Scholar Publishing Group
International Journal of Business Management and Economics and Trade
https://doi.org/10.38007/IJBMET.2025.060123
ISSN 2790-0959 Vol. 6, Issue 1:221-228

Data-driven Optimization of Capital Market Trading Strategies and Risk Management

Yilin Fu

Brandeis University, Brandeis International Business School, Waltham, Massachusetts, 02453, USA

Keywords: Data-driven trading strategy, Capital markets, Risk management, Optimization algorithm, Genetic algorithm

Abstract: This paper is an in-depth study of the data-based capital market trading strategy optimization and risk management methods. With the help of data acquisition, preprocessing, feature extraction and model construction, supplemented by optimization algorithms to enhance the performance of trading strategies. In this paper, the retracement rate, Sharpe ratio and other evaluation criteria of trading strategies are analyzed in detail, and a data-driven risk management framework is constructed, which involves the understanding of market risk, liquidity risk and derivative hedging strategies. The study found that data-based trading strategies significantly improved decision making in capital markets and enhanced risk management capabilities.

1. Introduction

Faced with the increasing complexity of the financial market, traditional trading strategies are difficult to meet the diversified needs of modern investment. Data-driven trading strategies based on big data and intelligent algorithms have advantages in improving decision-making efficiency and strengthening risk management that cannot be ignored. However, how to optimize strategy models and effectively manage risks remains a key issue. The core objective of this paper is to study the framework of data-driven trading strategy and its optimization methods, and analyze the internal mechanism of its risk management, in order to provide theoretical reference and support for the investment decision making in the capital market.

2. Framework and method of data-driven trading strategy

2.1 Core components of data-driven strategy

In the data-driven trading strategy, it covers data collection, feature extraction, model construction, optimization and evaluation. Relying on the market information collected by API channels or financial data service providers, preliminary data pre-processing and feature extraction (such as technical indicators and basic element information) are carried out to meet the input requirements of the model. In the process of model building, the model is trained using machine

Copyright: © 2025 by the authors. This is an Open Access article distributed under the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (https://creativecommons.org/licenses/by/4.0/).

learning or deep neural network techniques (such as random forests, long term memory networks) to explore the inherent patterns in the data. By adjusting the hyperparameters of cross-validation method and grid search, the generalization effect of the model is strengthened. On this basis, the effectiveness of the strategy is further evaluated through historical backtesting and simulated trading, and the strategy is deeply optimized with the help of key indicators such as sharp ratio and maximum retracement. Figure 1 below is a flowchart of the core components of a data-driven trading strategy:

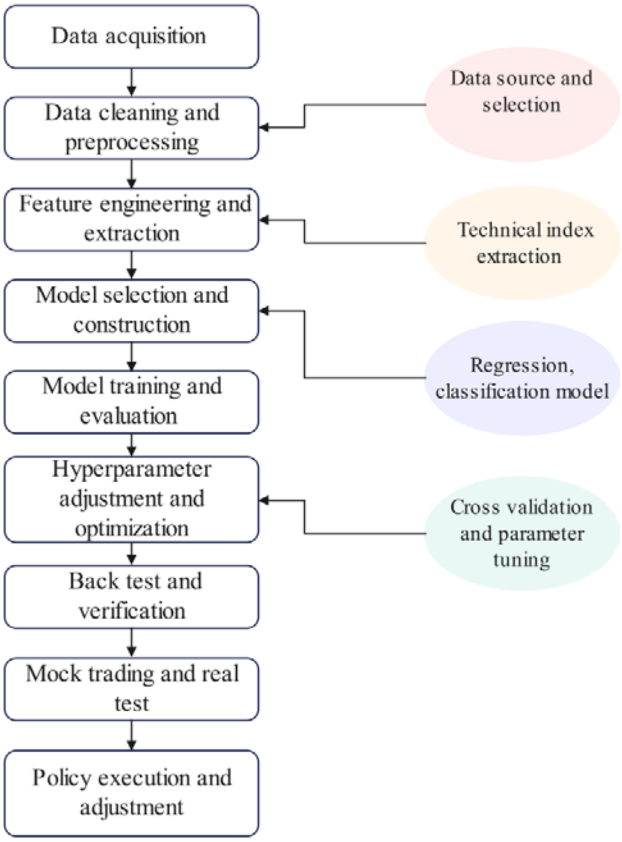

Figure 1. Flowchart of the core components of a data-driven trading strategy

2.2 Data acquisition and preprocessing technology

Data acquisition and pre-processing play a crucial role in the construction of data-centric trading strategies, which directly determine the efficiency of the model and the accuracy of the prediction. In the initial step of data collection, a variety of methods must be used to collect the necessary market data, including stock prices, volume, foreign exchange rates, futures contracts, and so on. At the same time, unstructured data such as corporate financial statements, overall economic indicators, media information, and public opinion on social platforms also provide valuable auxiliary data for strategy. Data can be collected through programmatic interfaces, web crawlers, or through specialized financial data providers such as Bloomberg and Reuters. Due to the large amount of data and the existence of noise, the pretreatment technology is particularly important. In the process of data cleaning, common techniques include the completion of lost data, the elimination of redundant records and the detection of irregular data to ensure the quality of the data set.

Conventional methods to deal with missing data include mean interpolation, forward filling or interpolation technology to complete, and for the processing of outliers, statistical analysis means (such as box chart, Z-score) are generally used to detect and adjust. Data normalization and unification can help to adjust data of different magnitude to the same scale and meet the requirements of various model training. In the process of data preprocessing, feature selection also plays an important role, and the most critical attributes for model prediction are selected through strategies such as correlation testing and principal component analysis (PCA). Through a series of data acquisition and preprocessing techniques, the data set created shows excellent structural quality, and strongly supports the subsequent model training and strategy optimization, ensuring the accuracy and reliability of the trading decision process.

2.3 Model optimization and parameter adjustment method

In data-driven trading strategies, model optimization and tuning play a crucial role in enhancing the accuracy of model predictions and their applicability in different scenarios. Commonly used optimization methods involve parameter tuning, feature screening and improvement, and regularization technology. Among them, parameter tuning is a common method to improve model performance. Machine learning models such as support vector machines (SVM), random forests, neural networks, etc., have multiple hyperparameters that need to be set, such as C and γ in SVM, number and depth of trees in random forests, learning rate in neural networks, batch size, etc. In order to find the optimal hyperparameters, the commonly used parameter tuning methods include Grid Search and Random Search. Grid search selects the best parameter by exhaustively searching all possible combinations of hyperparameters, while random search finds the best solution by randomly selecting combinations in the hyperparameter space. The goal is to minimize the loss function $L(\theta)$, i.e. :

$$\hat{\theta} = \arg\min_{\theta} L(\theta,\ X,\ y) \tag{1}$$

Where $L(\theta)$ is the loss function, θ is the model parameter, X is the feature, and y is the target variable.

One of the key steps of model optimization is feature selection. The removal of redundant or irrelevant features significantly increases the model's operational efficiency and accuracy. In the process of feature selection, some strategies such as decision tree feature importance scoring and recursive feature elimination (RFE) can be adopted. The selected models tend to be more compact and have better generalization performance. In addition, in order to avoid overfitting the model, regularization methods are widely used, including L1 regularization (Lasso) and L2 regularization (Ridge), two common methods. The objective function of Lasso regression is:

$$\min_{\theta}\left(\sum_{i=1}^{n}\left(y_i - f\left(x_i,\ \theta\right)\right)^2 + \lambda\sum_{j=1}^{p}\left|\theta_y\right|\right) \tag{2}$$

λ is a regularization parameter that controls the complexity of the model. Through these optimization and parameter adjustment methods, the stability and return rate of trading strategies can be significantly improved.

3. Optimization and performance evaluation of trading strategies

3.1 Multi-objective optimization and genetic algorithm

When optimizing trading strategy, it is necessary to take into account multiple objectives such as profit, risk control, strategy stability and capital utilization. The multi-objective optimization technique can optimize these contradictory objectives simultaneously, so as to find the excellent performance of the strategy scheme. The core objectives of optimization usually include return maximization, risk reduction, retracement reduction, and strategy robustness enhancement. Genetic algorithm is widely used in this field to effectively solve multi-objective optimization problems.

Genetic algorithm simulates the process of biological evolution and searches for the optimal solution through three steps: selection, crossover and variation. Each individual represents a possible trading strategy, and its fitness function is evaluated based on indicators such as return and risk. For multi-objective optimization problems, the fitness function needs to take into account multiple objectives, usually using weight allocation method or Pareto frontier technology integration objectives. Set a trading strategy optimization problem, assuming two optimization objectives: maximizing return R and minimizing retracement D, then a weighted fitness function can be constructed:

$$F(x) = \omega_1 R(x) - \omega_2 D(x) \tag{3}$$

Where $F(x)$ is the fitness function, $R(x)$ is the return of the strategy, $D(x)$ is the retracement of the strategy, ω_1 and ω_2 are the target weights, indicating the importance of the return and retracement. Through the iterative process of genetic algorithm, the optimal solution that can balance these two objectives is found step by step.

Through the steps of population initialization, individual fitness evaluation, selection, crossover and variation, the genetic algorithm continuously optimizes the strategy. The crossover process mimics gene binding and combines high-quality individual traits to create new strategies. Variation improves the variability of the population by randomly modifying parameters to prevent falling into local optimal solutions. With its excellent global search ability and expertise in dealing with multi-objective problems, genetic algorithm can effectively balance the relationship between return, retracement and risk, so as to create a durable and reliable trading strategy. Compared with the traditional single-objective optimization method, it shows higher flexibility and application value.

3.2 Retracement, Sharpe ratio and information ratio

In the process of optimization and evaluation of trading strategies, the retracement range, sharpe ratio and information ratio are the main indicators to evaluate its performance, which reflect the risk tolerance, profitability and stability of the strategy respectively. Among them, the extent of the retracement reflects the maximum decline suffered by the strategy, that is, the amount of money lost from peak to trough. The smaller the indicator, the better the strategy is at managing risk.The Sharpe ratio is used to assess the level of additional return per unit of risk. This measure measures the return performance after adjusting for risk by comparing the profitability of the strategy with its volatility (i.e., standard deviation). When the Sharpe ratio is high, it means that for the same level of risk, the strategy can achieve more substantial additional returns. Its calculation formula is as follows:

$$S = \frac{R_p - R_f}{\sigma_p} \tag{4}$$

Where R_p is the expected return of the strategy, R_f is the risk-free interest rate, and σ_p is the standard deviation of the strategy.

The information ratio is used to measure the excess return of a strategy relative to its volatility. It reflects the performance of the strategy relative to the benchmark index. When the information ratio value is large, it means that the strategy can obtain more stable excess returns while maintaining lower volatility. Its calculation formula is as follows:

$$IR = \frac{R_p - R_b}{\sigma_{Rp-R_b}} \tag{5}$$

Where R_p is the return of the strategy, R_b is the benchmark return, and σ_{Rp-Rb} is the standard deviation of the excess return. Table 1 below shows the performance of the three strategies on retracement, Sharpe ratio and information ratio:

Table 1. Comparison of retracement, Sharpe ratio and information ratio of three trading strategies

Tactics	Retracement(%)	Sharpe Ratio	Information ratio
Strategy A	12.5	1.45	0.75
Strategy B	8.3	1.75	1.1
Strategy C	15.2	1.2	0.6

Data is displayed in the table. Strategy A faces a high retracement (12.5%), and both Sharpe ratio and information ratio are not ideal. In contrast, Strategy B shows a low retracer (8.3%) and its sharpe ratio (1.75) and information ratio (1.10) are both prominent, showing its excellent performance. Although strategy C has a significant retracement (15.2%), the other indicators are relatively weak, and the overall performance is poor. On the whole, strategy B shows the best performance in the balance of benefits and risks.

3.3 Robustness and long-term performance of the model

The stability and lasting profitability of a trading strategy are the key factors to measure its reliability. Stability is the ability of a strategy to maintain stable performance and adapt to a variety of market conditions, such as bull, bear, and volatile markets. As for long-term profitability, it focuses on whether a strategy can consistently generate profits over a long period of time to avoid the risk of accidental success or over-fitting of data based solely on short-term performance. Testing stability usually involves stress testing and detailed analysis of market segments. Through the simulation test of the extreme market environment, we can observe the response effect of the strategy when the market fluctuates violently or falls rapidly. Market segmentation analysis divides the historical data into different market stages, and calculates the return and risk indicators of the strategy in each stage respectively. As for the long-term performance of the strategy, it is based on the comprehensive backtracking analysis, and the comprehensive evaluation of the key indicators such as the return situation, volatility and maximum retracement of the strategy in the entire historical data cycle. Table 2 below shows the performance of a model in different market environments:

Table 2. Long-term performance and robustness assessment of the model in different market environments

Market environment	Annualized return(%)	Annualized Volatility(%)	Maximum Retracement(%)
Bull market	15.8	10.2	8.3
Bear market	7.2	12.5	12.8
Volatile market	10.4	9.8	7.5
Full sample	12.1	10.8	9.5

The strategy has achieved the highest annualized return of 15.8% in a bull market with low volatility and retracements, outperforming. In a bear market, it remained positive despite the return falling to 7.2%, showing good risk management ability. In a volatile market, the yield was 10.4%. Combined analysis shows that the average annualized return of the strategy is 12.1% and the maximum retracement is 9.5%, which proves its robustness and long-term stability. The model is highly adaptable and has sustainable profitability, which can be further improved by subsequent optimization.

4. Risk management of data-driven trading strategies

4.1 Market risk and liquidity risk identification

In the process of making trading strategy, market volatility risk and capital liquidity risk occupy the core position, and the accurate identification and quantitative analysis of these risks are crucial to the stable execution of the strategy. Market volatility risk is the potential loss caused by the price changes of market prices (such as stocks, exchange rates, interest rates, etc.), and the detection of such risks is generally completed by the volatility analysis and value risk (VaR) evaluation model. By calculating the standard deviation and distribution of asset returns, VaR is able to quantify the maximum possible loss at a given confidence level:

$$VaR = \mu - z \cdot \sigma \tag{6}$$

Where μ is the expected return of the asset, σ is the standard deviation of the return, and z is the value of the standard normal distribution at the confidence level. Market liquidity risk is mainly caused by insufficient market depth or poor capital flow, which may cause delayed transactions or increase transaction costs. Observing bid-ask spreads, volume, and order thickness can detect a lack of liquidity in time. By dynamic analysis of sliding points in different market conditions, the cost of trading can be measured. By tracking market volatility and liquidity conditions, data-based strategies can optimize the efficiency of trade execution while reducing losses that can occur in extreme situations.

4.2 Risk hedging of insurance instruments and derivatives

As an important means of risk hedging, insurance instruments and financial derivatives (such as options, futures and swaps) are widely used in trading strategies, which can effectively reduce potential risk losses caused by market fluctuations. By properly configuring these instruments, investors can ensure the expected returns while limiting the maximum possible losses, thereby enhancing the stability of their investment strategy. As a key type of insurance, option gives investors the ability to hedge risks non-linear. For example, an investor may buy a protective put option while holding an asset to ensure that the value of the asset does not fall below a predetermined baseline. The profit and loss formula is as follows:

$$\text{Portfolio return} = \max(S - K, 0) - P \tag{7}$$

S is the price of the underlying asset, K is the option strike price, and P is the option premium. The core of this strategy is to ensure that investors can control their losses when asset prices fall, but still enjoy profits when assets appreciate. Futures and swaps are widely used to stabilize future market price movements. Futures contracts, for example, help investors reduce the risk of price changes by setting a price for trading at a certain time in the future. Swap contracts give investors a more diversified hedge against risk by swapping cash flows. Figure 2 below shows the overall process of risk hedging:

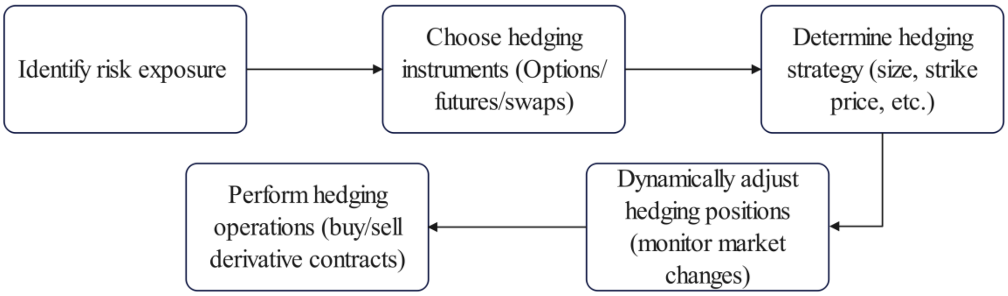

Figure 2. Risk hedging flow chart of insurance instruments and derivatives

Investors identify potential risk exposures, select matching insurance products or financial derivatives, set hedging strategy parameters (such as hedging scale, execution price, etc.) and implement related operations. After that, the position is dynamically managed according to the market fluctuations, and the market dynamics are continuously tracked to ensure the effectiveness of the hedging means.

4.3 Monte Carlo simulation in risk management

Monte Carlo simulation technology, in essence, is a calculation method based on random sampling, which is widely used in the field of risk assessment, aiming at evaluating the effect of trading strategies in different market environments. By simulating a large number of possible market scenarios, Monte Carlo simulation can quantify the potential risks of a strategy and predict its return distribution, providing a basis for strategy optimization. In trading strategies, Monte Carlo simulations are often used to calculate value risk (VaR) and conditional value risk (CVaR). Suppose that the evolution of asset price S_t follows the geometric Brownian motion, which is expressed as:

$$S_t = S_0 \cdot e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t} \tag{8}$$

Where S_0 is the initial price, μ is the expected return, σ is the volatility, and W_t is the standard Brownian motion. By randomly sampling the equation, a large number of asset price paths are generated, and the gain or loss of the strategy under each path is calculated. The simulation results can be represented as a return distribution or risk measure, using CVaR to measure the average loss in extreme cases outside the VaR range by calculating VaR values at the 95% confidence level at the 5th percentile. Monte Carlo simulation is known for its high adaptability and is particularly suitable for dealing with asset portfolios that contain complex elements and non-linear returns. The simulation can simulate market turbulence and extreme conditions, help investors to comprehensively assess risks, improve risk management strategies, and enhance the stability of investment strategies.

5. Conclusion

This paper comprehensively analyzes the optimization and risk management techniques of data-driven trading strategies, and constructs a comprehensive architecture covering strategy optimization, multi-objective balance adjustment and risk control. Using genetic optimization algorithm and Sharpe ratio and other technologies to improve the efficiency of the strategy, while integrating market risk identification, derivative hedging operations and Monte Carlo simulation to achieve risk management, to ensure the stability of the strategy and long-term benefits. The research points out that in the changeable market environment, data-driven trading strategies show obvious advantages, contributing to the process of capital market intelligence.

References:

[1] *Liu P, Zhang Y, Bao F, et al. Multi-type data fusion framework based on deep reinforcement learning for algorithmic trading. Applied Intelligence, 2023, 53(2):1683-1706.*

[2] *ZHOU Rongtian, XIONG Xiong, ZHANG Xiaoxuan. Research on the Short-Selling Rate in China's Capital Market——Based on the Perspective of the Artificial Stock Market. Journal of Systems Science and Mathematical Sciences, 2022, 42(8):2019-2039.*

[3] *Myers M. Leveraging Your Insurance Carrier's Resources for Risk Management. Engineering news-record, 2023(1):291.*

[4] *Pomaza-Ponomarenko A, Kryvova S, Hordieiev A, et al. Innovative risk management: identification, assessment and management of risks in the context of innovative project management. GeSec: Revista de Gestao e Secretariado, 2023, 14(10).*

[5] *Guo Q, Gao L, Chu X, et al. Parameter Identification for Static Var Compensator Model Using Sensitivity Analysis and Improved Whale Optimization Algorithm. CSEE Journal of Power and Energy Systems, 2022, 8(2):535-547.*

[6] *Wu Y. Software Engineering Practice of Microservice Architecture in Full Stack Development: From Architecture Design to Performance Optimization. 2025.*

[7] *Zhang, X. (2025). Optimization of Financial Fraud Risk Identification System Based on Machine Learning. Journal of Computer, Signal, and System Research, 2(6), 82-89.*

[8] *Wang, Y. (2025). Exploration and Clinical Practice of the Optimization Path of Sports Rehabilitation Technology. Journal of Medicine and Life Sciences, 1(3), 88-94.*

[9] *Li W. The Influence of Financial Due Diligence in M&A on Investment Decision Based on Financial Data Analysis. European Journal of AI, Computing & Informatics, 2025, 1(3): 32-38.*

[10] *Sheng, C. (2025). Innovative Application and Effect Evaluation of Big Data in Cross-Border Tax Compliance Management. Journal of Computer, Signal, and System Research, 2(6), 40-48.*

[11] *Sheng, C. (2025). Research on the Application of AI in Enterprise Financial Risk Management and Its Optimization Strategy. Economics and Management Innovation, 2(6), 18-24.*

[12] *Tu, X. (2025). Optimization Strategy for Personalized Recommendation System Based on Data Analysis. Journal of Computer, Signal, and System Research, 2(6), 32-39.*

[13] *Sun, Q. (2025). Research on Cross-language Intelligent Interaction Integrating NLP and Generative Models. Engineering Advances, 5(4).*

[14] *Liu, Y. (2025). Use SQL and Python to Advance the Effect Analysis of Financial Data Automation. Financial Economics Insights, 2(1), 110-117.*

[15] *Ye, J. (2025). Optimization of Neural Motor Control Model Based on EMG Signals. International Journal of Engineering Advances, 2(4), 1-8.*

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

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