Networks and Heterogeneous Media

https://www.aimspress.com/journal/nhm

NHM, 20(5): 1411–1436.
DOI: 10.3934/nhm.2025061
Received: 05 September 2025
Revised: 19 November 2025
Accepted: 04 December 2025
Published: 16 December 2025

Research article

Dual-Uncertainty modeling in financial time-series via VMD-LSTM with concrete dropout and VMD-WGAN

Jeonggyu Huh[1], **Dajin Kim**[1], **Minseok Jung**[1] **and Seungwon Jeong**[2,*]

[1] Department of Mathematics, Sungkyunkwan University, Suwon 16419, Republic of Korea

[2] Global-Learning & Academic research institution for Master's·PhD students, and Postdocs, Chonnam National University, Gwangju 61186, Republic of Korea

* **Correspondence:** Email: junior1492@jnu.ac.kr.

Abstract: Financial time-series forecasting requires not only accurate point predictions but also a principled characterization of the uncertainty around future outcomes. We proposed a unified dual-path framework that modeled both forms using a shared variational mode decomposition (VMD) backbone. VMD stabilized nonstationary signals, enabling the predictive path—a long short-term memory (LSTM) network integrated with concrete dropout—to quantify epistemic uncertainty, while the generative path used a conditional Wasserstein generative adversarial network (WGAN) to capture aleatoric risk. Empirical evaluations on the Standard & Poor's 500 (S&P 500) index and Financial Times Stock Exchange 100 (FTSE 100) index demonstrated superior predictive accuracy and distributional fidelity over strong baselines. Comprehensive ablation studies and regime-conditioned analyses revealed a clear frequency-wise division of labor: Low-frequency modes drove predictive accuracy, while the generative path successfully reproduced heavy-tailed, regime-dependent return distributions. These findings underscored the efficacy of decomposed uncertainty modeling for robust financial risk assessment.

Keywords: stock prediction; variational mode decomposition; deep learning; generative adversarial networks; uncertainty modeling

1. Introduction

A central goal in quantitative finance is to move beyond point forecasting and provide a principled account of the uncertainty surrounding future outcomes. Ignoring such uncertainty leads to flawed inference and exposes investors to substantial risks [1]. Uncertainty is not monolithic: It arises from model (epistemic) sources—limited data and model misspecification—and from market (aleatoric) sources—the intrinsic randomness of financial markets. A robust framework should therefore quantify, and importantly separate, these two components.

Prior research typically tackles only one side of this challenge. Bayesian-inspired methods such as Monte Carlo Dropout [2, 3] quantify epistemic uncertainty in neural predictors, while generative approaches (e.g., generative adversarial networks proposed by Goodfellow et al. [4]) learn the aleatoric distribution of returns [5]. When combined within a single predictor, the two sources often remain entangled, obscuring whether predictive risk stems from the model's limitations or from the market's inherent volatility. This ambiguity limits practical risk management and forms the research gap we aim to address.

A key ingredient in our approach is a shared signal decomposition that stabilizes noisy financial series in a way that is useful for both prediction and generation. We employ variational mode decomposition (VMD) [6] to obtain a set of frequency-localized intrinsic mode functions (IMFs) from each raw series. This decomposition mitigates nonstationarity and reduces cross-frequency interference, so that downstream models operate on simpler, band-limited components instead of a single broadband process. On the predictive side, we apply VMD to price levels, which facilitates straightforward recomposition into price trajectories. On the generative side, we apply VMD to log-returns, whose approximately stationary behavior is better suited to distributional modeling. Preliminary experiments confirmed that this shared decomposition significantly improves training stability and fidelity for both tasks.

Building on this common representation, we propose a unified dual-path framework that explicitly separates and models both types of uncertainty. The predictive path is implemented by VMD–LSTM–CD, which uses a long short-term memory network [7] for each IMF and represents model uncertainty as the dispersion across stochastic forward passes with concrete dropout (CD) [8], yielding one-step-ahead forecasts with epistemic intervals. In parallel, the generative path is implemented by VMD–Wasserstein generative adversarial network (WGAN), a conditional Wasserstein GAN [9, 10] trained on decomposed log-return components; it learns the conditional distribution of future returns and represents market uncertainty as the spread of samples drawn from the generator. The shared IMF decomposition ensures that both paths learn from the same stabilized structure while specializing in their respective objectives.

To validate our approach, we conduct comprehensive experiments on the Standard & Poor's 500 (S&P 500) and Financial Times Stock Exchange 100 (FTSE 100) indices. On the predictive side, VMD–LSTM–CD consistently outperforms strong baselines—a standard LSTM and a VMD-augmented LSTM—in terms of root mean square error (RMSE), and produces posterior predictive intervals that evolve sensibly over time. Regime-conditioned evaluations under high-return, low-return, and high-volatility days show that gains in log-return accuracy persist across upside, downside, and turbulent markets. An IMF-wise leave-one-out analysis further reveals a clear frequency-wise division of labor: Low- and medium-frequency modes drive both point accuracy and calibration, while the high-frequency modes contribute only limited additional forecastable content.

On the generative side, we benchmark VMD–WGAN against a standard WGAN trained directly on log-returns. In-sample histograms and unconditional moments demonstrate that VMD–WGAN more faithfully reproduces the heavy-tailed, asymmetric return distributions observed in the data, whereas the baseline especially underestimates kurtosis, in particular, for the FTSE index. Regime-wise Johnson–SU (the unbounded Johnson family) log-likelihoods, computed for high-return, low-return, and high-volatility states both in-sample and out-of-sample, show consistent improvements of VMD–WGAN over the baseline, with particularly pronounced gains in volatile and downside regimes. Monthly out-of-

sample likelihoods indicate that the heavy-tailed distributions learned from 2013–2023 remain stable and robust throughout the 2024 test period. An IMF-wise leave-one-out study on the generative path complements these findings by clarifying which frequency bands control the center and tails of the return distribution and which mainly behave as residual high-frequency components.

Taken together, these results demonstrate that our framework provides a comprehensive and practically relevant solution for risk-aware financial analysis: the predictive path offers calibrated, uncertainty-aware forecasts, while the generative path supplies realistic return scenarios with heavy tails and regime sensitivity. The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 details the methodology and design rationale. Section 4 describes the experimental setup and empirical results, including the mode-wise comparative analysis. Section 5 concludes the study and outlines directions for future work.

2. Related work

Our research builds upon four principal streams of literature in financial forecasting: Deep learning for time-series modeling, decomposition-based hybrid forecasting, uncertainty quantification in deep learning, and generative modeling for distributional analysis.

To capture the nonlinear and complex dependencies present in financial markets, recent research has increasingly turned to deep learning models, particularly LSTM networks [7]. These architectures, equipped with gating mechanisms to retain long-term dependencies, have demonstrated substantial improvements over traditional statistical models in stock forecasting tasks. For instance, Nelson et al. [11] emphasized that LSTM networks are particularly well-suited for capturing long-range dependencies in stock price movements, which shallow models often fail to represent. Fischer and Krauss [12] demonstrated that LSTMs outperform both linear models and feedforward networks by effectively learning complex patterns in cross-sectional equity data. Pang et al. [13] extended this by introducing a deep LSTM model with an embedding layer, which leverages high-dimensional stock vectors to enhance multi-stock forecasting performance. Moghar and Hamiche [14] showed that even standard LSTM models can effectively capture the nonlinear dynamics of financial time-series, underscoring the centrality of temporal structure in deep learning-based forecasting. However, an important limitation remains: most LSTM pipelines operate on broadband or only minimally preprocessed financial signals such as log returns, which are noisy, multi-scale, and nonstationary in higher moments. Without an explicit mechanism to separate scales, the network must learn across entangled frequencies and regimes, blurring meaningful structure with transient noise and increasing the risk of overfitting and weak out-of-sample generalization. We address this gap with VMD, which provides a data-adaptive, band-limited representation so that the LSTM models more tractable, locally short-memory dynamics rather than a raw multi-scale mixture.

To mitigate these issues, researchers have increasingly adopted decomposition-based hybrid models. These approaches first decompose raw signals into simpler subcomponents and then apply predictive models to each part. By separating the original series into frequency-specific or volatility-based modes, these methods allow learning algorithms to focus on more stable and interpretable patterns. Specifically, Liang et al. [15] demonstrates that wavelet-thresholding enhances the smoothness of input data, thereby improving an LSTM's performance on volatile series. Similarly, Cao et al. [16] and Lin et al. [17] adopt complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) in combination

with LSTM networks, demonstrating consistent improvements in predictive accuracy across different stock indices. These results highlight the value of signal decomposition in stabilizing learning and isolating meaningful temporal structures.

However, their limitations are material: Wavelets rely on a fixed dictionary and preset time–frequency tiling, while empirical mode decomposition (EMD)/CEEMDAN are prone to mode mixing, noise sensitivity, weak spectral control and a lack of mathematical rigor [18–20]. As a result, sub-bands can still bleed across scales, and downstream models inherit residual nonstationarity despite higher apparent smoothness. VMD offers a more principled alternative: It learns center frequencies and minimizes baseband bandwidth of demodulated analytic modes via a constrained variational problem, yielding nonoverlapping, frequency-localized IMFs, and exact linear recomposition. Niu et al. [21] constructed a hybrid VMD–LSTM model and found it fit time-varying stock data better than EMD-based or single LSTM models. Nasiri and Ebadzadeh [22] proposed a hybrid forecasting model combining VMD and multi-functional recurrent fuzzy neural networks (MFRFNN), where each decomposed component is modeled separately, resulting in more accurate multistep predictions. Zhang et al. [23] reported that a VMD-preprocessed LSTM significantly improved prediction accuracy and stability for Chinese equity indices, compared to an undecomposed approach.

Recent literature has further pushed this boundary by focusing on advanced optimization and uncertainty estimation techniques for VMD-based models. For instance, Wang et al. [24] introduced a hybrid WOA-VMD-LSTM model where a whale optimization algorithm (WOA) is used to optimize the VMD parameters, and nonparametric kernel density estimation (KDE) is subsequently applied to the forecast errors to construct probabilistic prediction intervals. In a similar vein, Xu et al. [25] employed a dung Beetle optimization (DBO) algorithm to adaptively select parameters for both the VMD decomposition and the downstream least squares support vector machine predictor, additionally incorporating an explicit error correction strategy to model residuals and enhance final forecast accuracy. These studies highlight an emerging trend of combining signal decomposition with meta-heuristic optimization and sophisticated error modeling.

Collectively, these studies confirm that VMD-based decomposition enhances both the interpretability and stability of deep learning models, but most remain point-forecast systems—they do not calibrate epistemic intervals nor evaluate distributional fit on held-out data. Understanding the confidence level of predictions is equally important—particularly in high-volatility financial markets. Model uncertainty stems from limited training data or model misspecification, and has been actively addressed through Bayesian-inspired methods. Gal and Ghahramani [3] showed that dropout can be viewed as a Bayesian approximation, providing a principled way to quantify epistemic uncertainty via Monte Carlo sampling. Recent works have applied this idea to stock price prediction, for instance, Serpell et al. [26] combined Monte Carlo dropout with mean-variance estimation in LSTMs, capturing both model and data uncertainty to improve prediction intervals. Asare et al. [2] used Monte Carlo dropout in an LSTM for U.S. stock forecasts, yielding well-calibrated interval predictions that capture the uncertainty in volatile periods. An alternative is CD, a technique proposed by Gal et al. [8] that learns optimal dropout rates as continuous parameters during training. CD also produces probabilistic forecasts with the added benefit of automating the dropout tuning. These methods allow deep learning models to express model uncertainty in a structured way, providing more informative outputs for risk-sensitive financial applications.

Complementing epistemic modeling approaches, another strand of research focuses on capturing the inherent randomness in asset returns—termed market or aleatoric uncertainty—by modeling the

full distribution of possible outcomes. GANs have emerged as a powerful tool to learn complex data distributions via an adversarial game between a generator and discriminator. While originally developed for images, GAN variants have rapidly progressed and found applications in sequential and financial data. Improvements like the WGAN and WGAN with gradient penalty approaches mitigate training instabilities and mode collapse by using a more stable divergence measure [9, 10]. Time-series specific GAN frameworks soon followed: for instance, Mogren [27] proposed continuous recurrent neural network–generative adversarial network (C-RNN-GAN) with recurrent generators and discriminators for sequence generation, and Esteban et al. [28] introduced recurrent generative adversarial network (RGAN) for continuous-valued medical time-series. Yoon et al. [29] developed TimeGAN, combining supervised learning objectives with GAN training to better preserve temporal dynamics. In finance, researchers have started to apply GAN-based models to capture the distributional characteristics of asset returns. Wiese et al. [5] demonstrated that WGANs can generate realistic synthetic financial time-series that mimic the statistical properties of market returns. Yet a critical limitation remains under-discussed: Training GANs directly on raw or broadband financial sequences—with regime shifts, clustered volatility, and multi-scale structure—exacerbates critic/generator imbalance, sensitivity to hyperparameters, and mode collapse, making calibrated distribution learning unreliable. Our VMD–WGAN addresses this by applying VMD to log returns: The decomposition reduces higher-order and spectral nonstationarities per sub-band and lowers the entropy/complexity the critic must learn. Each mode is closer to local stationarity within the training window, adversarial training becomes more stable, and the recomposed distribution achieves higher fidelity to held-out data.

3. Methodology

3.1. Overview of the proposed framework

Our framework is designed to explicitly quantify and, crucially, disentangle the two primary sources of uncertainty in financial forecasting: Model uncertainty and market uncertainty. This objective motivates our core design: A unified dual-path architecture where each path is specialized for one type of uncertainty. The fundamental natures of these uncertainties are distinct; epistemic uncertainty relates to confidence in a single point forecast, while aleatoric uncertainty concerns the inherent spread of potential outcomes. Therefore, our framework assigns these tasks to two parallel, specialized modeling pipelines: A predictive model and a generative model.

The foundation of this architecture is a shared signal decomposition step using VMD. Rather than a simple preprocessing trick, VMD serves as an essential enabling technology that provides a common, stabilized representation for both paths. It decomposes complex, nonstationary financial signals into multiple IMFs, each representing a distinct and more manageable frequency component. This shared representation ensures that both the predictive and generative models learn from the same underlying market structure while focusing on their complementary objectives, as illustrated in Figure 1.

The framework then proceeds along its two paths:

- **Predictive path (VMD-LSTM-CD)**: This path targets the raw stock price series. Its goal is to produce accurate one-step-ahead price forecasts. The dispersion of predictions from multiple stochastic forward passes through this model serves as our operational measure of epistemic uncertainty.

- **Generative path (VMD-WGAN)**: This path operates on the log-return series, which exhibits more desirable statistical properties for distributional modeling. Its goal is to learn the conditional distribution of future returns. The spread of an ensemble of generated samples provides our operational measure of aleatoric uncertainty.

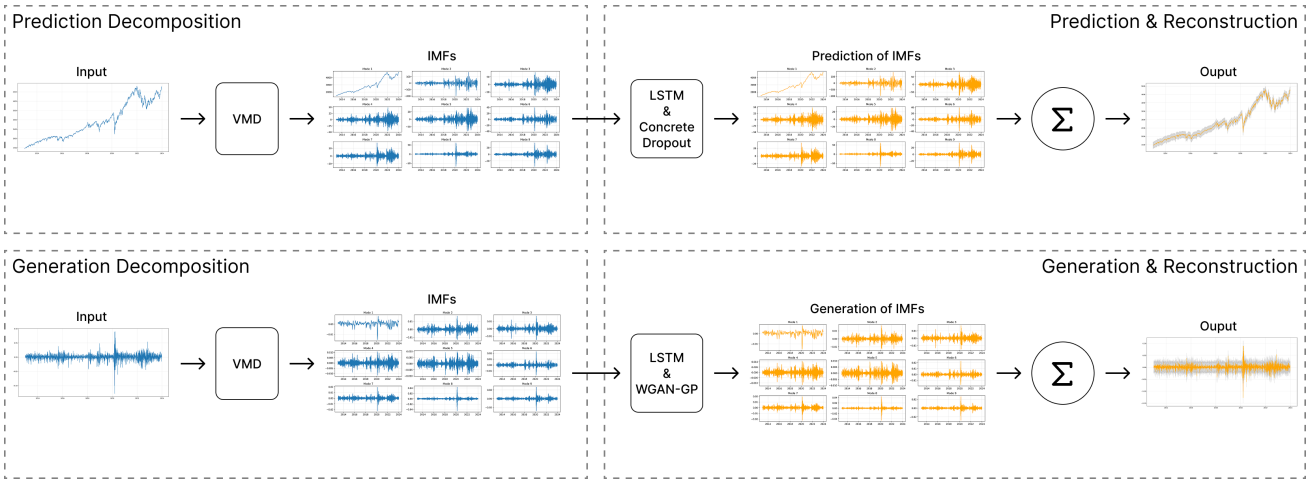


Figure 1. Overview of the dual-path uncertainty modeling framework.

This structure allows the framework to jointly capture predictive accuracy and distributional properties in a principled and structurally separated manner, offering a more robust and interpretable approach to risk-aware financial decision-making.

3.2. Variational mode decomposition

VMD expresses a time-series $X(t)$ as a linear superposition of a small number of modes $u_k(t)$, each occupying a compact, data-adaptive frequency band. Unlike empirical mode methods, VMD is posed as a well-defined constrained optimization directly in the frequency domain, which suppresses mode mixing and yields interpretable, nonoverlapping sub-bands [6].

In the standard formulation, the bandwidth of each mode is measured after converting it to its analytic representation and demodulating to baseband around a learned center frequency. Denoting $\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)$ as the analytic signal of $u_k(t)$ (with $*$ the convolution and δ the Dirac impulse), VMD seeks modes and center frequencies that minimize the aggregate baseband bandwidth:

$$\min_{\{u_k\},\{\omega_k\}} \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \quad \text{s.t.} \quad X(t) = \sum_{k=1}^{K} u_k(t). \tag{3.1}$$

Placing the demodulation factor $e^{-j\omega_k t}$ inside the time derivative ensures the objective depends on ω_k and concentrates each mode's spectrum around its learned center. The problem is efficiently solved via alternating direction method of multipliers (ADMM), returning narrow-band amplitude modulation-frequency modulation (AM-FM) like components with locally slowly varying envelopes and locally coherent phase; during shock episodes this local smoothness need not hold globally, yet band separation still reduces cross-band interference and stabilizes learning. Figure 2 illustrates a raw price series (left) and its frequency-specific VMD modes (right).

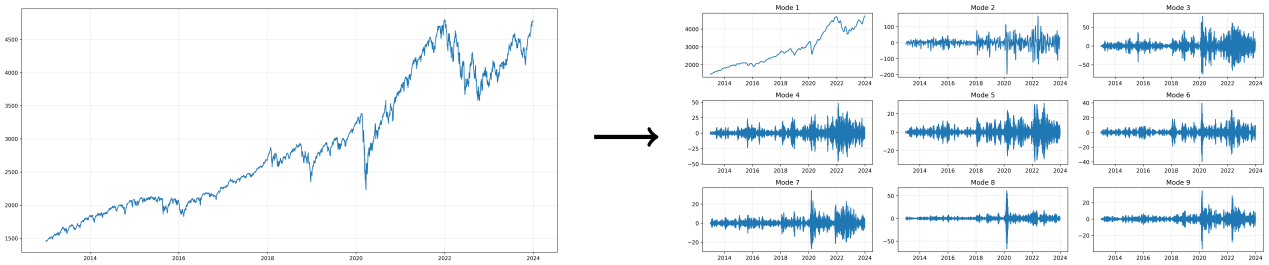


Figure 2. Raw price of Standard & Poor's 500 (S&P 500) index (left) and its intrinsic mode functions (right).

Relative to fixed-basis transforms, VMD occupies a distinct mathematical niche. For reference, the classical Fourier transform is defined by

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t}\, dt, \quad x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)\, e^{j\omega t}\, d\omega.$$

The short-time Fourier transform (STFT) applies a fixed analysis window w,

$$X(\tau, \omega) = \int_{-\infty}^{\infty} x(t)\, w^*(t-\tau)\, e^{-j\omega t}\, dt,$$

thereby imposing an a *priori* time–frequency resolution (* denotes complex conjugation); and the continuous wavelet transform (CWT) uses a chosen mother wavelet ψ to yield

$$W_x(a,b) = \int_{-\infty}^{\infty} x(t)\, \frac{1}{\sqrt{|a|}}\, \psi^*\Big(\frac{t-b}{a}\Big)\, dt,$$

with scale $a > 0$ and translation $b \in \mathbb{R}$. Each of these methods expands $x(t)$ over a fixed dictionary–global sinusoids, windowed sinusoids, or scaled/shifted wavelets–so band partitions are selected a priori. In contrast, VMD instead builds an analytic, data-adaptive filter bank by minimizing baseband bandwidth of demodulated modes. The Hilbert representation provides a one-sided spectrum and well-posed instantaneous amplitude/phase, and the learned centers ω_k align sub-bands to the data rather than to a fixed grid. In practice, this reduces cross-band interference and mode mixing while preserving exact linear reconstructability.

On the predictive side, the decomposition yields band-limited modes with nonoverlapping spectral support, which reduces cross-frequency interference in the inputs the network sees. Within the sliding input window used by the LSTM, these band-limited modes are well approximated by short-memory, low-order autoregressive (AR)/autoregressive moving average (ARMA) dynamics, because the local envelope and instantaneous frequency vary only gradually. As a result, the temporal dependency length the network must explain is shortened, optimization becomes more stable and sample-efficient, and the LSTM can focus its capacity on residual nonlinearities and cross-mode interactions rather than on long-range entanglement. This effect is also consistent with the spectral bias of deep networks, which tend to learn low-complexity structure first [30].

For the generative path, the log transform largely removes level nonstationarity, but higher-order and spectral nonstationarities persist such as time-varying volatility, skewness/kurtosis, regime mixing. Decomposing the series into narrow-band modes reduces cross-band interference and lowers the

distributional complexity each critic must learn; within the training window, each mode is closer to local stationarity and exhibits simpler, lower-entropy dynamics. The generator and critic are therefore trained per mode, and the aggregate signal is recovered by linear superposition $\sum_k u_k$. This sub-band conditioning also clarifies our uncertainty split: The predictive path quantifies model uncertainty via CD on mode inputs, whereas the generative path targets market randomness. We evaluate the latter with time-conditional diagnostics—moment profiles (mean, variance, skewness, kurtosis), Johnson-SU log-likelihood, and regime-wise summaries—and, when needed, mode-occlusion re-synthesis to attribute dispersion and tail behavior to specific frequency bands.

Taken together, VMD serves as a principled preprocessing step that simplifies temporal structure and mitigates nonstationarity and noise, enabling both predictive and generative modules to operate on stable, structured inputs typical of financial time-series.

3.3. Predictive model: VMD-LSTM with CD

This section details the architecture of our uncertainty-aware prediction model, VMD-LSTM-CD. We first provide a brief overview of the LSTM network, which serves as the core predictive engine. We then describe how we integrate LSTM with VMD and CD to construct a robust forecasting framework capable of generating model uncertainty intervals.

3.3.1. LSTM

LSTM networks are a specialized class of RNNs designed to capture long-range dependencies in sequential data. Unlike traditional RNNs, which often suffer from vanishing or exploding gradients and therefore struggle to retain information over extended time horizons, LSTMs address these limitations through the use of a dedicated memory cell and a set of gating mechanisms that dynamically regulate the flow of information.

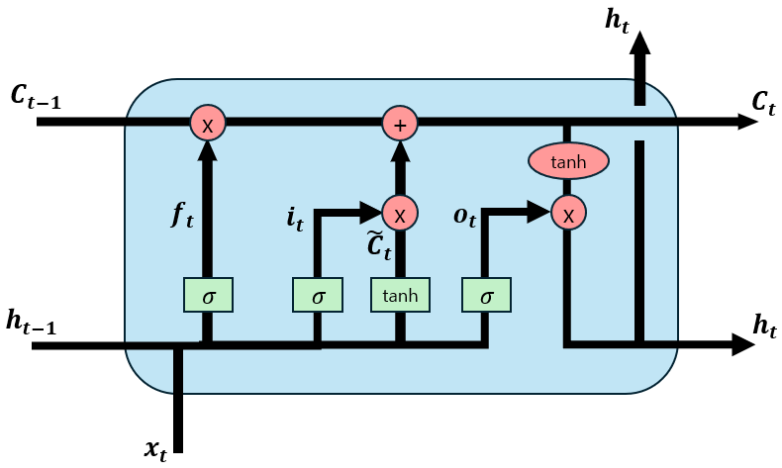


Figure 3. LSTM cell structure.

At each time step t, the LSTM cell receives the current input x_t and the previous hidden state h_{t-1}, and updates its internal cell state C_t and output h_t. This is achieved through three gates: the forget gate, the input gate, and the output gate. These gates control how much past information is retained, how new

input is incorporated, and how the final output is formed. The overall structure of this update process, including the interactions among gates, memory state, and output, is illustrated in Figure 3.

The full update equations governing these operations are as follows:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t \odot \tanh(C_t) \end{aligned}$$

Here, $\sigma(\cdot)$ denotes the sigmoid activation function, $\tanh(\cdot)$ is the hyperbolic tangent function, and $\odot$ represents element-wise multiplication.

This architecture allows LSTM networks to selectively remember and forget information over time, making them particularly effective in financial time-series applications where patterns evolve gradually or recur intermittently.

3.3.2. VMD-LSTM with CD

To model short-term price dynamics and quantify model-driven uncertainty, we construct a forecasting framework based on LSTM networks enhanced with CD. The entire process is deliberately applied to the raw stock price series, a choice motivated by its nonstationary and trend-dominant characteristics. Price data contains multi-scale trend and cyclical components that VMD is particularly adept at isolating into distinct IMFs. This decomposition effectively improves the signal-to-noise ratio for each component, allowing the downstream LSTMs to learn more stable and meaningful patterns. Furthermore, operating directly on prices enables a straightforward and stable recomposition of the IMF-level forecasts back into a final price trajectory.

We begin by applying VMD to the raw stock price series $P(t)$, decomposing it into IMFs, each denoted as $u_k(t)$, as described in Section 3.2. Each IMF captures a distinct frequency component of the original signal, simplifying the temporal structure and enabling more efficient learning.

To forecast future prices, we train a dedicated prediction model $g_k(\cdot)$ for each IMF. Specifically, for each k, the model takes as input a historical window of length N, represented as:

$$u_{k,t-N:t-1} = [u_k(t-N), u_k(t-N+1), \cdots, u_k(t-1)]$$

and produces a one-step-ahead forecast:

$$\hat{u}_k(t) = g_k(u_{k,t-N:t-1}).$$

The final price forecast is obtained by aggregating the forecasts from all IMF-specific models:

$$\hat{P}(t) = \sum_{k=1}^{K} \hat{u}_k(t).$$

To quantify model uncertainty, we adopt CD. Unlike standard dropout, which requires a fixed dropout rate, CD treats the dropout probability p as a learnable parameter and applies a continuous relaxation of the Bernoulli distribution to enable gradient-based optimization. The soft dropout mask m is sampled as:

$$m = \sigma\left(\frac{1}{\tau}\left(\log p - \log(1-p) + \log u - \log(1-u)\right)\right), \quad u \sim U(0,1)$$

where σ is the sigmoid function, τ is a temperature parameter controlling the smoothness of the approximation, and U indicates a uniform distribution. This allows the model to perform stochastic forward passes during inference, from which the sample mean is used as the point forecast and the variance reflects epistemic uncertainty.

The training objective for each g_k combines a prediction loss and a regularization term from CD. The prediction loss is defined as the mean squared error (MSE) between the true and predicted IMF values:

$$\mathcal{L}_{\text{pred}}(g_k) = \frac{1}{T}\sum_{t=1}^{T}(u_k(t) - \hat{u}_k(t))^2$$

The regularization term penalizes the learned dropout rate p, derived from the Kullback–Leibler (KL) divergence between the variational posterior and a log-uniform prior:

$$\mathcal{L}_{\text{reg}}(p) \approx (\log\tau - \log p) + p\log p + (1-p)\log(1-p).$$

The total loss to be minimized is given by:

$$\mathcal{L}_{\text{total}}(g_k) = \mathcal{L}_{\text{pred}}(g_k) + \lambda_p \mathcal{L}_{\text{reg}}(p),$$

where λ_p is a hyperparameter that balances accuracy and regularization. By combining VMD-based decomposition and CD-based uncertainty estimation, the VMD-LSTM-CD framework produces not only accurate forecasts but also well-calibrated intervals that adapt to the underlying model uncertainty.

3.4. Generative model: VMD-WGAN

The generative model is designed to learn the complete probability distribution of financial returns, thereby capturing market-driven uncertainty. This model is built upon the WGAN, an advanced generative architecture known for its training stability and high-quality sample generation.

3.4.1. WGAN

Traditional GANs are known to suffer from training instability and mode collapse, largely due to the use of divergence metrics such as the Jensen–Shannon divergence. To address these issues, the WGAN was introduced, which replaces the original GAN loss with the Earth Mover's (or Wasserstein) distance. This modification yields smoother gradients and more stable training dynamics.

The Wasserstein distance $\mathcal{W}(P_r, P_g)$ between the real data distribution P_r and the generated data distribution P_g can be expressed, via the Kantorovich–Rubinstein duality, as:

$$\mathcal{W}(P_r, P_g) = \sup_{\|C\|_L \le 1} \mathbb{E}_{x\sim P_r}\left[C(x)\right] - \mathbb{E}_{x\sim P_g}\left[C(x)\right],$$

where C is a 1-Lipschitz continuous critic function.

In the original WGAN formulation, the Lipschitz constraint was enforced via weight clipping, which often led to optimization difficulties and poor expressiveness. To improve this, Gulrajani et al. [10] proposed a gradient penalty technique, which softly enforces the Lipschitz condition by penalizing the norm of the gradient of the critic with respect to its input.

Let $x \sim P_r$ be a real sample, $\hat{x} \sim P_g$ be a generated sample from the generator $G(z)$, where $z \in \mathbb{R}^d$ is a latent noise sampled from standard multivariate Gaussian distribution, $z \sim \mathcal{N}(0, I_d)$. The interpolated sample $\tilde{x}$ is constructed as:

$$\tilde{x} = \epsilon x + (1 - \epsilon)\hat{x}, \quad \epsilon \sim U(0, 1).$$

Then, the critic loss function with gradient penalty is defined as:

$$\mathcal{L}_{\text{critic}} = \mathbb{E}_{\hat{x}\sim P_g}\left[C(\hat{x})\right] - \mathbb{E}_{x\sim P_r}\left[C(x)\right] + \lambda_g \mathbb{E}_{\tilde{x}}\left[(\|\nabla_{\tilde{x}} C(\tilde{x})\|_2 - 1)^2\right],$$

where λ_g is the penalty coefficient controlling the strength of the constraint. The generator aims to produce samples that receive high scores from the critic. Its objective is thus defined as:

$$\mathcal{L}_{\text{gen}} = -\mathbb{E}_{\hat{x}\sim P_g}\left[C(\hat{x})\right].$$

This formulation enables stable and effective training of WGANs in high-dimensional and noisy environments, such as financial time-series generation.

3.4.2. VMD-WGAN framework

The generative path is designed to capture the full conditional distribution of financial returns, thereby modeling the inherent market uncertainty. In contrast to the predictive path, we apply VMD to the log-return series for this task. This choice is critical for stable distributional modeling. Log-returns are closer to weak stationarity and are better suited for learning the complex, nonGaussian characteristics of market volatility, such as fat tails and skewness. Unlike price series with their persistent trends, applying VMD to near-stationary returns allows each WGAN to focus on modeling the conditional distribution of volatility within specific frequency bands, which is more effective than attempting to generate a nonstationary price series directly—A task at which GANs typically struggle.

We begin by applying VMD to the log-return series $r(t)$, which tends to exhibit stronger stationarity than raw prices. This yields K frequency-specific IMFs, denoted $v_k(t)$, each capturing distinct temporal dynamics. For each IMF $v_k(t)$, we construct a conditional generative model based on WGAN. The generator $G_k(\cdot)$ synthesizes plausible future outcomes conditioned on a historical sequence, while the critic $C_k(\cdot)$ evaluates the fidelity of these generated samples. Specifically, each generator takes as input a fixed-length historical window, $v_{k,t-N:t-1}$, along with a d-dimensional noise vector $z_k \sim \mathcal{N}(0, I_d)$, and outputs a one-step-ahead synthetic sample:

$$\hat{v}_k(t) = G_k(v_{k,t-N:t-1}, z_k).$$

The critic C_k is trained to assign higher scores to real samples $v_k(t)$ and lower scores to generated ones $\hat{v}_k(t)$, with the training objective including a gradient penalty to enforce the Lipschitz constraint. The loss functions for mode k are given as:

$$\mathcal{L}_{\text{critic}}(C_k) = \mathbb{E}\left[C_k\left(\hat{v}_k(t)\right)\right] - \mathbb{E}\left[C_k(v_k)\right] + \lambda_g \mathbb{E}_{\tilde{v}}\left[\left(\|\nabla_{\tilde{v}} C_k(\tilde{v})\|_2 - 1\right)^2\right]$$
$$\mathcal{L}_{\text{gen}}(G_k) = -\mathbb{E}\left[C_k\left(\hat{v}_k(t)\right)\right]$$

where $\tilde{v}$ denotes samples linearly interpolated between real and generated inputs.

After training, the final synthetic return $\hat{r}(t)$ is reconstructed by aggregating the generated outputs from all IMF-specific generators:

$$\hat{r}(t) = \sum_{k=1}^{K} \hat{v}_k(t).$$

This modular framework allows each generator to focus on frequency-localized behavior, while the aggregation reconstructs a coherent and distributionally rich forecast. The generation process produces an ensemble of possible return outcomes for each time step, from which a generation interval can be derived—Representing the full range of plausible returns driven by market uncertainty.

4. Experiments

This section details the experimental setup designed to evaluate the performance of our proposed dual-path uncertainty modeling framework. We describe the datasets and preprocessing steps, the specific architectures and training procedures for our models, and the metrics used for evaluation.

4.1. Experimental setup

4.1.1. Data and preprocessing

We use historical daily stock price data from the S&P 500 and FTSE 100 indices, sourced from Yahoo Finance. The dataset spans from 2013 to 2023 for training, while the full year of 2024 is reserved for out-of-sample testing.

We first apply VMD to the raw time-series—closing prices for the prediction model and log returns for the generative model—to obtain IMFs, as described in Section 3. After VMD, we normalize each IMF individually, depending on the downstream task. For the prediction model, which targets daily closing prices, each IMF is scaled to the [0, 1] range using min–max normalization. For the generative model, which learns the conditional distribution of log returns, each IMF is standardized using z-score normalization based on the training set statistics. After prediction or generation, the outputs from all IMFs are aggregated and then inverse-transformed to recover the original price or return scale, enabling evaluation in real-world units. All experiments were conducted using NVIDIA RTX 3090 GPU.

4.1.2. Implementation details

We describe the implementation details of the two core components of our framework: The predictive model (VMD-LSTM-CD) and the generative model (VMD-WGAN). Each follows a consistent training-testing schedule and leverages frequency-decomposed inputs obtained via VMD.

Predictive path. Each predictive model receives a 20-day historical window of daily closing prices and produces a one-step-ahead forecast. We employ a monthly expanding-window rolling forecast strategy. Models are initially trained on data spanning 2013–2023. From January 2024 onward,

the training set is expanded monthly by incorporating newly observed data. To simulate a realistic forecasting workflow, the model parameters from the previous month are used to initialize training for the current window, enabling fine-tuning with a reduced learning rate.

We compare three architectures: A standard LSTM, a VMD-LSTM, and our proposed VMD-LSTM-CD. For VMD-based models, the decomposition parameters $(K, \alpha, \tau, \text{DC}, \text{init}, \text{tol})$ were set to $(9, 450, 0, 1, 1, 1 \times 10^{-4})$ for the S&P 500 and $(4, 500, 0, 1, 1, 1 \times 10^{-4})$ for the FTSE 100.

In VMD-LSTM-CD, model uncertainty is quantified using Monte Carlo CD. The CD weight and dropout regularizers are set to $(1 \times 10^{-9}, 5 \times 10^{-7})$ for the S&P 500 and $(1 \times 10^{-9}, 9 \times 10^{-6})$ for the FTSE 100, respectively. At inference time, 100 stochastic forward passes are performed to obtain the uncertainty interval around the point forecast.

All models are implemented with a single LSTM layer and trained using the Adam optimizer. To ensure that each architecture is evaluated at its optimal performance, architecture-specific hyperparameters such as hidden units, learning rates, and batch sizes are individually tuned for each index, as summarized in Table 1.

Table 1. Hyperparameter settings for prediction models. Note: Only the learning rate has two values, corresponding to the initial training phase and subsequent monthly fine-tuning updates. All other hyperparameters remain fixed throughout.

Asset	Hyperparameter	LSTM	VMD-LSTM	VMD-LSTM-CD
S&P 500	Hidden nodes	200	200	200
	Learning rate	1×10^{-3}, 1×10^{-4}	3×10^{-3}, 4×10^{-4}	4×10^{-3}, 4×10^{-4}
	Batch size	64	16	16
FTSE 100	Hidden Nodes	200	100	100
	Learning rate	1×10^{-3}, 2×10^{-4}	4×10^{-3}, 4×10^{-4}	4×10^{-3}, 4×10^{-4}
	Batch size	64	16	16

Generative path. The generative model learns to produce one-step-ahead log-return samples conditioned on a 20-day history and driven by a 3-dimensional latent noise vector.

Both the baseline WGAN and the proposed VMD–WGAN follow the same evaluation timeline as the predictive model: Models are trained on data from 2013 to 2023, and monthly out-of-sample generations are performed throughout 2024. For each trading day in the test period, we draw 100 synthetic log-return samples, which are recomposed and used to construct a 95% generation interval representing the model's view of plausible outcomes.

The generator and critic share the same architecture in both WGAN variants. Each network consists of a two-layer LSTM with 64 hidden units, followed by a fully connected hidden layer and an output layer. Models are trained using the Adam optimizer with a learning rate of 1×10^{-3} and a batch size of 128 for 200 max epochs. As illustrated in Section 3, both models are trained using the WGAN objective with gradient penalty, where the Lipschitz constraint is enforced by using the penalty coefficient $\lambda_g = 10$. Specifically, for each training batch, we first create interpolated samples $\tilde{x}$ by sampling a random weight $\alpha \sim U(0, 1)$ and computing $\tilde{x} = \alpha x + (1 - \alpha)\hat{x}$, where x and $\tilde{x}$ are real and generated samples, respectively. The gradients of the critic's output with respect to these interpolated samples $(\nabla_{\tilde{x}} C(\tilde{x}))$ are

then calculated. The final gradient penalty is defined as the mean squared difference between the L_2 norm of these gradients and 1, corresponding to the term $\mathbb{E}_{\tilde{x}}\left[(\|\nabla_{\tilde{x}}C(\tilde{x})\|_2 - 1)^2\right]$. This penalty, scaled by λ_g, is added to the critic's loss function to enforce the 1-Lipschitz constraint.

For the VMD-WGAN model, the VMD parameters were set to $(9, 2500, 0, 0, 1, 1 \times 10^{-7})$, which were selected to ensure meaningful separation of signal modes in the log return domain.

4.1.3. Evaluation metrics

To comprehensively assess model performance, we adopt distinct metrics tailored to the objectives of our predictive and generative frameworks.

For deterministic point forecasting models, we evaluate accuracy using the RMSE. This metric is applied to both the closing price predictions and the log return predictions to provide a thorough assessment. Given the target values y_t and predictions from the models $\hat{y}$, the RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}.$$

For our generative models, we adopt a log-likelihood-based evaluation to quantify how well the generated distribution aligns with the observed outcomes. This metric is computed on a monthly basis for both training and test periods. Specifically, we generate 100 samples per trading day, aggregate the daily simulated returns, and fit a Johnson's SU distribution by maximum likelihood. This distribution was selected over the Gaussian due to its flexibility in capturing the skewness and heavy tails commonly observed in financial returns.

The probability density function (PDF) of the fitted distribution for each time t is given by:

$$f_t(x) = \frac{\delta_t}{\lambda_t\sqrt{2\pi}} \cdot \frac{1}{\sqrt{1 + \left(\frac{x-\xi_t}{\lambda_t}\right)^2}} \exp\left(-\frac{1}{2}\left[\gamma_t + \delta_t \sinh^{-1}\left(\frac{x-\xi_t}{\lambda_t}\right)\right]^2\right),$$

where the parameters $\gamma_t, \delta_t, \xi_t, \lambda_t$ are estimated for each month t in the test period. Using this fitted PDF, we compute the log-likelihood of each observed return and aggregate them to produce a log-likelihood score.

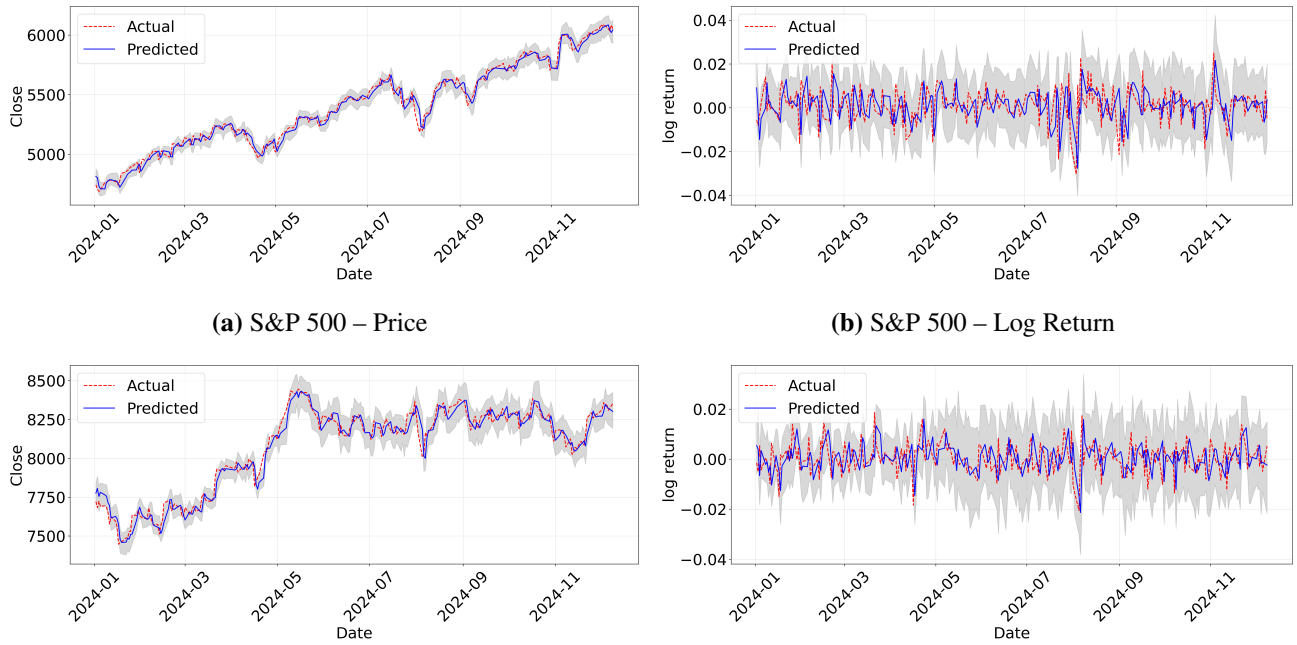
In addition, to analyze performance under different market conditions, we report regime-conditioned versions of these metrics. We partition trading days into three regimes based on quantiles of the training window: High-return days are defined as the top 25% of daily log-returns, low-return days as the bottom 25% of daily log-returns, and high-volatility days as the top 25% of daily realized volatility. RMSE and log-likelihood are evaluated within each regime, allowing us to compare model behavior across upside, downside, and turbulent markets.

By combining RMSE, log-likelihood, and their regime-conditioned counterparts, our evaluation protocol provides a balanced assessment of point prediction accuracy and distributional fidelity across both the predictive and generative paths of the framework. In addition to these primary metrics, we also report, for the predictive path, the time-averaged posterior predictive standard deviation and the realized 95% predictive coverage, and, for the generative path, the empirical 95% generation coverage in the leave-one-out ablation experiments in Section 4.2.3; these quantities serve as diagnostic measures of epistemic and aleatoric calibration, respectively.

4.2. *Experiment results*

4.2.1. Prediction result

This subsection presents the forecasting performance of our proposed predictive model, VMD-LSTM-CD, evaluated against baseline models on the S&P 500 and FTSE 100 indices for the 2024 test period.



(a) S&P 500 – Price

(b) S&P 500 – Log Return

(c) FTSE 100 – Price

(d) FTSE 100 – Log Return

Figure 4. One-step-ahead forecasts with daily 95% posterior predictive intervals via Monte Carlo-dropout (gray shaded region) for the S&P 500 and FTSE 100 using the VMD–LSTM–CD model.

Figure 4 provides a time-evolution view of one-step-ahead predictions from VMD–LSTM–CD together with the daily 95% posterior predictive intervals obtained via MC-Dropout (shaded bands). The bands quantify epistemic rather than raw market volatility: they widen precisely when input patterns depart from the training regime and contract in stable segments, indicating where the model is less confident while the mean forecast continues to track the realized path.

We next quantify performance under distinct market conditions in Table 2 for both the price level and the log return. Focusing on the S&P 500, VMD–LSTM–CD achieves the lowest overall RMSE for both price and log return. The advantage persists by regime. In high-return days, log-return RMSE falls from 0.0146 to 0.0134; in low-return days, from 0.0132 to 0.0128; and under high volatility, from 0.0121 to 0.0109. Price-level RMSE shows the same ordering overall and in upside days; in low-return and high-volatility subsets, the price-level gap between VMD–LSTM and VMD–LSTM–CD narrows, but the log-return accuracy—the economically diagnostic metric—remains strictly improved. For FTSE 100, the same story holds on log returns, while price-level RMSE can inflate in extreme regimes—an expected sensitivity of level-scale errors to cumulative drift and sharp swings. Taken together, the regime-conditioned evidence supports the main claim: the proposed model stabilizes return

forecasting across upside, downside, and turbulent conditions, even when price-scale errors can be more fragile.

Table 2. RMSE for S&P 500 and FTSE 100 over 2024, overall and by regime. Regimes are defined out-of-sample using the following quantiles: High Return = top 25% daily log-returns; Low Return = bottom 25%; High Volatility = top 25% daily realized-volatility.

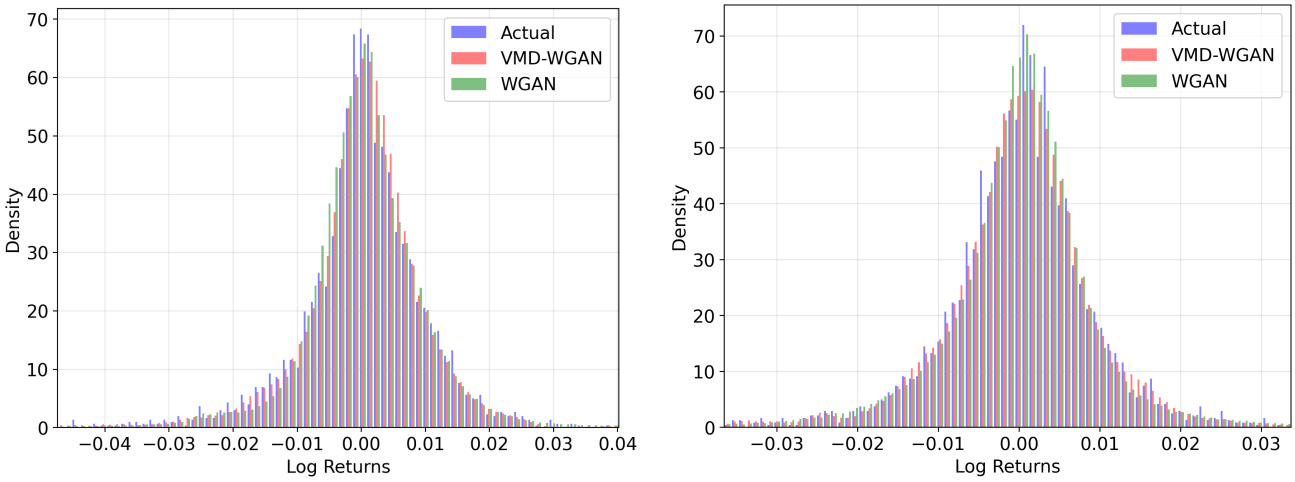
Regime	Model	S&P 500		FTSE 100	
		Price	Log return	Price	Log return
Overall	LSTM	51.06	0.0114	57.31	0.0093
	VMD–LSTM	45.41	0.0109	53.56	0.0082
	VMD–LSTM-CD	44.94	0.0102	51.87	0.0078
High return	LSTM	69.04	0.0146	81.57	0.0115
	VMD–LSTM	60.20	0.0139	72.64	0.0106
	VMD–LSTM-CD	59.27	0.0134	103.22	0.0105
Low return	LSTM	65.17	0.0132	78.16	0.0111
	VMD–LSTM	61.46	0.0129	71.20	0.0107
	VMD–LSTM-CD	63.58	0.0128	99.16	0.0096
High volatility	LSTM	57.96	0.0121	55.45	0.0073
	VMD–LSTM	52.47	0.0110	43.79	0.0064
	VMD–LSTM-CD	54.87	0.0109	79.76	0.0066

The substantial improvements in predictive accuracy are primarily attributed to the integration of VMD, which decomposes the original nonstationary series into band-limited IMFs with reduced cross-frequency interference. Each IMF exhibits simpler, locally short-memory dynamics, so the per-IMF LSTM learns a cleaner mapping; the final forecast is a linear recomposition. This preprocessing alleviates the burden of learning across entangled scales and regimes from raw inputs, which is precisely where plain LSTMs tend to overfit and generalize poorly.

The additional gains from VMD–LSTM to VMD–LSTM–CD stem from CD's adaptive regularization and model-averaging. During training, learned dropout rates suppress brittle co-adaptations; at inference, multiple stochastic passes are averaged, which reduces variance in the point forecast and tempers over-reaction to transient noise—particularly valuable in regime stress. Two features explain why net RMSE still improves here. First, after VMD has isolated the forecastable low/mid-frequency structure, the residual bias introduced by averaging is typically small relative to the variance reduction; the net effect on log returns—a local, de-trended scale—is positive across regimes. Second, price-level RMSE aggregates tiny return errors over time and can amplify them under steep trends; thus, in extreme FTSE regimes, variance reduction may not fully offset level-scale accumulation, even as the underlying return forecasts remain more stable. In short, VMD delivers the separability that makes the mapping learnable, and CD delivers the stability that makes it robust—producing consistent gains where it matters most for downstream decision-making, on the log-return scale.

4.2.2. Generation result

This subsection evaluates the generation capability of the proposed VMD–WGAN model by comparing it against a baseline WGAN trained directly on log returns. We assess both in-sample (2013–2023) and out-of-sample (2024) performance on the S&P 500 and FTSE 100 indices, focusing on how the VMD-driven conditional generation framework improves the modeling of market uncertainty across different market conditions.



(a) S&P 500 **(b)** FTSE 100

Figure 5. Unconditional histograms of daily log-returns for the actual series (blue) and synthetic samples generated by VMD–WGAN (red) and WGAN (green) over the in-sample period 2013–2023.

Table 3. Unconditional moments of daily log-returns for the S&P 500 and FTSE 100 over the in-sample period 2013–2023, comparing Actual, VMD–WGAN, and WGAN.

Asset	Model	Mean	Variance	Skewness	Kurtosis
S&P 500	Actual	4.36×10^{-4}	1.19×10^{-4}	-0.8186	16.4909
	VMD–WGAN	4.37×10^{-4}	1.16×10^{-4}	-0.6455	16.7258
	WGAN	3.82×10^{-4}	1.17×10^{-4}	-0.7444	12.9338
FTSE 100	Actual	9.76×10^{-5}	9.63×10^{-5}	-0.8477	12.6832
	VMD–WGAN	9.70×10^{-5}	9.40×10^{-5}	-0.6080	13.3032
	WGAN	7.72×10^{-5}	9.55×10^{-5}	-0.0955	7.7527

We first examine the unconditional distributional fit on the training window. Figure 5 displays in-sample histograms of realized log returns together with those generated by VMD–WGAN and the baseline WGAN, while Table 3 reports the corresponding moments. For both indices, VMD–WGAN closely matches the empirical mean and variance: on the S&P 500, the mean and variance under VMD–WGAN deviate from the actual values by only a few percent, whereas WGAN noticeably

underestimates the mean; a similar pattern appears for the FTSE 100, where the baseline's mean is substantially lower than the empirical one. More importantly, VMD–WGAN reproduces the pronounced negative skewness and extreme kurtosis of the training data, especially for the FTSE 100, where the baseline WGAN yields nearly symmetric returns with much lighter tails. These results indicate that the VMD-based decomposition allows the generative model to capture the heavy, asymmetric tails of equity returns without sacrificing the bulk of the distribution.

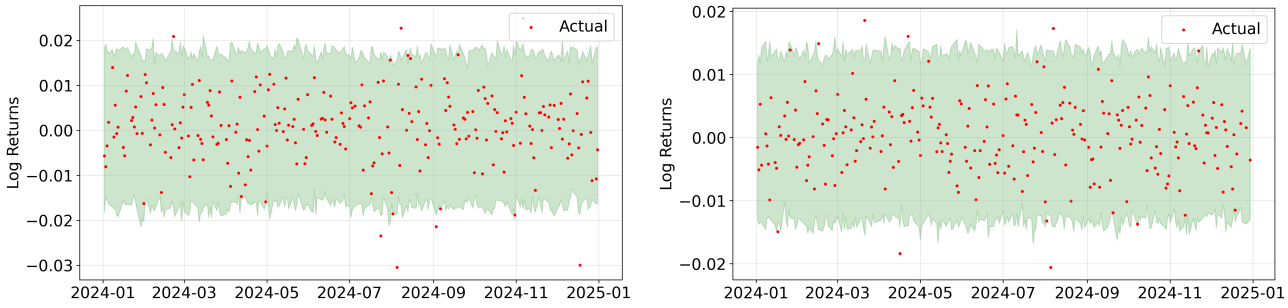
Table 4. Regime-specific log-likelihoods of daily log-returns for the S&P 500 and FTSE 100, reported for the in-sample period 2013–2023 and the out-of-sample period 2024. Regimes are defined out-of-sample using quantile cutoffs computed on the training distribution: High Return = top 25% of daily log-returns; Low Return = bottom 25%; High Volatility = top 25% of daily realized volatility. "Overall" aggregates all days without conditioning on regimes.

Regime	Model	S&P 500		FTSE 100	
		In-sample	Out-of-sample	In-sample	Out-of-sample
Overall	VMD–WGAN	312.90	334.89	318.88	355.15
	WGAN	305.60	282.05	317.69	295.43
High return	VMD–WGAN	276.99	313.05	283.23	338.01
	WGAN	249.76	255.46	238.49	246.77
Low return	VMD–WGAN	253.86	303.81	252.33	337.88
	WGAN	202.81	216.73	224.94	253.57
High volatility	VMD–WGAN	190.37	271.55	186.56	314.71
	WGAN	110.89	199.31	140.26	226.93

To investigate performance under distinct market regimes, Table 4 reports regime-specific Johnson–SU log-likelihoods for VMD–WGAN and WGAN, separately for the in-sample and out-of-sample periods. Across all regimes, VMD–WGAN achieves higher log-likelihoods than the baseline on both indices. The gains are modest in calm regimes but become substantial in stressed conditions: For example, during high-volatility periods the log-likelihood improvements amount to several dozen points in both the training window and the 2024 test year. This pattern shows that VMD–WGAN is not merely matching the unconditional distribution; it also maintains superior fit when the market is in upside, downside, or turbulent states.

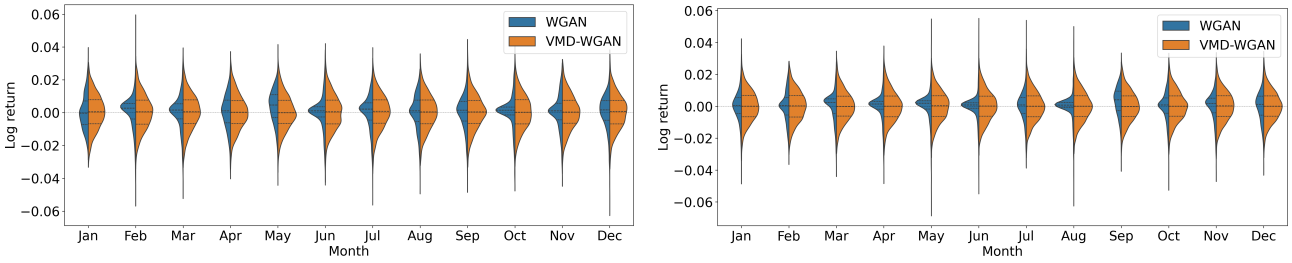
We next inspect the out-of-sample generation bands shown in Figure 6 for the S&P 500 and FTSE 100 during the 2024 test period. The empirical 95% bands produced by VMD–WGAN form a nearly stationary tube around zero, with only mild fluctuations in width over time. This behavior is consistent with interpreting daily log-returns as almost noise-like at the one-day horizon: most of the uncertainty comes from the heavy-tailed distribution of returns, while genuinely time-varying volatility plays a more limited role. Within these bands, the realized returns are spread throughout the interior, and only a small number of extreme shocks fall outside, indicating that the learned distribution maintains reasonable coverage on unseen data. Consistently, the monthly out-of-sample violin plots in Figure 7 show that VMD–WGAN produces wider, more heavy-tailed return distributions across

months, whereas the baseline WGAN samples remain much more tightly concentrated around zero with visibly lighter tails. This visual pattern is in line with the unconditional moments in Table 3, where VMD–WGAN closely follows the empirical kurtosis—especially for the FTSE—while WGAN substantially underestimates the tail thickness. Together with the superior out-of-sample log-likelihoods in Table 5, this visual evidence suggests that VMD–WGAN has learned a stable, heavy-tailed return distribution from 2013–2023 that generalizes robustly to 2024, in contrast to the baseline WGAN.



(a) S&P 500 **(b)** FTSE 100

Figure 6. Empirical 95% generation bands (green) and actual values (red) from the VMD–WGAN model. The green shaded area represents the empirical 95% generation band, computed from multiple samples generated by the proposed model. These bands capture the distributional uncertainty learned from the training data, rather than classical statistical confidence intervals.



(a) S&P 500 **(b)** FTSE 100

Figure 7. Monthly distributions of generated log-returns from WGAN and VMD–WGAN models for S&P 500 and FTSE 100.

Finally, the architectural rationale for these results aligns with our earlier design discussion. By applying VMD to the log-return series and training one conditional WGAN per IMF, the model assigns each generator a frequency-localized subproblem that is closer to local stationarity and has lower distributional complexity. The critics no longer need to learn a single broadband, regime-mixed return process; instead, they learn simpler sub-band dynamics whose recomposition recovers the heavy-tailed, regime-dependent distribution of returns. The in-sample histograms and moment matching, the regime-wise log-likelihood gains, and the adaptive test-period bands jointly support this explanation: VMD–WGAN achieves improved and more stable modeling of market uncertainty precisely because the decomposition disentangles slow and fast components before adversarial training, allowing the generative model to remain calibrated across time and across heterogeneous market conditions.

Table 5. Monthly out-of-sample log-likelihood scores and relative improvements (Improv.) of VMD–WGAN over WGAN for S&P 500 and FTSE 100 log returns.

	S&P 500			FTSE 100		
Month	WGAN	VMD–WGAN	Improv.(%)	WGAN	VMD–WGAN	Improv.(%)
Jan.	347.77	341.67	-1.75	362.98	354.99	-2.20
Feb.	338.73	334.76	-1.17	359.20	355.45	-1.04
Mar.	347.66	345.53	-0.61	71.12	355.67	400.09
Apr.	326.62	330.58	1.21	205.47	348.91	69.81
May	323.88	347.89	7.41	258.42	360.77	39.61
Jun.	234.91	356.03	51.56	281.83	360.86	28.04
Jul.	338.11	329.15	-2.65	363.99	356.58	-2.03
Aug.	300.64	296.78	-1.28	186.53	338.58	81.51
Sep.	342.37	331.80	-3.09	340.09	354.27	4.17
Oct.	141.29	343.37	140.03	365.03	357.24	-2.13
Nov.	324.48	336.06	3.57	364.35	356.27	-2.22
Dec.	339.59	328.36	-3.31	365.10	362.86	-0.61

4.2.3. Comparative analysis

To interpret how the VMD decomposition allocates information across frequencies on both paths, we conduct a leave-one-out (LOO) ablation at the composition layer of our framework for the S&P 500 index. For each IMF, we keep all per-IMF networks fixed and reconstruct the predictive outputs or generated returns while omitting that single mode, and then measure how the main evaluation metrics change relative to the full-IMF models.

Predictive path. On the predictive path, recall that each IMF is modeled by its own LSTM trained independently on the S&P 500 to forecast the next-day price level. Predictive uncertainty is approximated with Monte Carlo CD: for each date, we generate multiple stochastic forward passes through the network and treat the resulting ensemble of forecasts as samples from the posterior predictive distribution.

For each LOO configuration, we then recompute, on the 2024 S&P 500 test period, three out-of-sample metrics as explained in Section 4.1.3: (i) RMSE of the price level, (ii) the time-averaged posterior predictive standard deviation (Pred STD), and (iii) the realized 95% predictive coverage, defined as the fraction of days on which the realized price falls within the prediction interval. Table 6 reports these metrics and their relative changes with respect to the full-IMF predictive model.

Removing the lowest-frequency mode (IMF 1) is catastrophic for both accuracy and calibration: The price-level RMSE explodes, the posterior predictive STD contracts sharply, and realized 95% coverage collapses to 0. This is a textbook signature of over-confidence around a biased mean once the level/cycle component is removed: The predictor concentrates too tightly but around the wrong center, so intervals become unrealistically narrow and miss the truth almost always.

Table 6. Performance metrics of VMD–LSTM–CD (out-of-sample, 2024) and VMD–WGAN (in-sample, 2013–2023) on the S&P 500 by excluded IMF (leave-one-out). For VMD–LSTM–CD we report RMSE, Pred STD (mean posterior predictive standard deviation via MC-Dropout), and 95% predictive-interval coverage; for VMD–WGAN we report log-likelihood and 95% generation-band coverage. Values in parentheses are absolute changes relative to the full model with all IMFs.

Excluded IMF	VMD-LSTM-CD			VMD-WGAN	
	RMSE	Pred STD	Coverage	Log-likelihood	Coverage
1	3934.49 (86.55)	8.99 (-0.736)	0.000 (1.00)	300.36 (-4.01)	94.83 (0.08)
2	72.49 (0.61)	33.27 (-0.022)	0.558 (0.34)	309.97 (-0.94)	94.80 (0.04)
3	51.17 (0.14)	33.79 (-0.007)	0.796 (0.06)	303.42 (-3.03)	94.94 (0.19)
4	46.85 (0.04)	33.94 (-0.003)	0.829 (0.02)	303.10 (-3.13)	94.76 (0.00)
5	45.35 (0.009)	33.98 (-0.001)	0.854 (-0.010)	307.47 (-1.74)	94.54 (-0.23)
6	44.98 (0.001)	34.02 (-0.0003)	0.858 (-0.014)	312.89 (0.00)	94.98 (0.23)
7	44.71 (-0.005)	34.02 (-0.0005)	0.858 (-0.014)	312.71 (-0.06)	94.87 (0.12)
8	44.87 (-0.002)	34.03 (-0.0002)	0.842 (0.005)	307.74 (-1.65)	94.51 (-0.27)
9	44.95 (0.0001)	34.01 (-0.0006)	0.850 (-0.005)	309.96 (-0.94)	94.87 (0.12)

Ablating the next medium-frequency bands (IMF 2–5) still degrades accuracy and calibration. RMSE remains elevated, Pred STD stays below baseline, and coverage is well below the nominal 0.95 (0.558–0.854). These modes carry slowly varying, forecastable structure that anchors mean tracking and appropriate interval width.

By contrast, removing higher-frequency modes (IMF 6–9) has little impact on uncertainty and yields small improvements or minimal changes in RMSE with coverage close to baseline (0.842–0.858) and Pred STD almost unchanged. This indicates that the highest bands behave like near-white noise at the daily horizon: they add estimation noise rather than predictive signal, and strong regularization or exclusion can slightly improve generalization without harming calibration.

Altogether, the LOO analysis clarifies the division of labor across frequency bands in the predictive path: Low-frequency IMFs are essential for both point accuracy and well-calibrated posterior predictive STD/intervals, whereas the highest-frequency IMFs have only limited impact on predictive performance and behave largely as residual high-frequency components—precisely the separation VMD is intended to induce.

Generative path. In this part, we examine how each IMF in our model contributes to representing market uncertainty. We perform IMF-wise LOO ablations of the VMD-WGAN model and, for each excluded IMF, recompute the in-sample (2013–2023) log-likelihood of realized returns under the generated distribution and the corresponding 95% coverage, defined as the fraction of days on which the realized return lies inside the simulated prediction band. The right-hand columns of Table 6 report these LOO log-likelihoods and coverages as relative changes with respect to the full-IMF VMD-WGAN. Although log-returns are approximately stationary, the VMD still provides a decomposition into modes ordered from low to high center frequencies. In this sense, the interpretation of IMFs as low-, medium-, and high-frequency components remains valid, even if the separation is less pronounced than in nonstationary

stock level series as shown in the predictive model.

Removing the lowest-frequency IMF (IMF 1) produces by far the largest deterioration in in-sample distributional fit: the log-likelihood drops by about 4%, while the coverage remains close to its baseline level. This pattern means that the generative bands still contain a similar fraction of realized returns over 2013–2023, but the probability mass is allocated less efficiently across the sample. In line with its low frequency, IMF 1 therefore appears to capture the slow level and cycle component that anchors the center of the return distribution and controls a substantial part of its tail behavior in the generative path.

When IMFs 2–5 (medium-frequency bands) are excluded, the impact is milder but still systematic. Across these LOO experiments, the in-sample log-likelihood typically decreases by 1–3%, while the coverage stays close to its nominal level. This means that the model still reproduces how often large deviations occur over 2013–2023, but fits the detailed shape of the in-sample return distribution less accurately. In particular, the deterioration in log-likelihood suggests that the medium-frequency IMFs help the model allocate probability more appropriately between moderate and extreme returns. Hence, IMFs 2–5 can be interpreted as capturing multi-day fluctuations in return magnitudes and refining the center and tails of the distribution beyond what IMF 1 alone can provide.

Turning to the highest-frequency part of the decomposition, IMFs 6–9, the effect depends on the specific mode. For IMFs 6 and 7, removing each component induces only a very small change in in-sample log-likelihood, which is consistent with these modes behaving largely as residual high-frequency noise. In contrast, excluding IMFs 8 and 9 leads to a more noticeable deterioration, although the impact is still moderate compared with that of the medium-frequency IMFs. This pattern is consistent with viewing daily log-returns as an almost noise-like process: most of the high-frequency energy behaves as residual noise, but certain higher-frequency bands can still carry nontrivial structure that the generative model exploits to better match the empirical return distribution.

Overall, the in-sample LOO analysis of VMD-WGAN reveals a graded, frequency-wise allocation of roles in the generative model. The low- and medium-frequency IMFs account for most of the gains in log-likelihood and thus for the core shape of market uncertainty, in terms of the center, dispersion, and tails of the return distribution. The high-frequency IMFs act as residual components with limited marginal impact, but the behaviors of some IMFs illustrate that within this high-frequency range some bands can make a measurable contribution. This view is consistent with interpreting daily log-returns as an almost noise-like process while still allowing VMD to extract meaningful low- and medium-frequency structure.

5. Discussion and conclusion

This study presents a unified deep learning framework for financial time-series forecasting that explicitly models both model uncertainty and market uncertainty through a task-specialized dual-path architecture. By integrating uncertainty-aware prediction with generative modeling, our framework provides not only accurate forecasts but also a distributional understanding of financial risks for major equity indices such as the S&P 500 and FTSE 100.

In the predictive path, we develop VMD-LSTM-CD, which integrates variational mode decomposition and CD to generate one-step-ahead forecasts accompanied by model uncertainty intervals that reflect epistemic risk. In the generative path, we develop VMD-WGAN, a conditional adversarial model trained on decomposed log-return components, enabling the generation of diverse return paths and capturing distributional market uncertainty through generation intervals.

Empirical evaluations on the S&P 500 and FTSE 100 show that our approach outperforms conventional baselines in both predictive accuracy and distributional fidelity. On the predictive side, regime-conditioned experiments and an IMF-wise leave-one-out analysis reveal a clear frequency-wise division of labor, with low- and intermediate-frequency modes driving both point accuracy and calibration, while the highest-frequency modes behave largely as residual components with limited forecastable content. On the generative side, in-sample histograms, unconditional moments, and regime-wise log-likelihoods demonstrate that VMD–WGAN more faithfully reproduces heavy-tailed return distributions, especially in stressed market conditions, and that the learned distributions generalize robustly to out-of-sample periods. Taken together, these results suggest that modeling uncertainty through complementary predictive and generative pathways provides richer and more stable information than deterministic forecasts alone.

Looking ahead, we envision two complementary directions to advance the framework's utility and rigor. First, regarding practical viability, we aim to align the architecture with standard asset management workflows and governance protocols. By adopting a block update schedule—where computationally intensive retraining occurs offline on a monthly cadence—daily inference remains a lightweight, batchable operation suitable for time-sensitive decision-making. To further mitigate computational overhead in the generative path, we plan to investigate spectral normalization, parameter sharing, and model compression techniques such as pruning or distillation, alongside benchmarking nonadversarial likelihood-based sequence models for resource-constrained environments.

Second, to refine epistemic uncertainty quantification on the predictive path, we intend to move beyond the CD baseline toward more expressive Bayesian approaches. Future work will explore variational LSTM layers trained via the evidence lower bound (ELBO), deep ensembles, and Laplace approximations, validated through proper scoring rules and calibration diagnostics. Finally, a critical evolution of this framework is to synthesize the distinct predictive and generative paths into a fully unified architecture. By optimizing a joint objective function within a single end-to-end model, we aim to capture the interplay between epistemic and aleatoric uncertainty more effectively, thereby reducing architectural redundancy while maintaining the benefits of signal decomposition.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work received financial support from the National Research Foundation of Korea (No. RS-2025-00562904). This research was supported by Global-Learning & Academic research institution for Master's · PhD students, and Postdocs (LAMP) Program of the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. RS-2024-00442775).

Conflict of interest

All authors declare no conflicts of interest in this paper.

Author contributions

Jeonggyu Huh: Conceptualization, Methodology, Supervision, Funding acquisition; Dajin Kim: Software, Formal analysis, Data curation, Writing – original draft; Minseok Jung: Software, Formal analysis, Data curation, Writing – original draft; Seungwon Jeong: Methodology, Software, Validation, Writing – review & editing.

References

1. D. Avramov, Stock return predictability and model uncertainty, *J. Financ. Econ.*, **64** (2002), 423–458. https://doi.org/10.1016/S0304-405X(02)00131-9
2. C. Asare, D. Asante, J. F. Essel, Probabilistic LSTM modeling for stock price prediction with Monte Carlo dropout, *Int. J. Innov. Sci. Res. Technol.*, **8** (2023), 2316–2323. https://doi.org/10.5281/zenodo.8224141
3. Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: Representing model uncertainty in deep learning, preprint, arXiv: 1506.02142, 2016. https://doi.org/10.48550/arXiv.1506.02142
4. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, et al. Generative adversarial networks, *Adv. Neural Inf. Process. Syst.*, **63** (2014), 139–144. https://doi.org/10.1145/3422622
5. M. Wiese, R. Knobloch, R. Korn, P. Kretschmer, Quant GANs: Deep generation of financial time series, *Quant. Finance*, **20** (2020), 1419–1440. https://doi.org/10.1080/14697688.2020.1730426
6. K. Dragomiretskiy, D. Zosso, Variational mode decomposition, *IEEE Trans. Signal Process.*, **62** (2014), 531–544. https://doi.org/10.1109/TSP.2013.2288675
7. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.*, **9** (1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
8. Y. Gal, J. Hron, A. Kendall, Concrete dropout, preprint, arXiv: 1705.07832, 2017. https://doi.org/10.48550/arXiv.1705.07832
9. M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in *Proceedings of the 34th International Conference on Machine Learning*, **70** (2017), 214–223. Available from: `https://proceedings.mlr.press/v70/arjovsky17a.html`.
10. I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. Courville, Improved training of Wasserstein GANs, preprint, arXiv: 1704.00028, 2017. https://doi.org/10.48550/arXiv.1704.00028
11. D. M. Q. Nelson, A. C. M. Pereira, R. A. de Oliveira, Stock market's price movement prediction with LSTM neural networks, in *2017 International Joint Conference on Neural Networks (IJCNN),* Anchorage, AK, USA, 2017, 1419–1426. https://doi.org/10.1109/IJCNN.2017.7966019
12. T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, *Eur. J. Oper. Res.*, **270** (2018), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054
13. X. Pang, Y. Zhou, P. Wang, W. Lin, V. Chang, An innovative neural network approach for stock market prediction, *J. Supercomput.*, **76** (2020), 2098–2118. https://doi.org/10.1007/s11227-017-2228-y
14. A. Moghar, M. Hamiche, Stock market prediction using LSTM recurrent neural network, *Procedia Comput. Sci.*, **170** (2020), 1168–1173.

15. X. Liang, Z. Ge, L. Sun, M. He, H. Chen, LSTM with wavelet transform based data preprocessing for stock price prediction, *Math. Probl. Eng.*, **2019** (2019), 1–8. https://doi.org/10.1155/2019/1340174
16. J. Cao, Z. Li, J. Li, Financial time series forecasting model based on CEEMDAN and LSTM, *Physica A*, **519** (2019), 127–139. https://doi.org/10.1016/j.physa.2018.11.061
17. Y. Lin, Y. Yan, J. Xu, Y. Liao, F. Ma, Forecasting stock index price using the CEEMDAN-LSTM model, *N. Am. J. Econ. Finance*, **57** (2021), 101421. https://doi.org/10.1016/j.najef.2021.101421
18. M. H. Al-Badrawi, *Statistical Properties and Applications of Empirical Mode Decomposition*, Ph.D. thesis, University of New Hampshire, 2017.
19. M. Jabloun, Empirical mode decomposition revisited using ordinal pattern concepts, in *2022 30th European Signal Processing Conference (EUSIPCO)*, Belgrade, Serbia, 2022, 2186–2190. https://doi.org/10.23919/EUSIPCO55093.2022.9909668
20. B. Xu, Y. Sheng, P. Li, Q. Cheng, J. Wu, Causes and classification of EMD mode mixing, *Vibroeng. Procedia*, **22** (2019), 158–164. https://doi.org/10.21595/vp.2018.20250
21. H. Niu, K. Xu, W. Wang, A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network, *Appl. Intell.*, **50** (2020), 4296–4309. https://doi.org/10.1007/s10489-020-01814-0
22. H. Nasiri, M. M. Ebadzadeh, Multi-step-ahead stock price prediction using recurrent fuzzy neural network and variational mode decomposition, *Appl. Soft Comput.*, **148** (2023), 110867. https://doi.org/10.1016/j.asoc.2023.110867
23. Z. Zhang, Q. Liu, Y. Hu, H. Liu, Multi-feature stock price prediction by LSTM networks based on VMD and TMFG, *J. Big Data*, **12** (2025), 74. https://doi.org/10.1186/s40537-025-01127-4
24. W. C. Wang, B. Wang, K. W. Chau, D. M. Xu, Monthly runoff time series interval prediction based on WOA-VMD-LSTM using nonparametric kernel density estimation, *Earth Sci. Inform.*, **16** (2023), 2373–2389. https://doi.org/10.1007/s12145-023-01038-z
25. D. M. Xu, Z. Li, W. C. Wang, An ensemble model for monthly runoff prediction using least squares support vector machine based on variational modal decomposition with dung beetle optimization algorithm and error correction strategy, *J. Hydrol.*, **629** (2024), 130558. https://doi.org/10.1016/j.jhydrol.2023.130558
26. C. Serpell, I. A. Araya, C. Valle, H. Allende, Probabilistic forecasting using Monte Carlo dropout neural networks, in *Lecture Notes in Computer Science*, **11896** (2019), 387–397. https://doi.org/10.1007/978-3-030-33904-3_36
27. O. Mogren, C-RNN-GAN: Continuous recurrent neural networks with adversarial training, preprint, `arXiv:1611.09904`.
28. C. Esteban, S. L. Hyland, G. Rätsch, Real-valued (medical) time series generation with recurrent conditional GANs, preprint, `arXiv:1706.02633`.

29. J. Yoon, D. Jarrett, M. van der Schaar, Time-series generative adversarial networks, in *Advances in Neural Information Processing Systems*, **32** (2019).

30. N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, et al. On the spectral bias of neural networks, preprint, arXiv: 1806.08734, 2019. https://doi.org/10.48550/arXiv.1806.08734

AIMS Press