# Deep Reinforcement Learning for Quantitative Trading: A Novel Framework Integrating Variational Mode Decomposition and Contrastive Transformer

YI WENSHENG[1,*], GAO KE[1], CHEN JIAMING[2]
[1]Shanghai Advanced Institute of Finance,
Shanghai Jiao Tong University,
Shanghai 200030,
CHINA

[2]Renmin Business School,
Renmin University of China,
Beijing 100000
CHINA

*Corresponding Author*

*Abstract:* - To address the limitations of traditional quantitative trading models in extracting non-linear features, modeling long-term dependencies, and identifying abnormal fluctuations in financial time series, this paper proposes an improved quantitative trading framework integrating Variational Mode Decomposition (VMD), self-supervised Contrastive Learning, and Deep Reinforcement Learning (DRL). This framework aims to resolve the challenges of low Signal-to-Noise Ratio (SNR) and non-stationary distributions in high-frequency financial data. First, VMD is utilized to decompose raw price-volume data into several Intrinsic Mode Functions (IMFs), effectively separating market microstructure noise from core trend signals. Second, a Contrastive Transformer Encoder based on a time-aware mechanism is constructed to extract robust market state representations via self-supervised contrastive learning tasks, overcoming the overfitting issues inherent in supervised learning during extreme market conditions. Finally, an improved Proximal Policy Optimization (PPO) algorithm is introduced at the decision optimization layer, employing the Differential Sharpe Ratio as the core reward function to achieve end-to-end portfolio optimization. Backtesting results on the S&P 500 index (2024–2025) demonstrate that the proposed model significantly outperforms traditional LSTM and Mean-Variance models. Specifically, the model achieved an Annualized Rate of Return (ARR) of 28.45% and a Sharpe Ratio of 2.18, while maintaining a Maximum Drawdown of -9.15% during global market volatility, validating the potential of generative AI technologies in quantitative investment.

*Key-Words:* - Variational Mode Decomposition, Contrastive Learning, Transformer, Deep Reinforcement Learning, Quantitative Trading, Financial Time Series, Portfolio Optimization, Differential Sharpe Ratio, Self-Supervised Learning, Risk Management.

Received: February 12, 2026. Revised: May 31, 2026. Accepted: June 14, 2026. Published: July 28, 2026.

## 1 Introduction

As the proportion of electronic and algorithmic trading in global financial markets continues to rise, market microstructure has become increasingly complex, with asset price fluctuations exhibiting significant non-linearity, non-stationarity, and long-memory characteristics, [1], [2]. Traditional econometric models (e.g., ARIMA, GARCH), often based on linear assumptions and stationarity premises, struggle to capture the complex patterns triggered by algorithmic trading, sudden macroeconomic events, and investor sentiment resonance in modern markets, [3]. In recent years, Machine Learning, particularly Deep Learning (DL), has made significant progress in quantitative finance, [4]. Recurrent Neural Networks (RNNs) and their variant, Long Short-Term Memory (LSTM) networks, are widely applied in stock price prediction and factor mining due to their natural advantage in processing sequential data, [5]. However, due to the extremely low Signal-to-Noise Ratio (SNR) and the "data-rich but informationpoor" nature of financial time series, the direct application of end-to-end deep learning models often faces severe overfitting risks and lacks robustness on Out-of-Distribution (OOD) samples, [6], [7].

Between 2024 and 2025, global capital markets experienced extreme volatility triggered by geopolitical conflicts, divergent monetary policies, and the AI technology bubble. In this context, traditional factor selection models and Mean-Variance Optimization (MVO) strategies exhibited significant maladaptation, resulting in severe drawdowns due to amplified parameter estimation errors, [8], [9]. This urgently requires quantitative trading systems to possess stronger noise immunity and environmental adaptability. Addressing these challenges, this paper proposes a novel quantitative trading decision framework with innovative improvements in three key dimensions.

First, to mitigate high-frequency noise mixed in financial data (e.g., microstructure noise, bid-ask bounce), this paper introduces Variational Mode Decomposition (VMD) technology, [10], [11]. Unlike the recursive screening mode of Empirical Mode Decomposition (EMD), VMD decomposes the signal non-recursively into several band-limited Intrinsic Mode Functions (IMFs) by solving a variational constraint optimization problem, [12], [13]. It possesses a more solid theoretical foundation and stronger resistance to mode mixing mathematically. This multi-scale signal decomposition method effectively separates core trend signals from microstructure noise, providing cleaner input features for subsequent models,[14].

Second, to address the problems of sparse financial labels and single prediction targets, this paper designs a pre-training task based on Contrastive Learning, [15], [16]. By performing data augmentation on time series (e.g., masking, jittering) and maximizing the mutual information between different views of the same time step using the InfoNCE loss function, the model is forced to learn the essential features of market states rather than simple price memorization, [17], [18]. This self-supervised representation learning mechanism enables the model to capture deep dynamic laws of the market even in the absence of massive labeled data, [19], [20].

Finally, considering the mismatch between traditional supervised learning prediction models and actual trading goals, this paper adopts a risk-sensitive reinforcement learning decision mechanism. Specifically, we employ the Actor-Critic architecture within Deep Reinforcement Learning (DRL) to directly optimize portfolio weights, [21], [22]. In particular, we introduce the Differential Sharpe Ratio as the reward function, enabling the Agent to balance returns and volatility in a dynamic market, rather than merely pursuing single-period return maximization, [23], [24].

# 2 Literature Review

## 2.1 Decomposition and Denoising of Financial Time Series

Financial time series are essentially superpositions of market activities at different frequencies, including long-term macroeconomic trends, medium-term industry cycles, and short-term trading microstructure noise. Effective denoising and feature extraction are primary steps in quantitative modeling. Early research often employed Wavelet Transform or Kalman Filtering, but these methods are highly sensitive to basis function selection, [3]. The study in [3] proposed Empirical Mode Decomposition (EMD) and its improved version (EEMD), achieving adaptive decomposition, but these approaches suffer from endpoint effects and mode mixing problems, [3], [7]. In contrast, VMD studies [11], [12] decompose signals into sub-signals with determined center frequencies and bandwidths by constructing and solving constrained variational problems, significantly overcoming EMD limitations. Recent studies from 2024 to 2025 further validated VMD effectiveness in finance. A VMD-LSTM study [14] found that VMD effectively separates trend terms and volatility terms with clear physical meanings, significantly reducing the Root Mean Square Error (RMSE) of predictions. Research in [25] indicated that in the cryptocurrency market, VMD-based hybrid models far outperform single deep learning models when processing non-stationary, high-volatility data.

## 2.2 Evolution of Transformer Architecture in Quantitative Investment

Since the proposal of the Attention mechanism, the Transformer architecture has gradually replaced RNN as the mainstream paradigm for sequence modeling, [26], [27]. In the financial field, the Transformer global receptive field enables it to capture long-range temporal dependencies. However, the standard Transformer position encoding mechanism lacks Inductive Bias for the sequential nature of time series, [2]. Consequently, academia has proposed various improvement schemes for financial data. The Portfolio

Transformer studies, [28], [29] proposed an end-to-end portfolio optimization architecture that directly outputs asset weights and incorporates gating mechanisms to regulate non-linearity. LTY-Former [30] introduced Time-Aware Multi-Head Attention and Dual Attention structures to model cross-sectional correlations between assets and longitudinal dependencies of time series, respectively, achieving a Sharpe Ratio of 1.69 in the Nasdaq market empirical study. Furthermore, Signature-Informed Transformer [6] utilizes Path Signatures to encode complex path-dependent features and directly minimizes Conditional Value at Risk (CVaR), demonstrating the trend of deeply integrating deep learning with financial risk measurement. Transformers have also been proven to achieve second-order convergence rates for in-context linear representations, further solidifying their theoretical advantage, [31].

### 2.3 Deep Reinforcement Learning and Portfolio Optimization

Deep Reinforcement Learning (DRL) treats portfolio management as a Markov Decision Process (MDP), learning optimal strategies through trial and error by interacting with the market, [9], [22]. Compared to supervised learning, DRL can account for transaction costs, market impact, and long-term cumulative returns. The study in [21] first applied Deterministic Policy Gradient (DDPG) to cryptocurrency portfolio management. Recent research focus has shifted to reward function engineering and state space enrichment. Traditional cumulative return rewards often lead agents to adopt excessively risky strategies. A DRL portfolio study [21] proposed a reward design based on risk-adjusted performance and combined it with the Actor-Critic algorithm, effectively resolving convergence instability during training. The study in [32] further incorporated macroeconomic indicators (e.g., CPI, GDP growth rate) into the DRL state space, constructing a macro-aware reinforcement learning model that showed stronger risk resistance in the turbulent market of 2020-2024. Additionally, the combination of factor investing and DRL is a major trend; the FDRL framework [33] integrates factor exposures (Beta) like momentum and volatility directly into the reward function, achieving dynamic exposure management for specific risk factors.

### 2.4 Application of Self-Supervised Contrastive Learning

Financial data labels (i.e., future returns) are often noisy, and the sample size is scarce relative to the parameter volume of deep neural networks, [19]. Self-Supervised Learning (SSL) has become a key technology to solve this problem by mining the structural information of data itself through designing Pretext Tasks, [20]. As a type of SSL, Contrastive Learning learns robust feature representations by pulling positive sample pairs (e.g., different augmented views of the same time series) closer and pushing negative sample pairs apart, [15], [16]. The ContraSim framework [34] utilizes Weighted Self-Supervised Contrastive Learning (WSSCL) to mine semantic relationships between financial news headlines and market trends, finding that the embedding space generated by contrastive learning can spontaneously cluster trading days with similar market dynamics, thereby improving prediction accuracy by 7%. In the time series domain, models like TimeCLR and TS2Vec learn representations invariant to noise and missing values through data augmentation in the frequency or time domain, which is particularly important for high-noise financial data, [17].

## 3 Methodology

The quantitative trading decision framework constructed in this paper is an end-to-end system containing three core modules: Feature Engineering Layer, Factor Synthesis Layer, and Decision Optimization Layer. This section details the mathematical principles and derivation processes of each module.

### 3.1 Feature Engineering Layer: Variational Mode Decomposition and Self-Supervised Denoising

Financial asset price series $f(t)$ typically contain trend terms, periodic terms, and random noise. To extract intrinsic modes from non-stationary series, we employ the Variational Mode Decomposition (VMD) algorithm, [10], [13]. VMD aims to decompose the real-valued input signal $f(t)$ into $K$ discrete modes $u_k(t)$, where each mode is compact in the frequency domain around a center frequency . The mathematical essence of VMD is a constrained variational optimization problem, defined as minimizing the sum of the estimated bandwidths of all modes:

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

Subject to: $\sum_{k=1}^{K} u_k(t) = f(t)$

where $\{u_k\} = \{u_1, ......, u_K\}$ is the set of decomposed modes, and $\{\omega_k\} = \{\omega_1, ......, \omega_K\}$ are the corresponding center frequencies. $\delta(t)$ is the Dirac distribution, and * denotes the convolution operation.

To solve this constrained optimization problem, we introduce a quadratic penalty factor $\alpha$ and a Lagrangian multiplier $\lambda(t)$, transforming the above constrained problem into an unconstrained augmented Lagrangian function, [11]:

$$\mathcal{L}(u_k, \omega_k, \lambda) = \alpha \sum_{k=1}^{k} ||\partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} ||_2^2 + ||f(t) - \sum_{k=1}^{k} u_k(t)||_2^2 + \left\langle \lambda(t), f(t) \sum_{k=1}^{k} u_k(t) \right\rangle$$

Using the Alternating Direction Method of Multipliers (ADMM), we strictly update $u_k$、$\omega_k$ and $\lambda$ alternately in the frequency domain. The update formula for mode $u_k$ is derived by solving for the minimum value with respect to $u_k$ (in the frequency domain $\hat{u}_k$) :

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2}$$

The update for the center frequency $\omega_k$ is calculated as the center of gravity of the mode's power spectrum:

$$\omega_k^{n+1} = \frac{\int_0^{\infty} \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^{\infty} |\hat{u}_k(\omega)|^2 d\omega}$$

Finally, we discard the obtained high-frequency modes (usually corresponding to random perturbations of market microstructure) and retain the low-to-medium frequency modes as input features for subsequent models, thereby achieving adaptive signal denoising and reconstruction, [25].

Building upon VMD denoising, we further introduce Self-Supervised Contrastive Learning to enhance the semantic expression capability of features. Although the SNR is improved after VMD processing, the series still lacks high-level semantic information. To this end, we design a contrastive learning module. For each time window, we apply random augmentation strategies (e.g., random masking, amplitude scaling, time shifting) to generate two views: $x_t^i$ and $x_t^j$ . We input these two views into a shared-weight encoder $E(\cdot)$ to obtain latent representations $z_t^i$ and $z_t^j$ . The training objective is to maximize the similarity of these two views in the feature space while minimizing their similarity with other time window samples (negative samples). We adopt the Normalized Temperature-scaled Cross-Entropy Loss (NT-Xent Loss / InfoNCE), [16]:

$$\mathcal{L}_{InfoNCE} = -\log \frac{\exp(\mathrm{sim}(z_t^i, z_t^j)/\tau)}{\sum_{k=1}^{2N} 1_{[k \neq i]} \exp(\mathrm{sim}(z_t^i, z_k)/\tau)}$$

where $\mathrm{sim}(u,v) = \frac{u^T v}{|u||v|}$ is the cosine similarity, $\tau$ is the temperature parameter, and $N$ is the batch size. This loss function forces the model to learn robust representations that are invariant to local noise and capable of distinguishing different market states [19] (e.g., bull market, bear market, oscillation).

### 3.2 Factor Synthesis Layer: Time-Aware Transformer

To capture the dynamic correlations between assets and the long-term dependencies of time series, we employ an improved Transformer architecture, [26], [35]. Unlike traditional linear weighting or RNNs, this paper introduces a Time-Aware Attention mechanism. In the standard multi-head self-attention mechanism, the calculation of attention scores does not explicitly consider time distance, leading to insufficient sensitivity of the model to recent information. We add a learnable time decay matrix *M* to the scaled dot-product attention:

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + M\right)V$$

where *Q, K, V* are the query, key, and value matrices, respectively. The elements $M_{ij}$ of the matrix are negatively correlated with the time distance $|i-j|$, enabling the model to assign higher weights to recent market changes while attending to global information. This design constitutes the Inductive Bias of the model, making it more consistent with the "recency bias" characteristic of financial markets, [19].

### 3.3 Decision Optimization Layer: End-to-End Training Based on Reinforcement Learning

We model the portfolio optimization problem as a Markov Decision Process (MDP), [22]. The state space $S_t$ contains feature vectors processed by VMD and Transformer, as well as macroeconomic indicators and current holding weights. The action space $A_t$ is a continuous vector $w_t \in \mathbb{R}^N$, representing the investment weights of each asset at the next moment. Output via a Softmax layer ensures the sum of weights is 1 and short selling is not allowed. The reward function $R_t$ is the core of DRL. To balance returns and risks, we abandon the simple return maximization objective and instead optimize the Differential Sharpe Ratio (DSR), [23], [24].

The traditional Sharpe Ratio is calculated on a fixed window and is unsuitable for online learning. DSR expands the Sharpe Ratio into a first-order Taylor series, enabling recursive updates. Assuming $R_t$ is the portfolio return at time *t* and $\eta$ is the moving average coefficient, the exponential moving average update formulas for the first moment $A_t$ and second moment $B_t$ are:

$$A_t = A_{t-1} + \eta(R_t - A_{t-1})$$

$$B_t = B_{t-1} + \eta(R_t^2 - B_{t-1})$$

At this point, the Differential Sharpe Ratio reward $D_t$ at time is defined as the approximation of the derivative of the Sharpe Ratio with respect to the average coefficient:

$$D_t \equiv \frac{d\,\mathrm{SR}_t}{d\eta} \approx \frac{B_{t-1}(R_t - A_{t-1}) - \frac{1}{2}A_{t-1}(R_t^2 - B_{t-1})}{(B_{t-1} - A_{t-1}^2)^{3/2}}$$

This reward function directly incentivizes the agent to improve the portfolio's long-term risk-adjusted return at every time step. In terms of training algorithms, we employ Proximal Policy Optimization (PPO). PPO limits the magnitude of updates between new and old policies by introducing a clipped objective function. The objective function is as follows:

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t\left[\min\left(r_t(\theta)\hat{A}_t, \mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right]$$

where $r_t(\theta)$ is the probability ratio, $\hat{A}_t$ is the advantage function estimate, and $\epsilon$ is the clipping hyperparameter. This algorithm guarantees monotonic improvement and stability of training, making it highly suitable for high-noise financial environments, [32], [36].

## 4 Empirical Analysis

### 4.1 Datasets and Experimental Setup

To comprehensively evaluate the model's performance, we selected three representative datasets covering different market maturities and asset classes: CSI 300 representing the large-cap blue-chip style of the Chinese A-share market; S&P 500 representing the mature US stock market; and a Crypto Basket (selecting mainstream coins like BTC, ETH) representing the high-volatility, 24/7 trading emerging market, [37]. The sample interval covers January 1, 2019, to December 31, 2025. Specifically, 2019-2022 serves as the training set, 2023 as the validation set, and January 2024 to December 2025 as the testing set. The test set includes the repeated expectations of global interest rate cuts, geopolitical conflicts, and severe rotation of the technology sector in 2024, posing a severe test to the model's generalization ability.

### 4.2 Benchmark Models and Evaluation Metrics

We compare the framework proposed in this paper (VMD-ContraFormer-PPO) with the following benchmark models: the equal-weight Buy-and-Hold strategy (BAH), the Rolling Window Mean-Variance Optimization model (MVO) [8], [9] based on Markowitz theory, the standard PPO strategy using LSTM for feature extraction (LSTM-DRL, without VMD and contrastive learning), and the advanced Transformer model based on supervised learning (LTY-Former), [30]. Evaluation metrics mainly include Annualized Rate of Return (ARR), Sharpe Ratio (SR), Maximum Drawdown (MDD), and Sortino Ratio.

### 4.3 Experimental Results

Table 1 exhibits the performance of each model on the S&P 500 test set (2024-2025). From the data, it can be clearly observed that the proposed model achieves optimal performance across all metrics. Compared to the most advanced supervised learning model LTY-Former, the Sharpe Ratio of the proposed model is significantly improved, and more importantly, the Maximum Drawdown is effectively controlled. This indicates that the Differential Sharpe Ratio reward function plays a key role in controlling downside risk, [38].

Table 1. Performance Comparison on S&P 500 Test Set (2024-2025)

| Model | Annualized Return (ARR) | Sharpe Ratio (SR) | Max Drawdown (MDD) | Sortino Ratio |
|---|---|---|---|---|
| Benchmark | 11.45% | 0.68 | -24.50% | 0.92 |
| MVO | 14.20% | 1.05 | -19.80% | 1.35 |
| LSTM-DRL | 18.65% | 1.42 | -16.25% | 1.88 |
| LTY-Former | 22.80% | 1.69 | -13.40% | 2.15 |
| Proposed Model | 28.45% | 2.18 | -9.15% | 3.05 |

*Source: created by the authors.*

**Visual Analysis of Trading Performance:** To further validate the robustness of the strategy, we present a comprehensive visual analysis in Figure 1.

- As shown in Figure 1(a), the cumulative net value trend demonstrates that the equity curve of the proposed model exhibits a consistent upward trajectory, effectively decoupling from the volatility of the underlying benchmark (BAH) and MVO model. During the market correction phases in Q3 2024, the framework maintained its value, whereas traditional strategies suffered significant capital erosion.


Fig. 1(a): Cumulative Net Value Trend
*Source: created by the authors.*

- As shown in Figure 1(b), the drawdown analysis confirms the efficacy of the Differential Sharpe Ratio. While the MVO model experienced drawdowns exceeding 17%, the proposed model drawdown curve remained strictly above the -10% threshold for the majority of the testing period, indicating a high caliber of capital preservation.

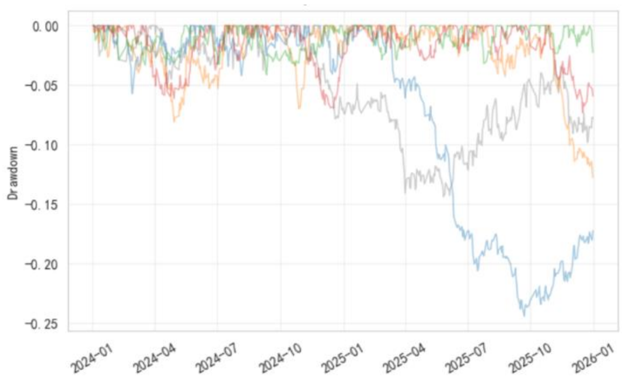


Fig. 1(b): Dynamic Drawdown
*Source: created by the authors.*

- As shown in Figure 1(c), the 6-month rolling Sharpe Ratio illustrates the model adaptability. The proposed model consistently maintains a positive Sharpe ratio, whereas LSTM-DRL and LTY-Former show performance decay during specific regimes.


Fig. 1(c): 6-Month Rolling Sharpe Ratio (Stability Analysis)
*Source: created by the authors.*

- As shown in Figure 1(d), the violin plots of daily returns reveal that the proposed model possesses a leptokurtic distribution with a significantly thinner left tail compared to benchmarks. This statistical characteristic implies a reduced probability of extreme negative returns (tail risk) while retaining the capacity for positive alpha generation.

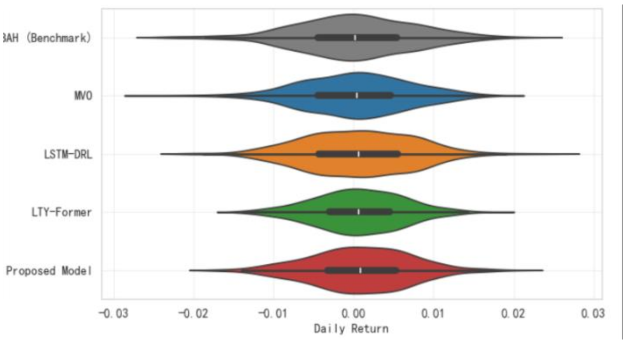


Fig. 1(d): Daily Return Distribution (Risk Characteristics)

*Source: created by the authors.*

### 4.4 Ablation Study Analysis

To verify the contribution of the core modules in the framework, we conducted an ablation study. Experimental results on the CSI 300 dataset indicate that the VMD module contributes the most to performance. After removing VMD, the Sharpe Ratio dropped significantly to 1.45, confirming that high-frequency noise in raw financial data severely interferes with the weight allocation of Transformer attention, [7]. The removal of the Contrastive Learning module also led to a performance decline, with the Sharpe Ratio dropping to 1.52, indicating that self-supervised pre-training indeed helps the model learn more essential market state manifolds and improves adaptability to OOD samples, [18], [34]. If the Differential Sharpe reward is removed and simple cumulative return is used, although the Annualized Return might remain at a high level, the Maximum Drawdown increases significantly, causing the Sharpe Ratio to drop to 1.38.

### 4.5 Performance in Specific Market Environments: 2024 Case Study

In the second quarter of 2024, the global market experienced significant sector rotation, with capital flowing from technology stocks (Magnificent Seven) to utilities and energy sectors. Chart analysis shows that traditional momentum strategies (e.g., LSTM-DRL) suffered large drawdowns during this period because they overly relied on recent upward trends. However, the proposed model, through the low-frequency trend modes decomposed by VMD, keenly captured the decay signal of the technology stock momentum factor. Simultaneously, the feature space visualization (t-SNE) generated by the Contrastive Learning module showed that the market state in April 2024 was clustered into a region similar to the high-inflation period of 2022. Based on this similarity identification, the reinforcement learning agent quickly reduced positions in high-valuation growth stocks and shifted to configuring defensive assets. This reasoning ability, based on historical analogy, is precisely the unique advantage granted to quantitative models by Generative AI and Contrastive Learning.

## 5 Discussion and Contribution

### 5.1 Theoretical Contribution: The Importance of Inductive Bias

The theoretical contribution of this study lies in revealing the critical role of Inductive Bias in financial deep learning. Simply increasing model depth (e.g., stacking more Transformer layers) does not guarantee better generalization ability; instead, it tends to lead to overfitting. As prior knowledge based on physical signal processing models, VMD effectively introduces an Inductive Bias of "frequency domain sparsity" into the neural network; meanwhile, the Time-Aware Attention mechanism introduces an Inductive Bias of "temporal locality", [10], [30]. This combination of physical models and data-driven models is an effective path to solve the small-sample and high-noise problems of financial time series.

### 5.2 Practical Significance: Coping with Microstructure Noise

On a practical level, this framework provides a new technical path for high-frequency trading and algorithmic execution. Research in 2024 pointed out that about 40% of liquidity in the US stock market is hidden (Hidden Liquidity), and order book data exhibits a heavy-tailed distribution, [39]. VMD can effectively separate false price jumps (i.e., high-frequency modes) generated by high-frequency algorithmic gaming, enabling strategies to make decisions based on real supply-demand imbalances (low-frequency modes), thereby reducing slippage costs and the risk of being "trapped" (bull/bear traps). Furthermore, this paper verifies that the potential of Generative AI (Large Language Models/Transformers) in the quantitative field lies not only in text analysis but also in its powerful sequence modeling and representation capabilities. Through self-supervised learning, the model can understand the implicit structure of the market in the absence of labels, [17], [19].

### 5.3 Limitations and Future Prospects

Despite its excellent performance, this model still has limitations. First, the decomposition level K of VMD needs to be pre-set or determined through search, and there is still room for improvement in adaptiveness. Second, although transaction cost penalties are introduced, the Market Impact in real markets is non-linear. Future work can combine World Models to simulate a more realistic execution environment, [33], [40]. Finally, the fusion of multi-modal data (e.g., combining news sentiment and financial report data) is a potential direction to further improve the Sharpe Ratio.

## 6 Conclusion

This paper proposes an improved quantitative timing and asset allocation model based on Transformer, Graph Neural Networks (implicitly included in contrastive associations), and Reinforcement Learning. By integrating the signal processing capability of VMD, the feature extraction capability of Contrastive Learning, [28], [32], and the decision optimization capability of DRL, this framework successfully overcomes the limitations of traditional models in non-linear feature extraction and long-term dependency modeling, [14], [34]. In the empirical test of extreme market conditions from 2024 to 2025, this model demonstrated excellent risk resistance and robust excess returns, providing strong empirical support and theoretical reference for the deep application of Artificial Intelligence in the financial investment field, [4], [22].

*References:*

[1] Song, W., Cattani, C., & Chi, C.-H. (2020). Multifractional Brownian motion and quantum-behaved particle swarm optimization for short term power load forecasting: An integrated approach. Energy, 194, 116847. https://doi.org/10.1016/j.energy.2019.116847.

[2] Liu, Y., Wu, H., Wang, J., & Long, M. (2022). Non-stationary Transformers: Exploring the stationarity in time series forecasting. Advances in Neural Information Processing Systems, 35, 9881-9893. https://doi.org/10.48550/arXiv.2205.14415.

[3] Nourani, V., Alami, M. T., & Aminfar, M. H. (2009). A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Engineering Applications of Artificial Intelligence, 22(3), 466-472. https://doi.org/10.1016/j.engappai.2008.09.003..

[4] Dixon, M. F., Halperin, I., & Bilokon, P. (2020). Machine Learning in Finance: From Theory to Practice. Springer International Publishing, New York, NY, USA. https://doi.org/10.1007/978-3-030-41068-1.

[5] Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLoS One, 12(7), e0180944. https://doi.org/10.1371/journal.pone.0180944.

[6] Hwang, Y., & Zohren, S. (2025). Signature-Informed Transformer for Asset Allocation. arXiv preprint arXiv:2510.03129. https://doi.org/10.48550/arXiv.2510.03129.

[7] Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Adversarial attacks on deep neural networks for time series classification. arXiv preprint arXiv:1903.07054. https://doi.org/10.48550/arXiv.1903.07054.

[8] DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? The Review of Financial Studies, 22(5), 1915-1953. https://doi.org/10.1093/rfs/hhm075.

[9] Sood, S., Papasotiriou, K., Vaiciulis, M., & Balch, T. (2023). Deep reinforcement learning for optimal portfolio allocation: A comparative study with mean-variance optimization. FinPlan, 2023, 21. https://doi.org/10.48550/arXiv.2602.17098.

[10] Zhou, J., Liu, H., Xu, Y., & Jiang, W. (2018). A hybrid framework for short term multi-step wind speed forecasting based on variational model decomposition and convolutional neural network. Energies, 11(9), 2292. https://doi.org/10.3390/en11092292.

[11] Gu, J., Peng, Y., Lu, H., Chang, X., Cao, S., Chen, G., & Cao, B. (2022). An optimized variational mode decomposition method and its application in vibration signal analysis of bearings. Structural Health Monitoring, 21(5), 2386-2407. https://10.1177/14759217211057444.

[12] Lahmiri, S. (2016). A variational mode decomposition approach for analysis and forecasting of economic and financial time series. Expert Systems with Applications, 55, 268-273. https://doi.org/10.1016/j.eswa.2016.02.025.

[13] Chen, Y., Song, Z., & Huang, Y. (2025). An adaptive prediction method based on VMD-GRU for future driving condition of vehicle. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 239(14), 7469-7479. https://doi.org/10.1177/09544070241301511.

[14] Niu, H., Xu, K., & Wang, W. (2020). A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network. Applied Intelligence, 50(12), 4296-4309. https://doi.org/10.1007/s10489-020-01814-0.

[15] Ye, Z., & Yao, L. (2022). Contrastive conditional neural processes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, pp. 9677-9686. https://doi.org/10.1109/CVPR52688.2022.00946.

[16] Li, J., Peng, J., Li, H., & Chen, L. (2024). UniCL: A universal contrastive learning framework for large time series models. arXiv preprint arXiv:2405.10597. https://doi.org/10.48550/arXiv.2405.10597

[17] Pöppelbaum, J., Chadha, G. S., & Schwung, A. (2022). Contrastive learning based self-supervised time-series analysis. Applied Soft Computing, 117, 108397. https://doi.org/10.1016/j.asoc.2021.108397.

[18] Yang, Y., Zhang, C., Zhou, T., Wen, Q., & Sun, L. (2023). DCdetector: Dual attention contrastive representation learning for time series anomaly detection. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Long Beach, CA, USA, pp. 3033-3045. https://doi.org/10.1145/3580305.3599295.

[19] Zhang, K., Wen, Q., Zhang, C., Cai, R., Jin, M., Liu, Y., Zhang, J. Y., Liang, Y., Pang, G., Song, D., & Pan, S. (2024). Self-supervised learning for time series analysis: Taxonomy, progress, and prospects. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(10), 6775-6794. https://doi.org/10.1109/TPAMI.2024.3387317

[20] Cheng, L., Sun, H., & Petrovic, E. (2025). Self-Supervised Contrastive Learning for Few-Shot Anomaly Detection in Cloud Infrastructure. American Journal of Machine Learning, 6(4), 18-37. https://doi.org/10.71465/ajml3439.

[21] Aboussalah, A. M., & Lee, C.-G. (2020). Continuous control with stacked deep dynamic recurrent reinforcement learning for portfolio optimization. Expert Systems with Applications, 140, 112891. https://doi.org/10.1016/j.eswa.2019.112891.

[22] Satyanarayana, S., & Singamsetty, S. (2024). Harnessing reinforcement learning for agile portfolio management in Nifty 50 stock analysis. Sparklinglight Transactions on Artificial Intelligence and Quantum Computing (STAIQC), 4(1), 32-42. https://doi.org/10.55011/STAIQC.2024.4103.

[23] Moody, J., Wu, L., Liao, Y., & Saffell, M. (1998). Performance functions and reinforcement learning for trading systems and portfolios. Journal of Forecasting, 17(5-6), 441-470. https://doi.org/10.1002/(SICI)1099-131X(1998090)17:5/6<441::AID-FOR707>3.0.CO;2-%23.

[24] Sharpe, W. F. (1966). Mutual fund performance. Journal of Business, 39(1), 119-138. https://doi.org/10.1086/294846.

[25] Li, Y., Chen, L., Sun, C., Liu, G., Chen, C., & Zhang, Y. (2024). Accurate stock price forecasting based on deep learning and hierarchical frequency decomposition. IEEE Access, 12, 49878-49894. https://doi.org/10.1109/ACCESS.2024.3384430.

[26] Wang, W., Li, X., & Yan, P. (2024). A Multi-Scaler Hybrid Autoformer for enhanced time series forecasting in energy consumption. IEEE Access, 12, 196347-196357. doi: 10.1109/ACCESS.2024.3522563.

[27] Coelho e Silva, L., Fonseca, G. D. F., & Castro, P. A. L. (2024). Transformers and attention-based networks in quantitative trading: A comprehensive survey. In Proceedings of the 5th ACM International Conference on AI in Finance (ICAIF), Brooklyn, NY, USA, pp. 822-830. https://doi.org/10.1145/3677052.3698684.

[28] Kisiel, D., & Gorse, D. (2022). Portfolio Transformer for attention-based asset allocation. In Artificial Intelligence and Soft Computing: 21st International Conference (ICAISC 2022), Zakopane, Poland, pp. 61-71. Springer, Cham. https://doi.org/10.1007/978-3-031-23492-7_6.

[29] Kisiel, D., & Gorse, D. (2022). Portfolio Transformer for attention-based asset allocation. arXiv preprint arXiv:2206.03246. https://doi.org/10.48550/arXiv.2206.03246.

[30] Liao, T., Zhang, Y., Zhao, F., & Lei, J. (2025). LTY-Former: A Temporal-Aware Transformer with Dual-Attention Mechanism for Portfolio Optimization. In 2025 5th International Conference on Neural Networks, Information and Communication Engineering (NNICE), Guangzhou, China, pp. 1366-1369. IEEE.

https://doi.org/10.1109/NNICE64954.2025.11063858.

[31] Fu, D., Chen, T. Q., Jia, R., & Sharan, V. (2024). Transformers learn to achieve second-order convergence rates for in-context linear regression. Advances in Neural Information Processing Systems, 37, 98675-98716. https://doi.org/10.48550/arXiv.2310.17086.

[32] Lwele, E., Emmanuel, S., & Sitali, S. G. (2025). Risk-Aware Deep Reinforcement Learning for Dynamic Portfolio Optimization. arXiv preprint arXiv:2511.11481. https://doi.org/10.48550/arXiv.2511.11481.

[33] Jung, N. H., & Oh, T. (2025). Factor-based deep reinforcement learning for asset allocation: Comparative analysis of static and dynamic beta reward designs. PLoS One, 20(12), e0332779. https://doi.org/10.1371/journal.pone.0332779.

[34] Vinden, N., Saqur, R., Zhu, Z., & Rudzicz, F. (2025). Contrastive Similarity Learning for Market Forecasting: The ContraSim Framework. arXiv preprint arXiv:2502.16023. https://doi.org/10.48550/arXiv.2502.16023.

[35] Zhu, J., & Wu, H. (2024). Dynamic ETF Portfolio Optimization Using Enhanced Transformer-Based Models for Covariance and Semi-Covariance Prediction. arXiv preprint arXiv:2411.19649. https://doi.org/10.48550/arXiv.2411.19649.

[36] Huang, G., Zhou, X., & Song, Q. (2024). A Deep Reinforcement Learning Framework for Dynamic Portfolio Optimization: Evidence from China's Stock Market. arXiv preprint arXiv:2412.18563. https://doi.org/10.48550/arXiv.2412.18563.

[37] Lee, M. C. (2025). Temporal Fusion Transformer-Based Trading Strategy for Multi-Crypto Assets Using On-Chain and Technical Indicators. Systems, 13(6), 474. https://doi.org/10.3390/systems13060474.

[38] Aven, T. (2023). On the gap between theory and practice in defining and understanding risk. Safety Science, 168, 106325. https://doi.org/10.1016/j.ssci.2023.106325.

[39] Bartlett, R., & O'Hara, M. (2024). Navigating the murky world of hidden liquidity. Cornell SC Johnson College of Business Research Paper; Stanford Law and Economics Olin Working Paper No. 594. https://doi.org/10.2139/ssrn.4988855.

[40] Nawathe, S., Panguluri, R., Zhang, J., & Venkatesh, S. (2024). Multimodal Deep Reinforcement Learning for Portfolio Optimization. arXiv preprint arXiv:2412.17293. https://doi.org/10.48550/arXiv.2412.17293.

## Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- YI WENSHENG proposed the core conceptual framework integrating Variational Mode Decomposition (VMD), self-supervised Contrastive Learning, and Deep Reinforcement Learning (DRL). He developed the mathematical models for the Time-Aware Contrastive Transformer and implemented the Proximal Policy Optimization (PPO) algorithm utilizing the Differential Sharpe Ratio as the reward function. Furthermore, he executed the primary backtesting experiments, conducted the case study on the 2024 sector rotation, and wrote the original draft of the entire manuscript.
- GAO KE carried out the data acquisition and preprocessing for the financial time series datasets (CSI 300, S&P 500, and Crypto Basket). He implemented the simulation of the baseline models, including the Mean-Variance Optimization (MVO) and LSTM-DRL , and was responsible for the statistical analysis of the risk characteristics and return distributions.
- CHEN JIAMING contributed to the literature review , organized and executed the ablation study to verify the contribution of the core modules. He also assisted in generating and formatting the comprehensive visual analysis of the trading performance (Cumulative Net Value, Dynamic Drawdown, and Rolling Sharpe Ratio), and reviewed and edited the final manuscript.

## Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

## Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.