sustainability

MDPI

Article

GPF-EVMoLE: An ETS-Driven Variable Selection and Mixture-of-Experts Framework for Multi-Step Garlic Price Forecasting

Xinran Yu [1], Ke Zhu [1,2,3,*], Honghua Jiang [1,2,3] and Ruofei Chen [4]

[1] College of Information Science and Engineering, Shandong Agricultural University, Tai'an 271018, China; x2172849@gmail.com (X.Y.); j_honghua@163.com (H.J.)
[2] Huang Huai Hai Key Laboratory of Intelligent Agricultural Technology, Tai'an 271018, China
[3] Agricultural Big-Data Research Center, Shandong Agricultural University, Tai'an 271018, China
[4] College of Science and Information, Qingdao Agricultural University, Qingdao 266109, China; 18053371654@163.com
* Correspondence: kezhu@sdau.edu.cn

Abstract

Predicting garlic prices is difficult because the crop behaves as both an agricultural commodity and a speculative asset. Unlike staple grains, which follow more predictable seasonal supply cycles, garlic can be stored for over a year, its production is geographically concentrated, and its demand remains inelastic. This industry structure makes it susceptible to speculative hoarding, where even minor harvest deficits may trigger sharp price spikes. A typical example is the "Suan Ni Hen" (crazy garlic) phenomenon in the Chinese market: during the 2009–2010 and 2016 periods, speculative capital repeatedly exploited expectations of harvest reduction to engage in large-scale hoarding. According to data released by China's National Development and Reform Commission (NDRC) at the end of October 2016, national wholesale garlic prices surged by 90% year-on-year, with purchase prices in some major producing areas doubling or multiplying within a short period. Such short-term price bubbles, together with severe volatility and abrupt regime shifts, can make standard forecasting models unreliable in this uncertain environment. Existing methods, ranging from traditional seasonal algorithms to deep learning networks, often overlook the need to decouple the local trend-weekly-seasonal baseline from the dynamic effects of multi-source external signals. This paper proposes GPF-EVMoLE, a compositional multi-step forecasting framework built on an explicit division of labor. The framework first extracts an interpretable local trend and weekly-seasonal baseline through an ETS decomposition module. Two specialized components then process the residual signal: a temporal fusion Transformer-style variable selection network (VSN) uses multi-source external features to identify informative macroeconomic and environmental signals at each forecasting step, while a Mixture of Linear Experts (MoLE) models phase-wise regime shifts within the residual series. Together, these modules adaptively integrate heterogeneous information. This study evaluates the framework on a custom daily evaluation dataset containing 17,685 records across six major producing regions in three provinces. At 7-day and 14-day forecasting horizons, GPF-EVMoLE consistently outperforms eight representative statistical, machine learning, and deep learning baselines across MAE, RMSE, and MAPE metrics. Ablation studies verify the necessity of each component, showing that structural separation of the forecasting tasks helps overcome the limitations of monolithic models and provides an accurate and interpretable solution for complex agricultural markets.

Check for updates

Academic Editor: Andreas Kanavos

Received: 20 March 2026
Revised: 26 April 2026
Accepted: 27 April 2026
Published: 30 April 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) license.

Keywords: garlic price forecasting; ETS decomposition; variable selection network; mixture of linear experts; multi-source features; multi-step forecasting

1. Introduction

Garlic occupies an unusual position between foodstuff and speculative asset. China accounts for roughly 80% of global garlic production and exports, with output concentrated in a handful of provinces—Shandong, Henan, and Jiangsu together supply the bulk of the domestic market. In practice, the price history of Chinese garlic often resembles a speculative commodity market rather than a staple-crop market. Demand is inelastic, supply is volatile, and the crop can be stored for more than a year under cold-storage conditions. These properties make hoarding economically attractive and distinguish garlic from wheat or maize. The phenomenon has a name in Chinese market parlance: *suàn nǐ hěn*, literally "garlic strikes back", coined to describe episodes in which prices surged tenfold or more within months.

The 2009–2010 spike was triggered by a reduced planted area after a period of depressed prices, amplified by H1N1-related panic buying, and then intensified by speculative capital that pushed garlic above pork prices in some regions. This episode was documented in *The Guardian*, *The Wall Street Journal*, and *China Daily*. A similar sequence recurred in 2015–2016, when heavy rains and cold waves damaged the Shandong harvest, post-stock-market investment moved into commodity storage, and prices again reached multi-year highs. Price watchdogs from the National Development and Reform Commission (NDRC) intervened, issued warnings, and fined hoarding companies, but prices remained erratic for months. These episodes show that garlic prices respond not only to supply fundamentals but also to expectations, momentum, and capital flows in ways that standard seasonal models do not anticipate [1,2].

Recent evidence from other Asian garlic and condiment-vegetable markets reinforces this policy relevance: garlic provisioning has been studied as a strategic food-commodity system under import-dependency pressure, and onion–garlic price volatility has been linked to food-security and market-stability concerns [3,4].

This combination of inelastic demand, weather sensitivity, cold-storage longevity, and speculation gives garlic price series a statistical profile that differs from many agricultural commodities. Abrupt regime shifts—prices doubling within weeks or collapsing as storage contracts expire—are not isolated outliers but recurring structural features. Seasonal ARIMA models can capture weekly periodicity, but fixed-coefficient linear models cannot readily adjust their behavior across regimes. They therefore tend to extrapolate from the last stable phase even after the market has shifted. End-to-end deep learning models provide nonlinear capacity, but without explicit structural decomposition, they must learn trend, seasonality, and sudden shocks under a single objective. This can destabilize multi-step forecasts when the training distribution spans multiple regimes [5].

A further difficulty is the need to integrate heterogeneous information channels. Garlic prices are affected by cold-storage inventory levels, supply–demand balance commentary from the Ministry of Agriculture and Rural Affairs (MOA), macro price indices, meteorological records from producing regions, holiday and solar term calendars, and cross-origin price spreads across provinces and grades. When these signals are ignored, the model may miss the dominant driver of the current regime. When they are included without selectivity, the model may instead confound informative signals with noise.

This concern is consistent with recent agricultural price-forecasting studies showing that weather and other exogenous variables can materially improve crop-price prediction,

while purely generic deep-learning capacity still struggles when abrupt anomalies, market uncertainty, and heterogeneous external drivers must be handled jointly [6–9].

We argue that the core difficulty is not simply model capacity, but the lack of a well-defined interface among structural extrapolation, exogenous-signal selection, and regime-dependent correction. Garlic prices contain a relatively persistent local trend and weekly-seasonal component, short-term disturbances triggered by weather and supply–demand information, and abrupt state changes driven by inventory release and speculative expectations. If these sources of variation are learned by a single black-box model, trend, shocks, and regime transitions can be confounded. If separate modules are fitted independently, the structural baseline and residual correction cannot adapt to each other. GPF-EVMoLE (Garlic Price Forecasting with ETS-driven Variable Selection and Mixture of Linear Experts) is designed around this interface problem. ETS constructs an interpretable local trend and weekly-seasonal baseline; the TFT-style variable selection network (VSN) with cross-attention temporal fusion selects informative and available exogenous signals at each forecast step; and the Mixture of Linear Experts (MoLE) models state-dependent residual corrections. The three components are trained under one objective, so structural extrapolation, external-driver filtering, and residual regime adaptation play complementary rather than merely sequential roles.

No dedicated public daily benchmark for multi-source garlic price forecasting existed prior to this work. We constructed an evaluation dataset by assembling daily price records from Mysteel (a leading Chinese steel and agricultural commodities data platform) and the China Meteorological Administration, enriched with MOA market-analysis report scores parsed from official publications, annual supply–demand balance sheets, the National Bureau of Statistics agricultural wholesale price index, and full holiday and solar term calendars. The result is 21 daily price series spanning six production origins across Henan, Shandong, and Jiangsu, running from 3 January 2023 to 15 May 2025 (864 days, 17,685 records), with each series identified by province, origin, variety, and grade.

Experiments across 7- and 14-day forecast horizons show that GPF-EVMoLE consistently outperforms eight representative baselines—including seasonal statistical models, machine learning, and deep learning methods—on MAE, RMSE, and MAPE. Ablation studies on both model components and weather inputs confirm that the gains are consistent with the intended roles of the proposed modules, and hyperparameter sensitivity analyses establish that the improvements are not tied to a narrow configuration range.

The contributions of this paper are as follows:

1. We formulate multi-step garlic price forecasting as a coupled structural-baseline and residual-correction problem rather than as a generic sequence-prediction task. The main methodological contribution is a constrained compositional interface: ETS extracts an interpretable local trend and weekly-seasonal baseline, VSN controls which available external drivers enter each forecast step, and MoLE assigns residual corrections to market-state-specific experts. This design prevents trend structure, exogenous shocks, and regime transitions from competing to explain the same price movement inside a single opaque model, while joint training allows the three roles to co-adapt instead of forming an independently fitted module stack.
2. We design a multi-source data integration scheme that handles the heterogeneous, multi-frequency nature of garlic market signals. Daily price and meteorological records, monthly MOA market-analysis reports, annual supply–demand balance sheets, and deterministically known calendar inputs are each assigned to appropriate input channels and aligned strictly by public release date. For decoder-side meteorological inputs, because historical forecast archives were unavailable, we use realized CMA observations as an oracle proxy for short-term forecasts; results should

therefore be interpreted as an upper-bound weather-input setting rather than fully deployment-realistic weather conditions.

3. We propose a TFT-style variable selection network with cross-attention temporal fusion that learns per-step input importance weights conditioned on the series' static identity. Rather than applying a fixed feature set across all forecast origins, the network assigns continuously varying weights to each exogenous signal—holidays, macro indices, supply–demand fields, and meteorological proxy inputs—as conditions change. The retained weights are interpretable by construction, providing a step-level record of which external drivers the model attributed each forecast adjustment to.
4. We provide systematic empirical evidence that the three-component design works as intended. Comparative experiments against eight baselines across 7- and 14-day horizons show consistent accuracy improvements under MAE, RMSE, and MAPE. Ablation studies isolate each component's measurable contribution and include a no-weather variant to evaluate the influence of the oracle-weather proxy. Hyperparameter sensitivity sweeps further show that the gains are stable rather than tuned to a narrow configuration.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 formalizes the forecasting problem. Section 4 details the GPF-EVMoLE architecture. Section 5 reports experimental results. Section 6 concludes.

2. Related Work

Garlic prices are shaped by seasonal harvest cycles, expectation-driven inventory decisions, and intermittent supply–demand shocks. This combination resists simple univariate extrapolation and requires forecasting models to handle structural patterns, exogenous feature selection, and regime changes. The literature relevant to this work therefore spans three areas: statistical and decomposition-based commodity forecasting, learning-based models with exogenous covariates, and studies of agricultural and garlic price dynamics under multi-source drivers.

Agricultural price volatility has attracted sustained research attention because it propagates across supply chains and feeds back into planting, storage, and trading decisions. Survey work has documented the co-occurrence of seasonality, non-stationarity, and external shocks as dominant features of agricultural markets [1]. Price instability at the food level also carries broader implications: elevated food prices have been linked to social unrest [2], and the call for more responsive information and early-warning systems grows as global agri-food value chains face increasing fragility [10]. Structural heterogeneity on the supply side—including variation in consumption patterns [11] and self-sufficiency gaps in staple food groups [12]—further complicates price dynamics, making robustness to exogenous shifts and interpretable evidence of causal mechanisms both practically necessary. Prior work on garlic specifically has addressed economic profitability and income risk [13], underscoring the stakes of accurate forecasting in markets where price swings materially affect growers and traders.

2.1. Statistical/Econometric and Decomposition-Based Forecasting

Statistical and econometric approaches remain foundational for understanding commodity price formation. Cointegration analysis, price transmission models, and tests of market integration have been applied to characterize spatial price linkages in markets including garlic [14,15], while structural models have been used to decompose supply and demand contributions to observed price variation—for example, in wheat markets [16]. On the forecasting side, distributed-lag specifications such as ARDL have been integrated with operational decision support [17], and there is recognition that price

processes may shift across regimes in ways that alter their predictable structure [18]. This body of work informs our treatment of cross-region heterogeneity: rather than training separate models per origin or province, we embed structural differences through static categorical context and introduce switchable components on the residual side to account for phase-dependent dynamics.

Decomposition-based methods—which separate a series into trend, seasonal, and irregular components before modeling each part—offer a principled way to reconcile stable structural regularities with erratic short-term disturbances. Variational mode decomposition in particular has seen application in agricultural price settings, with components fed into deep architectures and hyperparameters tuned via swarm-intelligence optimizers for vegetables [19] and for corn and wheat [20]. These frameworks establish the practical value of separating what can be modeled by simple rules from what cannot. Our design takes a different route: ETS decomposition is used to extract trend and seasonality into an explicit, interpretable baseline, keeping the residual module focused on the irregular and exogenous-driven component. Compared with more elaborate decomposition pipelines, this yields a cleaner separation between structural extrapolation and learned correction, with fewer moving parts to calibrate.

2.2. Learning-Based Forecasting with Exogenous Variables

Machine-learning approaches to time-series forecasting typically use lagged targets, rolling statistics, and exogenous covariates as supervised-regression features. In agricultural price settings, gradient-boosting models have been applied to chili price forecasting with feature-importance diagnostics [21], and boosting methods under weather-driven scenarios have been used for groundnut price prediction [7]. Systematic comparisons across crops and modeling setups show that the benefit of deep learning over machine learning is not universal and cannot be inferred from dataset characteristics alone [5]. These studies suggest that exogenous variables can provide meaningful incremental signal, but the relevance of any single variable changes across time, regions, and market conditions. This motivates the soft variable selection mechanism in our framework: instead of committing to a fixed feature set, we learn per-step importance weights that adapt as conditions change and remain interpretable to analysts.

Deep-learning methods—recurrent networks, convolutional encoders, attention-based architectures—offer greater capacity for nonlinear and long-range dependencies. A recurring pattern in recent agricultural price work is to pair these with a decomposition front-end (e.g., VMD or STL) and search hyperparameters with swarm-intelligence methods [19,20]. The resulting models can achieve strong in-sample fit, but their post hoc interpretability is limited to the component level and rarely extends to fine-grained attribution of which exogenous signals matter and when. A separately important design consideration is temporal data alignment: agricultural data arrive at multiple frequencies with publication lags (annual supply–demand reports, monthly MOA analyses, daily weather), and careless feature construction can allow future information to contaminate model inputs. Our framework addresses this by maintaining separate encoder and decoder input groups with explicit alignment by public release date.

Large-scale time-series foundation models have become an active research direction. Moirai and similar unified Transformer architectures pursue cross-dataset generalization through large-scale pre-training [22]. Chronos recasts forecasting as token prediction over quantized value sequences [23], while MOMENT learns multi-task representations over heterogeneous time-series corpora [24]. Mixture-of-experts architectures have also been incorporated into such models to scale capacity [25]. In addition, there is growing interest in reprogramming large language models for forecasting [26] and

multi-modal time-series reasoning [27]. Although these models achieve strong transfer performance on standard benchmarks, their suitability is less clear for problems with strict information-availability constraints, where future decoder inputs must be partitioned by release date. Their ability to provide per-step, per-variable importance outputs for operational interpretation is also limited. We do not rule out future integration; for the current task, a purpose-built compositional design offers more direct control over evaluation faithfulness and explainability.

2.3. Agricultural and Garlic Price Forecasting: Multi-Source Drivers and Regime Dynamics

Agricultural commodity price forecasting is inherently multi-source: weather, seasonal and calendar effects, regional supply structures, and macroeconomic conditions all contribute—sometimes jointly, sometimes in ways that shift across time—to price trajectories [1,5,7,21]. The practical use case for such forecasts is often risk management and supply-chain coordination [10], which puts a premium on robustness to exogenous shocks and on outputs that analysts can inspect and trust. Macroeconomic channels—input costs, credit conditions, cross-commodity price linkages—are a further source of indirect price pressure that machine-learning methods have begun to capture [28]. Building a dataset that accurately reflects what information is actually available at the forecast origin, accounting for multi-frequency publication lags and definitional differences across sources, is itself a non-trivial task; work on data governance and interoperability provides relevant guidance [29].

For garlic specifically, the literature has addressed spatial price transmission across production regions [14,15] and the economics of cultivation under price risk [13]. These studies point to substantial heterogeneity across provinces, origins, varieties, and grades—heterogeneity that a single aggregate model cannot capture. Our multi-series design addresses this by encoding structural differences through static categorical embeddings, allowing the model to share parameters across series while retaining origin- and grade-specific behavior.

Forecast combination research in agricultural economics has established that weighted ensembles of diverse models generally outperform individual methods, and that learned combination weights tend to outperform fixed schemes [30]. Decomposition-then-ensemble pipelines have been a common realization of this idea: a decomposition front-end separates structural and irregular components, which are then modeled separately and recombined [19,20]. The practical lesson is that, under strong seasonality and persistent non-stationarity, variance is better controlled by anchoring long-run structure in a stable baseline than by asking a single model to track both. Our use of ETS decomposition is grounded in this logic: the baseline absorbs trend and seasonal structure, leaving the residual module to handle the exogenous-driven and regime-dependent component.

That regime-dependent component is precisely where garlic markets are most unpredictable. Inventory release timing is governed by traders' expectations rather than physical supply alone, producing price transitions that are difficult to anticipate with fixed-coefficient models. Work on state-dependent commodity price dynamics has shown that predictable structure can vary substantially across regimes [18], and mixture-of-experts methods have demonstrated value for covering diverse temporal patterns [25], with ensembling also improving robustness for volatile price series [31]. In our framework, the residual head is a Mixture of Linear Experts (MoLE). A gating network allocates predictions across a small set of linear experts, each calibrated to a different local dynamic, so that the overall prediction can shift continuously as the market transitions between regimes. The gate weights are interpretable, providing a record of which dynamic mode the model relied on at each forecast origin.

Taken together, the literature points to three persistent gaps: difficulty reconciling structural regularities with short-term disturbances in a single model, limited interpretability when integrating high-dimensional exogenous information, and the risk of evaluation bias from inadequate information-availability controls. GPF-EVMoLE addresses each directly—ETS decomposition for structural stability, VSN variable selection for interpretable feature attribution, and MoLE for regime-adaptive residual correction—while explicitly reporting that decoder-side meteorological inputs are evaluated under an observation-based proxy setting due to unavailable forecast archives. Table 1 summarizes the comparison between representative forecasting studies and GPF-EVMoLE.

Table 1. Comparison of representative forecasting studies and GPF-EVMoLE.

Study or Method Family	Structural Decomposition	Exogenous Drivers	Dynamic Selection	Regime Adaptation/Validation
Statistical/econometric models	Partial	Limited	No	Partial
Machine-learning price models	No	Yes	Post hoc	Limited
Decomposition–deep learning hybrids	Yes	Partial	No	Implicit
Forecast combination/ensembles	Model-dependent	Model-dependent	Weight-based	Partial
Time-series foundation models	No	Limited	No	Implicit
GPF-EVMoLE (this study)	**ETS baseline + residual**	**Multi-source drivers**	**Step-wise VSN**	**MoLE + DM test**

3. Preliminaries

Let $\mathcal{S}$ denote the set of daily garlic price series, each identified by a combination of province, origin/market, variety, and grade. For series $s \in \mathcal{S}$, let $y_t^{(s)}$ denote the price (CNY/jin) on day t. Each series carries static context features $\mathbf{c}^{(s)}$ encoding its structural identity. At forecast origin t, we distinguish two covariate streams. The first is $\mathbf{x}_{t-L+1:t}^{(s)}$, the historical sequence of time-varying covariates (weather observations, macro price indices, report scores). The second is $\mathbf{z}_{t+1:t+H}^{(s)}$, the future covariate sequence used by the decoder, comprising deterministically known inputs such as holidays and solar terms, time encodings, and short-term meteorological inputs represented by observation-based forecast proxies under a *Perfect Forecast Assumption*. Throughout, $H \in \{7, 14\}$ denotes the forecast horizon.

In this study, multi-step-ahead forecasting refers to predicting the entire future price trajectory over the next H days from a forecast origin t, namely $\hat{\mathbf{y}}_{t+1:t+H}^{(s)}$, using only historical observations available up to t and covariates that are known or forecast-available at t. We evaluate two horizons, $H = 7$ and $H = 14$, corresponding to 7-day and 14-day ahead garlic price forecasting.

The forecasting task is to learn a mapping f_θ that, given the past L days of prices and covariates together with the available future covariates, produces an H-step price trajectory:

$$\hat{\mathbf{y}}_{t+1:t+H}^{(s)} = f_\theta\left(\mathbf{y}_{t-L+1:t}^{(s)}, \mathbf{x}_{t-L+1:t}^{(s)}, \mathbf{z}_{t+1:t+H}^{(s)}, \mathbf{c}^{(s)}\right), \quad s \in \mathcal{S}. \tag{1}$$

Parameters are estimated by minimizing empirical risk over training samples $\mathcal{D}$:

$$\theta^* = \arg\min_\theta \sum_{(s,t)\in\mathcal{D}} \mathcal{L}\left(\mathbf{y}_{t+1:t+H}^{(s)}, \hat{\mathbf{y}}_{t+1:t+H}^{(s)}\right). \tag{2}$$

The key constraint governing decoder inputs is information availability at forecast origin t: non-meteorological variables in $\mathbf{z}_{t+1:t+H}^{(s)}$ are either deterministically known at t or obtainable from forecasts issued at t. Because historical forecast archives were unavailable, decoder-side meteorological inputs use China Meteorological Administration (CMA) observed values over $[t+1, t+H]$ as an oracle proxy for accurate short-term forecasts under the *Perfect Forecast Assumption* (i.e., modern short-horizon forecasts have high accuracy). This setup does not use future prices, but it introduces a meteorological look-ahead relative

to real-time deployment and should be interpreted as an optimistic upper bound that isolates the framework's ability to exploit meteorological signals from errors in a specific forecast product.

4. Proposed Work

4.1. Framework

Garlic price series pose three interlocking structural difficulties. Harvest-related level changes, pronounced intra-week periodicity, and a persistent price trend make reliable multi-step extrapolation essential. Inventory-release timing and shifting market expectations produce abrupt regime transitions that are difficult to anticipate with fixed-coefficient models [18]. The exogenous signals available for forecasting—holidays and solar terms, macro price indices, supply–demand balance sheets, Ministry of Agriculture report scores, and meteorological variables—affect prices in ways that vary across time, region, and market state, making static feature selection inadequate.

Combining decomposition, variable selection, and expert routing is therefore not a simple matter of stacking modules. The first challenge is signal attribution: trend, weekly seasonality, and short-term shocks are superimposed in the observed price, so a residual learner may relearn structure already captured by ETS, while a structural baseline may be distorted by transient shocks if the interface is not constrained. The second challenge is information availability: calendar variables are known in advance, low-frequency reports become usable only after publication, and meteorological inputs depend on forecast availability, so exogenous covariates must be selected under multi-step timing constraints rather than appended indiscriminately. The third challenge is residual heterogeneity: inventory release, speculative expectations, and weather shocks can change the direction and scale of residual movements across market phases, making a single residual head prone to averaging distinct regimes. GPF-EVMoLE addresses these interface challenges through an explicit division of labor. ETS decomposition separates the local trend-weekly-seasonal structure as an interpretable baseline; VSN with temporal fusion determines which available exogenous signals are most informative at each forecast step; and MoLE maintains multiple local residual predictors whose contributions are gated by the inferred market state. The final forecast is the direct sum of the structural baseline and the residual correction, and all components are trained jointly under a single objective.

As shown in Figure 1, the framework consists of three functional components whose details are given in the following subsections.

4.2. ETS-Based Structural Decomposition

Garlic price series exhibit strong short-horizon periodic structure, especially intra-week variation, together with gradual level shifts as harvest and inventory conditions evolve. Asking a single learning module to simultaneously track this structure and respond to irregular disturbances creates a systematic tension that often erodes multi-step stability. This component addresses the tension by applying ETS decomposition to each series, extracting its local trend and weekly-seasonal structure as an explicit, interpretable baseline and passing only the residual—the portion the structural model cannot explain—to downstream learning.

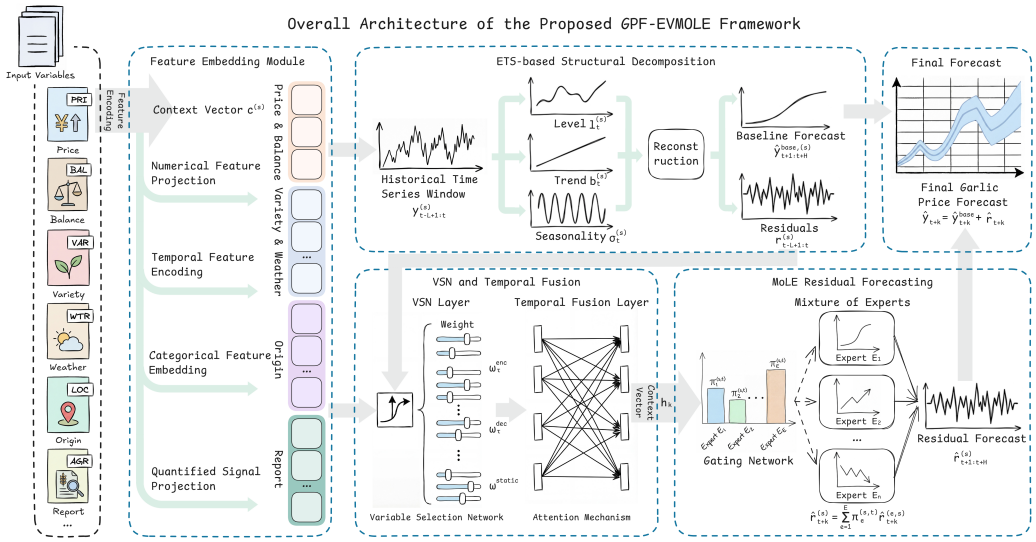

Figure 1. Overall framework of GPF-EVMoLE. Blue dashed boxes indicate major functional modules, colored input blocks denote different feature groups, light green arrows indicate information flow among modules, and gray arrows indicate sequential computation and reconstruction steps.

For series s, let $y_t^{(s)}$ denote the price on day t and m the seasonal length (set to $m = 7$ to capture weekly rhythms). Under an additive seasonal specification, the level $l_t^{(s)}$, trend $b_t^{(s)}$, and seasonal $\sigma_t^{(s)}$ components evolve as

$$l_t^{(s)} = \alpha\left(y_t^{(s)} - \sigma_{t-m}^{(s)}\right) + (1-\alpha)\left(l_{t-1}^{(s)} + b_{t-1}^{(s)}\right), \tag{3}$$

$$b_t^{(s)} = \beta\left(l_t^{(s)} - l_{t-1}^{(s)}\right) + (1-\beta)b_{t-1}^{(s)}, \tag{4}$$

$$\sigma_t^{(s)} = \gamma\left(y_t^{(s)} - l_t^{(s)}\right) + (1-\gamma)\sigma_{t-m}^{(s)}. \tag{5}$$

Based on the recursively estimated structural components, ETS produces an H-step structural baseline forecast:

$$\hat{y}_{t+k}^{\text{base},(s)} = l_t^{(s)} + k\, b_t^{(s)} + \sigma_{t-m+((k-1) \bmod m)+1}^{(s)}, \quad k = 1, \ldots, H. \tag{6}$$

Meanwhile, to capture short-term disturbances beyond the structural components, we define the ETS residual as the one-step-ahead structural forecast error:

$$r_t^{(s)} = y_t^{(s)} - \left(l_{t-1}^{(s)} + b_{t-1}^{(s)} + \sigma_{t-m}^{(s)}\right), \tag{7}$$

and the residual sequence $r_{t-L+1:t}^{(s)}$ serves as the primary learning target for the downstream module. By confining the learning task to this residual, the model is not required to rediscover seasonal periodicity or gradual trend from scratch; it can focus on the exogenous-driven and regime-dependent component that the structural model leaves unexplained.

The smoothing coefficients α, β, and γ are treated as learnable parameters and calibrated through end-to-end training, so that the baseline aligns with the downstream residual objective rather than being fixed by a separate statistical fitting step. They are constrained to the admissible ETS range $(0, 1)$ during optimization, preserving the interpretation of the recursive update equations. The fitted coefficients and the decomposed trend and seasonal trajectories are retained as interpretable outputs, providing a direct reading of structural update rates and weekly seasonal amplitude for each series.

4.3. TFT-Style Variable Selection and Temporal Fusion

The residual that ETS leaves behind is shaped by a heterogeneous set of exogenous forces. How strongly any given signal—a holiday indicator, a macro price index, a supply–demand balance figure, and meteorological input (forecast in deployment, observation-based proxy in this study)—affects residuals depends on the current time window, the active price regime, and the geographic-categorical identity of the series. A fixed feature set cannot accommodate this variability; a model that assigns uniform weight across all inputs risks diluting informative signals with noise.

4.3.1. Variable Selection Network

To learn the importance of each input at each time step, the framework employs a TFT-style variable selection network (VSN) applied separately to three input groups: static context, encoder-side historical observations, and decoder-side available-future covariates. For each group, every candidate variable is first transformed through its own Gated Residual Network (GRN) and Gated Linear Unit (GLU). A shared scoring GRN then receives the concatenation of all transformed representations and, conditioned on the static context $\mathbf{h}^{\text{static}}$, outputs softmax-normalized importance weights. The selected representation is the corresponding weighted sum.

At forecast origin t with historical window L, the encoder-side variable selection at time τ proceeds as

$$\boldsymbol{\omega}_{\tau}^{\text{enc}} = \text{softmax}\Big(\text{GRN}_{w}^{\text{enc}}\Big([\text{GRN}_1(\mathbf{v}_{\tau,1}^{\text{enc}}) \parallel \cdots \parallel \text{GRN}_{N_{\text{enc}}}(\mathbf{v}_{\tau,N_{\text{enc}}}^{\text{enc}})],\, \mathbf{h}^{\text{static}}\Big)\Big), \tag{8}$$

$$\tilde{\mathbf{x}}_{\tau}^{\text{enc}} = \sum_{i=1}^{N_{\text{enc}}} \omega_{\tau,i}^{\text{enc}} \cdot \text{GRN}_i(\mathbf{v}_{\tau,i}^{\text{enc}}), \tag{9}$$

where $\boldsymbol{\omega}_{\tau}^{\text{enc}} \in \mathbb{R}^{N_{\text{enc}}}$ is the time-varying importance weight vector and $\tilde{\mathbf{x}}_{\tau}^{\text{enc}}$ is the selected encoder representation. The decoder and static branches follow the same pattern, producing $\tilde{\mathbf{x}}_{k}^{\text{dec}}$, $\mathbf{h}^{\text{static}}$, and their respective weight outputs.

The static context vector $\mathbf{h}^{\text{static}}$ is obtained by applying embeddings and a static VSN to the province, origin/market, variety, and grade identifiers. It conditions all GRN modules in the encoder and decoder branches, allowing a single shared model to learn series-specific importance patterns without training a separate model per series. The encoder weights $\boldsymbol{\omega}_{\tau}^{\text{enc}}$ vary across time steps, while the decoder weights $\boldsymbol{\omega}_{k}^{\text{dec}}$ vary across forecast steps, capturing the changing relevance of each signal across different phases and horizons.

Decoder inputs are restricted to variables available at the forecast origin t: deterministically known covariates (holidays, solar terms, calendar encodings), low-frequency report indicators aligned by their public release dates, and short-term meteorological inputs. Because historical forecast archives were unavailable, decoder-side meteorological variables use CMA observed values as an oracle proxy for accurate short-term forecasts under the *Perfect Forecast Assumption* (assuming modern short-horizon forecasts have high accuracy). This setup does not leak future prices, but it introduces meteorological look-ahead by construction. Results should therefore be interpreted as a best-case meteorological-input setting (upper bound) relative to deployment with imperfect forecast products.

4.3.2. Temporal Fusion for Multi-Step Representation

After variable selection, a cross-attention mechanism integrates the encoder-side historical information with the decoder-side available-future information, constructing a distinct latent vector for each forecast step. GRNs conditioned on $\mathbf{h}^{\text{static}}$ project the selected encoder sequence $\tilde{\mathbf{x}}_{t-L+1:t}^{\text{enc}}$ into $\mathbf{H}^{\text{enc}} \in \mathbb{R}^{L\times d}$ and the selected decoder sequence $\tilde{\mathbf{x}}_{t+1:t+H}^{\text{dec}}$

into $\mathbf{H}^{\text{dec}} \in \mathbb{R}^{H \times d}$. Cross-attention then uses the future-step representations as queries and the historical representations as keys and values:

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right)\mathbf{V}, \tag{10}$$

where $\mathbf{Q} = \mathbf{H}^{\text{dec}}$ and $\mathbf{K} = \mathbf{V} = \mathbf{H}^{\text{enc}}$. The resulting fused vector $\mathbf{h}_k$ for step k encodes a step-specific summary of historical dynamics drawn from whichever portion of the window is most relevant, alongside the available-future signals and the series identity captured through the static context.

Two ETS outputs are injected into this pipeline as additional input variables. The historical residual sequence $\mathbf{r}^{(s)}_{t-L+1:t}$ enters on the encoder side, so that historical representations directly reflect the magnitude and pattern of past structural deviations. The structural baseline forecast $\hat{\mathbf{y}}^{\text{base},(s)}_{t+1:t+H}$ enters on the decoder side as an explicit extrapolation reference, anchoring the fusion module's learning around correcting the baseline rather than reconstructing the full price trajectory from scratch.

4.4. MoLE Residual Forecasting and Final Prediction

The residual process is not homogeneous across time. Inventory-release episodes, expectation reversals, and phase-wise demand surges each impose locally predictable linear structure on the disturbances, but these structures alternate in ways that are not known in advance [18]. A single regression head must average over all such regimes, introducing systematic bias at transition points. The Mixture of Linear Experts (MoLE) head avoids this by maintaining E separate linear predictors, each calibrated to a different local dynamic, and letting a gating network continuously shift their relative contributions as market conditions evolve.

Let $\mathbf{h}_k \in \mathbb{R}^d$ ($k = 1, \ldots, H$) be the fused representation from Component B. The e-th expert produces a residual prediction at step k as:

$$\hat{r}^{(e,s)}_{t+k} = \mathbf{w}_e^\top \mathbf{h}_k + b_e, \quad e = 1, \ldots, E, \tag{11}$$

and the gating network outputs an expert-weight vector $\boldsymbol{\pi}^{(s,t)} \in \mathbb{R}^E$ that satisfies $\pi^{(s,t)}_e \geq 0$ and $\sum_e \pi^{(s,t)}_e = 1$. We construct the gating input using both the static context vector and summary information of future-step representations, so that the weights reflect which volatility "state" the series is closer to at the forecast origin. The H-step residual forecast is then

$$\hat{r}^{(s)}_{t+k} = \sum_{e=1}^{E} \pi^{(s,t)}_e \hat{r}^{(e,s)}_{t+k}, \quad k = 1, \ldots, H. \tag{12}$$

The final price forecast adds the ETS structural baseline and the MoLE residual correction:

$$\hat{y}^{(s)}_{t+k} = \hat{y}^{\text{base},(s)}_{t+k} + \hat{r}^{(s)}_{t+k}, \quad k = 1, \ldots, H, \tag{13}$$

forming the complete forecast vector $\hat{\mathbf{y}}^{(s)}_{t+1:t+H}$. The additive decomposition is transparent by construction: $\hat{y}^{\text{base},(s)}$ carries the trend and seasonal contribution, and $\hat{r}^{(s)}$ carries the data-driven correction for exogenous shocks and regime shifts.

The gate weights $\boldsymbol{\pi}^{(s,t)}$ are retained as an interpretable output and can be inspected at any forecast origin to identify which dynamic mode the residual model is relying on—providing a post hoc record of inferred market states alongside the standard ETS parameters and per-step variable importance weights. All three components—ETS, VSN with temporal

fusion, and MoLE—are trained jointly under a single empirical risk objective, so structural decomposition and residual correction are co-adapted rather than optimized in isolation.

5. Experimental Evaluation

We report three experiments: a comparative evaluation against eight baselines (Experiment 1), ablation studies quantifying both module-level contributions and weather-input sensitivity (Experiment 2), and a two-dimensional hyperparameter sensitivity sweep (Experiment 3). The dataset, baselines, and evaluation metrics are introduced first.

5.1. Experimental Datasets

Garlic prices are driven simultaneously by harvest-induced seasonal rhythms and by inventory releases, expectation reversals, and episodic supply–demand shocks—dynamics that interact differently across provinces, origins, varieties, and grades. A single statistical assumption or end-to-end model may not capture both components well, which is precisely the problem GPF-EVMoLE is designed to address.

We assembled a multi-source daily garlic price dataset from industry data platforms, agricultural authorities, and the China Meteorological Administration, supplemented with holidays and solar terms, macro price indices, market-analysis report scores from the Ministry of Agriculture and Rural Affairs of China (MOA)—published on a regular official schedule—and annual supply–demand balance sheets. The dataset covers six origins across Henan, Shandong, and Jiangsu (Jinxiang, Laiwu, Lanling, Qixian, Zhongmu, and Pizhou Market), spanning 3 January 2023 to 15 May 2025 (864 days, 17,685 records). Series are defined by province, origin/market, variety, and grade, yielding 21 price series across three varieties and ten grades.

Inputs are organized into three groups. Static context (province, origin, variety, grade) encodes structural identity differences across series. Time-varying categorical features (weather condition, wind category, holiday name, solar term name) provide discrete event signals. Time-varying numerical features include price quantiles, meteorological observations (temperature, sunshine duration, monthly precipitation, and extreme temperatures), holiday window and distance variables, solar-term cycle encodings, macro price indices, MOA market-analysis scores, and supply–demand balance sheet fields.

Information availability is enforced for all non-meteorological decoder inputs. Low-frequency inputs—annual balance sheets and MOA market-analysis reports—take effect only after their public release dates, aligned by publication date and forward-filled. Holidays, solar terms, and calendar encodings are deterministically known and enter as decoder-side future covariates. For meteorological drivers, because historical forecast archives were unavailable, we use CMA observed values over $[t+1, t+H]$ as an oracle proxy for short-term forecasts under the *Perfect Forecast Assumption* (modern short-horizon forecasts have high accuracy). This oracle-weather design does not use future prices, but it introduces meteorological look-ahead. It isolates the model's use of meteorological drivers, and results should be interpreted as a best-case setting with respect to weather-forecast errors.

To assess the influence of this assumption, Experiment 2 includes a no-weather ablation that completely removes meteorological inputs from the proposed framework while keeping the same data split, training protocol, and remaining covariates.

All samples are constructed using a sliding window of length $L = 60$ over a strict chronological split (approximately 70/10/20): training from 3 January 2023 to 31 August 2024 (607 days), validation from 1 September 2024 to 30 November 2024 (91 days), and test from 1 December 2024 to 15 May 2025 (166 days). Samples are assigned to each split by their forecast-origin date, and each sample is retained only when the required historical window, horizon-specific target values, and aligned covariates are available for

the corresponding series. Sample counts are $H = 7$, train/val/test = 11,487/1932/2880; $H = 14$, train/val/test = 11,487/1932/2733. Dataset and split statistics are summarized in Tables 2 and 3.

Table 2. Dataset summary.

Item	Description
Target	Daily garlic price (CNY/jin)
Time span	3 January 2023∼15 May 2025 (864 days)
Regions	Henan, Shandong, Jiangsu (three provinces)
Origins/Markets	6 (Jinxiang, Laiwu, Lanling, Qixian, Zhongmu, Pizhou Market)
Varieties	3
Grades	10
Records	17,685
Series definition	(province, origin/market, variety, grade)
Number of series	21
External signals	Holidays/solar terms, macro price indices, supply–demand balance sheets, MOA market-analysis signals, weather variables, etc.

Table 3. Time split and sample sizes.

Split	Date Range	Number of Dates	Number of Samples ($H = 7$/$H = 14$)
Train	3 January 2023∼31 August 2024	607	11,487/11,487
Val	1 September 2024∼30 November 2024	91	1932/1932
Test	1 December 2024∼15 May 2025	166	2880/2733

5.2. Baseline Methods

All baselines share the same data split, sample construction, and evaluation metrics as GPF-EVMoLE:

- **Seasonal Naive:** A weekly-seasonality ($m = 7$) predictor that repeats observations from the same position in the previous season; specifically, $\hat{y}_{t+k} = y_{t+k-7}$, with no trainable parameters. It constitutes a natural lower-bound reference given the strong periodicity in garlic prices.
- **SARIMA (1970):** A seasonal ARIMA model capturing linear autoregressive structure and periodic patterns through seasonal differencing. We set `order = (1,1,1)`, `seasonal_order = (1,1,1,7)`, and `trend = 'n'`, and fit one model for each of the 21 price series.
- **Prophet (2018):** An additive decomposition model with trend and seasonality components. We use a linear growth term, additive seasonality, weekly and yearly seasonalities, no daily seasonality, `changepoint_prior_scale = 0.05`, `seasonality_prior_scale = 10.0`, and `n_changepoints = 25`. To limit evaluation cost across 21 series, we use a rolling refit strategy with a stride of 7 days.
- **Random Forest (2001):** A tree-based ensemble trained on lagged features that casts multi-step forecasting as supervised regression. The model uses 500 trees, `max_depth = 10`, `min_samples_split = 2`, `min_samples_leaf = 2`, `max_features = 1.0`, bootstrap sampling, squared-error splitting, and `random_state = 42`; supervised samples are built from the same 60-day input window and directly output the H-step forecast through multi-output regression.
- **SVR (1997):** Support vector regression with an RBF kernel, representing a medium-capacity non-linear baseline. We use `C = 1.0`, `epsilon = 0.1`, `gamma = 'scale'`, and `tol` $= 10^{-3}$; inputs are standardized using training-set statistics, and multi-step

prediction is implemented with a multi-output wrapper, equivalent to fitting one SVR for each forecast step.

- **LSTM (1997):** A recurrent network that learns temporal dependencies from historical windows, included as a standard deep-learning baseline. It uses an input length of 60, hidden size 64, two recurrent layers, dropout 0.1, no bidirectionality, a direct H-dimensional output head, Adam optimization with learning rate 10^{-3}, batch size 64, MSE loss, a maximum of 100 epochs, and early stopping patience of 10 epochs.
- **iTransformer (2024):** An inverted-dimension Transformer forecaster that applies attention over variate tokens to capture multivariate correlations [32]. We set `seq_len = 60`, `pred_len`$\in \{7, 14\}$, `d_model = 512`, 8 attention heads, 2 encoder layers, `d_ff = 2048`, dropout 0.1, GELU activation, learning rate 10^{-4}, batch size 32, 10 training epochs, and early stopping patience of 3 epochs.
- **Chronos (Large) (2024):** A pretrained time-series foundation model that tokenizes values via scaling and quantization and performs forecasting as token prediction; we use the `amazon/chronos-t5-large` checkpoint [23]. It is evaluated in a zero-shot setting without fine-tuning, with context length 60, prediction length $H \in \{7, 14\}$, 20 sampled trajectories, and the median trajectory as the point forecast.

5.3. Evaluation Metrics

Forecasting accuracy is measured by MAE, RMSE, and MAPE. Let N denote the number of test samples and $y_{i,k}$, $\hat{y}_{i,k}$ the ground-truth and predicted values at step k of sample i (CNY/jin). The three metrics are:

MAE is

$$\mathrm{MAE} = \frac{1}{NH} \sum_{i=1}^{N} \sum_{k=1}^{H} \left| y_{i,k} - \hat{y}_{i,k} \right|. \tag{14}$$

RMSE is

$$\mathrm{RMSE} = \sqrt{\frac{1}{NH} \sum_{i=1}^{N} \sum_{k=1}^{H} (y_{i,k} - \hat{y}_{i,k})^2}. \tag{15}$$

By squaring errors before averaging, RMSE penalizes large deviations more heavily than MAE and better reflects robustness under high-volatility periods.

MAPE is

$$\mathrm{MAPE} = \frac{1}{NH} \sum_{i=1}^{N} \sum_{k=1}^{H} \left| \frac{y_{i,k} - \hat{y}_{i,k}}{y_{i,k} + \varepsilon} \right|. \tag{16}$$

MAPE is scale-free, facilitating comparisons across series with different price levels. Since all target prices are positive, the metric is well defined; we set $\varepsilon = 10^{-8}$ as a numerical safeguard in denominators, with negligible effect on our data.

To test whether the observed accuracy differences are statistically significant, we apply the Diebold–Mariano (DM) test for equal predictive accuracy [33]. For each baseline, the loss differential is computed from the horizon-averaged absolute error:

$$d_i = \frac{1}{H} \sum_{k=1}^{H} \left| e_{i,k}^{\mathrm{base}} \right| - \frac{1}{H} \sum_{k=1}^{H} \left| e_{i,k}^{\mathrm{ours}} \right|. \tag{17}$$

The corresponding DM statistic is calculated as

$$\mathrm{DM} = \frac{\bar{d}}{\sqrt{\widehat{\sigma}_{\bar{d},H}^2 / n}}, \qquad \bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i, \tag{18}$$

where $\widehat{\sigma}^2_{d,H}$ denotes a heteroskedasticity- and autocorrelation-consistent long-run variance estimate using lag H. A positive DM statistic indicates lower loss for GPF-EVMoLE, and two-sided p-values are reported.

5.4. Performance Comparison

5.4.1. Experiment 1

Figure 2 reports test-set results for both horizons; GPF-EVMoLE attains the lowest MAE, RMSE, and MAPE across all baselines, including iTransformer and Chronos (Large).

The DM statistics in Figure 3 are positive for every pairwise comparison, indicating that the horizon-averaged absolute loss of each baseline is higher than that of GPF-EVMoLE. The pairwise tests show statistically significant differences across all baseline comparisons, including the strongest neural comparator at the longer horizon. These tests support the conclusion that the reported improvements are not only larger in aggregate error metrics, but also statistically significant under paired forecast-comparison tests.

Seasonal Naive is the strongest classical baseline, reflecting the dominant weekly periodicity in garlic prices. GPF-EVMoLE achieves consistent reductions in MAE, RMSE, and MAPE relative to Seasonal Naive under both horizons. Compared with the strongest neural competitor iTransformer, GPF-EVMoLE also maintains clear accuracy gains across the evaluated metrics. Errors grow as the horizon extends from 7 to 14 for all methods, consistent with empirically documented patterns of uncertainty accumulation in multi-step agricultural price forecasting [5]; meanwhile, GPF-EVMoLE shows the smallest growth in error and preserves its lead at the longer horizon.

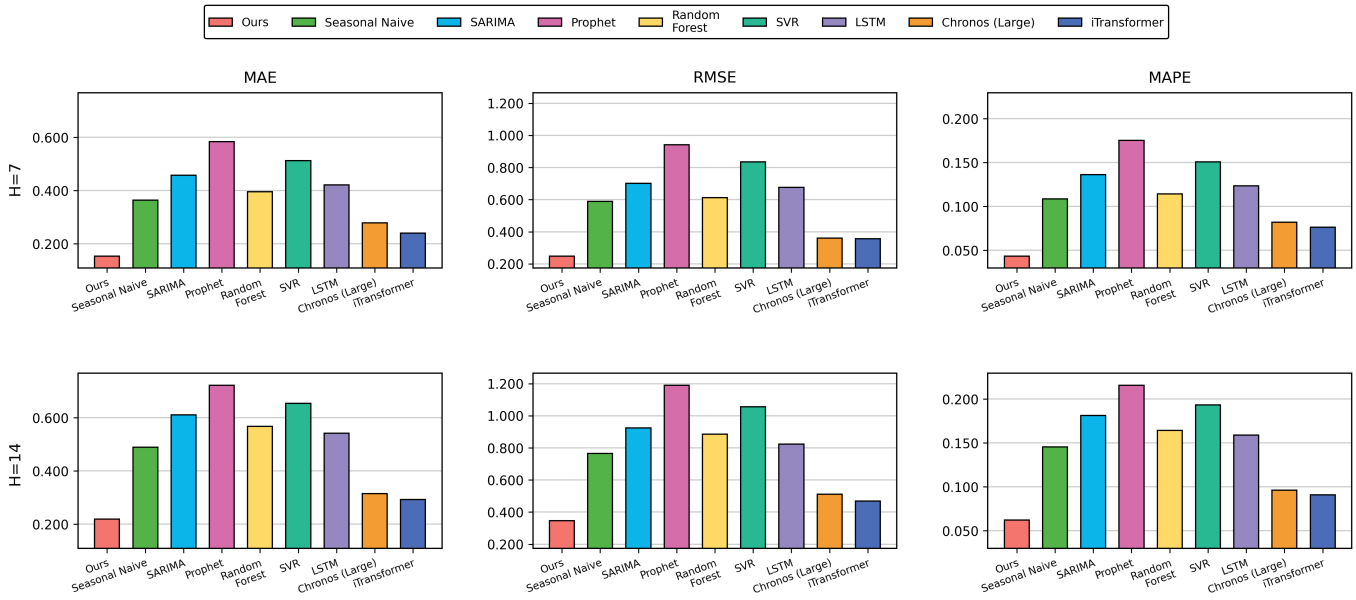

Figure 2. Main performance comparison of GPF-EVMoLE and baseline methods on the test set.

The pattern across baseline categories is informative. Seasonal Naive and SARIMA perform well because weekly seasonality is the strongest predictable signal in this market, yet univariate seasonal extrapolation cannot anticipate abrupt transitions triggered by inventory releases, expectation reversals, or supply–demand shocks. Prophet's additive structure fares worse when the series exhibits strong non-linearity and state changes. Random Forest, SVR, and LSTM bring non-linear capacity, but without explicit structural decomposition or access to the heterogeneous exogenous signals fed to GPF-EVMoLE, they struggle to maintain long-run structural consistency while tracking short-term disturbances, and their errors compound more rapidly over longer horizons. The iTransformer and

Chronos (Large) baselines provide substantially stronger neural comparators: iTransformer is the most competitive comparator, and Chronos benefits from large-scale pre-training, yet both remain consistently behind GPF-EVMoLE under all three metrics and both horizons. GPF-EVMoLE avoids these tensions by construction. ETS anchors the local trend and weekly periodicity as an interpretable baseline, freeing the downstream modules from relearning stable short-horizon structure; the VSN dynamically weights which signals among holidays, macro indices, MOA reports, and weather variables are most informative at each step; and MoLE distributes residual prediction across a small set of regime-specific linear experts, sharpening correction during phase transitions. The result is consistent improvement across all three metrics and both horizons in a real market where strong weekly periodicity, cross-series heterogeneity, and regime switching coexist.

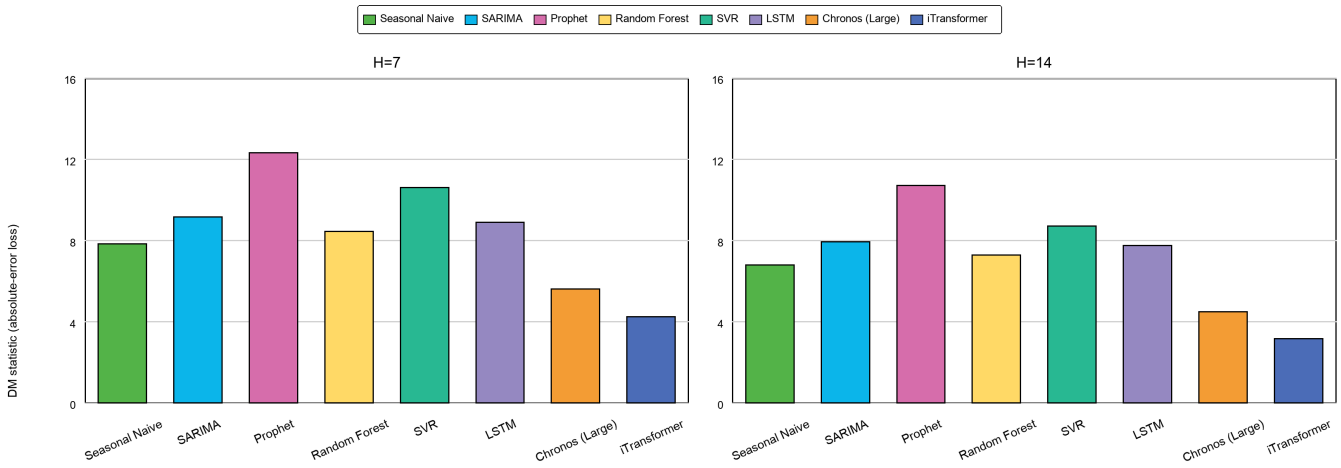

Figure 3. Diebold–Mariano test statistics comparing GPF-EVMoLE against each baseline.

5.4.2. Experiment 2

Four ablation variants test the contribution of each learned module and the sensitivity to weather inputs. ETS-only retains only the structural baseline; the abbreviation w/o denotes "without" and is retained for consistency with Figure 4; w/o MoLE replaces the expert mixture with a single regression head; w/o VSN removes variable selection and applies a uniform linear transformation to all inputs; and w/o weather completely removes meteorological variables, including the decoder-side oracle-weather proxy. All variants are trained and evaluated under identical splits and training settings.

Removing each tested component or input group degrades MAE, RMSE, and MAPE under both horizons (Figure 4), confirming that the full configuration is not dominated by a single isolated design choice. The w/o weather variant examines the oracle-weather setting: removing meteorological inputs increases error at both horizons, showing that weather information provides a useful incremental signal for garlic price forecasting. At the same time, w/o weather remains clearly better than w/o MoLE and ETS-only, indicating that the model's advantage does not rely solely on the oracle-weather proxy. Dropping VSN also causes a moderate but consistent deterioration, which supports the role of dynamic per-step variable reweighting in suppressing redundant or noisy external signals. Removing MoLE has a larger effect than removing weather inputs, especially at the longer horizon, consistent with the interpretation that expert switching primarily reduces large deviations during regime transitions. Finally, ETS-only remains the weakest variant, confirming that structural extrapolation alone cannot track short-term disturbances shaped by heterogeneous external drivers.

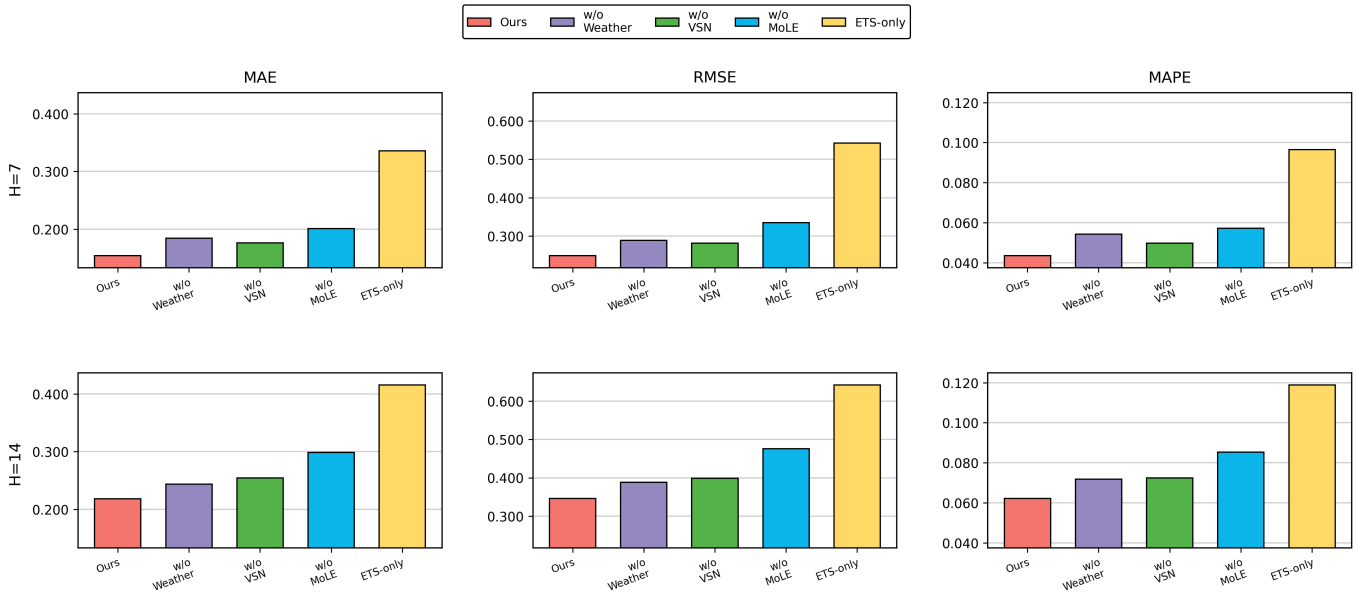

Figure 4. Ablation results of GPF-EVMoLE variants on the test set.

5.5. Sensitivity Analysis of Key Hyperparameters

We examine the sensitivity of GPF-EVMoLE to two architectural hyperparameters: the hidden dimension d and the number of MoLE experts E. A 5×5 grid over $d \in \{16, 32, 64, 128, 256\}$ and $E \in \{2, 3, 4, 6, 8\}$ is evaluated under identical splits and training settings.

Figure 5 shows the resulting error surfaces; the default $(d, E) = (64, 4)$ lies near the global minimum under both horizons.

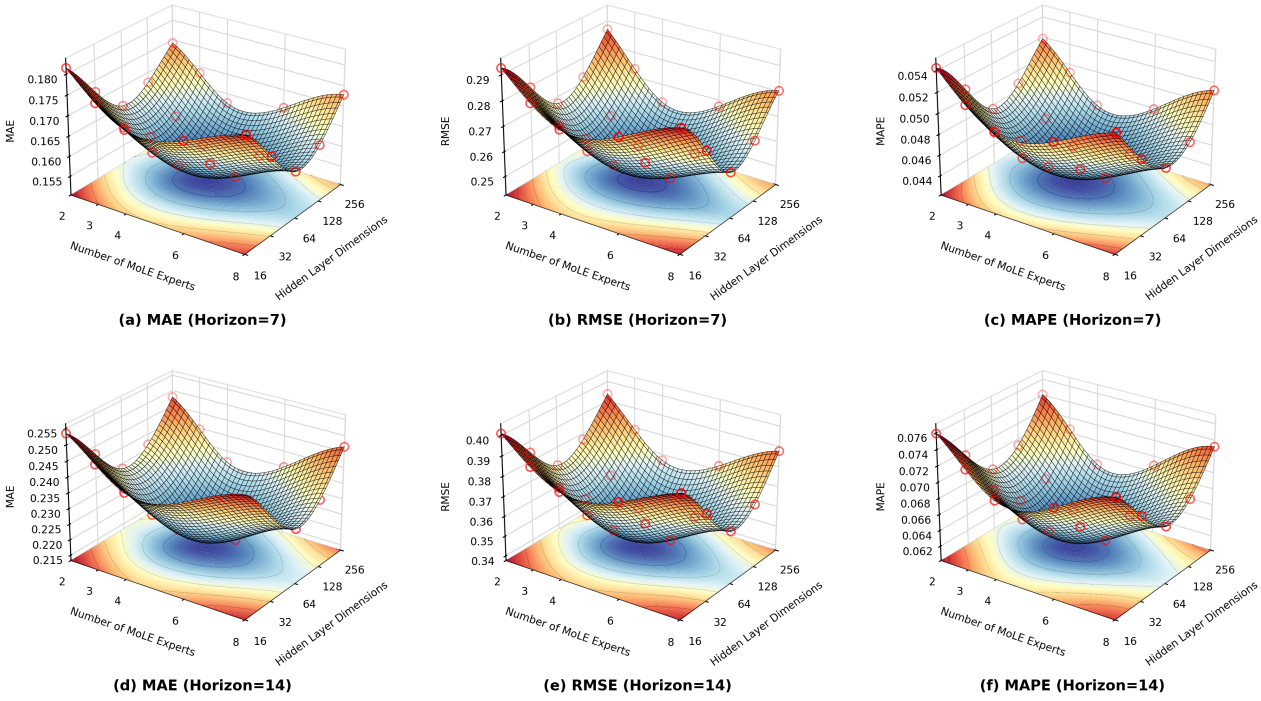

Figure 5. Sensitivity surfaces of GPF-EVMoLE with respect to hidden dimension d and the number of MoLE experts E. Surface colors and projected contours indicate error magnitude, with blue denoting lower errors and yellow/orange denoting higher errors. Red circles mark the evaluated hyperparameter grid points.

https://doi.org/10.3390/su18094404

The hidden dimension d exhibits a clear capacity trade-off. When d is too small, the VSN and temporal fusion modules have insufficient representational capacity. Increasing d improves performance up to a moderate range, after which further enlargement no longer provides consistent gains and can slightly degrade accuracy. This pattern suggests that the finite-sample regime of a real-world agricultural dataset cannot support unlimited model expansion without overfitting or optimization instability [5,19].

The number of MoLE experts E follows a similar inverted-U pattern. Too few experts provide insufficient diversity to cover the range of residual regimes in garlic markets, whereas too many experts make the routing problem harder and reduce the effective data available to each expert. A moderate number of experts therefore provides the best balance between flexibility and stability.

Overall, the loss surfaces are smooth and bowl-shaped, with the default $(d, E) = (64, 4)$ located in a near-optimal region under both forecast horizons. Neighboring configurations remain competitive, indicating that the model is not brittle to a single hyperparameter choice. The surfaces under $H = 14$ are elevated and slightly steeper, consistent with greater sensitivity to capacity decisions as uncertainty accumulates over longer horizons.

6. Conclusions and Discussion

This paper investigates multi-step garlic price forecasting through structural decomposition, dynamic feature selection, and regime-adaptive residual modeling. Existing methods typically overlook the conflict between stable seasonal patterns and abrupt price transitions driven by speculation and inventory releases, and they also struggle to dynamically assign appropriate weights to heterogeneous external signals. GPF-EVMoLE addresses these challenges through a compositional architecture.

Experiments conducted over an 864-day period on 21 price series from six major production origins confirm that GPF-EVMoLE significantly reduces MAE, RMSE, and MAPE. It outperforms statistical, machine learning, and deep learning baselines, with its advantages becoming particularly pronounced at longer forecasting horizons (7 and 14 days) and during severe regime shifts. These findings validate a constrained divide-and-conquer strategy for complex commodity forecasting rather than a loose combination of existing modules: ETS stabilizes the local trend and weekly-seasonal baseline, VSN controls how external information enters each forecast step, and MoLE adapts residual corrections to inferred market states. By defining these interfaces and training them jointly, the framework reduces the instability that arises when a monolithic model must handle long-run structure, short-term shocks, and regime shifts at the same time.

Although the proposed method demonstrates strong performance in multi-step garlic price forecasting, several limitations warrant careful interpretation. First, the treatment of meteorological drivers remains an optimistic evaluation setting. Because historical weather forecast archives were unavailable, this study uses realized CMA observations rather than archived forecasts that would have been strictly available at each prediction origin. The no-weather ablation shows that meteorological inputs provide useful incremental information, but it also reinforces that deployment performance may depend on the accuracy and availability of real-time weather forecasts. Future work should therefore evaluate the framework with archived forecast products or operational forecast feeds when such data become available.

Second, the generalizability of the framework should be validated beyond the current empirical setting. The dataset is focused on Chinese garlic prices from six major production origins over 2023–2025, and the model relies on external signals such as MOA reports, balance sheets, weather records, and calendar variables. Other crops, countries, or markets with different storage conditions, policy reporting systems, or data availability may

require re-tuning of the input design and validation protocol. Finally, the current study produces point forecasts, whereas practical risk management often requires uncertainty estimates. Future research may explore interval or probabilistic forecasting methods, such as conformal prediction or probabilistic mixtures of experts, to strengthen early warning, risk identification, and decision-support applications.

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

Funding: This research was funded by the Key Research and Development Plan of Shandong Province (2022CXGC010609), the Key Research and Development Plan of Shandong Province (2022TZXD0030), and the Special Funds for Centralized Guidance of Local Science and Technology Development (YDZX2022073).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgments: Thanks to the Agricultural Big Data Research Center, Shandong Agricultural University, for its support to scientific research.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

$\mathcal{S}$	Set of daily garlic price series grouped by province, origin/market, variety, and grade
s	Series index, $s \in \mathcal{S}$
t	Time index (daily), representing a forecast origin or an observation time
L	Historical input window length (encoder length)
H	Forecast horizon (next H days), with $H \in \{7, 14\}$ in this paper
$y_t^{(s)}$	Target value of series s on day t (price, CNY/jin)
$\mathbf{y}_{t-L+1:t}^{(s)}$	Historical target vector of series s over $[t-L+1, t]$
$\mathbf{c}^{(s)}$	Static context features of series s (province, origin/market, variety, grade)
$\mathbf{x}_{\tau}^{(s)}$	Time-varying covariate vector of series s at time τ (observed in history)
$\mathbf{x}_{t-L+1:t}^{(s)}$	Historical covariate sequence available at forecast origin t (encoder inputs)
$\mathbf{z}_{t+1:t+H}^{(s)}$	Known-in-advance or forecast-available future covariate sequence at forecast origin t (decoder inputs)
$\hat{\mathbf{y}}_{t+1:t+H}^{(s)}$	Forecast vector of series s for the next H days from origin t
$f_\theta(\cdot)$	Multi-step forecasting model parameterized by θ
$\mathcal{D}$	Dataset of samples, consisting of pairs (s, t) (series and forecast origin)
$\mathcal{L}(\cdot, \cdot)$	Loss function (e.g., MAE/MSE) measuring forecast errors
m	Seasonal length in ETS decomposition
α, β, γ	ETS smoothing coefficients for level, trend, and season updates
$\hat{\mathbf{y}}_{t+1:t+H}^{\text{base},(s)}$	Structural baseline forecast produced by ETS (for H steps)
$\mathbf{r}_{t-L+1:t}^{(s)}$	Historical residual sequence from ETS decomposition (short-term disturbance)
$\hat{\mathbf{r}}_{t+1:t+H}^{(s)}$	Residual forecast vector produced by MoLE (for H steps)
E	Number of experts in MoLE

$\boldsymbol{\pi}^{(s,t)}$	Expert-weight vector output by the MoLE gating network (nonnegative and sums to one)
$\boldsymbol{\omega}^{\text{static}}, \boldsymbol{\omega}_{\tau}^{\text{enc}}, \boldsymbol{\omega}_{k}^{\text{dec}}$	VSN variable-importance weights for static/encoder/decoder inputs (softmax-normalized)

References

1. Mustafa, Z.; Vitali, G.; Huffaker, R.; Canavari, M. A systematic review on price volatility in agriculture. *J. Econ. Surv.* **2024**, *38*, 268–294. [CrossRef]
2. Bellemare, M.F. Rising Food Prices, Food Price Volatility, and Social Unrest. *Am. J. Agric. Econ.* **2015**, *97*, 1–21. [CrossRef]
3. Kiloes, A.M.; Puspitasari; Sulistyaningrum, A.; Khaririyatun, N.; Mulyono, D.; Prabawati, S.; Anwarudin Syah, M.J.; Devy, N.F.; Hardiyanto. Unravelling the provisioning system of a strategic food commodity to minimise import dependency: A study of garlic in Indonesia. *Food Policy* **2024**, *123*, 102604. [CrossRef]
4. Hinlo, J.; Lee, Y. From price stability to food security: What are the drivers of price volatilities in philippine vegetable markets? *Agric. Food Secur.* **2026**, *15*, 28. [CrossRef]
5. Paul, R.K.; Yeasin, M.; Tamilselvi, C.; Paul, A.K.; Sharma, P.; Birthal, P.S. Can Deep Learning Models Enhance the Accuracy of Agricultural Price Forecasting? Insights From India. *Intell. Syst. Account. Financ. Manag.* **2025**, *32*, e70002. [CrossRef]
6. Nayak, G.H.H.; Alam, M.W.; Singh, K.N.; Avinash, G.; Kumar, R.R.; Ray, M.; Deb, C.K. Exogenous variable driven deep learning models for improved price forecasting of TOP crops in India. *Sci. Rep.* **2024**, *14*, 17203. [CrossRef]
7. Muthuswamy, K.; Dolli, S.; Khandeparkar, K.; Gowda, C.; Narasimma Bharathi, V.P.; Shivakumar, K.M.; Sumathi, C.S.; Appavu, S.; Natarajan, B.; Sivakumar, K. Weather-driven groundnut price forecasting and profitability assessment of cropping patterns in Tamil Nadu using boosting algorithms. *Sci. Rep.* **2025**, *15*, 33880. [CrossRef]
8. Manogna, R.L.; Dharmaji, V.; Sarang, S. Enhancing agricultural commodity price forecasting with deep learning. *Sci. Rep.* **2025**, *15*, 20903. [CrossRef] [PubMed]
9. Sediyono, E.; Hartomo, K.D.; Arthur, C.; Utami, I.; Prabowo, R.; Chiong, R. An integrated framework for multi-commodity agricultural price forecasting and anomaly detection using attention-boosted models. *J. Agric. Food Res.* **2025**, *22*, 102021. [CrossRef]
10. Awokuse, T.; Lim, S.; Santeramo, F.; Steinbach, S. Robust policy frameworks for strengthening the resilience and sustainability of agri-food global value chains. *Food Policy* **2024**, *127*, 102714. [CrossRef]
11. Liang, W.; Sivashankar, P.; Hua, Y.; Li, W. Global food expenditure patterns diverge between low-income and high-income countries. *Nat. Food* **2024**, *5*, 592–602. [CrossRef]
12. Stehl, J.; Vonderschmidt, A.; Vollmer, S.; Alexander, P.; Jaacks, L.M. Gap between national food production and food-based dietary guidance highlights lack of national self-sufficiency. *Nat. Food* **2025**, *6*, 571–576. [CrossRef]
13. Berevoianu, R.L.; Ilie, D.M.; Giucă, A.D. The Economic Profitability of Garlic Cultivation through the Application of the "De Minimis" Aid Scheme, Estimates for the Year 2024 in Romania. In Proceedings of the Paradigm, Bucharest University of Economic Studies, Bucharest, Romania, 21–23 March 2024; Volume 18, pp. 1552–1562. [CrossRef]
14. Shankar, S.V.; Chandel, A.; Nandhini, C.; Yadav, A.K.; Gupta, P. Statistical analysis of market cointegration and price dynamics for garlic in key northern Indian markets. *J. Appl. Hortic.* **2025**, *27*, 108–112. [CrossRef]
15. Afriani, B.D.; Fernandez, E.; Sudjatmiko, D.P. Analysis of Market Integration and Price Transmission of Garlic in West Nusa Tenggara Province. *Int. J. Manag. Commer. Innov.* **2025**, *13*, 280–291. [CrossRef]
16. Valenti, D.; Bertoni, D.; Cavicchioli, D.; Olper, A. Understanding the role of supply and demand factors in the global wheat market. *Eur. Rev. Agric. Econ.* **2026**, *53*, 102–126. [CrossRef]
17. Madziwa, L.; Pillalamarry, M.; Chatterjee, S. Integrating stochastic mine planning model with ARDL commodity price forecasting. *Resour. Policy* **2023**, *85*, 104014. [CrossRef]
18. Crespo Cuaresma, J.; Fortin, I.; Hlouskova, J.; Obersteiner, M. Regime-dependent commodity price dynamics: A predictive analysis. *J. Forecast.* **2024**, *43*, 2822–2847. [CrossRef]
19. Wang, G.; Xu, S.; Chen, Z.; Li, Y. A Hybrid Model Integrating Variational Mode Decomposition and Intelligent Optimization for Vegetable Price Prediction. *Agriculture* **2025**, *15*, 919. [CrossRef]
20. Sun, K.; Yao, Q.; Li, Y. A novel agricultural commodity price prediction model integrating deep learning and enhanced swarm intelligence algorithm. *PLoS ONE* **2025**, *20*, e0337103. [CrossRef]
21. Sugiartawan, P.; Wardani, N.W. Enhancing Price Classification of Chili Using Gradient Boosting Machines. *J. Sist. Inf. Dan Komput. Terap. Indones. (JSIKTI)* **2024**, *6*, 165–174. [CrossRef]
22. Woo, G.; Liu, C.; Kumar, A.; Xiong, C.; Savarese, S.; Sahoo, D. Unified Training of Universal Time Series Forecasting Transformers. *arXiv* **2024**, arXiv:2402.02592. [CrossRef]
23. Ansari, A.F.; Stella, L.; Turkmen, C.; Zhang, X.; Mercado, P.; Shen, H.; Shchur, O.; Rangapuram, S.S.; Arango, S.P.; Kapoor, S.; et al. Chronos: Learning the Language of Time Series. *arXiv* **2024**, arXiv:2403.07815. [CrossRef]

24. Goswami, M.; Szafer, K.; Choudhry, A.; Cai, Y.; Li, S.; Dubrawski, A. MOMENT: A Family of Open Time-series Foundation Models. *arXiv* **2024**, arXiv:2402.03885. [CrossRef]
25. Shi, X.; Wang, S.; Nie, Y.; Li, D.; Ye, Z.; Wen, Q.; Jin, M. Time-MoE: Billion-Scale Time Series Foundation Models with Mixture of Experts. *arXiv* **2025**, arXiv:2409.16040. [CrossRef]
26. Jin, M.; Wang, S.; Ma, L.; Chu, Z.; Zhang, J.Y.; Shi, X.; Chen, P.Y.; Liang, Y.; Li, Y.F.; Pan, S.; et al. Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. *arXiv* **2024**, arXiv:2310.01728. [CrossRef]
27. Guan, T.; Meng, Z.; Li, D.; Wang, S.; Yang, C.H.H.; Wen, Q.; Liu, Z.; Siniscalchi, S.M.; Jin, M.; Pan, S. TimeOmni-1: Incentivizing Complex Reasoning with Time Series in Large Language Models. *arXiv* **2025**, arXiv:2509.24803. [CrossRef]
28. Zhang, L.; Padhan, H.; Singh, S.K.; Gupta, M. The impact of renewable energy on inflation in G7 economies: Evidence from artificial neural networks and machine learning methods. *Energy Econ.* **2024**, *136*, 107718. [CrossRef]
29. Hutchins, J.; Hueth, B. 100 years of data sovereignty: Cooperative data governance and innovation in US dairy. *Appl. Econ. Perspect. Policy* **2023**, *45*, 1551–1576. [CrossRef]
30. Ramsey, A.F.; Adjemian, M.K. Forecast combination in agricultural economics: Past, present, and the future. *Appl. Econ. Perspect. Policy* **2024**, *46*, 1450–1478. [CrossRef]
31. Yuan, F.; Huang, X.; Jiang, H.; Jiang, Y.; Zuo, Z.; Wang, L.; Wang, Y.; Gu, S.; Peng, Y. An xLSTM–XGBoost Ensemble Model for Forecasting Non-Stationary and Highly Volatile Gasoline Price. *Computers* **2025**, *14*, 256. [CrossRef]
32. Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; Long, M. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. *arXiv* **2024**, arXiv:2310.06625. [CrossRef]
33. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [CrossRef]

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.