

Portfolio risk management model based on machine learning

Liu Yizheng

University of International Business and Economics, Beijing, China

Keywords: Machine learning, financial risk management, portfolio optimization, big data technology

Abstract: This paper comprehensively analyzes many domestic and foreign literatures related to machine learning, financial risk management, the investment portfolio, etc. Domestic literature covers flood forecasting, portfolio construction and optimization, financial intelligence, and application of big data technology in hospital archives management. The foreign literature involves multi-factor semi-parameter distribution of investment portfolio, international practice and countermeasures of customs risk management, autonomous navigation technology based on machine learning and so on. From these literatures, it can be found that machine learning has shown a wide range of application prospects in flood forecasting, hospital archives management, e-commerce marketing and other fields. In terms of financial risk management and control, the research discusses internal control, financial fraud identification model, enterprise information construction and other strategies, as well as the application of big data technology in risk management. In addition, the research on portfolio construction and optimization focuses on the perspective of genetic algorithm and portfolio selection of multi-risk assets. At the same time, emerging fields based on big data technology are also mentioned in literature, such as the research status of plant factories, and the application of smart finance in internal control optimization and risk management in universities.^[1]

1. Introduction

Portfolio risk management plays a vital role in modern finance. The optimization of investment portfolio and risk control is the core issues of investment decision-making, which is related to the asset appreciation and wealth preservation of investors. However, the volatility and non-linear characteristics of the financial market make the traditional risk management methods seem limited, and it is difficult to accurately reflect the real situation of the market. In the face of emerging financial instruments and diversified investment strategies, investors urgently need innovative risk management methods to cope with the changing market environment.^[2]

The purpose of this study is to build an innovative portfolio risk management model based on machine learning technology to better solve the challenges of risk management in financial markets. By introducing machine learning algorithms and portfolio theory in the field of finance, we aim to build an efficient, intelligent and controllable portfolio management framework with risk exposure. Through the learning of historical market data, machine learning models will help to more

accurately predict the risks and returns of assets, thus providing more targeted risk management strategies.

To achieve the above research objectives, this study adopted a systematic research method and framework. First, we will collect historical data from the financial markets and conduct data preprocessing and sample selection to ensure the quality and reliability of the data.

2. Literature review

Portfolio theory is one of the important theories in modern finance, which aims to minimize the risk or maximize the return at a given risk level by reasonably dispersing the investment funds. In the development of portfolio theory, modern portfolio theory (MPT) and effective market hypothesis and market factor model and APT model is two important milestones.

The core idea of modern portfolio theory (Modern Portfolio Theory, MPT) MPT is to form an effective frontier by combining different assets together, that is, a portfolio set that can achieve the maximum return at a given risk level. MPT stresses that investors should focus on the risks and returns of the entire portfolio, rather than on individual assets. Related to the MPT is the effective market hypothesis (EMH). According to EMH, investors should adopt passively's investment strategy, which tracks the entire market through tools such as index funds.

Market factor models and APT models are multiple regression models used to explain asset returns, which tries to find factors affecting asset returns. Similar to CAPM, arbitrage pricing theory (Arbitrage Pricing Theory, APT) is also a model used to explain asset returns. Unlike CAPM, the APT model allows multiple factors to affect asset returns, which can be macroeconomic indicators, interest rate levels, industry factors, etc. The APT model emphasizes that investors should take multiple factors into account and build more integrated risk management strategies.

3. Data collection and processing

In finance majors, the study of the portfolio risk management model based on machine learning requires a large amount of historical market data and macroeconomic data. There are two key steps for data collection and processing.

First, you need to obtain historical price data for multiple assets (e. g., stocks, bonds, commodities, etc.). This data can be obtained from financial data providers, exchanges, or financial databases. Then, the raw data were cleaned and preprocessed, including removing missing values, outliers, and the data frequency was adjusted to accommodate the time period of the study, ensuring the integrity and accuracy of the data.

When constructing the portfolio risk management model, we also need to consider the macroeconomic data and the financial factor data. Macroeconomic data can include indicators such as gross domestic product (GDP), inflation rate and unemployment rate, which can be obtained through economic research institutions or government departments.

After the data is ready, this is followed by feature engineering to extract meaningful features from the raw data for use by machine learning algorithms. The goal of feature engineering is to provide more informative input to machine learning algorithms to improve the predictive accuracy and generalization ability of the model.

When selecting fundamental factors, it is necessary to select appropriate factors according to the research objectives and finance theory. Fundamental factors can include the company's financial indicators (such as revenue, profit, etc.), profitability indicators, solvency indicators, etc.

Machine learning algorithms are very important tools to help investors discover patterns and patterns from massive amounts of data, and make more accurate predictions and decisions.

The SVM is a supervised learning algorithm, often used for classification and regression

problems. In portfolio risk management, SVM can be applied to risk forecasting to help investors identify high-risk assets and low-risk assets for risk control and asset allocation.

Random forest is an ensemble learning algorithm for classification and regression by building multiple decision trees. In portfolio risk management, random forest can be used to classify assets, dividing different assets according to their characteristics, so as to better understand the risk attributes of different assets.

In portfolio risk management, deep neural networks can be used to predict the returns of assets and help investors to make more promising investment decisions.

4. A portfolio risk management model based on machine learning

4.1 Model framework and design

To accurately predict the risk of the portfolio, we employ multiple machine learning algorithms to construct the risk prediction models. First, we use supervised learning methods, such as support vector machines (SVM) and deep neural networks (DNN), to learn asset volatility and risk characteristics from historical market data.

Second, we also use unsupervised learning methods, such as clustering algorithms, to classify the assets. Meanwhile, we also introduce reinforcement learning algorithms, such as Q-learning, to dynamically adjust portfolios. Through continuous learning and adjustment, the portfolio can be more adapted to market changes, reduce systemic risk and improve return on investment.

On the basis of the risk prediction model, we construct the portfolio by using the modern portfolio theory (MPT) and the effective market hypothesis, as well as the ideas of the market factor model and the APT model. Based on the risk level and expected return predicted by the model, we find the optimal asset allocation scheme through the optimization algorithm.

When constructing a portfolio, we consider the investor's risk appetite and investment objectives. By setting different risk preference parameters, we can build personalized investment portfolios for different types of investors to meet the needs of different investors. The raw data to ensure the integrity and accuracy of the data. This may involve removal of missing values, processing of outliers, adjustment of data frequency to accommodate the time period of the study, etc. After the data is ready, the researchers will train and evaluate the model using the previously designed risk prediction model and portfolio construction method. First, historical market data is used to train machine learning algorithms. Meanwhile, a clustering algorithm is used to classify assets to identify the correlations and differences between assets. Then, the trained model is used to predict and assess the risk of the future market data. By comparing the risk level predicted by the model and the actual market performance, researchers can evaluate the predictive ability and risk control effect of the model.

To further verify the effectiveness of the machine learning-based portfolio risk management model, the researchers will also conduct a comparative analysis with the traditional portfolio risk management methods. Traditional methods may include the variance-covariance approach, minimize the risk portfolio, etc. By comparing with traditional methods, researchers can evaluate the advantages and differences of machine learning-based models in risk prediction and portfolio construction.^[3]

4.2 Model empirical study

First, in order to conduct empirical research, researchers need to collect and process real financial market data. In the data processing phase, the researcher needs to clean and preprocess.

In the empirical study, the researchers will also build a personalized investment portfolio based

on the investors' risk preferences and goals. By setting different risk preference parameters, researchers can construct investment portfolios of different risk levels according to the needs of investors. In this way, investors can choose the most suitable portfolio strategy according to their own risk tolerance and investment objectives. Finally, the investigators will conduct a comprehensive evaluation of the results of the model empirical study. They will compare the performance of different models in risk prediction and portfolio construction and analyze the strengths and weaknesses of the models. At the same time, they will evaluate the performance of the model in different market environments to verify the model robustness and adaptability.

5. Empirical study

5.1 Data sample and experimental design

In the empirical study of this paper, we will combine finance knowledge and machine learning technology to explore the application effect of machine learning-based portfolio risk management model. To ensure the reliability and validity of the empirical study, we will select data samples and conduct experimental design.

The selection of the data samples is the basis of the empirical research. We will obtain historical price data for multiple asset classes from reliable financial data sources, covering different asset types, such as stocks, bonds, and commodities. These data will be the basis for building our risk management model. When selecting the data samples, we will note the following points:

Data representativeness: to ensure that the selected historical data has a certain representativeness, can reflect the real market situation, and avoid the selection of specific biased data sets.

Data integrity: Try to select complete and continuous historical data to avoid excessive missing values or outlier values in the data.

Data time range: We will choose the appropriate time range, including historical data from different market states, to examine the performance of the model in different market environments, such as bull, bear, volatile markets, etc.

In the experimental design, we will carry out the following steps according to the research methods and framework in the outline of the paper: preprocessing the selected historical market data, including removing missing values, outliers, smoothing data, etc., to ensure the reliability and stability of the data. We will train the risk management model using a machine learning algorithm.

In order to evaluate the generalization ability of the model, we will test and verify the model by using cross-validation. The trained model is used to predict the risk of the selected future market data. Based on the prediction results, the optimization algorithm is used to make asset allocation and build the investment portfolio. The constructed portfolio is backtested and simulated for trading to evaluate the performance of the model in the real market. At the same time, the comparative experiment with the traditional portfolio construction method is conducted to analyze the advantages and improvements of the machine learning model.

5.2 Model implementation and result analysis

During the model implementation, we will actually apply the machine learning-based portfolio risk management model according to the experimental design and parameter setting described above. First, we will train the models using historical market data, including the training of supervised learning models (e. g., support vector machines and deep neural networks) and the classification of unsupervised learning models (e. g., clustering algorithms). After the model implementation, we will evaluate the predictive ability and risk control effect of the model. We will

compare the difference between the risk level predicted by the model and the actual market performance, and analyze the prediction accuracy and stability of the model.^[4]

We will compare the risk level predicted by the model with the actual market fluctuations, and calculate the indicators such as prediction error and accuracy, to evaluate the risk prediction ability of the machine learning model.

Through the above empirical results and performance analysis, we can draw conclusions about the effectiveness and practicability of the machine learning-based portfolio risk management model in practical application.

5.3 Robustness analysis and sensitivity test

In finance majors, the changes in the market environment and the volatility of the data may affect the performance of the model, so the robustness of the model needs to be evaluated. At the same time, the sensitivity test can help us understand the impact of the parameter setting of the model on the empirical results and provide a basis for model tuning.

Robustness analysis aims to verify the performance of the model in different market conditions. We will select historical data from different time periods, covering different market states, and we can evaluate the robustness of the model by comparing the prediction ability and portfolio performance of the model in different market environments.

Sensitivity tests aim to assess the impact of model parameter setting on the empirical results. We will adjust the parameters of the model to observe the impact of changes in these parameters on the performance of the portfolio. Through sensitivity tests, we can determine which parameters have a large influence on the performance of the model, providing a reference for the tuning and parameter setting of the model.

In addition to the robustness and sensitivity tests, robustness analysis is also important for empirical research. In the robustness analysis, we will consider the impact of different market data quality and anomalous data on the model. By introducing some noise or anomalous data, we can test the robustness of the model to noise and evaluate the model performance in a noisy environment.

6. Discussion

6.1 Advantages and limitations of the model

The machine learning-based portfolio risk management model presents many advantages in empirical research that makes it a powerful tool in finance:

Machine learning models can more accurately predict the risk levels of assets by learning the patterns and trends in historical data. Compared with traditional statistical models, machine learning is able to handle more complex nonlinear relationships, improving the accuracy of risk prediction.

Machine learning models allow for personalized portfolio strategies based on investors' risk preferences and goals. This allows investors to find the best balance between risk and return based on their own needs.

Machine learning models are highly adaptable and can automatically adjust their investment strategies according to changes in market conditions and data. They can capture the dynamic characteristics of the market and make timely adjustments to adapt to the changes in the market.^[5]

Financial market data is usually very large, and machine learning models have the ability to process large-scale data, which can mine useful information from large amounts of data and improve the effectiveness of investment decisions.

However, machine learning-based portfolio risk management models have some limitations that

need to be considered in practical applications:

Over-fitting: Machine learning models may overfit historical data during training, resulting in poor prediction performance on unknown data. Therefore, techniques such as suitable regularization methods and cross-validation are needed to prevent overfitting problems.

Data quality and stability: The predictive ability and performance of the model is affected by the data quality and stability. If outliers or missing values are included in the historical data, the outcome of the model. Therefore, the quality and integrity of the data should be guaranteed in practical applications.

Market irrational behavior: there are a large number of irrational behaviors and emergencies in the financial market, which may lead to the prediction error of the model. Machine learning models still have some challenges in dealing with market irrationality.

Parameter selection and interpretation: Machine learning models usually need to set some parameters, and the choice of these parameters may have a great impact on the performance of the model. Moreover, the results of some machine learning algorithms are difficult to interpret, which may cause some trouble in Practical applications.

6.2 Practical application considerations

When applying a machine learning-based portfolio risk management model to real investment decisions. We need to consider the following factors:

Data quality and stability: We can ensure that the selected historical market data has high quality and stable, and avoid the impact of outliers and missing values on the model.

Risk preference and goal: We can set the model parameters according to the investors' risk preference and investment goal, and build a personalized investment portfolio.

Market environment and cycle: We can consider the different states and cycles of the market, applies the model to the forecast and asset allocation of different market conditions.

Risk control: When constructing the portfolio, the risk control strategy should be fully considered to avoid the excessive concentration and exposure of the portfolio.

Model interpretability: We can choose machine learning models with good interpretability as far as possible, so that investors can understand the decision-making process and prediction results of the model.

Transaction cost: In the practical application of the model, the impact of the transaction cost on the investment portfolio should be considered to ensure the feasibility and effectiveness of the actual operation.

7. Conclusion

7.1 Study summary

This paper aims to combine the finance major portfolio theory with machine learning technology to build a comprehensive and intelligent portfolio risk management model. Through empirical research, we validate the advantages of machine learning-based models in risk prediction and asset allocation, and explore the model limitations and practical application considerations. In the selection of data samples and experimental design, we pay attention to the quality and stability of the data to ensure the reliability of the research results. In terms of the advantages of the model, machine learning models show the advantages of high predictive accuracy, personalized risk management and flexibility. However, we also need to pay attention to the limitations such as overfitting, data quality and parameter selection, as well as transaction costs and irrational market behavior in practical applications.

7.2 Conclusion and contributions

The portfolio risk management model based on machine learning shows some advantages in this study, and makes the following contributions to the risk management research in the field of finance:

Research method: This paper builds a new portfolio risk management model by combining finance theory and machine learning technology. This method provides new research ideas and methods for the field of finance.

Empirical research: Through empirical research, we verify the effectiveness of machine learning-based models in risk prediction and asset allocation. The results provide investors with more accurate risk forecasting and portfolio building recommendations.

Theoretical expansion: This study also expands the traditional portfolio theory, introduces machine learning algorithms into the field of risk management, and provides investors with more diversified and personalized portfolio management solutions.

7.3 Opinions and suggestions on financial risk

Here are suggestions on the financial risk management practice:

Risk prediction optimization: Machine learning-based portfolio risk management model can provide more accurate risk prediction results, which investors can use to conduct risk control and optimize asset allocation.

Personalized risk management: The model allows investors to develop personalized portfolio strategies based on investors' risk preferences and goals, helping investors find the best balance between risk and return.

Real-time decision support: We can apply the model to the real-time decision support system to provide investors with real-time and accurate risk management suggestions and help investors adjust their investment portfolio in time.

Risk control strategy: In practical application, it is necessary to fully consider the risk control strategy to avoid the excessive concentration and exposure of the investment portfolio and reduce the investment risk.

Continuous improvement and optimization: In practical application, the model should be continuously improved and optimized, combined with the new market conditions and data, to improve the adaptability and effect of the model.

References

- [1] Song Lujun. *Portfolio construction and optimization: Investor preference and genetic algorithm perspective* [J]. *Economic Issues*, 2023 (8): 60-66.
- [2] Huang Guanglin, Lu Wanbo. *Study of high-dimensional dynamic higher moments based on semi-parameter distribution of variable coefficients* [J]. *Management science in China*, 2023 (19): 16-18.
- [3] Yang Liuqing. *Building enterprise internal control system and improving enterprise risk management practice thinking* [J]. *Tax payment*, 2023, 17 (10): 106-108.
- [4] Fan Jiaqi. *Research on Financial Fraud Identification Model Based on Privacy Protection Machine Learning* [D]. *Chinese Academy of Fiscal Sciences*, 2023.
- [5] Ye Changhong. *Analysis on the financial risk control of medical device enterprises* [J]. *Accounting study*, 2023 (16): 28-30.