

A portfolio investment strategy of financial products with statistical machine learning

Xiaorong Liu *

Environmental planning and management, The City Vocational College of Jiangsu, Nanjing, China

* Corresponding Author Email: 1035143965@qq.com

Abstract. At present, portfolio investment is an investment strategy widely used in practice. Statistical machine learning has good performance and accuracy for portfolio investment return and risk calculation. Most current researches focus on the price rise and fall of a single stock or risky securities. This method uses machine learning technology to calculate the portfolio return of multiple financial products, including the portfolio return of high-risk securities, medium-risk securities and low-risk securities and the related risk degree. The simulation results show that the proposed method has good reference value.

Keywords: Portfolio investment, financial investment, microeconomics, Machine learning

1. Introduction

In modern financial theory, portfolio investment is an important investment strategy, and portfolio investment plays an important role in financial investment, which diversifies investors' funds into a variety of different assets in order to obtain the highest possible return while reducing risks, that is, to spread the risk by diversifying investments.

The theoretical basis of portfolio investment mainly comes from Modern Portfolio Theory, investors should consider the expected return and risk of assets, and through reasonable allocation of different assets when making investment decisions, in order to achieve the best investment effect. In this process, investors need to predict various possible investment outcomes and calculate their expected returns and risks.

In practice, the methods of portfolio investment mainly include two kinds: one is equal weight method, that is, investors allocate funds equally to each asset; The other is the optimization method, that is, investors according to the expected return and risk of assets, as well as their own risk tolerance, through a mathematical model to calculate the optimal proportion of asset allocation.

In recent years, with the development of statistical machine learning technology, more and more researchers began to try to use this method to make portfolio investment. They collect a lot of historical data, build a forecast model, and then make investment decisions according to the forecast results of the model. This method can not only improve the investment efficiency, but also reduce the investment risk to a certain extent.

The importance of portfolio investing lies in its ability to reduce the risk of specific investments and increase the stability of the overall return of the portfolio. It has the following features.

1) Spread your risk: Investing your money in a variety of assets reduces the risk of losing money if a single investment underperforms. If one asset or asset class falls, its impact in the overall portfolio can be offset by other assets that have performed well.

2) Enhance returns: Different asset classes within a portfolio may perform differently in different market cycles. This diversification can help investors maintain stable returns amid market volatility.

3) Risk tolerance management: By building a diversified portfolio, investors can adjust asset allocation according to their individual risk tolerance. For example, investors with a high risk tolerance could add stocks to their portfolios, while those with a low risk tolerance might opt for more bonds or money-market funds.

4) Reach long-term goals: Through portfolio investments, clients can balance daily volatility with opportunities for long-term growth, which helps reach long-term financial goals, such as retirement savings plans.

5) Capture uncorrelated returns between assets: In different market environments and economic cycles, some asset classes may perform better than others. By investing in a portfolio of assets with a low correlation, the volatility of the overall portfolio may be reduced, thus increasing the safety of the investment.

6) Adaptability: As the market environment and personal circumstances change, the portfolio can adapt to these changes through asset rebalancing to maintain the original investment strategy and risk appetite of investors.

7) Leveraging expertise: Portfolio investing allows individual investors to take advantage of professional fund managers or market opportunities to manage their portfolios by purchasing investment vehicles such as mutual funds, index funds, or ETFs.

Transaction fees are especially important when buying multiple stocks for small amounts. While stocks are always a prediction of future profitability, the earnings of a particular stock we rely on are a sequence of past earnings. History may repeat or be similar, but history will never be the same due to the unpredictability of the future. Therefore, there is a certain risk in using this concept in practice. Although portfolio theory has some flaws, when we apply it to the stock market, it provides us with a whole new perspective on the issue. Technical analysis and fundamental analysis are the main tools for stock market investors to make investment decisions. However, most of these two analysis methods focus on calculating the return rate of a single security, and ignore the correlation between security returns. Therefore, this paper studies the method of portfolio investment income.

2. Methodology

2.1. Processing Scheduler

2.1.1. Data acquisition and preprocessing

Data collection and pre-processing is a crucial part in constructing the portfolio investment strategy model of financial products based on statistical machine learning. Firstly, we need to obtain relevant wealth management product data from multiple channels, including historical returns, risk indicators, market environment and other information. This data can be accessed through financial market databases, provided by financial institutions, or third-party data providers.

Secondly, we need to pre-process the collected data to facilitate subsequent analysis and modeling. The steps of pre-processing include data cleaning, missing value processing, outlier detection and feature selection. Data cleaning refers to the format conversion of data, the removal of duplicate values and outliers and other operations to ensure the quality and consistency of data. Missing value processing refers to the filling or deletion of missing values in the data to avoid the impact of subsequent analysis. Outlier detection refers to the identification and processing of outliers in data through statistical methods or machine learning algorithms to improve the accuracy and stability of the model. Feature selection refers to the selection of features with high correlation and predictive ability to target variables from the original data to reduce redundant information and improve the efficiency of the model. In addition, in the data acquisition and preprocessing, we also need to consider the timeliness and reliability of data. Timeliness is the frequency and timeliness of index data updates to ensure that the model can capture the latest market changes and trends. Reliability is the source and accuracy of data, and it is necessary to ensure that the data source is reliable and the data quality is high, and to avoid misleading results of the model due to data errors or biases.

To sum up, data collection and pre-processing is an important part of building a financial product portfolio investment strategy model based on statistical machine learning. Through data collection, cleaning, missing value processing, outlier detection, feature selection and other operations, we can obtain high-quality, accurate and reliable data sets, which provide strong support for subsequent analysis and modeling.

2.1.2. Feature engineering

Feature engineering and model selection are two key steps in constructing the portfolio investment strategy model of financial products based on statistical machine learning. This section describes these two aspects in detail.

Feature engineering refers to the extraction of feature variables with high predictive ability for target variables by transforming, screening and dimensionality reduction of original data in the process of data pre-processing. In the research of financial product portfolio investment strategy, feature engineering mainly includes the following aspects:

(1) Data cleaning: Missing value processing, outlier processing, duplicate value processing, etc. are carried out on the original data to ensure the integrity and accuracy of the data.

(2) Feature extraction: Extract feature variables with high predictive ability for target variables from the original data. These characteristic variables may include yield, risk indicators, market factors, etc.

(3) Feature transformation: The extracted feature variables are transformed, such as logarithmic transformation, standardization, etc., in order to eliminate the dimensional influence between different features and improve the prediction ability of the model.

(4) Feature selection: Through correlation analysis, principal component analysis and other methods, the feature variables with high predictive ability for target variables are selected to reduce the complexity of the model and improve the generalization ability of the model.

2.1.3. Model selection

After the completion of the feature project, it is necessary to select a suitable statistical machine learning model to carry out the research on the investment strategy of financial product portfolio. In the statistical machine learning model, support vector machine model, linear regression model, random forest model, decision tree model, neural network model, etc., will be used. There are several factors to consider when choosing a model:

(1) Predictive ability of models: Cross-validation and other techniques are used to evaluate the predictive performance of various models on training sets and test sets, and the model with strong predictive ability is selected.

(2) Model complexity: Considering the number of parameters of the model, structural complexity and other factors, choose a model with moderate complexity to avoid overfitting or underfitting.

(3) Explain of the model: According to the research purpose and actual needs, choose a model with strong explain to better understand the forecast results and investment strategy of the model.

To sum up, feature engineering and model selection are two key links in the research of financial product portfolio investment strategy based on statistical machine learning. Through feature engineering processing of the original data, the feature variables with high predictive ability for target variables are extracted. Then select the appropriate statistical machine learning model for modeling and forecasting, in order to achieve the research objectives of financial product portfolio investment strategy.

a) Model training and optimization

After building a financial product portfolio investment strategy model based on statistical machine learning, we need to train and optimize the model. This step is key to ensuring that the model can accurately predict future returns from wealth management products.

First, we need to select the appropriate training data set. This data set should contain the historical data of financial products in the past period of time, including the yield, risk level, investment period and other information of various financial products. Through the analysis of these data, we can find out the key factors affecting the return of financial products, and provide a basis for the subsequent model training.

Next, we need to choose the right machine learning algorithm to train our model. According to the literature information, the commonly used algorithms include support vector machine (SVM), decision tree (DT), random forest (RF) and so on. Each of these algorithms has advantages and disadvantages, and we need to choose the appropriate algorithm according to the actual problem and

data characteristics. For example, if there are nonlinear relationships in the data, we can choose nonlinear algorithms such as support vector machines or random forests; If the amount of data is large and there are many features, we can choose algorithms such as decision trees or random forests that can handle high-dimensional data. After selecting the appropriate algorithm, we need to train the model. During the training process, we need to adjust the parameters of the model to obtain the best prediction effect. This process typically requires multiple iterations, each of which uses a portion of the data as a validation set to evaluate the performance of the model. By comparing the performance of the model under different parameter Settings, we can find the optimal combination of parameters.

Finally, after the model is trained, we need to optimize it. The purpose of optimization is to further improve the prediction accuracy and generalization ability of the model. Common optimization methods include regularization, ensemble learning, etc. Regularization is a method to prevent overfitting by adding regular terms to the loss function to limit the model complexity. Ensemble learning is a method that combines multiple models to get the final prediction result by voting or weighting, etc., so as to improve the stability and accuracy of the model.

In short, model training and optimization is a key link in the research of financial product portfolio investment strategy based on statistical machine learning. We need to select appropriate algorithms and optimization methods according to actual problems and data characteristics to improve the prediction accuracy and generalization ability of the model.

b) Model application and verification

After building a financial product portfolio investment strategy model based on statistical machine learning, we need to carry out practical application and verification to ensure that the proposed strategy can achieve good investment results in the actual market. This section describes the application and verification process of the model in detail. Firstly, we need to collect a large amount of historical financial product data, including the return rate, risk indicators, investment period and other information of each product. This data can be obtained from major financial institutions or financial information providers. At the same time, we also need to collect market environment data to better analyze market trends, such as macroeconomic indicators and interest rate levels. Next, we will preprocess the collected data, including data cleaning, missing value processing, outlier processing, etc., to ensure the quality and integrity of the data. In addition, we need to feature engineering the data to extract key features that can help predict and evaluate investment strategies.

After data preprocessing, we can divide the data set into a training set and a test set. The training set is used to build and optimize our investment strategy model, while the test set is used to evaluate the model's performance on unknown data. We can use cross-validation and other methods to ensure the generalization ability of the model. Next, we will use the training set data to build and optimize our statistical machine learning-based investment strategy model. In this process, we can choose the appropriate machine learning algorithms, such as support vector machines, random forests, neural networks, etc., and the corresponding parameter Settings. Through repeated iteration and tuning, we can make the model achieve better prediction effect on the training set. After the model is built, we will evaluate the model using the test set data. We can calculate the return rate, Sharpe ratio and other indicators of the model under different investment periods, risk preferences and other conditions to fully understand the advantages and disadvantages of the model. In addition, we can further verify the effectiveness of our proposed model by comparing it with other investment strategies.

2.2. A framework and process for machine learning to calculate the returns of portfolio investments

The framework for calculating portfolio returns using machine learning can be broken down into the following steps:

1. Data collection and pre-processing: Firstly, relevant data, such as stock price, volume, market value, etc. are obtained from the financial market, as well as factors that may affect investment behavior, such as macroeconomic data and market sentiment. Pre-processing includes filling in missing values, processing outliers, standardizing data, etc.

2. Feature selection: Based on the collected data, various techniques (such as correlation analysis, principal component analysis (PCA) or automatic feature selection methods) are used to select features that are relevant to investment returns.

3. Divide training set and test set: Divide the entire data set into training set and test set. The training set is used to train the machine learning model, and the test set is used to evaluate the model performance.

4. Select the right machine learning model: Select the right machine learning model according to the collected data and features, such as linear regression, decision tree, support vector machine (SVM), or deep learning model such as neural network.

5. Training model: Use the training data to train on the selected model. For models that require parameter tuning, methods such as grid search can be used to optimize the parameters.

6. Model evaluation: Evaluation of trained models using test data sets. Compare the forecast results with the actual realized portfolio return, and calculate the forecast performance indicators (such as accuracy, mean square error).

7. Portfolio optimization: Use trained models to predict asset returns and apply optimization algorithms (e.g., greedy algorithms, simulated annealing, genetic algorithms, etc.) to redistribute assets to maximize returns and reduce risk.

8. Backtest strategy: Simulate the constructed investment strategy using historical data to check the performance of the strategy in the past period of time. Models and strategies are adjusted and optimized based on backtest results to improve forecast accuracy and return on investment.

In short, continuous iteration and optimization are required throughout the process to improve models and strategies to more accurately predict portfolio returns and achieve higher investment returns in reality.

2.3. Machine learning models to calculate the risk degree of portfolio investments

In fact, portfolio investment has a certain degree of risk, including but not limited to operational risk market, liquidity risk, credit risk and market risk. Through investment diversification, portfolio investment can reduce the non-systematic risk of a single investment, but it still brings some major risks. There are several ways to calculate the risk degree of portfolio investment:

2.3.1. standard model

The standard of regulatory capital, value at risk as a measure of risk has become more and more popular. It means that within a certain holding period and at a certain level of confidence, a particular asset portfolio of a financial institution is likely to suffer the most losses in the face of normal market volatility. Because this method of quantifying risk involves calculating specific numbers to measure the risk of an asset, the application of value at risk was popularized. According to the difference of volatility model and valuation model, the calculation methods can be divided into three types: Monte Carlo simulation method, historical simulation method and variance-covariance method. Among them, the historical simulation method assumes that the future trend of variables is the same as the historical trend; Monte Carlo simulation assumes that the variation of variables follows a given random model, of which the most common model for stock movements is the random walk; The variance-covariance method assumes that the variable's tendency to fluctuate has a specific regular distribution, such as a normal distribution.

2.3.2. Volatility analysis

Contrary to conventional wisdom, asset volatility is not constant and markets are not inefficient. For example, although it is commonly believed that the distribution of asset returns follows a normal distribution, the actual distribution contains the characteristics of a thick tail of spikes, a slightly wider tail, and points around the mean that are higher than those around the normal distribution. In addition, shocks from external markets can have a persistent impact on asset values, a phenomenon known as volatility clustering. There are two methods to estimate volatility: static method and dynamic method. The static method assumes that volatility remains constant for a specified period of time through the

variance of the sample period. On the other hand, dynamic estimation based on historical data can use GARCH model and moving average method.

3. Experimental results

This paper mainly studies the portfolio investment strategy of financial products based on statistical machine learning. Portfolio investment is a widely used investment strategy, and statistical machine learning has good performance and accuracy in portfolio investment. Previous studies mainly focused on the price rise and fall prediction of a single stock or risk security. This paper proposes a financial product portfolio investment strategy based on statistical machine learning, which calculates the return and risk degree of multiple financial products through machine learning technology and performs portfolio optimization. The effectiveness of the proposed method is verified by simulation experiments. The importance of portfolio investing, that is, reducing risk and achieving higher returns through diversification. Modern portfolio theory holds that investors should consider the expected return and risk of assets, and achieve the optimal investment effect through reasonable asset allocation. With the development of statistical machine learning technology, more and more researchers begin to apply it to portfolio investment, to improve investment efficiency and reduce risks by establishing predictive models for investment decisions. In this paper, random forest algorithm and XGBoost prediction are used to collect and pre-process data, and relevant financial product data, including historical return rate and risk indicators, are obtained from multiple channels. Then, the machine learning technology is used to calculate the income and risk degree of multiple financial products, and the combination optimization is carried out. Through simulation experiments, the effectiveness of the financial product portfolio investment strategy based on statistical machine learning is verified. The results show that this method has good reference significance, which can improve the investment efficiency and reduce the risk.

4. Related work

The financial market at home and abroad has experienced great fluctuations and changes, and the financial crisis has had a serious impact on the global financial market. Over time, financial markets recovered. There are many risks in the financial market. The risks are not evenly distributed in the financial market. The risk of investment can be reduced to improve the stability of global finance. By using minimum spanning tree and plane maximum filter graph, diversified stocks, increased the ratio between average return and standard deviation, and reduced the probability of negative return. And this approach has the added advantage of visually seeing portfolio selection rather than the graphical layout of the network [1]. The problem of systemic financial risk and analyzed the mechanism of financial systemic risk. They used a risk prediction model, RNN and LSTM to establish a reference model for exchange rate risk assessment in the financial investment market and realized the application of deep learning technology in systemic financial risk [2]. Nicolau Martin-Bassols then elaborated on the impact of financial investment from the perspective of cognitive ability in old age. He believed that cognitive level was positively correlated with risky and non-risky financial investment, even when combined with other important factors (such as education, labor status, age and family wealth). The explanatory power of cognition is also significant at the significance level of 0.1%. What's more, cognitive decline was only significantly associated with risky financial investments. This means that when individuals begin to suffer from age-related cognitive decline, they reduce their risky financial investments while not significantly changing their holdings of non-risky financial investments [3].

Machine learning can better solve the portfolio investment stock and reduce the investment risk. Many scholars have conducted in-depth research in this field. The random forest algorithm is proposed, which is a classifier composed of multiple decision trees, each of which depends on the value of a random vector and is equally distributed among all trees in the forest. Random forests can

be used for classification and regression analysis [4]. The price and quantity of each commodity traded in the market was an independent random variable at every moment, and he proposed a statistical extension of the microeconomic model, which was a continuation of the research on microeconomics [5].

With the development of machine learning, current machine learning and deep learning have become popular methods for financial data analysis, and future stock prediction is one of the most popular and complex deep learning. By using new methods, they can truly predict the future stock price trend more effectively and help investors make correct investment decisions [6].

Scholars argue that with advances in machine learning techniques, it is possible to incorporate predictive concepts into portfolio selection. They used a hybrid approach as a unique portfolio building technique to predict the value of stocks in the next period. The results show that the mean-VAR model predicted by AdaBoost is superior to other models [7].

The model HT-XGB combines the Hilbert-Yellow transform (HHT) as the feature engineering part and extreme Gradient Lift (XGBoost) as the closing price trend classifier. The categorical output is a rise and fall sequence that is used to optimize the stock portfolio weight with the best trading performance. Finally, it is shown that the performance of HT-XGB strategy is better than the benchmark [8]. Decision trees and discrete choice problems in economics, and proposed a new model tree Bayes that can solve economic considerations and model uncertainty problems in discrete choice problems, which is of great significance in promoting the field of statistics and machine learning [9]. A new method of using MTL in deep neural network architecture, which can effectively solve the momentum model of multi-task time series, and MTL can become a powerful tool in the financial field [10].

5. Conclusion

This paper mainly combines portfolio investment strategy with machine learning model. Portfolio investment is a widely used investment strategy, and statistical machine learning plays an important role in financial investment. Machine learning models are used to study portfolio investment, risk assessment and stock forecasting. Compared with traditional investment methods, this method has the following advantages:

(1) Data-driven: Statistical machine learning-based methods can use a large amount of historical data for training and forecasting, so as to more accurately analyze and predict the returns and risks of financial products.

(2) Portfolio optimization: This method can automatically select the best portfolio through the optimization algorithm to achieve the goal of minimizing risk and maximizing return.

(3) Consider multiple factors: Statistical machine learning methods can consider multiple factors at the same time, such as historical returns, risk indicators, market factors, etc., so as to evaluate and select a portfolio more comprehensively.

To sum up, the advanced feature of the method in this paper is that statistical machine learning is applied to the investment strategy of financial product portfolio, and more accurate, comprehensive and stable investment decisions are realized through data-driven and multi-factor consideration.

Although the paper has so much significance, but the machine learning technology still has a long way to go, we can continue to study the statistical machine learning financial products

(1) The return rate of portfolio investment is volatile, and the maximum test of return indicates that the stop loss point is not set prudently. To ensure the robustness and applicability of the model, we can reduce the test by refining the stop loss conditions and optimizing the model parameters on this basis.

(2) Application of Reinforcement Learning in portfolio optimization: Reinforcement learning is an algorithm that analyzes whether they have successfully reached the preset goal according to the previous decisions, so as to adjust the decision-making strategy. In portfolio optimization, the method

based on reinforcement learning can build adaptive dynamic investment strategies. Reinforcement learning can adjust current investment strategies when necessary to maximize risk-adjusted returns

(3) Model interpretability and stability research: While machine learning may have good predictive power, its "black box" nature brings unpredictable risks. How to improve the interpretability of models is an important topic in the application of machine learning in the financial field. At the same time, the stability of the model is also an important issue, and how to maintain the stability of the model performance under the changing market environment needs to be further studied.

Reference

- [1] Pozzi F, Di Matteo T, Aste T .Spread of risk across financial markets: better to invest in the peripheries [J].Scientific Reports, 2013.
- [2] Sun Y, Li J. Deep Learning for Intelligent Assessment of Financial Investment Risk Prediction [J]. Computational Intelligence and Neuroscience, 2022.
- [3] Bassols N M. Risky and non-risky financial investments and cognition [J]. Journal of Economic Psychology, 2023.
- [4] Statistics L B, Breiman L .Random Forests [J].Machine learning, 2001:5-32. Belal E.
- [5] Baaquie B E. Statistical microeconomics [J]. Physica A: Statistical Mechanicsand its Applications, 2013, 392(19): 4400-4416.
- [6] Li Y, Pan Y. A novel ensemble deep learning model for stock prediction basedon stock prices and news [J]. International Journal of Data Science and Analytics, 2022: 1-11.
- [7] Behera J, Pasayat A K, Behera H, ET al.Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets [J].Engineering Applications of Artificial Intelligence: The International Journal of Intelligent Real-Time Automation, 2023.
- [8] Dezhkam A, Manzuri M T.Forecasting stock market for an efficient portfolio by combining XGBoost and Hilbert-Huang transform [J].Engineering Applications ofArtificial Intelligence: The International Journal of Intelligent Real-Time Automation, 2023.
- [9] Brathwaite T, Vij A, Walker J L .Machine Learning Meets Microeconomics: The Case of Decision Trees and Discrete Choice.2017.
- [10] Ong J, Herremans D. Constructing time-series momentum portfolios with deep multi-task learning [J]. Expert Systems with Applications, 2023.