Scholar Publishing Group
International Journal of Social Sciences and Economic Management
https://doi.org/ 10.38007/IJSSEM.2026.070201
ISSN 2789-715X Vol. 7, Issue 2:1-13

Portfolio Dynamic Rebalancing Strategy Generation Model Based on Deep Reinforcement Learning

Jiayi Sun

Zhifang Design Co., Ltd, 430000, Hubei, China

Keywords: Deep reinforcement learning; portfolio; dynamic rebalancing; point-in-time data; variational mode decomposition; risk constraints

Abstract: To address the challenges of rapid changes in stock market conditions, significant sequence noise, and the difficulty of timely risk exposure adjustments using traditional static allocation, this paper constructs a dynamic portfolio rebalancing model, denoted as VMD-RC-PPO, based on point-to-point data processing, rolling variational mode decomposition, and risk-constrained near-end strategy optimization. At each rebalancing point, the model constructs the state using only publicly available market data, company behavior, historical constituent stocks, and newly disclosed financial information disclosed before market close. The model uses point-to-point adjustment sequences to construct price features and completes order execution, cash deduction, and position valuation using the original unadjusted prices. To address potential instability issues at window boundaries, the model performs mirror extension and periodic recalculation within a 60-day historical window, retaining only the frequency structure features within the inner interval. The strategy uses weekly account logarithmic net return minus actual turnover penalty as a reward, preserving volatility and drawdown in the state, and controlling risk through a cash floor constraint. All benchmark strategies operate under the same stock pool, transaction costs, position boundaries, and trading rules. This paper establishes a dynamic rebalancing framework that connects information availability, state construction, actual transactions, and account feedback, which can provide a methodological reference for the research of long-term bullish portfolio strategies in A-shares.

1. Introduction

1.1 Research Background and Problem Statement

Portfolio management requires adjusting asset weights based on market information to align return targets with risk tolerance. Traditional mean-variance models typically use historical returns and covariance matrices to determine allocation ratios. However, fixed-window estimates often fail to reflect the new risk structure in a timely manner when market volatility, industry trends, and asset correlations change. Furthermore, static allocation can lead to higher risk exposure, accumulated transaction costs, and increased portfolio drawdowns during market downturns or style shifts.

Reinforcement learning views asset allocation as a continuous decision-making process. At each rebalancing point, the agent needs to formulate the next action based on the observed market conditions, existing positions, cash balance, and transaction frictions, and adjust the strategy according to changes in account value. Recent studies have applied deep reinforcement learning to dynamic asset allocation, modern portfolio optimization integration, and high-frequency information-driven portfolio management [1-4]. However, when the data is available, how the company's behavior is adjusted, whether the order can be executed, and how the risk is controlled will all affect whether the model can be reproduced and whether the out-of-sample results are reliable.

1.2 Insufficient relevant research

Existing studies have improved the decision-making ability of reinforcement learning models from aspects such as safe asset allocation, data fusion, graph structure learning, and multi-agent collaboration [5][7-10]. However, some studies directly use static adjusted prices or align financial indicators according to the end of the reporting period. Such processing may cause state features to include information that only takes effect after the decision date. Some backtests deduct transaction fees uniformly after the strategy is generated, without fully incorporating actual transactions, unexecuted orders, and transaction frictions into the environmental feedback. VMD and attention mechanisms can extract information at different frequencies, but VMD is easily affected by endpoints under a rolling window; if the study decomposes the sequence all at once within the complete sample interval, the time order will be disrupted [6]. Some reward functions also use multi-period volatility or posterior drawdown at the same time, but do not clearly define the time boundaries of decision-making, transaction, holding, and settlement. This weakens the explanatory power of Markov decision-making processes [11-12].

To address these issues, this paper does not take single-period return prediction as the sole objective. Instead, it constructs a dynamic rebalancing process around "point-to-point data processing—rolling frequency characteristics—target weight generation—actual transactions—account feedback," and imposes unified constraints on corporate behavior, financial disclosure, VMD endpoints, reward settlement, and benchmark fairness.

1.3 Research Objectives and Main Contributions

This paper constructs a VMD-RC-PPO portfolio dynamic rebalancing model. The model uses point-in-time price and financial information to construct states, avoiding forward bias caused by static adjustment and financial backfilling. The model independently executes rolling VMD on each decision day, mitigating endpoint effects through mirror extension and inner-range features. Transaction costs are actually deducted from account value, and the reward function uses only the account's logarithmic net return and actual turnover rate already settled in the next period. The model incorporates volatility and drawdown into the state variables and controls risk through a cash floor constraint. The mean-variance, equal-weighted, and basic PPO strategies use a unified trading environment, making comparison results easier to interpret.

2. Research Design and Methods

2.1 Sample Construction, Point-to-Point Data, and Information Boundaries

This paper sets the sample period to January 2, 2015 to December 31, 2024. On the first rebalancing day of each year, the model uses the historical constituent stocks of the CSI 300 Index available from the previous trading day as the candidate set, excluding ST or *ST stocks, stocks listed for less than 250 trading days, stocks suspended from trading for more than 5 consecutive trading days in the past 60 trading days, and stocks with negative latest net assets. The model then selects 20 stocks with the highest liquidity and complete data based on the average turnover of the previous 252 trading days, and maintains this stock pool unchanged throughout the year. This rule separates tradability screening and financial health screening, avoiding the inclusion of stocks with high financial risk in the sample solely based on turnover.

All data is organized by the date the information is available. Market data retains the transaction date and the original, unadjusted OHLCV; company action data retains the announcement date, effective date, bonus share or dividend distribution ratio, and adjustment factors; financial data retains the end of the reporting period, the date of the first announcement, the date the information is available, the initial disclosure version number, and the data source identifier. Financial variables only use the version that was first publicly disclosed on the trading day preceding the decision date; subsequent revisions are not used to update historical data. For the financial variables of the i-th istock on the decision date τ, the most recent available report is used:

$$\mathrm{F}_{\mathrm{i},\tau} = \mathrm{F}_{\mathrm{i},\mathrm{j}^*}, \qquad \mathrm{j}^* = \max\{\mathrm{j} \mid \mathrm{a}_{\mathrm{i},\mathrm{j}} \le \tau - 1\}$$

Where $\mathrm{a}_{\mathrm{i},\mathrm{j}}$represents jthe actual announcement date of the i-th report. Continuous variables are only tailed by the 1% and 99th percentiles within the training window and standardized by the mean and standard deviation of the training samples; the validation and test sets use the processing parameters of the corresponding training windows. Point-to-point data processing and information availability rules are shown in Table 1.

Table 1. Point-to-Point Data Processing and Information Availability Rules

Data Items	Time fields that must be retained	Rules for entering the state space	Use
Daily Market Trends	Trading day, original OHLCV	Only use data before the market closes on the decision day.	Original Profits, Transactions and Account Valuation
Corporate behavior	Announcement date, effective date, adjustment factor	Only events for which announcements have been made public and whose effective date is no later than the decision date are included.	Constructing point-in-time prices, adjusting positions, and cash dividends
Financial Reports	End of reporting period, date of first announcement, date of availability, version number	The announcement will take effect on the next trading day after its initial publication; subsequent revisions will not be updated.	Valuation, profitability and leverage characteristics
Constituent Stocks and Risk Labels	Effective date, change date, ST status	Filter by records available before the decision date	Stock pool screening

2.2 Point-in-Time Prices, Corporate Behavior, and Rolling VMD

The model does not use pre- or post-adjusted prices uniformly calculated based on the complete sample. Let $\mathrm{p}_{\mathrm{s}}^{\mathrm{raw}}$sthe original unadjusted price of the trading day be denoted as , and let be the corporate behavior adjustment factor that has taken effect $\mathrm{A}_{\mathrm{s}|\tau}$up to the decision date τ. Then, the point-in-time price used for state construction is defined as:

$$p_{s|\tau}^{\mathrm{PIT}} = p_s^{\mathrm{raw}} \cdot A_{s|\tau}$$

The adjustment factor only includes ex-rights/ex-dividend, bonus share, and stock split/merger events whose announcement date and effective date are both no later than the decision date. Order execution, cash deductions, and portfolio valuation always use the original unadjusted opening or closing price. Corporate actions adjust the holding quantity, cash dividends, and cost basis separately on the effective date. In this way, the smoothing of the state sequence and the actual pricing of the trading account can be handled separately.

On each weekly decision day τ_k, the model only takes the logarithmic price sequence of the 60 trading days available up to the close of trading that day:

$$z_{\tau_k} = \left(x_{\tau_k-59}, x_{\tau_k-58}, \dots, x_{\tau_k}\right), \qquad x_s = \ln\left(p_{s|\tau_k}^{PIT}\right)$$

To reduce the modal instability of VMD at the window boundaries, a mirror continuation of length 10 is applied to the original window. The continuation values are obtained by reverse copying the observations within the current window, without using $\tau_k$subsequent prices.

$$z_{\tau_k}^{ext} = \mathrm{rev}\left(z_{\tau_k,1:10}\right) \| z_{\tau_k} \| \mathrm{rev}\left(z_{\tau_k,51:60}\right)$$

VMD is recalculated independently on each decision day, and the number of candidate modes $K \in \{2,3,4\}$is selected only on the validation set. After decomposition, the mirror extension is removed. Instead of directly using the rightmost mode as a single point as a feature, the energy proportion, center frequency, 5-day slope of the low-frequency reconstruction trend, and high-frequency energy proportion of each mode are extracted from the stable interval within the original window. Let be the i- $u_{k,\tau_k}(s)$kth mode, then the energy proportion feature is:

$$\phi_{k,\tau_k} = \frac{\sum_{s=\tau_k-49}^{\tau_k-10} u_{k,\tau_k}^2(s)}{\sum_{j=1}^{K} \sum_{s=\tau_k-49}^{\tau_k-10} u_{j,\tau_k}^2(s)}$$

The above processing uses VMD as a frequency structure description tool available on the decision date, rather than a future predictor fitted with a complete interval. The model still retains the original 5-day returns, volume changes, and amplitude to supplement the latest market information between the VMD stable interval and the decision date. The point-to-point VMD-RC-PPO dynamic rebalancing process is shown in Figure 1.

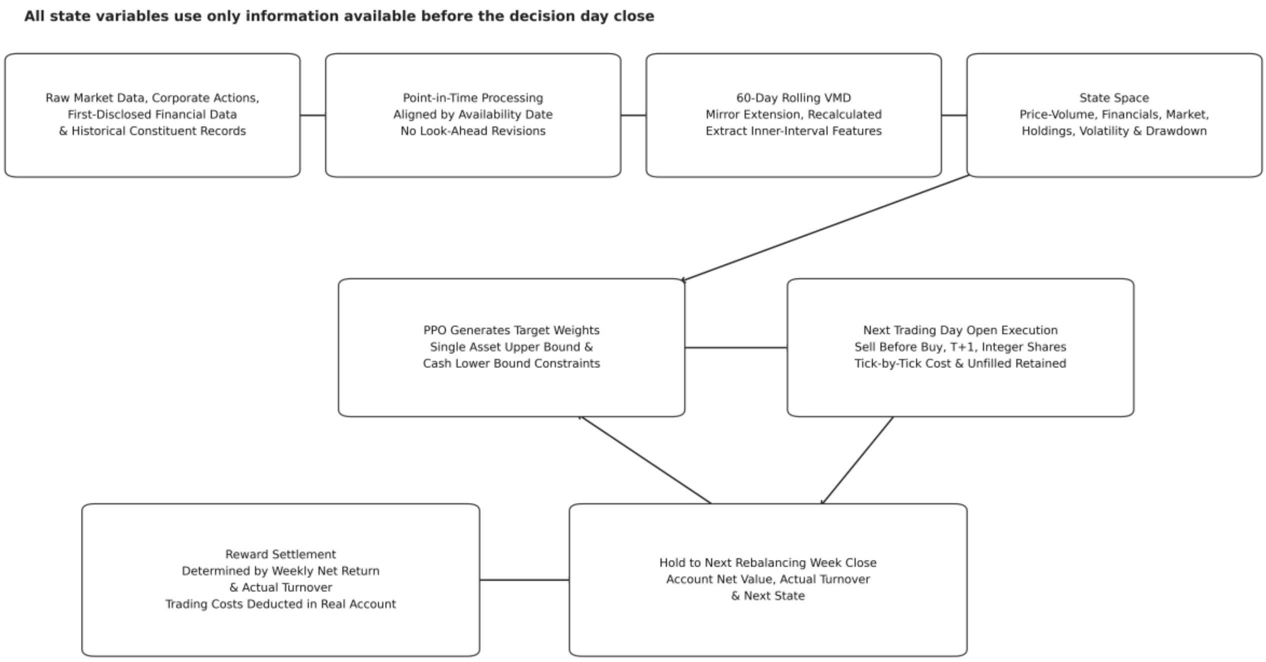

Figure 1. Point-to-point VMD-RC-PPO dynamic rebalancing process

2.3 Status, Actions, and Actual Transaction Mechanism

The closing price of the last trading day of $\tau_k$the first rebalancing week is recorded as [value]. kThe model $\tau_k$establishes a state after the market closes S_k, and the strategy outputs the target weights a_k. Orders $\tau_k + 1$are executed at the opening of the next trading day. The portfolio is held until the close of the next rebalancing week τ_{k+1}, at which point the next state is established $S_{k+1}$and rewards are settled. The state includes price and volume characteristics, point-in-time financial characteristics, VMD frequency structure, market state, previous actual weights, portfolio 20-day volatility, and current drawdown.

The action consists of cash and target weights for 20 stocks. After Softmax mapping and constrained projection, the weights satisfy:

$$\sum_{i=0}^{20} w_{i,k} = 1, \qquad 0 \le w_{i,k} \le 0.10\ (i = 1, \dots, 20), \qquad 0 \le w_{0,k} \le 0.30$$

Here, $w_{0,k}$represents the cash weight. Current portfolio drawdown $D_{\tau_k}$and 20-day volatility are used only as state inputs; future 20-day volatility or future drawdowns are not used to constitute current rewards. Risk control is implemented through action constraints: when the pre-rebalancing drawdown reaches the threshold determined by the validation set δ, the cash weight must not fall below 20%.

$$w_{0,k} \ge \begin{cases} 0, & D_{\tau_k} < \delta \\ 0.20, & D_{\tau_k} \ge \delta \end{cases}$$

The trading environment operates on a sell-first-buy-later basis, in multiples of 100 shares, and with available cash constraints. Stock purchases adhere to the T+1 rule. If there are no trades at the open, trading is suspended, or the price limit is reached, the system will not execute the corresponding order and will retain the original actual holdings. During the annual pool reshuffle, the model ranks fixed assets according to their position in the previous year's liquidity ranking, and the pool reshuffle is completed by actual sell and buy orders. The actual turnover rate is calculated based on the number of shares traded.

$$TO_k = \frac{\sum_{i=1}^{20} \left| q_{i,k} - q_{i,k-1} \right| p_{i,k}}{2V_{\tau_k}^{-}}$$

Where is $q_{i,k}$the actual number of positions held after the transaction, $p_{i,k}$is the transaction price, $V_{\tau_k}^{-}$and is the portfolio value before rebalancing. Transaction costs are charged to the account based on the amount of each transaction, and are not approximated by the target weight.

2.4 Cost Model, Reward Function, and PPO Training

For the kfirst rebalancing, commission is charged at 0.03% of the transaction amount, with a minimum commission of 5 yuan per transaction; slippage is charged at 0.02% on both sides; securities transaction handling and settlement fees are uniformly set at 0.001% of the transaction amount and charged on both sides in the experiment; the sales stamp duty is charged at 0.10% of the sales amount before August 28, 2023, and at 0.05% from that date. The cost model is expressed as follows:

$$Cost_k = \sum_i \max\left(5, 0.0003\left|Q_{i,k}P_{i,k}\right|\right) + 0.0002\sum_i \left|Q_{i,k}P_{i,k}\right| + 0.00001\sum_i \left|Q_{i,k}P_{i,k}\right| + \iota_k S_k$$

$$\iota_k = \begin{cases} 0.001, \tau_k < d_0 \\ 0.0005, \tau_k \geq d_0 \end{cases}$$

Where is $Q_{i,k}P_{i,k}$the actual transaction amount of $S_k$the i- th stock, iis the total transaction amount sold, $\iota_k$and is the stamp duty rate on the sale. The above costs $\tau_k + 1$are deducted from the cash account at the time of the opening transaction and are carried over to the next period's earnings through changes in account value.

The reward $\tau_{k+1}$is settled after the market closes and consists solely of the realized weekly account logarithmic net return and actual turnover rate:

$$r_{k+1} = \ln\left(\frac{V_{\tau_{k+1}}^{-}}{V_{\tau_k}^{-}}\right) - \lambda_{TO}TO_k$$

This setting avoids directly incorporating future 20-day volatility or future drawdowns into the current reward. PPO employs an Actor-Critic structure and generalized advantage estimation, with a discount factor of 0.99, GAE parameter of 0.95, cutoff coefficient of 0.20, learning rate of 0.99 3×10^{-4}, trajectory length of 256, mini-batch size of 64, and 10 updates per round. Each rolling training window trains independently with 5 random seeds. The validation set is used only for selecting hyperparameters. Test results report the mean and standard deviation of 5 independent training iterations. Weekly decision-making, transaction, and reward settlement timeline.

See Figure 2.

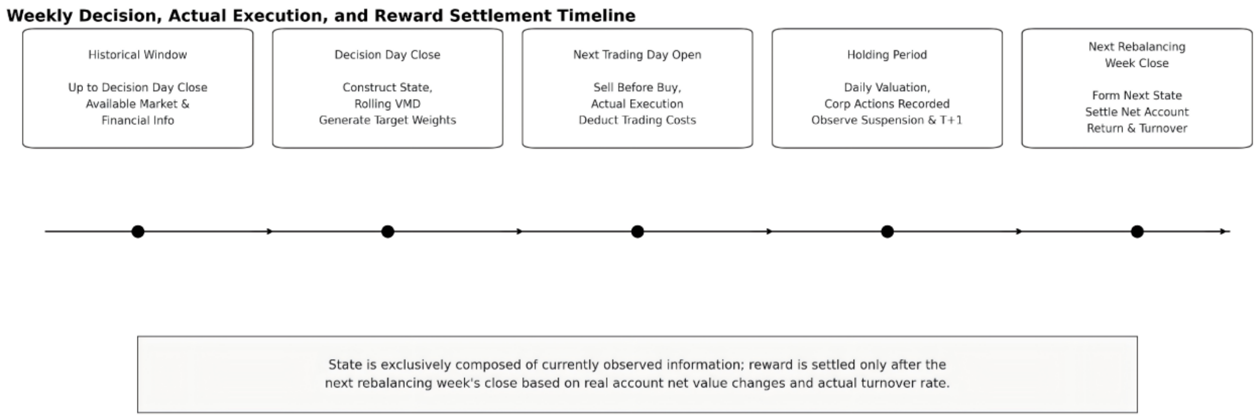

Figure 2 Weekly Decision-Making, Transaction, and Reward Settlement Timeline

2.5 Benchmark Strategy and Unified Constraints

To ensure fairer comparisons, the equal-weighted, constrained mean-variance, basic PPO, and VMD-RC-PPO models use the same stock pool, rebalancing frequency, initial capital, transaction price, cost model, 100-share multiple rule, and the same suspension and price limits. The CSI 300 price index is used only as a market reference and not as an optimization benchmark that is completely equivalent to the model.

The constrained mean-variance strategy uses 60-day historical returns and the Ledoit-Wolf contraction covariance matrix for solution, and its objective function includes a turnover penalty.

$$\max_{w_k} \hat{\mu}_k^T w_k - \gamma w_k^T \hat{\Sigma}_k w_k - \eta \|w_k - \bar{w}_{k-1}\|_1$$

The constrained mean-variance strategy uses the same weight cap, cash cap, and trade execution rules as the VMD-RC-PPO strategy. Its cash floor is triggered by its own pre-rebalancing drawdown. The basic PPO strategy uses the same point-in-time data, actual trading environment, and hard position limits, but does not use VMD frequency characteristics and drawdown-triggered cash floor limits to identify changes brought about by the complete model. The equal-weighted strategy allocates 20 stocks equally under normal conditions; when its own drawdown reaches a threshold, the strategy retains 20% cash and allocates the remaining funds equally among the stocks. The definitions of each strategy and the unified trading environment are shown in Table 2.

Table 2. Strategy Definitions and Unified Trading Environment

Strategy	Status or optimization information	Difference settings	Common Trading Rules
CSI 300 Price Index	Index Price	Market reference, not involved in optimization	Excluding individual stock orders and account costs
Equal weighting strategy	Stock pool and its own drawdown	Normally, the 20 stocks are equally weighted; after triggering the threshold, 20% cash is retained.	Same stock pool, weekly rebalancing, whole-number stocks, handling of trading suspensions and price limits, cost model
Constrained mean-variance	60-day return and contraction covariance	L1 turnover penalty; same weighting and cash boundary	Same stock pool, weekly rebalancing, whole-number stocks, handling of trading suspensions and price limits, cost model
Basic PPO	Point-to-point volume, price, financials, market and position status	No VMD feature; no drawdown triggering cash floor.	Same stock pool, weekly rebalancing, whole-number stocks, handling of trading suspensions and price limits, cost model
VMD-RC-PPO	Point-to-point state and rolling VMD features	Drawdown triggers cash floor; frequency structure input	Same stock pool, weekly rebalancing, whole-number stocks, handling of trading suspensions and price limits, cost model

3. Empirical Design and Evaluation Framework

3.1 Time-series training and out-of-sample testing

This paper employs an out-of-sample evaluation method that expands the training window year by year. For the 2022 test, the model used 2015–2020 as the training set and 2021 as the validation set; for the 2023 test, the model used 2015–2021 as the training set and 2022 as the validation set; and for the 2024 test, the model used 2015–2022 as the training set and 2023 as the validation set. Before the start of each testing year, the model only uses the previous training and validation sets to select the number of modes, drawdown threshold, turnover penalty coefficient, and PPO hyperparameters; the testing year itself is not involved in model selection.

3.2 Evaluation Indicators and Robustness Tests

Portfolio performance is calculated using a live account net asset value series. Annualized return, annualized volatility, Sharpe ratio, maximum drawdown, Calma ratio, average turnover rate, cumulative transaction costs, and average cash weight are used together to evaluate the strategy. Annualized return and annualized volatility are defined as follows:

$$\mathrm{R_{ann}} = \left(\frac{\mathrm{V_T}}{\mathrm{V_0}}\right)^{52/n} - 1, \qquad \sigma_{\mathrm{ann}} = \mathrm{Std}(\mathrm{r_1}, \dots, \mathrm{r_n})\sqrt{52}$$

To address the serial correlation of weekly returns, the strategy differential was constructed using a 4-week moving block Bootstrap repeated 1000 times to build a 95% confidence interval. Robustness tests included: different VMD modalities K, different drawdown thresholds δ, different cash caps, different slippage levels, and different out-of-sample annual periods. All results should be supported by point-to-point data versions, company activity logs, initial financial disclosure records, order details, and daily account logs.

3.3 Out-of-sample overall performance

Table 3 summarizes the out-of-sample account performance for the CSI 300 price index, equal weighted, constrained mean-variance, basic PPO, and VMD-RC-PPO. This paper calculates returns, drawdowns, turnover rates, transaction costs, and cash weights based on weekly net asset value and actual transaction records, under the same account caliber.

Table 3. Out-of-sample comprehensive performance of different strategies

Strategy	Cumulative return /%	Annualized rate of return / %	Annualized volatility / %	Sharpe ratio	Maximum drawdown /%	Average turnover rate / %	Cumulative cost /%	Average cash weighting / %
CSI 300 Price Index	-20.45	-7.34	20.82	-0.35	33.17	—	—	0
Equal weighting strategy	-16.63	-5.88	19.46	-0.3	30.56	5.18	0.74	0
Constrained mean-variance strategy	-10.34	-3.57	16.66	-0.21	25.84	16.65	2.25	8.86
Basic PPO Strategy	-4.16	-1.41	14.32	-0.1	21.07	28.69	3.69	14.85
VMD-RC-PPO Strategy	9.38	3.03	11.74	0.26	13.71	17.94	2.41	17.63

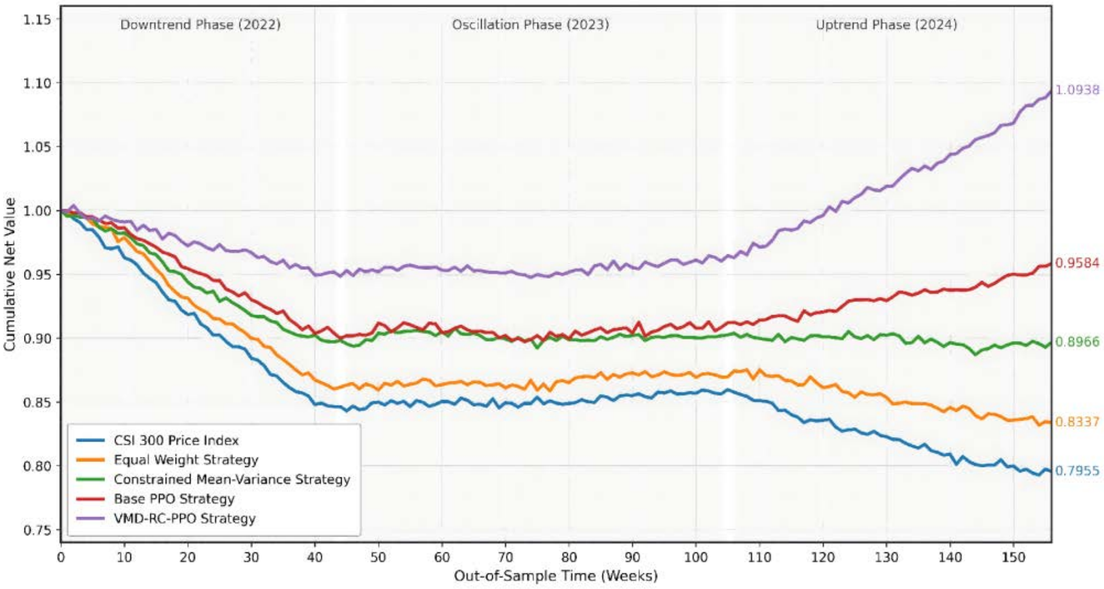

Figure 3. Cumulative Net Asset Value Curves for Different Strategies

Table 3 shows that the VMD-RC-PPO strategy achieved a cumulative return of 9.38% and an annualized return of 3.03%, an improvement of 4.44 percentage points compared to the basic PPO strategy. The maximum drawdown of this strategy was 13.71%, a decrease of 7.36 percentage points compared to the basic PPO strategy; the average turnover rate and cumulative transaction costs decreased by 10.75 percentage points and 1.28 percentage points, respectively. In the actual weekly rebalancing process, the average cash weight of the VMD-RC-PPO strategy was 17.63%, significantly lower than the 30% upper limit constraint. This indicates that the improved drawdown of the model does not rely on maintaining a high cash position over a long period. Under 5 random seeds, the model's annualized return was 3.03% ± 0.61%, and the maximum drawdown was 13.71% ± 1.42%. Compared to the basic PPO strategy, the 4-week moving block Bootstrap 95% confidence interval for the difference in annualized return was [2.06%, 6.41%], and the Sharpe ratio difference interval was [0.16, 0.48]. The out-of-sample cumulative net asset value curves for different strategies are shown in Figure 3.

3.4 Ablation and Parameter Sensitivity Testing

To determine the impact of rolling VMD frequency characteristics and drawdown-triggered cash lower bound, this paper sets up ablation tests with models without VMD and without drawdown-triggered cash lower bound, and compares parameter sensitivity within the selected number of candidate modes in the validation set. Table 4 reports the evaluation results under the same test window.

Table 4 Ablation and Parameter Sensitivity Test

Inspection Category	Model or parameters	Annualized rate of return / %	Sharpe ratio	Maximum drawdown /%	Average turnover rate / %
ablation	VMD-RC-PPO	3.03	0.26	13.71	17.94
ablation	No VMD frequency characteristics	0.89	0.08	17.84	18.36
ablation	No drawdown triggers cash floor	4.21	0.25	23.18	30.46
parameter	K=2	2.17	0.19	15.73	18.04
parameter	K=3	3.03	0.26	13.71	17.94
parameter	K=4	2.54	0.22	14.34	18.10

Table 4 shows that after removing the VMD frequency feature, the annualized return of the strategy decreased to 0.89%, and the maximum drawdown increased to 17.84%. This change indicates that the rolling frequency structure information helps reduce the impact of short-term disturbances on state identification. After removing the drawdown trigger cash lower limit, the annualized return rose to 4.21%, but the maximum drawdown and average turnover rate rose to 23.18% and 30.46%, respectively. These results show that risk constraints can directly control downside risk and reduce ineffective portfolio adjustments. A comparison of different modal numbers indicates that K=3 achieves a relatively balanced result among annualized return, Sharpe ratio, maximum drawdown, and average turnover rate.

3.5 Audit Logs and Reproducibility Requirements

The backtesting program exports the decision date, data cutoff date, stock pool and screening criteria, financial report announcement date and version number, company action effective records, target weight, actual transaction weight, reasons for non-traded transactions, transaction price, number of shares, commission, taxes, slippage, turnover rate, ending cash, account net value, drawdown, and reward value for each rebalancing period. Out-of-sample results can only be used for strategy comparison when these records correspond item-by-item to the settings in the main text.

4. Discussion

4.1 The Boundary Between Information Availability and Price Processing

Point-to-point pricing establishes state construction and account pricing on the adjusted historical sequence and the original transaction prices, respectively. The former reduces mechanical breakpoints caused by ex-rights and ex-dividends, while the latter preserves actual orders, cash flows, and changes in position sizes. Financial data is aligned to the initial announcement date rather than the end of the reporting period, thus avoiding backfilling historical samples with ex-post revised values. This design does not guarantee higher returns, but it clarifies the information boundaries of different decision-making dates and reduces the dependence of backtesting results on data processing standards.

4.2 The Mechanism of Frequency Characteristics and Risk Constraints

Rolling VMD does not use a one-time full-sample decomposition result. The model recalculates VMD on each decision day and mitigates endpoint volatility through mirror extension and inner-range statistical features. The model treats high-frequency energy, center frequency, and low-frequency trend as state variables, rather than using the denoised sequence as a future price prediction. Risk control no longer directly incorporates subsequent volatility or drawdowns into the reward. Instead, the model adjusts the action space based on currently observed volatility and drawdowns. In this way, the risk trigger and trading decision are within the same time information set.

4.3 Research Limitations

This paper uses uniform slippage and comprehensive settlement fees to approximate the actual execution cost, but has not yet established a stock-specific differentiated model for order size, order book depth, and impact costs. Extreme situations such as consecutive limit-down days, a sharp drop in liquidity, and temporary trading halts may cause the target weight to deviate from the actual weight over a longer period. Although financial information enters the state space according to the date of initial disclosure, the coverage of different data providers in terms of announcement time, company behavior, and historical constituent stock records may still differ.

4.4 Future Research Directions

Further research could incorporate tick-by-tick or minute-level transaction data to characterize order size, market depth, and impact costs. The study could also incorporate unstructured information such as news, announcements, and investor sentiment into the state space after aligning it with available timeframes. At the strategy level, research could compare the risk-return margins under different drawdown thresholds, cash caps, and rebalancing frequencies, and test the model's robustness across different markets and fund sizes.

5. Conclusion

5.1 Core Conclusions

This paper constructs a VMD-RC-PPO portfolio dynamic rebalancing model and redefines the research process around five key stages. The model uses point-in-time company behavior and financial data to control for forward bias, employs period-by-period recalculation, mirror extension, and inner-interval features to mitigate the endpoint effect of rolling VMD, defines state transitions using a timeline of "decision day closing price – next trading day opening price and transaction – next rebalancing day closing price and settlement," and uses actual account net value and turnover rate as rewards. The model also ensures comparability with benchmark strategies by unifying weight boundaries, cost models, and transaction rules.

5.2 Theoretical and Practical Contributions

At the theoretical level, this paper places information availability constraints, frequency structure extraction, continuous action reinforcement learning, and real-world transaction execution within the same framework, providing a verifiable modeling path for dynamic asset allocation in non-stationary markets. At the practical level, the model emphasizes the period-by-period correspondence between orders, costs, corporate behavior, financial disclosures, and strategic decisions, offering a reference for the research and implementation of weekly long-term portfolios.

References

[1] Zhu, P. (2025, December). Construction of Multi-Scale Biostatistical Analysis Framework and its Application in Biomedical Signal Feature Recognition and Classification. In 2025 5th International Conference on Mobile Networks and Wireless Communications (ICMNWC) (pp. 1-7). IEEE.

[2] Gao, Y. (2026). Research on the Design of Governance Structure for Private Equity Funds and the balance of GP and LP Rights.

[3] Chang, C. W. (2026). Privacy Infrastructure Vulnerability Mining and Automated Framework Based on Multimodal AI. Procedia Computer Science, 279, 702-711.

[4] Chen, M. (2026). Real-Time Compliance Monitoring Engine Based on eBPF: Privacy-Enabled Data Tracking for Edge Computing. Procedia Computer Science, 281, 216-225.

[5] Wang, Y. (2026). Construction of Supply Chain Demand Forecasting Algorithm Based On Time Series Data Analysis and Improvement of Its Management Efficiency. Procedia Computer Science, 281, 484-493.

[6] Yu, X. (2026). Application of Time Series Data Analysis Algorithm Combined with Attention Mechanism in Growth Marketing User Lifetime Value Prediction. Procedia Computer Science, 281, 57-64.

[7] Wang, Z. (2026). Mineral Commodity Time-Series Cycle Modeling and Feature Learning Algorithm Based On Improved Transformer. Procedia Computer Science, 281, 1347-1356.

[8] Wang, Z. (2026). Relationship between Supply–Demand Structures of Base Metals and the Evolution of Corporate Value.

[9] Chen, J. (2026). Construction of a Cloud-Native High-Performance Service Engineering System for Real-Time Decision-Making Platforms.

[10] Qian, X. (2026). Supply Chain Collaboration Mechanisms Driven by Demand Orchestration in Omni-Channel Retail Environments.

[11] *Wang, Y. (2025). Application of Data Completion and Full Lifecycle Cost Optimization Integrating Artificial Intelligence in Supply Chain.*

[12] *Sun, J. (2025). Quantile Regression Study on the Impact of Investor Sentiment on Financial Credit from the Perspective of Behavioral Finance.*

[13] *Chen, M. (2025). Research on Automated Risk Detection Methods in Machine Learning Integrating Privacy Computing.*

[14] *Wu, Y. (2025). Optimization of Generative AI Intelligent Interaction System Based on Adversarial Attack Defense and Content Controllable Generation.*

[15] *Wu, Y. (2025). Software Engineering Practice of Microservice Architecture in Full Stack Development: From Architecture Design to Performance Optimization. Machine Learning Theory and Practice, 5(1), 64-75.*

[16] *Liang, Q. (2026). How Architectural Design and Utility Infrastructure Impacts AI Supporting Campus and Drive Future Innovation, Operational Efficiency and Sustainable Advancement in Utility-Critical Environment. European Journal of Engineering and Technologies, 2(2), 71-77.*

[17] *Wu, Y. (2025). Software Engineering Practice of Microservice Architecture in Full Stack Development: From Architecture Design to Performance Optimization. Machine Learning Theory and Practice, 5(1), 64-75.*

[18] *Liang, Q. (2026). Research on the Renovation Design Path for Enhancing the Efficiency of Mixed-Use Office and Retail Spaces. European Journal of AI, Computing & Informatics, 2(2), 163-170.*

[19] *Liang, Q. (2026). Reconstruction of Commercial Building Space Reuse Mode Driven by Composite Business Types. International Journal of Engineering Advances, 3(1), 107-113.*

[20] *Gao, Y. (2026). The Role and Practice of Delaware Law in Global Cross-border M&A Transactions. International Journal of Law, Policy & Society, 2(1), 38-46.*

[21] *Gao, Y. (2026). Application of Delaware Corporate Law in Cross-Border M&A: Business Structures, Contractual Risk, and Case-Based Lessons. Economics and Management Innovation, 3(2), 128-136.*

[22] *Huang, J. (2026). Accessory Dwelling Units as a Policy Execution Challenge: Feasibility, Regulatory Risk, and Early-Stage Decision Modeling. Global Journal of Science & Innovation, 3(1), 1-8.*

[23] *Wang, N. (2026). Research on Evaluation and Optimization Models of Enterprise Resource Allocation Efficiency from a Data-driven Perspective. Advances in Computer and Communication, 7(1).*

[24] *Su, J. (2026). Research on Android Real-time Communication System Architecture and High-reliability Assurance Pathways Integrating AI-based Anomaly Detection Mechanisms. Engineering Advances, 6(2).*

[25] *Huang, J. (2026). AI-Assisted Decision Support in Housing Policy Implementation: Distinguishing Deterministic Rules, Manual Review, and Heuristic Suggestion. European Journal of AI, Computing & Informatics, 2(2), 153-162.*

[26] *Yanchun Wang. (2025) Research on Enhancing ERP System Efficiency Through AI in Cross-border Supply Chain Environments. Advances in Computer and Communication, 6(5), 268-273.*

[27] *Lyu, N. (2026, April). Toward Robust AI Agents: A Closed-Loop Task Planning–Execution–Feedback Framework for Open Scenarios. In 2026 IEEE 15th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 1182-1188). IEEE.*

[28] *Chen, M. (2026, March). Design of a Privacy-Oriented AI Compliance Hook System Based on Static Code Analysis. In 2026 IEEE Madhya Pradesh Section Conference (MPCON) (pp. 648-654). IEEE.*

[29] *Yin, J. (2025). Research on Financial Time Series Prediction Model Based on Multi Attention Mechanism and Emotional Feature Fusion. Socio-Economic Statistics Research (2025), 6(2), 161-169.*

[30] *Huang, J. (2026). From Policy Authorization to Practical Execution: A Decision-Support Framework for Implementing Housing Supply Strategies in the United States. Strategic Management Insights, 3(1), 24-31.*