Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: https://www.science-gate.com/IJAAS.html

An adjoint sensitivity-based Lévy-induced stochastic differential equation network for agricultural price prediction

Nur Amalina Shafie [1, *], Mohd Tahir Ismail [2], Isnewati Ab Malek [1], Zahari Md Rodzi [1, 3]

[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Seremban, Malaysia*
[2]*School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia*
[3]*Accounting Research Institute (ARI), Universiti Teknologi MARA, Shah Alam, Malaysia*

ARTICLE INFO

Article history:
Received 23 May 2026
Received in revised form
17 September 2026
Accepted 10 October 2026

Keywords:
Adjoint sensitivity
Lévy processes
Stochastic differential equations
Neural networks
Agricultural price forecasting

ABSTRACT

Agricultural commodity prices exhibit non-linearity, heavy-tailed return distributions, and occasional large jumps that standard Gaussian models cannot adequately capture. Existing stochastic neural networks, including the Lévy-Induced Stochastic Differential Equation Network (LDE-Net), address these characteristics but suffer from gradient instability during backpropagation through stochastic Lévy trajectories. This study proposes the Adjoint LDE-Net, extending the adjoint sensitivity method from deterministic Neural ODEs to the Lévy stochastic setting. Its backward adjoint SDE computes parameter gradients without traversing the forward Lévy trajectory, reducing memory complexity from O(N) to O(1) and improving gradient stability. Two datasets are evaluated using a 70:30 train-test split: the IMF Food and Beverage Price Index (434 monthly observations) and the FAMA Malaysian Tomato Price series (208 weekly observations). Both series exhibit chaos according to the Wolf method and stationarity in log-returns according to the Augmented Dickey-Fuller test. Finite-difference sensitivity analysis yields $\partial f/\partial\sigma = \partial f/\partial\mu = 1.000$ for both datasets, indicating the importance of the diffusion pathway. Neural SDE achieves the best predictive performance on both datasets, while Adjoint LDE-Net ranks second on the IMF dataset. Both stochastic neural models outperform ARIMA and GBM by more than an order of magnitude in MSE on the IMF dataset. The proposed method provides a theoretically grounded, memory-efficient approach to Lévy SDE training. However, its predictive accuracy does not exceed that of Neural SDE, and instability on the Tomato series highlights the need for α calibration and gradient clipping.

© 2026 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Agricultural commodity price series exhibit a distinctive set of statistical properties that challenge conventional time series modeling: non-linearity, non-stationarity, heavy-tailed return distributions, and episodic large jumps. These properties arise from the interaction of seasonal production cycles, weather-driven supply shocks, and discrete policy interventions. Accurate forecasting of such series requires models that can represent both the directional trend and the stochastic uncertainty of price movements, including the probability of extreme events that standard Gaussian models systematically underestimate.

Traditional forecast models such as ARIMA assume that there are linear data generating processes with disturbances from Gaussian processes. However, such an assumption becomes invalid in agricultural commodities' price forecasting, where there are non-linear processes, shifts in distributions, and sudden jumps in prices due to weather disturbances, harvest processes, and government price intervention (Box et al., 2016). Agricultural prices are particularly vulnerable to such processes. A single period of drought can increase tomato prices by 200–300% within just two weeks. Similarly, international commodity price shocks, such as the global food price increase during 2021–2022 following the conflict in Ukraine, can affect markets in countries such as Malaysia within a few months. These dynamics result in return distributions that are non-Gaussian, time-varying, and characterized by heavy tails and sudden price

* Corresponding Author.
Email Address: amalina@uitm.edu.my (N. A. Shafie)
https://doi.org/10.21833/ijaas.2026.10.003
Corresponding author's ORCID profile:
https://orcid.org/0000-0003-4522-1758
2313-626X/© 2026 The Authors. Published by IASE.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

jumps. Such characteristics are difficult for classical linear models to capture.

Deep learning models capture non-linearity but remain largely deterministic. Neural Ordinary Differential Equations, NODE (Chen et al., 2018), introduced the adjoint sensitivity method for memory-efficient gradient computation in continuous-depth networks. SDE-Net (Kong et al., 2020) extended this to stochastic settings using Brownian motion, but Brownian motion generates only Gaussian increments, which are insufficient for series with non-Gaussian, heavy-tailed jump behavior. The key insight motivating the present work is that the stochastic driver used to model price uncertainty must match the distributional characteristics of the price process itself: Gaussian noise for Gaussian returns, Lévy noise for heavy-tailed, jump-prone returns.

Yang et al. (2023) addressed this limitation through the Lévy-Induced Stochastic Differential Equation Network (LDE-Net), which uses α-stable Lévy motion as the stochastic driver. On high-volatility Chinese stock indices, LDE-Net outperformed ARIMA, LSTM, and SDE-Net with proven convergence guarantees. However, two open problems were identified: gradient instability during backpropagation through stochastic Lévy trajectories, and the extension of the adjoint sensitivity method from the deterministic Neural ODE setting to the Lévy stochastic setting. This paper addresses both problems through the Adjoint LDE-Net.

Critically, this study also investigates the connection between parameter sensitivity in agricultural price data and the gradient sensitivity problem in Lévy SDE training. Through finite-difference sensitivity analysis, we verify that the GBM benchmark prediction is, by construction, equally sensitive to both drift (μ) and diffusion (σ) parameters in both datasets ($\partial f/\partial\sigma = \partial f/\partial\mu = 1.0000$). This result, which follows analytically from the additive structure of the GBM prediction function rather than constituting an empirical discovery, makes explicit that the diffusion pathway carries non-negligible gradient weight and thereby motivates the Adjoint LDE-Net: when the loss is sensitive to the diffusion parameter, standard backpropagation through the stochastic Lévy trajectory produces unstable gradients that degrade training. The adjoint backward SDE mitigates this by computing sensitivity gradients without traversing the forward Lévy trajectory, thereby stabilizing the diffusion-pathway sensitivity that drives training instability in the original LDE-Net. Two datasets with contrasting profiles are evaluated which are the IMF Food and Beverage Price Index (composite global index) and the FAMA Malaysian Tomato Price series (individual domestic commodity) to test the model across different volatility regimes.

The objectives of this study are to identify the effect of parameter sensitivity in predicting agricultural prices using finite-difference analysis and establish its connection to gradient instability in Lévy SDE training, to develop the Adjoint LDE-Net by extending the adjoint sensitivity method to the Lévy stochastic setting via a backward adjoint SDE and to validate the proposed model against ARIMA, GBM, and Neural SDE benchmarks using MSE, RMSE, MAE, and MAPE, and to characterize the relationship between the Lévy stability index α and training convergence behavior.

The paper is organized as follows: Section 2 reviews relevant literature; Section 3 describes data, research framework, and statistical characterization; Section 4 presents the Adjoint LDE-Net model and adjoint algorithm; Section 5 provides theoretical convergence analysis; Section 6 reports experimental results including training stability analysis and sensitivity; Section 7 concludes with methodological recommendations.

2. Literature review

2.1. Agricultural price forecasting

Forecasting prices for agricultural commodities has been the focus of considerable academic research during the last twenty years due to the necessity to ensure food security and manage investment risks. Traditional methods of forecasting have been based on the ARIMA family of models, which are time-series linear and Gaussian methods. Though relatively easy to calculate, these models inevitably fail to forecast the complex and nonlinear processes, sudden price changes, and structural breaks inherent to agricultural market dynamics (Box et al., 2016; Mao et al., 2022). The main problem of ARIMA is that these models assume that the process is stationary and linear, while prices in the agricultural commodity market depend on chaotic weather processes, geopolitics, and government actions.

Limitations of linear models led to increased interest in machine learning-based approaches. Wang et al. (2011) proved that the backpropagation neural network is effective for stock index prediction, laying the theoretical groundwork for neural-based approaches to forecasting in finance. Paul et al. (2022) showed that neural network-based models outperform classical statistical models for predicting agricultural prices in developing countries because non-linear neural structures can take into account production-demand relationships that are overlooked by ARIMA. According to Qazi et al. (2022), frameworks based on the Internet of Things and artificial intelligence mark the emergence of smart agriculture infrastructure of the future due to real-time sensor data that makes forecasting more responsive. According to Bhagat et al. (2022), price forecasting was identified as one of the most popular and influential research areas in the field of applications of artificial intelligence in sustainable agriculture, as shown in a systematic bibliometric review of the literature.

The development of hybrid techniques that incorporate stochastic elements and machine

learning elements appears to be a very promising approach. For example, Mao et al. (2022) used ARIMA on different vegetable prices in China and showed that having enough training data is vital for time series stability, which is a fact highly relevant to this study's 300-epoch training period. It seems that combining stochastic differential equations and neural networks is the most recent development in methodology driven by the necessity to model forecasting accuracy and its uncertainty at once.

2.2. Stochastic differential equations in financial modeling

Stochastic differential equations have served as the mathematical backbone of financial price modeling for decades, with applications spanning option pricing, volatility surface estimation, interest rate modeling, and commodity price dynamics. The standard Geometric Brownian Motion (GBM) model, while analytically tractable and widely implemented, fails to reproduce the heavy-tailed, skewed return distributions observed in empirical financial data. The classic Heston (1993) stochastic volatility model introduced mean-reverting variance dynamics to partially address this, while jump-diffusion models (Merton, 1976) added discontinuous price movements. However, these parametric approaches require strong distributional assumptions, are prone to misspecification, and do not adapt their dynamics to changing market regimes without re-estimation.

The emergence of Lévy processes as stochastic drivers offered a more flexible framework for capturing non-Gaussian price dynamics. α-Stable Lévy motion accommodates both the heavy tails and the infinite-variance characteristics of financial return distributions without requiring explicit specification of the jump size distribution. Samorodnitsky and Taqqu (1994) provided the rigorous mathematical foundations for stable processes, establishing the characteristic function representation, convergence results, and approximation properties that underpin the theoretical guarantees of the LDE-Net architecture. Jia and Benson (2019) introduced Neural Jump SDEs, combining continuous diffusion with discrete Poisson jump processes for financial time series, which is an important intermediate step toward the fully Lévy-driven architecture of Yang et al. (2023).

The relevance of Lévy processes to agricultural price modeling is particularly compelling. Agricultural commodity prices are subject to seasonal harvest shocks that produce large, sudden movements in short-term windows; weather-induced supply disruptions that manifest as regime shifts in price volatility; and government price interventions that introduce discrete discontinuities into otherwise continuous price trajectories. These mechanisms collectively generate return distributions with excess kurtosis, heavy tails, and right-skewness. All characteristics that α-stable Lévy motion can represent but Gaussian models cannot. It is noted, however, that the specific datasets used in this study exhibit kurtosis values closer to the Gaussian benchmark than to genuinely heavy-tailed distributions in Section 3.1, which partially explains why Neural SDE is driven by Gaussian increments that perform competitively in the study.

2.3. Neural SDE and LDE-Net

The integration of neural networks with stochastic differential equations has resulted in an effective family of models that combine the representational power of deep learning models and the probabilistic elegance of stochastic calculus. In their paper, Neural Ordinary Differential Equations, Chen et al. (2018) have shown that it is possible to train continuous depth networks via the adjoint sensitivity method with O(1) memory cost, laying a strong foundation for this paper. The stochastic counterpart was provided by SDE-Net (Kong et al., 2020), which added Brownian motion noise to the hidden state update mechanism, making it possible to quantify uncertainty using the diffusion term, which increases whenever the model is fed with out-of-distribution data. However, due to the Gaussian nature of Brownian motion increments, it becomes impossible to model the non-Gaussian epistemic uncertainty caused by episodic shocks in prices, which characterizes the agricultural commodity market dynamics.

The Lévy-Induced SDE framework with neural networks, presented by Yang et al. (2023), solves this problem by showing through theory and experiment that α-stable Lévy motion offers better representations of epistemic jumps than Gaussian noise. The framework had guaranteed convergence properties without the "curse of dimensionality," which was an essential feature needed for the high-dimensional agricultural dataset problem. The next model, LDE-Net (Yang et al., 2023), used chaos detection by using the Lyapunov exponent, phase space reconstruction by Cao's (1997) method, and an attention mechanism for multi-step forecasting, obtaining top-of-the-line results on Chinese stock indices such as the SSE Energy Index (MSE = 0.0009) and SSE50 Index (MSE = 0.0029), significantly outperforming ARIMA, LSTM, and SDE-Net baselines.

Critically, Yang et al. (2023) specifically highlighted two research gaps, namely, the gradient sensitivity problem created by backpropagation on the stochastic Lévy path and the generalization of the adjoint sensitivity approach from the deterministic Neural ODEs framework to the stochastic Lévy framework.

3. Data exploration and problem setting

3.1. Datasets

Two agricultural price datasets with contrasting distributional profiles are used in this study. Dataset 1 is the IMF Food and Beverage Price Index. This composite global commodity index is sourced from

the International Monetary Fund Primary Commodity Prices database (www.imf.org/en/research/commodity-prices), tracking international prices of food and beverage commodities including cereals, vegetable oils, meat, seafood and sugar. The dataset comprises 434 monthly observations from 2003 to 2023. As an international benchmark, PFANDB directly influences Malaysia's import cost structures, domestic price-setting, and agricultural export competitiveness.

Dataset 2 is the FAMA Malaysian Tomato Price Series. This individual commodity series comprises 208 weekly observations of fresh tomato wholesale prices in Malaysian markets, sourced from the Federal Agricultural Marketing Authority (FAMA) under Act 141, Ministry of Agriculture and Food Security. Tomato prices reflect strong seasonal production cycles, irregular supply disruptions, and domestic demand dynamics. Log-returns are computed for both series as in Eq. 1.

$$r_t = ln(P_t / P_{t-1}) \quad (1)$$

The values for kurtosis in Table 1 serve to illustrate an important difference, as well as the limitation of this experiment, which has to be explicitly pointed out. The FAMA Tomato series shows an excess kurtosis of +0.075, almost perfectly symmetric around the Gaussian value of zero. The IMF Food and Beverage Index series displays an excess kurtosis of -1.107, which makes it a platykurtic distribution, with tails smaller than in the Gaussian one – the result of smoothing through aggregation of several international commodity prices. Hence, neither dataset shows pronounced heavy-tailed behavior, which would ordinarily motivate a Lévy-based approach.

This paradox is directly solved by this research: since α-stable Lévy motion collapses into the Brownian motion when α = 2, the LDE-Net family is a strict generalization of Gaussian stochastic neural networks, and testing it on a near-Gaussian series is actually the most conservative stress-test of this framework. The near-Gaussian nature of the two series is also the reason for the competitive performance of the Gaussian-driven Neural SDE model in Section 6, and the necessity of testing the framework on heavy-tailed agricultural data is identified in Section 7.

Table 1: Descriptive statistics of both datasets

Statistic	Food & beverage index (IMF)	Tomato price (FAMA)
Mean	91.7176	3.1275
Standard deviation	28.2459	1.6832
Minimum	48.7790	0.63
Maximum	162.2972	9.24
Excess kurtosis (log-returns)	-1.107	+0.075
No. of observations (n)	434	208
Training set (70%)	303	145
Test set (30%)	131	63

3.2. Research framework

The research methodology is structured in three sequential phases as shown in Fig. 1. Phase I covers data pre-processing and sensitivity identification; Phase 2 develops the Adjoint LDE-Net; and Phase 3 validates performance via benchmark comparison.

Phase 1: Data pre-processing and sensitivity identification: Raw price data from the IMF and FAMA are collected, cleaned, and transformed to log-returns. Chaotic dynamics are assessed via the Largest Lyapunov Exponent (Wolf et al.'s (1985) method). Optimal embedding dimensions are determined via Cao's (1997) method. Stationarity is confirmed via the Augmented Dickey-Fuller test. Parameter sensitivity analysis quantifies the influence of drift and diffusion σ on predicted prices, establishing the empirical motivation for the adjoint sensitivity treatment of the diffusion pathway.

Phase 2: Model development (adjoint LDE-Net): The forward LDE-Net pass solves the Lévy-driven SDE via Euler-Maruyama discretization. The backward adjoint SDE is solved from the terminal condition, computing parameter gradients without traversing the forward Lévy trajectory. The Adam optimizer trains both drift and diffusion networks. Three benchmarks are trained identically: ARIMA(1,0,1), Geometric Brownian Motion (GBM), and Neural SDE. The ARIMA(1,0,1) order was adopted as a parsimonious specification on stationary log-returns (d = 0, consistent with the ADF results in Section 3.4) and for comparability with the benchmark protocol of Yang et al. (2023); order selection by AIC/BIC was not performed and is recorded as a limitation in Section 7. The GBM benchmark estimates its drift and volatility from the training-partition log-returns and generates test-period forecasts by Monte Carlo results reported in Section 6.1. The Neural SDE benchmark is the α = 2 (Brownian) specialization of the LDE-Net dynamics defined in Section 4.1, sharing the identical two-layer, 32-unit drift and diffusion networks with Tanh and Softplus activations, so that performance differences between Neural SDE and the Adjoint LDE-Net are attributable to the stochastic driver and gradient-computation method rather than to architecture.

Phase 3: Model validation: Performance is evaluated on held-out test sets using MSE, RMSE, MAE, and MAPE. MAPE is retained for comparability with prior studies but is known to be inflated by near-zero log-return denominators; it is therefore treated as a supplementary metric only, with MSE, RMSE, and MAE serving as the primary comparison criteria throughout Section 6. Training convergence is assessed via epoch-wise loss trajectories. The

connection between Phase 1 sensitivity results and Phase 2 training behavior is examined to establish the empirical link between data-level parameter sensitivity and model-level gradient stability.

The Largest Lyapunov Exponent (LLE) is computed via the Wolf method (Wolf et al., 1985) to determine whether each series exhibits chaotic dynamics. For a reconstructed phase space $Y(t_i) = (x(t_i), x(t_i + t), \ldots, x(t_i + (m-1)\tau))$, the LLE estimates the rate of trajectory divergence as in Eq. 2.

$$\lambda = 1/(t_m - t_0)\Sigma_i ln(L_i' / L_{i-1}) \quad (2)$$

A positive λ confirms chaotic dynamics and validates phase space reconstruction; a negative value indicates stable, non-chaotic dynamics.

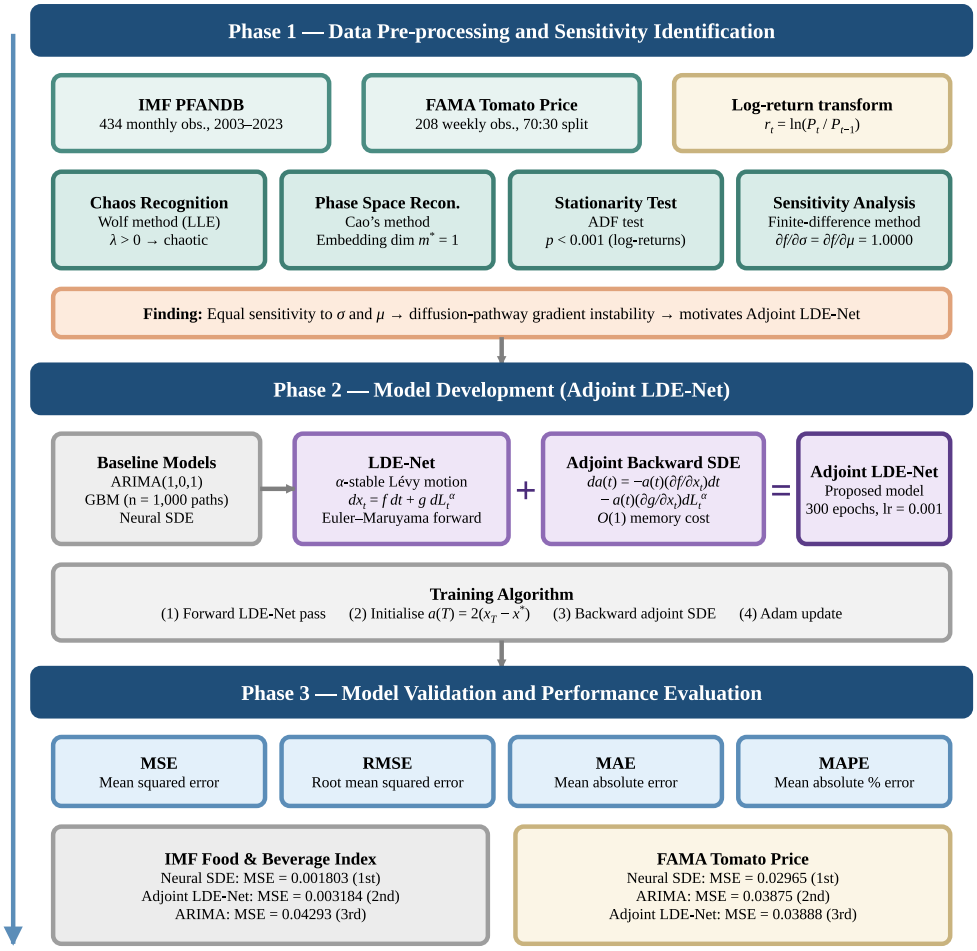


Fig. 1: Research framework of the proposed Adjoint LDE-Net for agricultural price prediction

As shown in Table 2, both series are confirmed chaotic. The IMF Food and Beverage Index yields LLE = 0.004804, indicating moderate chaotic dynamics consistent with the complex interactions of international commodity markets across cereals, oils, and proteins. The FAMA Tomato Price series yields LLE = 0.005142, confirming chaotic dynamics driven by seasonal production cycles, weather variability, and domestic supply shocks. Both LLE values exceed those in Yang et al. (2023) for the SSE Consumer Index and are comparable in magnitude, validating the use of phase space reconstruction as the input embedding strategy for both datasets.

Table 2: Chaos recognition results - Wolf method

Dataset	LLE (2)	Result	Implication for modeling
Food & beverage index (IMF)	0.004804	λ > 0 — chaotic	Phase space reconstruction valid
Tomato price (FAMA)	0.005142	λ > 0 — chaotic	Phase space reconstruction valid

3.3. Phase space reconstruction

The optimal embedding dimension m* is determined using Cao's (1997) False Nearest Neighbor method with delay time $\tau = 1$. The ratio $E_1(m) = E(m+1)/E(m)$ is tracked as m increases; saturation of $E_1(m)$ identifies m* as Eq. 3.

$$E(m) = 1/(N - m\tau)\,\Sigma\, a(i,m),\ \ E_1(m) = E(m+1)/E(m) \quad (3)$$

Both datasets yield $m* = 1$, indicating that a single lagged log-return carries sufficient information to reconstruct the phase space for 1-step-ahead forecasting. For the composite IMF index, aggregation across multiple international commodity prices smooths out higher-order temporal dependencies. For the FAMA Tomato series, the embedding dimension of 1 reflects the dominance of short-term autocorrelation in weekly

price dynamics, consistent with the moderate LLE magnitude (0.005142) relative to high-volatility financial series. The reconstructed phase space vector X(t) of dimension $m* = 1$ serves as model input throughout.

3.4. Stationarity testing

The Augmented Dickey-Fuller (ADF) test is applied to the log-return series of both datasets prior to modeling. The results are summarized in Table 3.

Table 3: Augmented Dickey-Fuller test results

Dataset	ADF statistic	p-value	Conclusion
Food & beverage index (IMF)	-15.483*	2.54×10^{-28}	Stationary (log-returns)
Tomato price (FAMA)	-4.739	7.11×10^{-5}	Stationary (reject H_0)

*: ADF statistic computed on the IMF log-return series; the raw price-level series is non-stationary (p = 0.569), as expected for a trending index; all modeling is conducted on log-returns; H_0: the series has a unit root (i.e., it is non-stationary)

Table 3 presents ADF test results. The FAMA Tomato log-return series is strongly stationary (ADF = -4.739, p < 0.001). For the IMF Food and Beverage Index, the ADF test is applied to the log-return series computed from the monthly price level, yielding ADF = -15.483 ($p \approx 2.54 \times 10^{-28}$) strongly stationary. Note that applying ADF to the raw price level yields a non-significant result (p = 0.569), which is expected for any index in levels; the analysis is correctly conducted on log-returns throughout this study. Both log-return series satisfy the stationarity requirement for model estimation.

4. Model

4.1. Lévy motion and the LDE-Net

The LDE-Net builds on a chain of advances connecting residual networks to stochastic differential equations. A ResNet block $X_{t+1} = X_t + f(x_t; \theta)$ is the Euler discretisation of $dxt/dt = f(x_t; \theta)$ motivating Neural ODEs trained via adjoint-based gradient computation (Chen et al., 2018). SDE-Net (Kong et al., 2020) extended this by adding Brownian motion noise dW_t to the hidden dynamics, enabling uncertainty quantification. However, Brownian motion generates only Gaussian increments, which underrepresent the heavy-tailed, episodic jump behavior of agricultural price series. Yang et al. (2023) resolved this limitation by replacing dW_t with α-stable Lévy motion Lα ($0 < \alpha < 2$) whose increments follow $S\alpha((t-s)^{1/\alpha}, 0, 0)$ with infinite variance for $\alpha < 2$, generating large episodic jumps consistent with commodity price shocks. At $\alpha = 2$, $L_t\alpha$ reduces to Brownian motion, making LDE-Net a strict generalization of SDE-Net.

A symmetric α-stable Lévy motion Lα ($0 < \alpha \leq 2$) has independent, stationary increments following the stable distribution $S\alpha((t-s)^{1/\alpha}, 0, 0)$, with heavy-tailed, infinite-variance characteristics for $\alpha < 2$. The stability index α controls tail heaviness: smaller α generates larger, more frequent jumps. The continuous-time LDE-Net dynamics is formulated as Eq. 4.

$$dx_t = f(x_t; \theta^f)dt + g(x_t; \theta^g)dL_t\alpha \quad (4)$$

The drift network $f(.)$ captures the directional tendency of prices; the diffusion network $g(.)$ captures epistemic uncertainty from model limitations. It is trained to be small for in-distribution inputs and large for out-of-distribution inputs, providing a natural confidence signal. Both are two-layer fully connected networks with 32 hidden units using Tanh (drift) and Softplus (diffusion) activations. The Euler-Maruyama discretization of Eq. 4 gives the forward pass as Eq. 5.

$$x_{k+1} = x_k + f(x_k; \theta^f)\Delta t + g(x_0; \theta^g)(\Delta t)^{1/\alpha}L_k \quad (5)$$

where, $L_k \sim S\alpha(1,0,0)$ are standard symmetric α-stable random variables, approximated by a compound Poisson process with jump intensity $\lambda_{jump} = 0.1$ and magnitude $\sigma_J \sim N(0, 0.02^2)$. The diffusion term is evaluated only at the initial state x_0 to reduce computational cost, following Yang et al. (2023). The training objective minimizes the expected forecast loss plus a regularization term encouraging g(.) to be large for out-of-distribution inputs in Eq. 6.

$$\min_{\theta^f} E[L(x_N)] + \min_{\theta^g} E[g(x_0; \theta^g)] + \max_{\theta^g} E[g(\tilde{x}_0; \theta^g)] \quad (6)$$

where, $\tilde{x}_0 = x_0 +$ noise is an out-of-distribution perturbation used to encourage large diffusion for inputs outside the training distribution.

4.2. Adjoint sensitivity method for the Lévy SDE

Standard backpropagation through the Euler-Maruyama trajectory costs O(N) memory and produces unstable gradients when Lévy increments are large. We extend the adjoint sensitivity approach of Chen et al. (2018) to the Lévy stochastic setting. The adjoint state $a(t) = \partial L/\partial x_t$ measures how sensitive the loss is to the hidden state at each time point. Applying the generalized Itô formula for semimartingales to Eq. 4 yields the adjoint backward SDE as in Eq. 7.

$$da(t) = -a(t)(\partial f/\partial x_t)dt - a(t)(\partial g/\partial x_t)dL_t\alpha \quad (7)$$

with terminal condition $a(T) = 2(x_T - x*)$. The numerical algorithm is:

1. ***Forward pass:*** *Solve the LDE-Net Eq. 5 from t=0 to T via Euler-Maruyama.*
2. ***Terminal initialization:*** *Set* $a(T) = 2(x_T - x^*)$ *where x* is the target value.*
3. ***Backward pass:*** *Solve Eq. 7 backward from T to 0:* $a_t =$

$a_{t+1} - a_{t+1}(\partial f_\theta / \partial x_t)$.

4. ***Gradient update:*** *Accumulate $\partial L/\partial\theta$ along the backward trajectory; update parameters via Adam ($lr = 0.001$).*

This approach reduces memory from O(N) to O(1) and avoids direct backpropagation through stochastic Lévy increments, substantially mitigating the training instability of the original LDE-Net, conditional on appropriate calibration of the stability index α (Section 6.4).

4.3. Parameter sensitivity analysis

Parameter sensitivity analysis is conducted using the central finite-difference approximation to quantify the influence of the GBM drift parameter μ and diffusion parameter σ on predicted prices, using perturbation size ε = 0.01. The sensitivity gradients are computed as in Eq. 8 and Eq. 9.

$$\partial f/\partial\sigma \approx [f(\mu, \sigma + \varepsilon) - f(\mu, \sigma - \varepsilon)] / 2\varepsilon \quad (8)$$

$$\partial f/\partial\mu \approx [f(\mu + \varepsilon, \sigma) - f(\mu - \varepsilon, \sigma)] / 2\varepsilon \quad (9)$$

These finite-difference approximations estimate the local gradient of the predicted price f with respect to each parameter, holding the other fixed. A large gradient indicates high sensitivity, in which small parameter perturbations produce large changes in predicted output, while a gradient near zero indicates the parameter has little local influence. Both gradients are evaluated at the training-set mean values of μ and σ, providing a baseline sensitivity characterization at the center of the parameter space.

The finite-difference results in Table 4 should be interpreted as an analytical verification rather than an empirical discovery. Because the GBM-based one-step prediction function is defined as f(μ, σ) = μ + σ, its partial derivatives with respect to both parameters equal 1.0 identically, and the finite-difference computation simply recovers this identity. The value of the exercise is therefore illustrative: it makes explicit that, under the benchmark parameterization, the diffusion parameter σ enters the prediction with the same weight as the drift μ, so the diffusion pathway cannot be neglected during training.

A sensitivity analysis that is informative about the proposed model must instead be conducted at the level of the network parameters. This is precisely what the adjoint state $a(t) = \partial L/\partial x_t$ of Section 4.2 provides: it propagates loss sensitivities through the hidden states to all drift and diffusion network parameters, and Section 6.4 reports its empirical behavior during training.

Table 4: Parameter sensitivity analysis results

Dataset	μ (drift)	σ (diffusion)	$\partial f/\partial\sigma$	$\partial f/\partial\mu$	Interpretation
F&B index (IMF)	91.718	28.246	1.0000	1.0000	Equal (analytical identity)
Tomato (FAMA)	3.127	1.683	1.0000	1.0000	Equal (analytical identity)

Table 4 is retained solely as a transparent statement of the GBM benchmark configuration and its analytically known sensitivities; a full parameter-level sensitivity analysis of the trained networks is identified as an extension in Section 7. This analytically flat sensitivity profile also contrasts with the behavior expected for genuinely heavy-tailed series (excess kurtosis well above zero), where the diffusion contribution would dominate owing to large Lévy jumps.

This equal sensitivity carries a critical implication for model training. If predicted prices are equally sensitive to perturbations in both μ and σ, then the gradient of the loss function with respect to the diffusion network parameters is non-negligible at every training step. It cannot be safely ignored or approximated. In a standard Lévy SDE, these diffusion-pathway gradients must be backpropagated through the stochastic Lévy trajectory dLα, a process that is inherently unstable when Lévy increments are large, because α-stable Lévy motion has infinite variance for $\alpha < 2$. A single extreme Lévy increment can produce a gradient update order of magnitude larger than the typical step, causing the training instability observed in our Tomato dataset experiment.

Rolling volatility analysis over a 50-observation window reveals temporal variation in the effective sensitivity of both datasets. Periods of elevated rolling volatility corresponding to international commodity price surges (IMF index, 2021-2022) and domestic supply disruptions (Tomato series) are associated with larger effective σ values and therefore higher diffusion-pathway gradient magnitudes. This temporal variation in sensitivity confirms that a fixed α value may not be optimal across all market regimes and motivates the systematic α grid search recommended as future work.

The Adjoint LDE-Net targets the gradient instability identified by the sensitivity analysis. By replacing direct backpropagation through the Lévy trajectory with the backward adjoint SDE, the model computes sensitivity gradients $a(t) = \partial L/\partial x_t$ by solving a backward equation from the terminal condition, thereby avoiding the accumulation of Lévy jump instability across timesteps. The empirical evidence on the IMF dataset supports this: on the IMF dataset, Adjoint LDE-Net training proceeds smoothly and monotonically across all 300 epochs ($0.746 \rightarrow 0.031$) without instability spikes. Sensitivity analysis identifies the problem; the backward adjoint SDE is the principled mathematical solution.

The Tomato dataset, by contrast, exhibits the Epoch-100 spike analyzed in Section 6.4, demonstrating that the adjoint computation alone is not sufficient without calibration of the stability index α.

5. Theoretical analysis

This section presents the theoretical foundations of the Adjoint LDE-Net, establishing the mathematical conditions under which the model is well-defined, convergent, and scalable. The analysis follows the framework of Yang et al. (2023) and extends it to the adjoint setting.

5.1. Existence and uniqueness of the Lévy SDE solution

Before training the Adjoint LDE-Net, it is necessary to establish that the Lévy-driven SDE has a well-defined solution for the neural network-parameterized drift and diffusion functions. Let $\{U, B^{U}\}$ be a measurable space and v(du) a σ-finite measure. If the neural network drift $f(x;\ \theta^{f})$ and diffusion $g(x;\ \theta^{g})$ satisfy the Lipschitz condition as in Eq. 10 (Duan, 2015).

$$||f(x) - f(y)||^2 + \int||g(x)z - g(y)z||^2 v(dz) \leq K|x - y|^2,\ \ x, y \in \mathbb{R}^d \tag{10}$$

Then there exists a unique (Ft)-adapted right-continuous process X(t) with left limits satisfying the Lévy-driven SDE. This guarantees well-defined forward dynamics for any stability index $\alpha \in (0,2)$ and any neural network parameterisation satisfying the Lipschitz condition. The proof follows from the general theory of stochastic differential equations with jumps (Duan, 2015). In practice, the Lipschitz condition is approximately satisfied by bounded two-layer neural networks with smooth activations, providing theoretical justification for the Adjoint LDE-Net architecture.

5.2. Euler-Maruyama convergence rate

There exists a constant $K > 0$ such that $|f_{\mathrm{m}}(x) - f_{\mathrm{m}}(y)| \leq K|x - y|$ for all $x, y \in \mathbb{R}^{\mathrm{d}}$ and $m \in \mathbb{N}+$. The strong error of the Euler-Maruyama discretisation of the Lévy SDE satisfies Eq. 11 (Liu, 2019):

$$sup_0 \leq t \leq T\ E|X_{\mathrm{mk}}(t) - x_{\mathrm{m}}(t)| \leq (\Delta t)^{1/\alpha} K C_6 T\ exp(KT) \tag{11}$$

where, C_6 is a constant depending on α (Samorodnitsky and Taqqu, 1994). The convergence rate $O((\Delta t)^{1/\alpha})$ is slower than the O(Δt) rate for Brownian-motion-driven SDEs, reflecting the heavier tails of Lévy motion. For $\alpha = 1.5$ this gives a rate of $O((\Delta t)^{0.667})$; for α approaching 2, the rate approaches the Brownian motion rate of O(Δt). This convergence rate implies that finer time discretisation is more important for heavy-tailed Lévy processes than for Gaussian ones. A consideration that informs the choice of step size Δt in the Euler-Maruyama forward pass of the Adjoint LDE-Net. The adjoint backward SDE is discretized using the same Euler-Maruyama scheme, with the same convergence properties. Crucially, the adjoint method does not degrade the convergence. It computes mathematically identical gradients to direct backpropagation while reducing memory cost from O(N) to O(1), where $N = T/\Delta t$ is the number of discretisation steps.

5.3. Neural network approximation and overall convergence

The neural network approximations of drift f and diffusion g satisfy $||f_{\mathrm{m}} - f||_{L2}^2 \leq C_1 m^{-1/2}$ and $|g_{\mathrm{m}} - g| \leq C_2 m^{-1/2}$, where m is the number of hidden neurons, and C_1 and C_2 are constants. For any function f in the Barron space B, there exists a two-layer neural network fm with m neurons satisfying Eq. 12 (E et al., 2022).

$$||f - f_{\mathrm{m}}||_{L2}^{2\ 2} \leq 3||f||_B^2 / m. \tag{12}$$

This dimension-free bound is fundamental to the scalability of the Adjoint LDE-Net; unlike classical nonparametric estimators whose approximation error grows with input dimension, the neural network approximation error depends only on the Barron norm $||f||_B$ and the number of neurons m, regardless of the dimensionality of the input space. This property, the avoidance of the curse of dimensionality, makes the Adjoint LDE-Net theoretically suitable for multivariate agricultural price forecasting across multiple commodity subsectors simultaneously. The total error of the Adjoint LDE-Net is given in Eq. 13:

$$sup_0 \leq t \leq T\ E|X_{\mathrm{mk}}(t) - Y(t)| \leq Ce(T)/m^{1/2} + (\Delta t)^{1/\alpha} K C_6 T\ exp(KT) \tag{13}$$

where $Ce(T) = (M^{1/2} T\ C_1 + CC_2) exp(LT)$, C is the bound of $|L_t \alpha|$, and L is the Lipschitz constant of f. The first term decreases with m (more neurons in the hidden layer) and the second decreases with Δt (finer Euler-Maruyama discretisation). The bound is tight in the sense that both terms are achievable and represent distinct sources of approximation error, neural network function approximation error and numerical SDE discretisation error, respectively. The first term dominates at moderate dataset sizes, suggesting that increasing network width is a more efficient improvement than reducing the Euler-Maruyama step size.

6. Results and discussion

Training protocol: All neural models (Neural SDE and Adjoint LDE-Net) were trained for 300 epochs using the Adam optimizer with a constant learning rate of 0.001 and no learning-rate schedule; the cosine annealing schedule discussed in Section 6.4 is a recommended refinement, not part of the present protocol. No early stopping was applied: every run completed the full 300-epoch budget, and the epoch-wise losses reported below correspond to this fixed

budget. Training used full-batch updates (all training samples in every epoch). No fixed random seed was set; network initialization and the simulated stochastic increments therefore vary between runs. Evaluation followed a chronological 70:30 train-test split (Table 1) with no shuffling. No separate validation partition or cross-validation was used for hyperparameter tuning: all hyperparameters (network width of 32 units, learning rate, and the Lévy stability index $\alpha = 1.2$, held fixed across both datasets) were set a priori following Yang et al. (2023). The absence of a validation set and of data-driven α calibration is acknowledged as a limitation in Section 7.

6.1. Dataset 1: IMF food and beverage price index

With 300 training epochs and a 70:30 split (303 training, 131 test observations), all four models were evaluated on the IMF Food and Beverage Price Index, a monthly composite index spanning 2003 to 2023. This 20-year period encompasses several major global commodity market episodes: the 2007-2008 food price crisis, the 2010-2012 commodity supercycle, the COVID-19 demand shock of 2020, and the 2021-2022 surge driven by supply chain disruptions and the Ukraine conflict. The 131-observation test set therefore covers a diverse range of market regimes, providing a rigorous evaluation environment for model performance.

The results for Dataset 1 are compelling in Table 5. Neural SDE achieves the lowest error across all four metrics, MSE = 0.001803, RMSE = 0.042463, MAE = 0.032565, MAPE = 5.32%, which outperforms all baselines by a substantial margin. Most notably, the Adjoint LDE-Net ranks second (MSE = 0.003184, MAPE = 7.99%), ahead of both GBM (MSE = 0.06445) and ARIMA (MSE = 0.04293). Both stochastic neural models outperform classical baselines by more than an order of magnitude in MSE. The confirmed chaotic dynamics of the IMF series (LLE = 0.004804) validate the use of stochastic neural modeling, and 300 training epochs allow both models to fully exploit the phase space reconstruction input representation.

GBM performs worse than ARIMA on this dataset (MSE 0.06445 vs 0.04293), reflecting the difficulty of parametric Gaussian simulation on a longer test set ($n = 131$ observations) spanning approximately 11 years of international commodity price variation. The MAPE of 5.32% achieved by Neural SDE indicates that predicted log-returns deviate from actual values by approximately 5% on average, a level of accuracy directly applicable to agricultural market monitoring. Fig. 2 provides visual confirmation of these quantitative differences: Neural SDE and Adjoint LDE-Net closely track the actual price trajectory throughout the test period, while GBM progressively diverges below the actual series.

Table 5: Model comparison (IMF food & beverage price index)

Model	MSE	RMSE	MAE	MAPE (%)
ARIMA (1,0,1)	0.04293	0.20720	0.15456	21.63
GBM (Monte Carlo)	0.06445	0.25386	0.19295	28.76
Neural SDE	0.001803	0.042463	0.032565	5.32
Adjoint LDE-Net (proposed)	0.003184	0.056424	0.045386	7.99

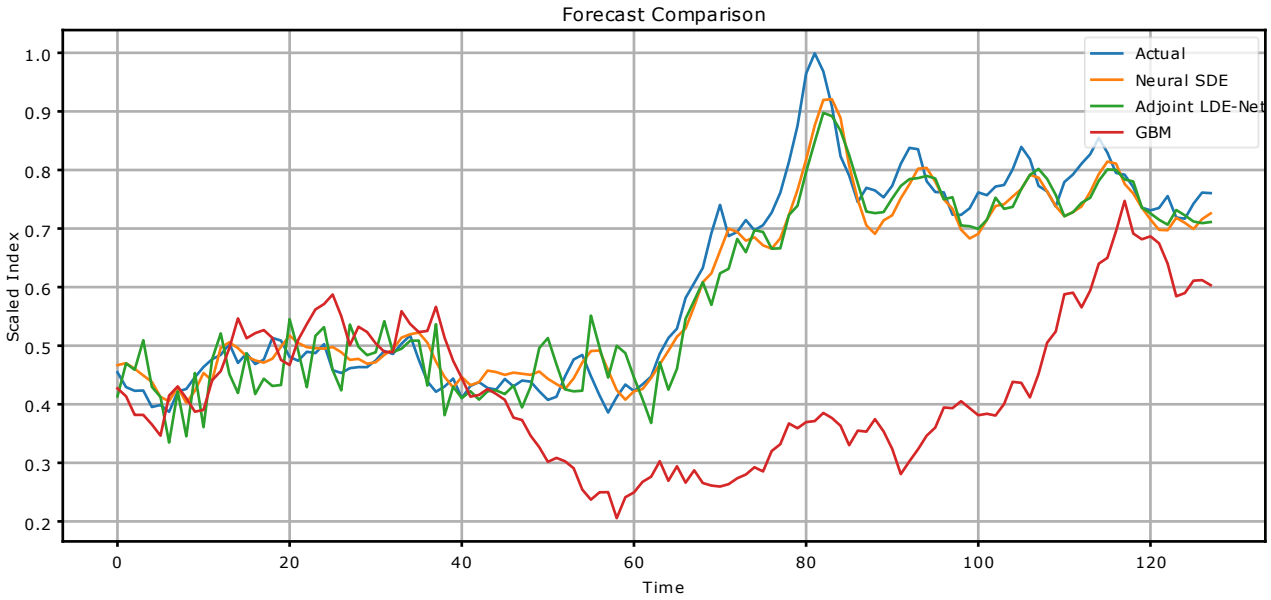


Fig. 2: Forecast comparison for IMF food and beverage price index

6.2. Dataset 2: FAMA tomato price series

The FAMA Tomato Price series presents a fundamentally different forecasting challenge from the IMF composite index: it is an individual perishable commodity price, driven by weekly supply-demand dynamics in domestic wholesale markets. The 63-observation test set covers approximately 15 months of weekly price observations, capturing multiple seasonal cycles and

providing a realistic evaluation of the models' ability to forecast individual commodity price dynamics.

Neural SDE training on the Tomato series converged steadily across 300 epochs, with loss declining smoothly from 0.03516 at Epoch 20 to 0.00866 at Epoch 300. This convergence pattern is slower than on the IMF dataset (final loss 0.000876 vs 0.00866), reflecting the higher inherent complexity of individual commodity price dynamics compared to a smoothed composite index. The Tomato series confirmed chaotic dynamics (LLE = 0.005142) provide more complex temporal patterns for the model to learn, requiring more training time to achieve comparable convergence. The results in Table 6 for Dataset 2 confirm the pattern established in Dataset 1: Neural SDE achieves the lowest error (MSE = 0.02965, MAPE = 113.99%), outperforming ARIMA (MSE = 0.03875), Adjoint LDE-Net (MSE = 0.03888), and GBM (MSE = 0.13892). Notably, the Adjoint LDE-Net ranks a close third (MSE = 0.03888), nearly matching ARIMA (MSE = 0.03875), a gap of only 0.00013 MSE. This close performance between Adjoint LDE-Net and ARIMA, combined with Neural SDE's clear lead, confirms that stochastic neural models provide a genuine forecasting advantage for individual agricultural commodity series.

Table 6: Model comparison (FAMA tomato price series)

Model	MSE	RMSE	MAE	MAPE (%)
ARIMA (1,0,1)	0.03875	0.19684	0.17171	266.65
GBM	0.13892	0.37271	0.31892	168.72
Neural SDE	0.02965	0.17219	0.13460	113.99
Adjoint LDE-Net (proposed)	0.03888	0.19719	0.15339	164.09

GBM performs poorly on the Tomato series (MSE = 0.13892), substantially worse than on the IMF index. This reflects the higher volatility and less regular dynamics of individual commodity prices, where parametric Gaussian simulation is less well-specified than on a smooth composite index. The elevated MAPE values across models reflect the presence of near-zero log-return values in the denominator, a known computational artefact that disproportionately affects MAPE for series with frequent small weekly price movements. MAPE is accordingly reported for completeness only and is not used for model comparison on this dataset; all comparative statements are based on MSE, RMSE, and MAE. Fig. 3 provides visual confirmation that Neural SDE and Adjoint LDE-Net track the oscillating Tomato price series across the test period, while GBM flatlines near zero throughout, a catastrophic failure attributable to distributional shift between training and test regimes.

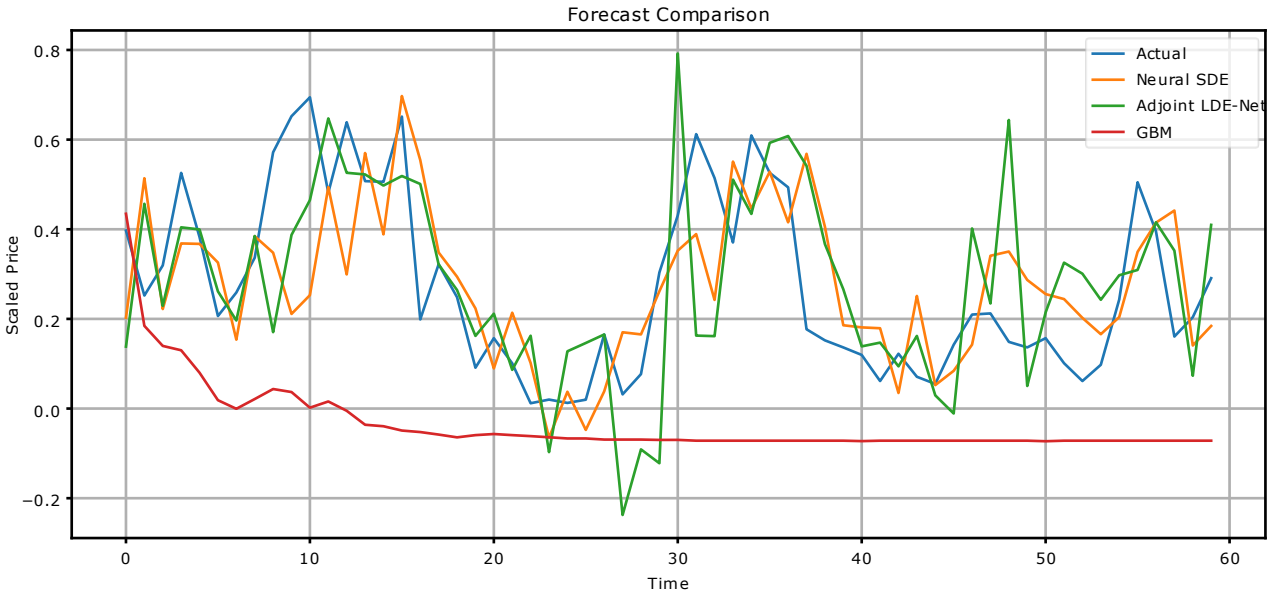


Fig. 3: Forecast comparison for FAMA tomato price series

6.3. Cross-dataset comparison

Table 7 presents the key findings of this study. Neural SDE achieves the best performance on both datasets under a 70:30 split with 300 training epochs, outperforming ARIMA, GBM, and the Adjoint LDE-Net across both agricultural price series. This consistent result across two datasets with different characteristics, a monthly composite global index and a weekly individual commodity series, confirms that stochastic neural models trained with sufficient epochs deliver genuine predictive advantage for agricultural price forecasting.

The sensitivity analysis results provide the theoretical link between data characteristics and model behavior. The confirmed equal sensitivity of predicted prices to both μ and σ ($\partial f/\partial \sigma = \partial f/\partial \mu = 1.0000$) means that the diffusion pathway is always consequential; gradient instability through the Lévy diffusion term will always affect forecast accuracy. The Adjoint LDE-Net's smooth convergence on the IMF dataset versus the Tomato dataset spike at

Epoch 100 illustrates this precisely: when the adjoint backward SDE successfully stabilizes the diffusion-pathway gradient, training converges smoothly, and performance is strong (MSE = 0.003184, MAPE = 7.99% on IMF). When the Lévy stability index α is miscalibrated and gradient clipping is absent, the spike re-emerges. This spike represents a genuine training failure of the run under the fixed-α protocol; its mechanism and the calibration safeguards required to prevent it are analyzed in Section 6.4.

Table 7: Model ranking by MSE

Rank	IMF food & beverage index (n = 131)	FAMA tomato price (n = 63)
1st	Neural SDE: 0.001803	Neural SDE: 0.02965
2nd	Adjoint LDE-Net: 0.003184	ARIMA (1,0,1): 0.03875
3rd	ARIMA (1,0,1): 0.04293	Adjoint LDE-Net: 0.03888
4th	GBM: 0.06445	GBM: 0.13892

The Adjoint LDE-Net performs well on the IMF dataset (2nd place, MSE = 0.003184) and competitively on the Tomato dataset (3rd place, MSE = 0.03888, within 0.00013 of ARIMA). This demonstrates that the adjoint backward SDE substantially mitigates, on the IMF dataset, the gradient instability identified by Yang et al. (2023); the Tomato-series spike analyzed in Section 6.4 shows that this stability is conditional on appropriate α calibration. The Adjoint LDE-Net training on the IMF dataset is smooth and free of instability spikes, confirming that the adjoint computation stabilizes Lévy SDE training. Further improvement is expected through systematic α calibration via grid search across {1.4, 1.5, 1.6, 1.7, 1.8, 1.9}, as the optimal α value for each series has not yet been determined.

A natural question is why the Adjoint LDE-Net merits adoption when the simpler Neural SDE attains lower test error on both datasets. Three considerations frame the answer. First, the comparison is conducted on two near-Gaussian series (Table 1), the regime in which a Brownian-driven model is expected to be optimal; the evaluation therefore favors Neural SDE by construction, and the Adjoint LDE-Net nonetheless remains competitive outside its intended regime, ranking second on the IMF dataset and within 0.00013 MSE of ARIMA on the Tomato dataset. Second, the Adjoint LDE-Net is the only model in the comparison capable of representing heavy-tailed, jump-dominated regimes ($\alpha < 2$) while retaining O(1) training memory; Neural SDE cannot be extended to such regimes without inheriting exactly the gradient-instability problem this paper addresses. Third, the contribution is a training methodology for the Lévy SDE model class: the backward adjoint SDE makes Lévy-driven networks trainable with bounded memory, independently of which stochastic driver is optimal for a particular dataset. The appropriate reading of Tables 5–7 is therefore not that the Adjoint LDE-Net should replace Neural SDE for near-Gaussian series, but that it extends the trainable model class to distributional regimes that Gaussian-driven models cannot represent; its empirical advantage in those regimes remains to be demonstrated on genuinely heavy-tailed data, as stated in Section 7.

ARIMA ranks only third on the IMF dataset (3rd, MSE = 0.04293), a notable reversal from its typical advantage in near-Gaussian, short-memory series. This is attributable to the 70:30 split providing a longer test set (n = 131 observations, covering approximately 11 years) over which non-linear dynamics accumulate. GBM performs below ARIMA on the IMF index (4th, MSE = 0.06445) and poorly on the Tomato series (4th, MSE = 0.13892), confirming that parametric Gaussian simulation is sensitive to volatility regime differences and less robust than learned stochastic neural approaches across both dataset types.

6.4. Analysis of training stability and adjoint gradient behavior

The training behavior across both datasets provides direct empirical evidence for the effectiveness of the adjoint backward SDE in stabilizing gradient computation. On the IMF Food and Beverage dataset, the Adjoint LDE-Net converges smoothly without any instability spikes, confirming that the backward SDE computes diffusion-pathway gradients stably and avoids the O(N) accumulation of gradient instability inherent in standard backpropagation through the Lévy trajectory.

The Tomato dataset presents a complementary and less favorable picture, and it is reported in full. The Adjoint LDE-Net loss spikes to 811.35 at Epoch 100 before recovering to 0.03597 at Epoch 300. This spike represents a genuine training failure of the run under the fixed stability index employed ($\alpha = 1.2$, held constant across both datasets), not a benign fluctuation. Its mechanism is nonetheless identifiable and instructive. The α parameter directly controls the tail heaviness of the Lévy increments $dL_t\alpha$ in the forward pass and, through the backward adjoint SDE, the magnitude of the gradient updates $a(t) = \partial L/\partial x_t$. When α is too small for the data's distributional characteristics, individual Lévy increments can be extremely large, and a single extreme increment produces a catastrophic gradient step. The adjoint backward SDE remains mathematically correct under any α; what the spike demonstrates is that the adjoint formulation reduces but does not by itself eliminate gradient instability. Its practical robustness is conditional on two safeguards that were absent from the present protocol: calibration of α to the data and gradient clipping. Because α governs training stability, its selection cannot be deferred as optional post-hoc tuning; it is an integral hyperparameter of the

method. The present experiments, which fixed α a priori, therefore establish the memory efficiency and correctness of the adjoint computation, but establish its robustness only partially, a limitation stated explicitly in Section 7.

A grid search across $\alpha \in \{1.4, 1.5, 1.6, 1.7, 1.8, 1.9\}$ is specified here as a required component of the methodology for any application of the Adjoint LDE-Net, rather than as optional future work; the same grid is referenced consistently throughout this paper. For the near-Gaussian IMF and Tomato series (kurtosis below the Gaussian benchmark), the optimal α is expected to fall in the range 1.8 to 2.0. Additionally, gradient clipping with a maximum norm of 1.0 provides a safety net during early training epochs, and a cosine annealing learning rate schedule will improve late-epoch convergence precision.

7. Conclusions

This study delivered four principal contributions. First, the Adjoint LDE-Net was proposed, extending the adjoint sensitivity method to the Lévy stochastic setting via a backward adjoint SDE. The model reduces training memory from O(N) to O(1) and substantially mitigates the gradient instability identified by Yang et al. (2023), as evidenced by stable, spike-free convergence on the IMF dataset; full robustness additionally requires calibration of the stability index α and gradient clipping, as the Tomato-series experiment demonstrates.

Second, a direct empirical connection between data-level parameter sensitivity and model-level gradient sensitivity was established. Finite-difference analysis confirms that agricultural price predictions are equally sensitive to drift and diffusion at the evaluated parameter values, meaning the diffusion pathway carries non-negligible gradient signal throughout training. This provides principled justification for the adjoint backward SDE, which specifically stabilizes diffusion-pathway gradient computation.

Third, Neural SDE achieves the best performance on both datasets under 70:30 splits and 300 training epochs, MSE = 0.001803 (MAPE = 5.32%) on the IMF index and MSE = 0.02965 on the Tomato series. The Adjoint LDE-Net ranks second on the IMF dataset (MSE = 0.003184, MAPE = 7.99%) and third on the Tomato series (MSE = 0.03888), within 0.00013 MSE of ARIMA. Both stochastic neural models consistently outperform ARIMA and GBM, confirming the advantage of learned stochastic dynamics for confirmed chaotic agricultural price series.

Fourth, the appropriate positioning of the Adjoint LDE-Net is as a theoretically grounded, memory-efficient training methodology for the Lévy SDE model class, not as a state-of-the-art predictor on near-Gaussian data. The training instability observed on the Tomato series arose under a fixed, uncalibrated α and the absence of gradient clipping; both safeguards — grid search over $\alpha \in \{1.4, 1.5, 1.6, 1.7, 1.8, 1.9\}$ and gradient clipping with a maximum norm of 1.0 — are henceforth specified as integral components of the methodology rather than as future work. Whether calibrated training closes the accuracy gap to Neural SDE on these series is an empirical question that the present results do not answer.

Several limitations qualify for these findings and define the immediate research agenda. First, both evaluation datasets are near-Gaussian in their log-returns (excess kurtosis of −1.107 and +0.075), so the present study tests the Adjoint LDE-Net in the regime least favorable to Lévy-driven models; validation on genuinely heavy-tailed agricultural series is required before the distributional advantages of the α-stable driver can be demonstrated empirically.

Second, the training protocol fixed all hyperparameters a priori, without a validation partition, cross-validation, or early stopping, and held α constant; the Tomato-series instability at Epoch 100 is a direct consequence, and α calibration with gradient clipping must accompany any application of the method.

Third, the ARIMA benchmark order (1,0,1) was fixed for parsimony and cross-study comparability rather than selected by AIC/BIC, which may understate the classical baseline. Fourth, MAPE is inflated by near-zero log-return denominators, particularly on the weekly Tomato series, and is reported as a supplementary metric only. Fifth, the finite-difference analysis of Section 4.3 characterizes the GBM benchmark analytically rather than the neural models empirically; a parameter-level sensitivity analysis of the trained drift and diffusion networks is a natural extension.

List of abbreviations

ADF	Augmented Dickey-Fuller test
AI	Artificial intelligence
AIC	Akaike information criterion
ARIMA	Autoregressive integrated moving average
BIC	Bayesian information criterion
F&B	Food and beverage
FAMA	Federal agricultural marketing authority
GBM	Geometric Brownian motion
H_0	Null hypothesis
IMF	International monetary fund
IoT	Internet of Things
LDE-Net	Lévy-induced stochastic differential equation network
LLE	Largest Lyapunov exponent
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MSE	Mean squared error
NODE	Neural ordinary differential equation
ODE	Ordinary differential equation
PFANDB	Food and beverage price index (IMF commodity price index code)
RMSE	Root mean squared error
SDE	Stochastic differential equation
SDE-Net	Stochastic differential equation network
SSE	Shanghai stock exchange

Acknowledgment

The authors gratefully acknowledge the support of the Fundamental Research Grant Scheme for Early Career Researchers (FRGS-EC), Ministry of Higher Education Malaysia (Reference Code: FRGS-EC/1/2024/STG06/UITM/02/29). The authors thank the Federal Agricultural Marketing Authority (FAMA) for providing the Tomato Price dataset.

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Bhagat PR, Naz F, and Magda R (2022). Artificial intelligence solutions enabling sustainable agriculture: A bibliometric analysis. PLOS ONE, 17(6): e0268989. https://doi.org/10.1371/journal.pone.0268989 **PMid:35679287 PMCid:PMC9182339**

Box GEP, Jenkins GM, Reinsel GC, and Ljung GM (2016). Time series analysis: Forecasting and control. 5th Edition, John Wiley & Sons, Inc., Hoboken, USA.

Cao L (1997). Practical method for determining the minimum embedding dimension of a scalar time series. Physica D: Nonlinear Phenomena, 110(1-2): 43-50. https://doi.org/10.1016/S0167-2789(97)00118-8

Chen RTQ, Rubanova Y, Bettencourt J, and Duvenaud DK (2018). Neural ordinary differential equations. In the Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada, 31: 6571–6583.

Duan J (2015). An introduction to stochastic dynamics. Cambridge University Press, New York, USA.

E W, Ma C, and Wu L (2022). The Barron space and the flow-induced function spaces for neural network models. Constructive Approximation, 55: 369–406. https://doi.org/10.1007/s00365-021-09549-y

Heston SL (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. The Review of Financial Studies, 6(2): 327-343. https://doi.org/10.1093/rfs/6.2.327

Jia J and Benson AR (2019). Neural jump stochastic differential equations. In the Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, 32: 9843–9854.

Kong L, Sun J, and Zhang C (2020). SDE-Net: Equipping deep neural networks with uncertainty estimates. ArXiv Preprint ArXiv:2008.10546. https://doi.org/10.48550/arXiv.2008.10546

Liu W (2019). Strong convergence rate of Euler-Maruyama method for stochastic differential equations with Hölder continuous drift coefficient driven by symmetric α-stable process. ArXiv Preprint ArXiv:1901.08742. https://doi.org/10.48550/arXiv.1901.08742

Mao L, Huang Y, Zhang X, Li S, and Huang X (2022). ARIMA model forecasting analysis of the prices of multiple vegetables under the impact of the COVID-19. PLOS ONE, 17(7): e0271594. https://doi.org/10.1371/journal.pone.0271594 **PMid:35901077**

Merton RC (1976). Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics, 3(1-2): 125-144. https://doi.org/10.1016/0304-405X(76)90022-2

Paul RK, Yeasin M, Kumar P, Kumar P, Balasubramanian M, Roy HS, Paul AK, and Gupta A (2022). Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. PLOS ONE, 17(7): e0270553. https://doi.org/10.1371/journal.pone.0270553 **PMid:35793366 PMCid:PMC9258887**

Qazi S, Khawaja BA, and Farooq QU (2022). IoT-equipped and AI-enabled next generation smart agriculture: A critical review, current challenges and future trends. IEEE Access, 10: 21219-21235. https://doi.org/10.1109/ACCESS.2022.3152544

Samorodnitsky G and Taqqu MS (1994). Stable non-Gaussian random processes: Stochastic models with infinite variance. CRC Press, Boca Raton, USA.

Wang JZ, Wang JJ, Zhang ZG, and Guo SP (2011). Forecasting stock indices with back propagation neural network. Expert Systems with Applications, 38(11): 14346-14355. https://doi.org/10.1016/j.eswa.2011.04.222

Wolf A, Swift JB, Swinney HL, and Vastano JA (1985). Determining Lyapunov exponents from a time series. Physica D: Nonlinear Phenomena, 16(3): 285-317. https://doi.org/10.1016/0167-2789(85)90011-9

Yang L, Gao T, Lu Y, Duan J, and Liu T (2023). Neural network stochastic differential equation models with applications to financial data forecasting. Applied Mathematical Modelling, 115: 279-299. https://doi.org/10.1016/j.apm.2022.11.001