, with the 90% bands achieving realized coverage rates of 89.3% at $h = 1$ and 78.0% at $h = 4$ —near-nominal at short horizons with expected degradation at longer horizons where calibration samples are limited.

Tomato: Rolling forecast with adaptive conformal prediction intervals

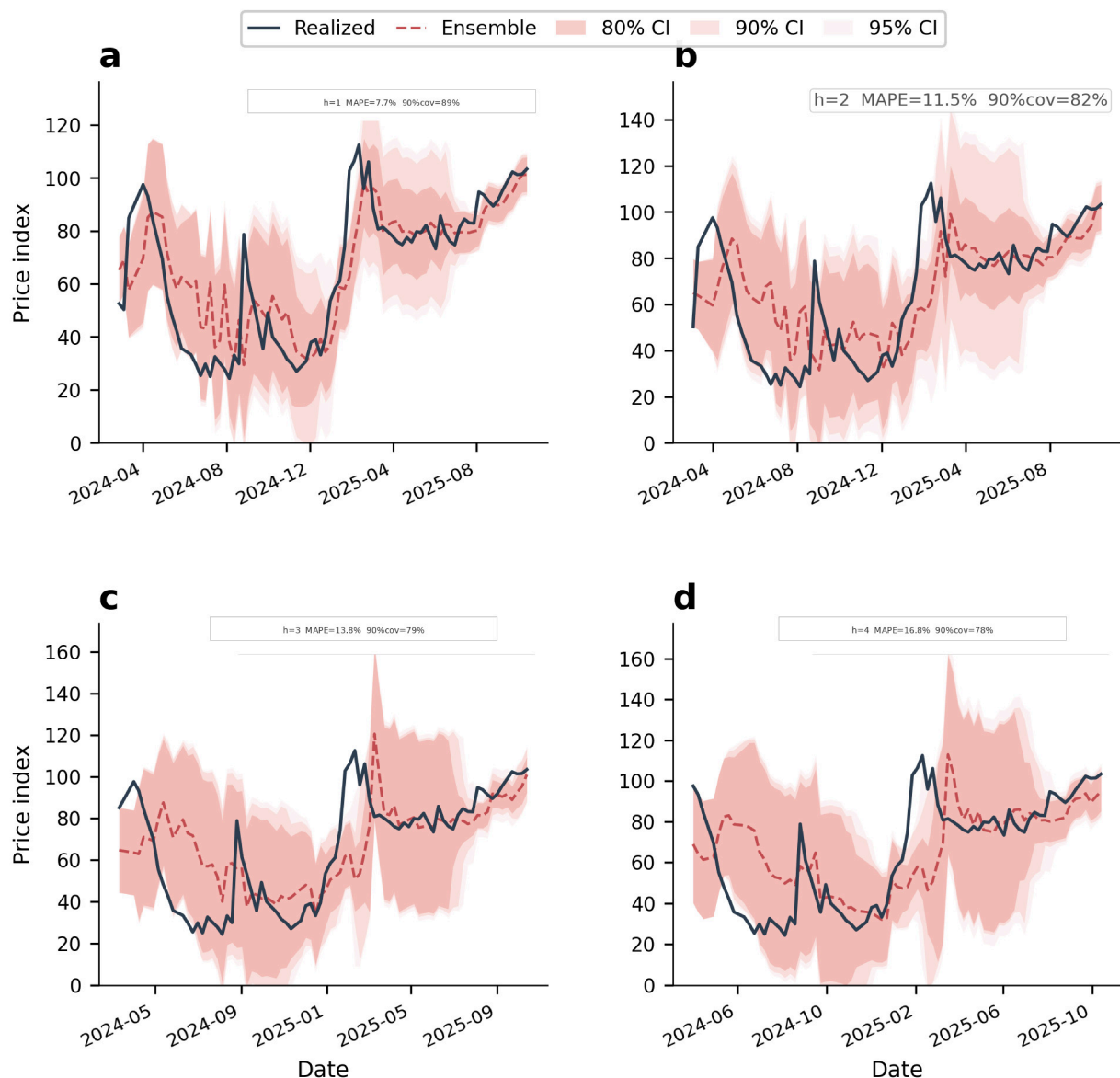


Figure 9. Rolling forecast with conformal prediction intervals for tomato. Panels show horizons $h = 1$ (a), $h = 2$ (b), $h = 3$ (c), and $h = 4$ weeks (d). Black solid lines show realized values; colored dashed lines show ensemble forecasts. Nested bands represent 80% (lightest), 90%, and 95% (darkest) adaptive conformal intervals. MAPE values are displayed in each panel.

Figure 10 and Table 2 provide a comprehensive comparison of forecast performance across models, crops, and horizons. The heatmap (Figure 10a) displays MAPE values for

each combination, revealing a horizon-dependent dominance pattern. Machine-learning models substantially outperform SARIMAX at short horizons: at $h = 1$, the best ML model reduces MAPE by 42% for tomato (TFT: 7.5% vs. SARIMAX: 13.0%), 58% for pepper (LightGBM: 6.8% vs. SARIMAX: 16.1%), and 52% for lettuce (LightGBM: 3.5% vs. SARIMAX: 7.3%). This advantage reflects the capacity of tree-based and attention-based architectures to exploit nonlinear dependencies in lagged features, moving averages, and product-mix indicators. The ML improvement chart (Figure 10b) further shows that this advantage narrows at longer horizons: at $h = 4$, SARIMAX becomes competitive with or outperforms individual ML models for pepper (SARIMAX: 23.8% vs. TFT: 26.3%, LightGBM: 28.1%), consistent with the well-documented finding that simple methods are difficult to beat at longer horizons where signal-to-noise ratios decline. The inverse-MAPE weighted ensemble provides robust cross-horizon performance: while it does not uniformly dominate at every horizon, it achieves the best MAPE at $h = 4$ for tomato (16.8%) and lettuce (10.1%), demonstrating that the weighted combination attenuates model-specific weaknesses precisely where uncertainty is greatest.

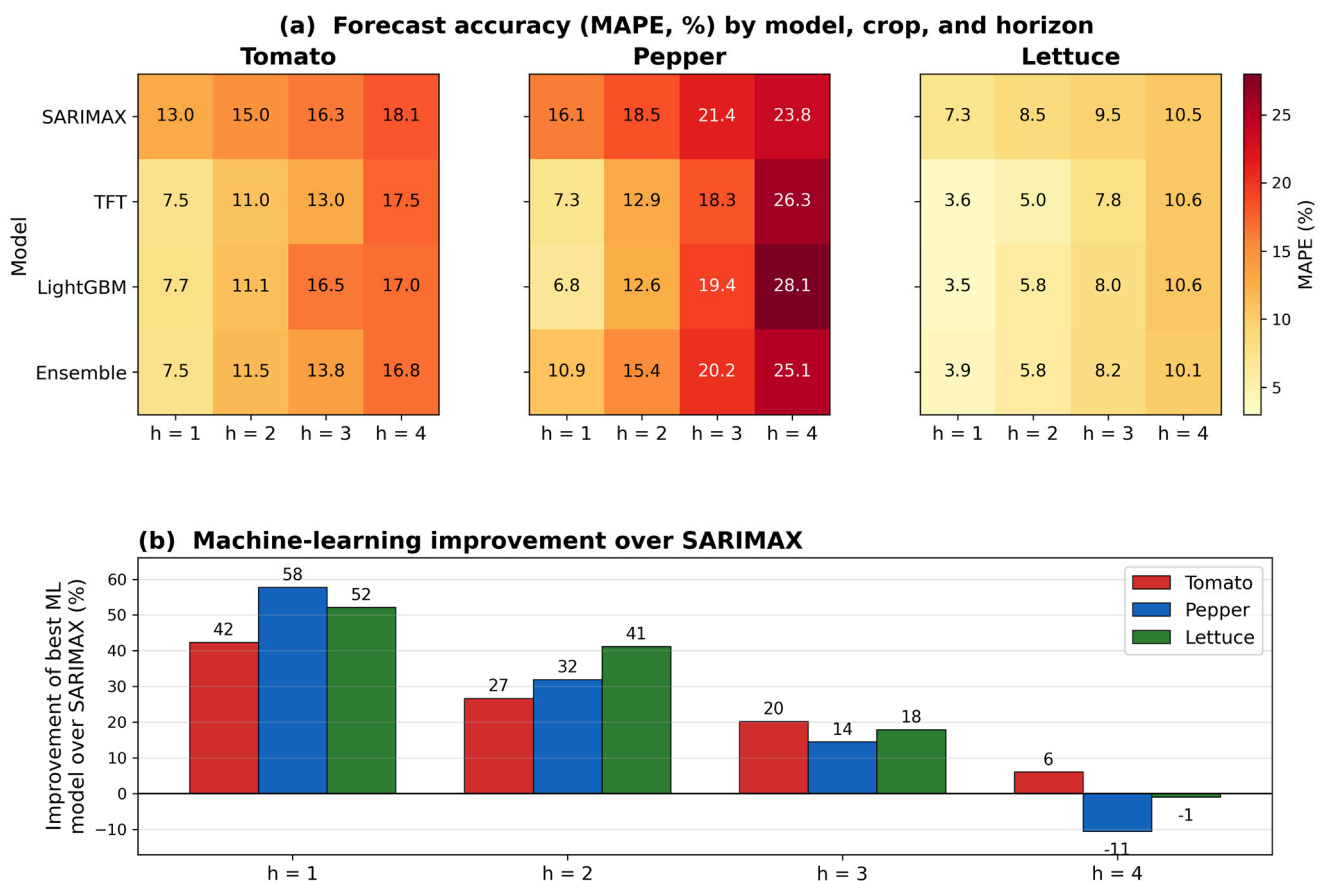


Figure 10. Model performance comparison. (a) MAPE heatmap across four models (SARIMAX, TFT, LightGBM, and Ensemble), three crops, and four forecast horizons. Cell values show MAPE (%). Lower values (yellow) indicate better performance. (b) ML improvement (%) over SARIMAX for each crop–horizon combination.

The comparatively high pepper errors at the longest horizons (Figure 10; $h = 3$: 18.3–21.4%, $h = 4$: 23.8–28.1%) coincide with the sharp, promotion-driven trough in the pepper index during 2024 (Figure 5b)—a low-signal, high-variance episode in which all learners struggle to extrapolate. At these horizons SARIMAX becomes competitive ($h = 4$: 23.8%), consistent with the M4 competition finding that simple models are difficult to beat when the signal-to-noise ratio is low [16].

Table 2 reports the MAPE for every model–crop–horizon combination together with Diebold–Mariano p -values comparing the ensemble with the best individual learner. The ensemble attains the lowest MAPE at $h = 4$ for tomato (16.8%) and lettuce (10.1%) and is within 0.4–1.5 percentage points of the best learner at shorter horizons. The DM tests are significant ($p < 0.05$) mainly for pepper at short horizons, where the gap between SARIMAX and the machine-learning learners is largest; elsewhere the differences are not statistically significant, which—given test-period samples of only 24–40 weeks—reflects limited statistical power rather than evidence of genuinely equal accuracy.

3.5. Sensitivity Analysis

Table 3 summarizes sensitivity checks across four design dimensions. Varying the base year from 2022 to 2024 produces mean index differences of less than 3 percentage points, confirming temporal robustness. The choice between fixed and annually rolling baskets has a larger effect: annual rolling increases the Top-10 representativeness by 4–7 percentage points while maintaining Kolmogorov–Smirnov test p -values above 0.15, indicating that the distributional shift is modest. The robustization method (Winsor 2–98% vs. Huber M-estimator) primarily affects high-promotion crops (lettuce and pepper), where Huber weighting reduces kurtosis by an additional 8–12% relative to winsorization. Enabling TDH quality adjustment reduces index volatility by 5–9% (measured by standard deviation of weekly returns) for pepper and lettuce, with a negligible effect on tomato where specification heterogeneity is lower.

Table 3. Sensitivity analysis across four design dimensions.

Dimension	Variant	Tomato Δ MAPE	Pepper Δ MAPE	Lettuce Δ MAPE
Base year	2022 avg	+1.2 pp	+0.8 pp	+1.5 pp
	2024 avg	−0.6 pp	−0.3 pp	−0.9 pp
Basket policy	Fully fixed (2022–2024)	+2.8 pp	+1.9 pp	+3.4 pp
	Quarterly rolling	−0.4 pp	−0.2 pp	−0.7 pp
Robustization	No robustization	+4.5 pp	+6.2 pp	+7.8 pp
	Huber M-estimator	−0.3 pp	−0.8 pp	−1.1 pp
TDH adjustment	TDH on	−0.5 pp	−1.2 pp	−1.8 pp
	TDH off (baseline)	—	—	—

Δ MAPE is the change in one-week-ahead forecast MAPE (percentage points) relative to the baseline configuration (annual rolling basket, Winsor 2–98%, TDH off, 2023 base). Positive values indicate degradation; negative values indicate improvement.

Overall, the index is most sensitive to the choice of robustization: disabling winsorization inflates one-week-ahead MAPE by 4.5–7.8 percentage points (Table 3), confirming that promotion-robust price treatment is the single most consequential design decision, particularly for the high-promotion lettuce and pepper categories. Base-year and basket-policy choices have only modest effects (at most 3.4 percentage points), and enabling the time-dummy hedonic adjustment yields small but consistent improvements for the more specification-heterogeneous pepper and lettuce series. Taken together, these checks support the baseline configuration (annual rolling basket, winsorization at the 2nd–98th percentiles, 2023 base period) adopted throughout the analysis.

4. Discussion

Accurate measurement and short-term prediction of agricultural input prices in China’s rapidly digitizing markets have become pressing methodological challenges, with implications for farmer purchasing decisions, agribusiness planning, and government policy monitoring. Using 452,489 de-identified seed transactions from major Chinese e-

commerce platforms across five years, this study proposes and validates an integrated framework combining a robust weekly price index with an ensemble forecasting system and calibrated uncertainty quantification. The findings, discussed below, offer methodological insights relevant to agricultural price index construction, machine-learning-based price forecasting, and operational deployment of monitoring tools.

The Paasche-robust formulation proved effective at suppressing promotion-driven noise while preserving economically meaningful price dynamics. The close alignment with superlative indices (Fisher and Törnqvist) confirms that the robustization does not introduce systematic bias, a concern that has been raised in the scanner data literature [42]. Meanwhile, the substantial divergence from the Laspeyres index—particularly during promotional episodes—underscores the importance of current-period weighting when product churn is high and promotional activity is frequent.

The crop-specific seasonal patterns observed in the price indices (Figures 4–6) reflect fundamental agronomic demand dynamics: tomato prices peak during the August transplanting season and again in December, pepper peaks in the September autumn planting window, and lettuce shows a September peak. These differentiated seasonal signatures have direct implications for agricultural policy: crop-specific seasonal adjustment could complement the crop-level price monitoring already provided by the Ministry of Agriculture and Rural Affairs of China [43]. The structural shift observed in 2025—driven by the entry of seedling products into e-commerce channels, with tomato transactions surging from 3736 in 2024 to 155,977 in the first seven months of 2025—highlights the importance of dynamic basket management and suggests that future index updates should explicitly accommodate product-category expansion.

The forecasting results reveal a horizon-dependent dominance pattern with practical implications for operational monitoring. Machine-learning models (TFT and LightGBM) achieve substantial MAPE reductions over SARIMAX at short horizons (42–58% at $h = 1$), leveraging their capacity to capture nonlinear dependencies in lagged features, moving averages, and product-mix indicators. However, this advantage attenuates at longer horizons, where signal-to-noise ratios decline and tree-based models face increased extrapolation risk. At $h = 4$, SARIMAX becomes competitive for pepper—a finding consistent with the M4 competition evidence that simple methods are difficult to beat at longer horizons. The inverse-MAPE weighted ensemble provides a practical compromise: while it does not uniformly dominate individual learners, it achieves the best or near-best MAPE at the longest horizon for two of three crops (tomato and lettuce), suggesting that the combination attenuates model-specific weaknesses in the most uncertain regime. This robustness property is operationally valuable for real-time monitoring systems that must provide reliable forecasts across multiple planning windows simultaneously. The conformal prediction intervals provided well-calibrated uncertainty estimates at short horizons without requiring distributional assumptions, a feature that is particularly valuable for operational dashboards where non-technical stakeholders need interpretable risk communication.

Placing these results in the context of prior work, our one-week-ahead errors (MAPE 3.5–7.7% across crops) are broadly comparable to—and in places better than—recent vegetable-price forecasting studies based on offline wholesale data. Optimized neural-network models achieve single-digit weekly errors on wholesale series [22]; LSTM models on Beijing wholesale prices report a high daily fit (coefficient of determination of roughly 0.96) on smoother, less promotion-distorted series [23]; recent reviews document the broader shift toward such nonlinear learners [24]; and mixture-of-experts ensembles improve robustness across price regimes [25], echoing our finding that combination outperforms single learners precisely at the most uncertain horizons. Unlike these studies, which target offline point forecasts, the present framework operates on noisier e-commerce

micro-data, integrates the upstream index-construction step, and supplies calibrated conformal intervals, thereby addressing the reproducibility and uncertainty-quantification gaps in agricultural price forecasting. The horizon-dependent dominance pattern we observe—machine-learning learners excelling at short horizons while parsimonious models remain competitive as the signal-to-noise ratio falls—is consistent with the broader forecasting literature [16,17].

4.1. Limitations

Several limitations warrant acknowledgment. First, our test-period sample sizes (24–40 observations per crop) limit the statistical power of Diebold–Mariano tests, as reflected in the non-significant p -values for several crop–horizon combinations; longer evaluation windows enabled by accumulating additional transaction data would strengthen inferential conclusions. Second, our analysis covers three crops; generalization to the broader seed market requires additional specification audits and platform harmonization. Third, the e-commerce channel captures only a portion of total seed market activity, and our indices may not fully represent over-the-counter or cooperative-mediated transactions. Fourth, geographic disaggregation is limited by the reliability of destination fields in order records and is recommended as a supplementary extension.

Finally, data quality and quantity directly shape the forecasting results. The short and partially gapped pepper and lettuce series (163 and 184 weeks, with later online entry) reduce the number of training and evaluation folds, widen the conformal intervals, and likely inflate longer-horizon errors—most visibly the elevated pepper MAPE at $h = 3$ –4. The forecast comparisons should therefore be read as conditional on currently available coverage; as additional weeks accumulate, both point accuracy and the statistical power of the Diebold–Mariano tests are expected to improve.

4.2. Future Research Directions

Future work will pursue four directions: (i) semi-parametric TDH models leveraging text embeddings for product descriptions, (ii) multivariate frameworks coupling seed prices with fertilizer and agrochemical inputs, (iii) nowcasting with state-space TFT hybrids, and (iv) counterfactual experiments on promotional intensity to quantify price elasticity and welfare effects.

5. Conclusions

We presented a reproducible framework for constructing weekly e-commerce seed price indices and generating short-term forecasts with calibrated uncertainty. The Paasche-robust index effectively attenuates promotional noise while maintaining alignment with superlative formulas. Machine-learning models (TFT and LightGBM) substantially outperform SARIMAX at short forecast horizons (42–58% MAPE reduction at $h = 1$), while the inverse-MAPE weighted ensemble provides robust cross-horizon performance, achieving the best accuracy at the four-week horizon for two of three crops. The framework is extensible to additional crop categories and can be integrated into agricultural monitoring dashboards for real-time decision support.

Beyond its methodological contribution, the framework has concrete practical value. For farmers and agribusinesses, weekly promotion-robust indices with short-horizon forecasts and explicit uncertainty bands support better-timed seed purchasing and inventory decisions; for monitoring agencies, the reproducible pipeline offers a low-latency complement to traditional wholesale-price reporting that captures the rapidly growing e-commerce input channel. The crop-specific seasonal signatures documented here further argue for differentiated, rather than uniform, seasonal-adjustment templates in official monitoring.

Three limitations temper these conclusions: the analysis covers three crops over evaluation windows of 24–40 weeks; the e-commerce channel represents only part of total seed-market activity; and exact platform and geographic coverage vary over time. Future work will extend the basket to additional crops, lengthen the evaluation period to strengthen inference, couple seed prices with fertilizer and agrochemical inputs in a multivariate setting, and package the pipeline as an interactive dashboard (for example, a Streamlit application) so that real-time monitoring becomes accessible to non-technical users. In sum, the proposed system provides a reproducible, deployable, and uncertainty-aware foundation for monitoring agricultural e-commerce input markets.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/agriculture16151620/s1>. Figure S1: Forecast residual diagnostics for tomato, pepper, and lettuce: (a–c) residual distributions with fitted normal densities and kernel density estimates, and (d–f) corresponding Q–Q plots; Figure S2: Adaptive conformal prediction diagnostics: (a–c) realized versus nominal coverage across forecast horizons $h = 1$ –4 for tomato, pepper, and lettuce, respectively, and (d) the realized-coverage–mean-interval-width trade-off at $h = 1$; Figure S3: Geographic distribution and spatial concentration of transactions: (a) combined transaction-density map, (b) regional distribution, (c–e) crop-specific transaction maps for tomato, pepper, and lettuce, (f) Lorenz curves, (g) the 15 provinces with the highest transaction volumes, and (h) the price–volume relationship; Figure S4: Provincial mean seed prices: (a–c) provincial mean-price maps for tomato, pepper, and lettuce, respectively, and (d) comparison of mean unit prices for the top 15 provinces by transaction volume; Table S1: Weekly crop-level price indices, moving averages, and active-product counts; Table S2: Annual quality-control summary statistics by crop; Code S1: Python code for index construction, comparator formulas, ensemble forecasting, and conformal prediction.

Author Contributions: Conceptualization, Q.L. and X.Q.; methodology, Q.L. and X.Q.; software, Q.L.; validation, Q.L., X.W. and X.Q.; formal analysis, Q.L.; investigation, Q.L.; resources, X.W. and X.Q.; data curation, Q.L., F.Y. and R.L.; writing—original draft preparation, Q.L.; writing—review and editing, Q.L., X.W., X.Q., F.Y., R.L., Y.P., W.Y. and L.Z.; visualization, Q.L. and Y.P.; supervision, X.W. and X.Q.; project administration, X.W. and X.Q.; funding acquisition, X.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Beijing Academy of Agriculture and Forestry Sciences Reform and Development Program “Research and Application of Agricultural Data Science Technology” (grant number GGFZSJS2026) and the Beijing Rural Revitalization Agricultural Science and Technology Program “Research and Application of Agricultural Product Production-Sales Information Prediction and Intelligent Service Technology” (grant number NY2502270025). The APC was funded by the Beijing Digital Agriculture and Rural Affairs Promotion Center.

Institutional Review Board Statement: Not applicable. This study analyzed de-identified commercial transaction records and did not involve human subjects or animal experimentation.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from major Chinese e-commerce platforms and are available from the authors with the permission of these platforms, subject to commercial confidentiality agreements. The aggregated weekly price index series, quality control statistics, and reproducibility code supporting the conclusions of this study are included in the Supplementary Materials. Further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors thank their colleagues at the Beijing Academy of Agriculture and Forestry Sciences and the Beijing Digital Agriculture and Rural Affairs Promotion Center for discussions on index methodology and data governance. During the preparation of this manuscript, the authors used Claude Sonnet 4 for English language polishing, reference-format checks, and template adjustment. The authors reviewed and edited all AI-assisted output and take full responsibility for

the scientific content, data analysis, interpretation, and conclusions of the publication. No AI tool was used for data generation, statistical analysis, or scientific content creation.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. FAO. *Seeds Toolkit—Module 5: Seed Marketing*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2018.
2. Spielman, D.J.; Smale, M. *Policy Options to Accelerate Variety Change Among Smallholder Farmers in South Asia and Africa South of the Sahara*; IFPRI Discussion Paper 01666; International Food Policy Research Institute: Washington, DC, USA, 2017.
3. Cavallo, A.; Rigobon, R. The Billion Prices Project: Using Online Prices for Measurement and Research. *J. Econ. Perspect.* **2016**, *30*, 151–178. [[CrossRef](#)]
4. Cavallo, A. Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers. *Am. Econ. Rev.* **2017**, *107*, 283–303. [[CrossRef](#)]
5. Liu, Y.; Chen, Y.; Wu, S.; Peng, G.; Lv, B. Composite Leading Search Index: A Preprocessing Method of Internet Search Data for Stock Trends Prediction. *Ann. Oper. Res.* **2015**, *234*, 77–94. [[CrossRef](#)]
6. International Monetary Fund. *Consumer Price Index Manual: Concepts and Methods*; International Monetary Fund: Washington, DC, USA, 2020. [[CrossRef](#)]
7. Eurostat. *Harmonised Index of Consumer Prices (HICP) Methodological Manual—2024 Edition*; Publications Office of the European Union: Luxembourg, 2024. [[CrossRef](#)]
8. Diewert, W.E. Exact and Superlative Index Numbers. *J. Econom.* **1976**, *4*, 115–145. [[CrossRef](#)]
9. Balk, B.M. *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference*; Cambridge University Press: Cambridge, UK, 2008. [[CrossRef](#)]
10. Ivancic, L.; Diewert, W.E.; Fox, K.J. Scanner Data, Time Aggregation and the Construction of Price Indexes. *J. Econom.* **2011**, *161*, 24–35. [[CrossRef](#)]
11. de Haan, J.; Hendriks, R.; Scholz, M. Price Measurement Using Scanner Data: Time—Product Dummy Versus Time Dummy Hedonic Indexes. *Rev. Income Wealth* **2021**, *67*, 394–417. [[CrossRef](#)]
12. Chessa, A.G. A New Methodology for Processing Scanner Data in the Dutch CPI. *EURONA* **2016**, *1*, 49–69.
13. de Haan, J.; Krsinich, F. Scanner Data and the Treatment of Quality Change in Nonrevisable Price Indexes. *J. Bus. Econ. Stat.* **2014**, *32*, 341–358. [[CrossRef](#)]
14. de Haan, J.; van der Grient, H.A. Eliminating Chain Drift in Price Indexes Based on Scanner Data. *J. Econom.* **2011**, *161*, 36–46. [[CrossRef](#)]
15. Wolpert, D.H. Stacked Generalization. *Neural Netw.* **1992**, *5*, 241–259. [[CrossRef](#)]
16. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: 100,000 Time Series and 61 Forecasting Methods. *Int. J. Forecast.* **2020**, *36*, 54–74. [[CrossRef](#)]
17. Petropoulos, F.; Apiletti, D.; Assimakopoulos, V.; Babai, M.Z.; Barrow, D.K.; Taieb, S.B.; Bergmeir, C.; Bessa, R.J.; Bijak, J.; Boylan, J.E.; et al. Forecasting: Theory and Practice. *Int. J. Forecast.* **2022**, *38*, 705–871. [[CrossRef](#)]
18. Zhang, G.P. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing* **2003**, *50*, 159–175. [[CrossRef](#)]
19. Angelopoulos, A.N.; Bates, S. Conformal Prediction: A Gentle Introduction. *Found. Trends Mach. Learn.* **2023**, *16*, 494–591. [[CrossRef](#)]
20. Romano, Y.; Patterson, E.; Candès, E.J. Conformalized Quantile Regression. In *Advances in Neural Information Processing Systems 32, Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 3543–3553.
21. Zaffran, M.; Féron, O.; Goude, Y.; Josse, J.; Dieuleveut, A. Adaptive Conformal Predictions for Time Series. In *Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; Proceedings of Machine Learning Research*; PMLR: Cambridge, MA, USA, 2022; Volume 162, pp. 25834–25866.
22. Hou, Y.; Sun, T.; Cui, Y.; Wang, X.; Zhao, A.; Wang, T.; Wang, Z.; Yang, W.; Gu, G.; Wu, S. Vegetable Price Prediction Based on Optimized Neural Network Time Series Models. *Smart Agric.* **2025**, *7*, 78–87. [[CrossRef](#)]
23. Zhang, Q.; Yang, W.; Zhao, A.; Wang, X.; Wang, Z.; Zhang, L. Short-Term Forecasting of Vegetable Prices Based on LSTM Model—Evidence from Beijing’s Vegetable Data. *PLoS ONE* **2024**, *19*, e0304881. [[CrossRef](#)] [[PubMed](#)]
24. Sun, F.; Meng, X.; Zhang, Y.; Wang, Y.; Jiang, H.; Liu, P. Agricultural Product Price Forecasting Methods: A Review. *Agriculture* **2023**, *13*, 1671. [[CrossRef](#)]

25. Zhao, C.; Wang, X.; Zhao, A.; Cui, Y.; Wang, T.; Liu, J.; Hou, Y.; Wang, M.; Chen, L.; Li, H.; et al. A Vegetable-Price Forecasting Method Based on Mixture of Experts. *Agriculture* **2025**, *15*, 162. [\[CrossRef\]](#)
26. ISO 8601-1:2019; Date and Time—Representations for Information Interchange—Part 1: Basic Rules. International Organization for Standardization: Geneva, Switzerland, 2019.
27. Huber, P.J. Robust Estimation of a Location Parameter. *Ann. Math. Stat.* **1964**, *35*, 73–101. [\[CrossRef\]](#)
28. Dixon, W.J.; Tukey, J.W. Approximate Behavior of the Distribution of Winsorized t (Trimming/Winsorization 2). *Technometrics* **1968**, *10*, 83–98. [\[CrossRef\]](#)
29. Diewert, W.E.; Fox, K.J. Substitution Bias in Multilateral Methods for CPI Construction. *J. Bus. Econ. Stat.* **2022**, *40*, 355–369. [\[CrossRef\]](#)
30. Rao, D.S.P. On The Equivalence of Weighted Country-Product-Dummy (CPD) Method and the Rao-System for Multilateral Price Comparisons. *Rev. Income Wealth* **2005**, *51*, 571–580. [\[CrossRef\]](#)
31. Krsinich, F. The FEWS Index: Fixed Effects with a Window Splice. *J. Off. Stat.* **2016**, *32*, 375–404. [\[CrossRef\]](#)
32. Zhen, C.; Taylor, J.L.; Muth, M.K.; Leibtag, E. Understanding Differences in Self-Reported Expenditures Between Household Scanner Data and Diary Survey Data: A Comparison of Homescan and Consumer Expenditure Survey Data. *Rev. Agric. Econ.* **2009**, *31*, 470–492. [\[CrossRef\]](#)
33. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; Wiley: Hoboken, NJ, USA, 2015.
34. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Advances in Neural Information Processing Systems 30. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3146–3154.
35. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: San Francisco, CA, USA, 2016; pp. 785–794. [\[CrossRef\]](#)
36. Lim, B.; Arik, S.Ö.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [\[CrossRef\]](#)
37. Bates, J.M.; Granger, C.W.J. The Combination of Forecasts. *J. Oper. Res. Soc.* **1969**, *20*, 451–468. [\[CrossRef\]](#)
38. Timmermann, A. Forecast Combinations. In *Handbook of Economic Forecasting*; Elliott, G., Granger, C.W.J., Timmermann, A., Eds.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 1, pp. 135–196. [\[CrossRef\]](#)
39. Stock, J.H.; Watson, M.W. Combination Forecasts of Output Growth in a Seven-Country Data Set. *J. Forecast.* **2004**, *23*, 405–430. [\[CrossRef\]](#)
40. Hyndman, R.J.; Koehler, A.B. Another Look at Measures of Forecast Accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [\[CrossRef\]](#)
41. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [\[CrossRef\]](#)
42. Diewert, W.E. Index Number Theory Using Differences Rather Than Ratios. *Am. J. Econ. Sociol.* **2005**, *64*, 311–360. [\[CrossRef\]](#)
43. Ministry of Agriculture and Rural Affairs of China. Agricultural Product Wholesale Price 200 Index. Available online: <https://zdscxx.moa.gov.cn/nyb/pc/index.jsp> (accessed on 12 July 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.