International Journal of Information Technology and Web Engineering
Volume 18 • Issue 1

Exploration on Portfolio Selection and Risk Prediction in Financial Markets Based on SVM Algorithm

Xinyu Han, Harbin University of Commerce, China
Dianqi Yao, The Hong Kong Polytechnic University, China*

ABSTRACT

In order to cope with the complex risk environment of the current financial market, achieve portfolio optimization and accurate risk prediction, this paper conducts effective research using SVM algorithm. This article uses stock data as a sample to empirically analyze the risk return and risk prediction performance of investment portfolio strategies based on SVM algorithm. Compared with traditional index fund investment strategies, the risk resistance of investment portfolio strategies is significantly improved, and the risk return is also stable at a high level. In addition, with the support of SVM algorithm, the risk prediction error level in the financial market remains within a relatively low range. From the perspective of practical applications, the financial market investment portfolio selection and risk prediction based on SVM algorithm has strong feasibility.

KEYWORDS

Financial Market, Portfolio Selection, Risk Prediction, Support Vector Machine Algorithm

INTRODUCTION

As economic globalization deepens, financial markets have evolved to offer diversified trading mechanisms, and a broad array of financial products have flourished. However, with this evolution comes increased complexity and heightened investment risks for investors. Traditional portfolio strategies often overlook the uncertainties of the financial market, hindering optimal asset allocation. Consequently, navigating these risks and optimizing portfolio choices are paramount challenges for investors. The rise of artificial intelligence has heralded significant advancements in machine learning. Among these, the support vector machine (SVM) algorithm stands out for its precision and stellar performance. In the realm of financial market portfolio selection and risk assessment, SVM adeptly discerns market trends and detects anomalies, thus facilitating effective risk mitigation and aiding investors in making optimal portfolio choices.

To address the intricate risk landscape of today's financial market and to achieve refined portfolio optimization and precise risk forecasting, this study employs the SVM algorithm. Through empirical analysis using stock datasets, we evaluate the risk-return profiles and predictive accuracy of investment

DOI: 10.4018/IJITWE.332777 *Corresponding Author

strategies guided by the SVM algorithm. When contrasted with conventional index fund strategies, our approach enhances the resilience of investment portfolios, ensuring a consistently elevated risk-return performance. Furthermore, the SVM algorithm aids in maintaining the risk prediction errors at a minimal range. Considering real-world implications, the utilization of the SVM algorithm for financial market portfolio selection and risk prediction proves highly viable.

LITERATURE REVIEW

With the development of the market economy, portfolio selection and risk prediction in financial markets have increasingly become the focus of scholars' attention. Through retrospective analysis, Aouni et al. (2018) applied methods and procedures to solve problems in portfolio selection with precision, and analyzed methods for solving problems in a multi-standard context, thereby expanding the characteristic impact of portfolio theory on mean variance. Zhang et al. (2020) proposed a cost-sensitive portfolio selection method with deep reinforcement learning, and analyzed, from a theoretical standpoint, the approximate optimal reward function proposed. Finally, based on empirical evaluation of the dataset, the effectiveness and superiority of the proposed method in terms of profitability, cost sensitivity, and presentation ability were proved. Miao (2018) employed a partial linear regression model to assess financial market stock data of varying lengths and forecasted risks associated with high-tech investments by enhancing the return and covariance matrix. Numerical examples indicated that the investment model, utilizing partial linear regression for risk prediction under inequality constraints, performs more effectively when confronted with inconsistent stock data lengths. Abensur and de Carvalho (2022) proposed an a priori classification of liquidity based on bid-ask spreads and a mathematical optimization model using liquidity as a defined participation constraint, and conducted simulation experiments using almost 20 years of data from the US and Brazilian stock exchanges. The results showed that the proposed method can lead to more successful investment portfolios. In recent years, portfolio selection and risk forecasting have made significant progress. However, with the complexity of today's financial markets, continued improvement and optimization remain necessary in portfolio selection and risk forecasting. Existing research inadequately addresses the dynamic nature and predictive accuracy challenges within the financial market.

The SVM algorithm can flexibly handle low-dimensional and high-dimensional data, with high prediction accuracy. Endri et al. (2020) developed an early warning system model that can predict the delisting risk of Islamic stocks through the SVM algorithm. To verify the effectiveness of the algorithm, they analyzed 102 companies registered in the Islamic financial market between 2012 and 2017. The results showed that the accuracy of the SVM model can reach 100%. Kurani et al. (2023) discussed issues such as financial risk management and stock prediction, and proposed SVMs and artificial neural networks as prediction algorithms. Finally, experiments have proven that SVMs and artificial neural networks have played a prominent role in risk management and risk prediction (Kurani et al., 2023). With the assistance of the SVM algorithm, portfolio selection and risk prediction in financial markets have achieved significant development, yet most studies have not taken into account practical issues in investment decision-making in order to provide more effective guidance for portfolio optimization and risk avoidance.

FINANCIAL MARKET PORTFOLIO SELECTION AND RISK FORECASTING UNDER THE SVM ALGORITHM

Portfolio Selection and Risk Prediction

Core Content

The *portfolio selection problem* refers to the asset selection and allocation activities in which investors engage to achieve an effective balance between income and risk under uncertain conditions (Breunig et al., 2021). In the financial realm, choosing and allocating assets is pivotal for investor decision-making. Time and random uncertainties play a crucial role in the use of assets. The objective is to fine-tune an investment portfolio and consumption strategy for participants' assets, ensuring it aligns with the maximization of investors' best interests.

There are several common methods of portfolio selection:

Balanced Portfolio

This is a simple investment strategy where funds are evenly distributed among different assets or asset classes. This method can help reduce the risk of specific assets, but may not maximize returns.

Maximizing Sharpe Ratio Portfolio

Sharpe ratio measures the balance between the return and risk of an investment portfolio. The strategy of maximizing Sharpe ratio aims to achieve the highest return at a given level of risk.

Minimizing Variance Portfolio

This strategy aims to minimize the variance or volatility of the portfolio in order to achieve relatively stable returns. It is suitable for investors who wish to reduce portfolio volatility.

Capital Market Line

Capital Market Line is a portfolio selection method that combines risk-free assets (usually Treasury bills, also known as T-bills) with high-risk assets (such as stocks) to achieve specific risk and return goals.

Factor Investment Strategy

This strategy constructs an investment portfolio based on specific market factors such as market value, risk, quality, and momentum. The factor investment strategy aims to capture the return premium of these factors.

Machine Learning and Quantitative Investment Strategies

Quantitative investment strategies based on machine learning algorithms and mathematical models are becoming increasingly popular. These strategies utilize big data and algorithms to identify patterns and trading opportunities. The research method of this article is a form of quantitative investment strategy, which is a method that utilizes machine learning and mathematical models to guide investment decisions.

Risk is a fundamental attribute of financial activities and the financial system. It is defined as events and behaviors that have a negative impact on an organization's achievement of its set goals or implementation of its set strategies. Various portfolio activities in the financial market come with inherent financial risks (Kouaissah & Hocine, 2021). How to accurately forecast risks, effectively mitigate them, optimize transaction costs, and refine investment portfolios are the primary concerns of investors (Thakur et al., 2018).

Risk prediction is one of the key tasks in the financial and investment fields, and multiple methods can be used to predict risk. The following are common risk prediction methods:

Historical Volatility

This method calculates volatility by analyzing historical price data of assets or investment portfolios, in order to estimate future risk levels. The historical volatility method is simple and easy to use, but it ignores potential non-randomness.

Monte Carlo Simulation

This method uses probability distributions to simulate various future market scenarios and estimate the level of risk in these scenarios. Monte Carlo simulation is suitable for risk analysis that considers uncertainty and complexity.

Generalized Autoregressive Conditional Heteroskedasticity Simulation

This is a time series model used to model the volatility of asset prices. It can capture the temporal changes in volatility and provide estimates of future risks.

Machine Learning Methods

Machine learning algorithms such as support vector machines (SVM), random forests, and neural networks can be used for risk prediction. They can handle nonlinear relationships and a large number of features, providing more accurate risk estimates.

Factor Models

Factor models use market factors and other macro or micro factors to explain and predict the risks of assets or investment portfolios. Common factor models include the capital asset pricing model and the Fama French three factor model.

These methods can be used individually or in combination, depending on research or investment needs, as well as the performance of available data and models. Different methods may be applicable to different types of assets and market environments.

Realistic Demand for Portfolio Selection

Financial Market Development

With the rapid development of the social economy, the scale of the financial market is likewise constantly expanding, which in turn promotes economic prosperity, and also creates a positive and profound impact on people's social lives (Li et al., 2018). At present, trading venues represented by the stock market, bond market, fund market, and various other financial asset markets have formed a relatively sound financial system encompassing a diversity of products (He & Strub, 2022). At the same time, the interaction and impact between various financial products constitute a close and complex relationship in the financial market, which puts forward new requirements for the formulation of financial market management policies and the scientific investment behavior of investors. Therefore, for managers and investors, it is crucial to formulate a correct response to the interaction and risk changes between financial products in the financial market (Nalepa & Kawulok, 2019).

The Random Time-Variance of Investment Opportunity Set

In the financial market, participants' investment is frequently of a dynamic nature with multiple stages. Most investors tend to focus only on the average rate of return and its dispersion for a certain asset during a certain period, while ignoring the random time-varying nature of the investment opportunity set during that period (Liu et al., 2020). They are more inclined to focus on the impact of risk on current wealth and the impact of risk on investment opportunities. In fact, investment opportunities change over time. Investors need to improve their portfolios to avoid the intertemporal risks they bring (Raj & Ananthi, 2019).

Limitations of Portfolio Selection Strategies

The existing dynamic portfolio selection methods in the current market suffer from limitations. There are many representative empirical factual features in the allocation of asset returns in the financial market, but there remains a lack of dynamic portfolio selection strategies based on these empirical factual features (Ghaddar & Naoum-Sawaya, 2018). Additionally, in practical scenarios, institutional investors should factor in their entire life cycle of investment and consumption while striving for an optimal portfolio, by means of momentum strategies, buying winners' portfolios, and selling losers' portfolios (Padhi et al., 2022). Most of the existing portfolio selection strategies only focus on the selection behavior of a single investor when facing a group of investment opportunities (Behera et al., 2023). Few strategies examine the selection preferences of various investors within identical investment opportunities, leading to a lack of comprehensive approaches (Song et al., 2017).

Specific Performance of Investment Risks

In the financial market, the specific performance of portfolio selection risk can be divided into five aspects:

Objectivity

Investment risk arises from uncertainties in the financial market and is an inherent characteristic of the ongoing development and evolution of assets in the market. Investors can adjust the conditions for the generation and development of investment risks in a specific region by selecting investment risks, thereby reducing the probability of risk events and reducing the degree of loss, but they cannot completely eliminate the risks (Chen et al., 2021).

Uncertainty

For individual investors, the risk events of investment activities carry a degree of uncertainty, which depends on the random characteristics of their financial assets within a constantly shifting market (Thakkar & Chaudhari, 2021). The uncertainty of risk is mainly reflected in various accidental factors—that is, whether, when, and how investment risks may occur, as well as how much loss they may cause.

Measurability

Based on the statistical data of a series of evolutions and risk events that have occurred during the historical period of the financial market, and through effective analysis of these data, it is possible to measure the frequency of investment risks and the degree of losses incurred, thereby predicting and avoiding possible risks. Examples of risk events include the Great Depression of 1929, Black Monday of 1987, and the financial crisis of 2008 caused by the stock market crash. By analyzing historical stock market data, the frequency and degree of losses of market crashes can be observed. These data analyses help investors to understand the potential risks of stock market crashes and take appropriate risk management measures. In addition, investors can analyze historical bond market data to reveal the occurrence of bond default events. By studying the default rates and loss levels of different types of bonds, investors can better evaluate the risks of bond investments and choose appropriate credit quality bonds.

Dichotomy

The risks of portfolio selection may cause investors to bear significant losses, but these risks may also enable investors to obtain valuable risk returns. Such a situation, where risks and opportunities coexist, demonstrates that investors cannot only use negative methods when facing risks, but must also integrate social, economic, and other factors at multiple levels (Yu et al., 2005). Analysis and evaluation of possible adverse factors can transform them into favorable factors. Only in this way can risks and opportunities be fully utilized in the investment process to achieve maximum returns.

Relevance

Relevance refers to the fact that the risks borne by investors in the investment process are closely related to their investment behavior and decision-making. Risk varies with investors, investment decisions, and time. Essentially, risk space includes decision space and state space. State space is an objective necessity (Yu et al., 2008). In the decision-making space, individuals can make decisions independently, thus determining the type and magnitude of risks faced by individuals.

SVM Algorithm

Support Vector Machine (SVM) is a machine learning algorithm used for classification and regression, which separates data points of different categories or establishes prediction models by finding the best hyperplane (Gadre-Patwardhan et al., 2016). Since the method was first proposed in the 1990s, it has received much attention from scholars. In recent years, the SVM algorithm has made great breakthroughs both theoretically and algorithmically, and has gradually become an effective means to solve problems such as "dimension disaster" and "over learning" (Patalay & Bandlamudi, 2020). This method has strong computing power and high computational efficiency, and has become a hot topic in machine learning research. As one of the most widely used methods in the field of machine learning, it has strong advantages in dealing with classification problems and regression problems. The biggest feature of the SVM algorithm is that it breaks through the empirical minimization principle and is established based on the structural minimization principle, so its adaptability to fresh samples is robust. In machine learning, the data distribution of many problems is nonlinear, meaning that data points cannot be perfectly separated by a single linear decision boundary (hyperplane) in the feature space. Traditional linear classifiers may not be able to effectively solve these problems. One of the strengths of SVM is that it can handle nonlinear problems and transform them into high-dimensional linear problems to achieve better classification performance. SVM maps data points from the original feature space onto a higher-dimensional feature space using kernel function techniques. This mapping can transform nonlinear problems into linear problems in high-dimensional feature spaces. In high-dimensional feature spaces, the decision boundary of SVM can be represented as the inner product (dot product) operation of feature vectors. This inner product operation allows SVM to effectively handle nonlinear problems without explicitly calculating data points in high-dimensional feature spaces. This allows it to flexibly solve complex problems, while effectively avoiding local minimum problems caused by "dimensional disasters" (Yang et al., 2020).

In addition to the SVM algorithm, many other machine learning techniques can be applied to portfolio selection and risk prediction in financial markets. The following are several common machine learning technologies:

Decision Trees

Decision trees are a very intuitive and easy-to-understand machine learning model. They predict and classify by constructing a series of decision rules. In the financial field, decision trees can be used to select investment portfolios or predict the trend of stock prices.

Random Forests

Random forests are ensemble learning methods based on decision trees. They improve overall prediction performance by combining the prediction results of multiple decision trees. In financial markets, random forests can be used to create portfolio models, as well as for risk prediction and trading decisions.

Neural Networks

Neural networks are machine learning models that simulate the workings of neurons in the human brain. In the financial field, neural networks can be used to capture complex nonlinear relationships and predict stock prices and market trends.

Support Vector Machines

Support vector machines, the tool applied in this article, are widely used in financial markets. They can be applied to classification problems, such as predicting stock price fluctuations or industry classification. SVM can handle high-dimensional data and nonlinear relationships, and they display good generalization performance.

Bayesian Networks

Bayesian networks model complex systems by specifying conditional dependencies between variables. In the financial market, Bayesian networks can be used to infer the potential risk and return relationship of investment portfolios, and can be applied in risk prediction and decision-making.

In the process of portfolio selection and risk prediction, a large number of complex uncertain factors are involved, and there is no clear functional relationship between each factor. Factors such as market volatility, company performance, macroeconomic environment, political risk, currency risk, black swan events, market sentiment, and changes in interest rates are all uncertain factors in portfolio selection and risk prediction. Therefore, in the process of solving the problem of risk, there are often significant requirements for the accuracy of the mathematical model constructed and the adequacy of learning samples. The SVM algorithm is very suitable for solving such problems. It is a data-driven empirical modeling method, and the structure of the system is completely unknown. It primarily uses a large quantity of input and output data for modeling. In establishing the functional form and error distribution type of the model, it does not excessively rely on a great amount of prior knowledge and settings. In summary, the authors chose SVM as the basis for our study by considering the following factors:

Data Characteristics

We used stock market data, which is usually high-dimensional, nonlinear, and highly noisy. The SVM algorithm performs well in handling this type of data, especially in nonlinear classification and regression problems.

Explanatory

The SVM algorithm has strong model interpretability, which is of great importance for financial market research. Investors and decision-makers are generally more willing to employ methods that provide transparency regarding the reasons for model decisions.

Data Efficiency

Financial market data typically contains a large number of samples and features, but we may only be interested in a small subset of these features. The SVM algorithm can selectively focus on important features through kernel functions, improving computational efficiency.

This paper applied the regression theory of the SVM algorithm to financial markets in order to establish a portfolio selection and risk prediction model. We posit that there are m stocks in a financial market with n trading cycle. The closing price of m stocks in the T trading period is represented by vector $\mathrm{p_T} = \left(\mathrm{p_T^1}, \mathrm{p_T^2}, \cdots, \mathrm{p_T^m}\right)$. It is assumed that element $\mathrm{p_T^j}$ represents the closing price of the j-th stock in period T. Before the beginning of the T-th trading period, the SVM algorithm allocates the total investment amount to m stocks based on the proportion of the portfolio vector $\mathrm{v_T} = \left(\mathrm{v_T^1}, \mathrm{v_T^2}, \cdots, \mathrm{v_T^m}\right)^{\mathrm{T}} \in \mathrm{R_+^m}$. It is presumed that element $\mathrm{v_T^j}$ represents the proportion of total wealth invested in the j-th stock in period T, and the sum of all elements of $\mathrm{v_T}$ is 1. The overall process of online portfolio selection can be expressed as $v_T : \left(R_+^m\right)^{T-1} \rightarrow \Delta_m, T = 1, 2, \cdots$. The goal of the SVM algorithm portfolio is to design an investment risk prediction function that maximizes cumulative returns.

We used a computer equipped with an Intel Core i7 processor and 16GB of memory. The operating system running on this computer is Windows 10 and uses the Python programming language to implement SVM algorithms and other data analysis tasks. We used the Scikit-Learn library as a machine learning tool and utilized its support vector machine classifier for empirical analysis. Before using the Scikit-Learn library, this article standardized the data and used the StandardScaler class in the preprocessing module for feature scaling operations. Then, using the model _ Train in the selection module _ Test_, the split function divides the dataset into training and testing sets. It is assumed that a training sample set for investment risk prediction has been established in a financial market—this paper represents it as $\left\{\left(\mathrm{m_i}, \mathrm{n_i}\right), \mathrm{i} = 1, 2, \cdots, \mathrm{s}\right\}$, and $\mathrm{m_i} \in \mathrm{T^d} = \left(\mathrm{m_i^1}, \mathrm{m_i^2}, \cdots, \mathrm{m_i^d}\right)$, $\mathrm{n_i} \in \mathrm{T}$. The definitions of each parameter are shown in Table 1.

The insensitive loss function is set to:

Table 1. Interpretation of training sample set parameters

Sequence	Parameter	Meaning
1	$\mathrm{m_i}$	Risk indicator value vector of the item i investment project (input values for training SVM)
2	D	Index dimension of investment risk influencing factors
3	$\mathrm{n_i}$	Risk value of item i investment project (target output value of training SVM)

$$\left|\mathrm{n}-\mathrm{g}\left(\mathrm{m}\right)\right|_{\prime} = \begin{cases} 0 & \left|\mathrm{n}-\mathrm{g}\left(\mathrm{m}\right)\right| \le \delta \\ \left|\mathrm{n}-\mathrm{g}\left(\mathrm{m}\right)\right| - \delta & \left|\mathrm{n}-\mathrm{g}\left(\mathrm{m}\right)\right| > \delta \end{cases} \tag{1}$$

$\prime$ is a standard normal number, which is a randomly given value.

Secondly, the appropriate kernel function is defined as:

$$\mathrm{k}\left(\mathrm{m_i}, \mathrm{m_j}\right) = \mathrm{f}\left(\mathrm{m_i}\right) \cdot \mathrm{f}\left(\mathrm{m_j}\right) \tag{2}$$

where $\mathrm{f}\left(\mathrm{m_i}\right)$ and $\mathrm{f}\left(\mathrm{m_j}\right)$ are nonlinear functions. Under these conditions, the investment risk prediction function h(m) can be expressed as:

$$\mathrm{h}\left(\mathrm{m}\right) = \mathrm{w} \cdot \mathrm{f}\left(\mathrm{m_i}\right) + \mathrm{b} \tag{3}$$

In Formula (3), the definitions of each parameter are shown in Table 2.

Among them, $\mathrm{w} \in \mathrm{T^d}$, and $\mathrm{b} \in \mathrm{T}$. After introducing the relaxation variable σ_i, σ_i^*, the optimization problem is obtained:

$$\min_{w,b,\sigma_i,\sigma_i^*} w^2 / 2 + L \sum_{i=1}^{s} \left(\sigma_i + \sigma_i^*\right) \tag{4}$$

$$s.t. n_i - \left(w \cdot f\left(m_i\right) + b\right) \le \delta + \sigma_i \tag{5}$$

$$\left(w \cdot f\left(m_i\right) + b\right) - n_i \le \delta + \sigma_i^* \tag{6}$$

Table 2. Parameter definition of Formula (3)

Sequence	Parameter	Meaning
1	w	Weight vector
2	b	Threshold
3	$\left(\cdot\right)$	Inner product

Among them, $\sigma_i, \sigma_i^* \geq 0$, and $i = 1, 2, \cdots, s$.

The dual optimization problem can be obtained by using Lagrange transformation, namely:

$$Max\left[-\frac{1}{2}\sum_{i=1}^{s}\sum_{j=1}^{s}\left(u_i - u_i^*\right)\left(u_j - u_j^*\right)k\left(m_i, m_j\right) - \delta\sum_{i=1}^{s}\left(u_i + u_i^*\right) + \sum_{i=1}^{s}n_i\left(u_i - u_i^*\right)\right] \tag{7}$$

$$s.t.\sum_{i=1}^{s}\left(u_i - u_i^*\right) = 0 \tag{8}$$

and there are:

$$0 \leq u_i, u_i^* \leq L \tag{9}$$

Among them, L is the penalty parameter. Before solving the model, a positive number can be given to solve the dual optimization problem. Finally, $w = \sum_{i=1}^{s}\left(u_i - u_i^*\right)f\left(m_i\right)$ can be obtained, and then Formula (10) is calculated:

$$b = \frac{1}{N_{Nr}}\left\{\sum_{0 \leq u_i^* \leq L}\left[n_i - \sum_{m_j \subset r}\left(u_j - u_j^*\right)k\left(m_j, m_i\right) - \delta\right] + \sum_{0 \leq u_i^* \leq L}\left[n_i - \sum_{m_j \subset r}\left(u_j - u_j^*\right)k\left(m_j, m_i\right) + \delta\right]\right\} \tag{10}$$

The threshold value b can be obtained; and the investment risk prediction function obtained through learning is:

$$h\left(m\right) = \sum_{m_j \subset r}\left(u_i - u_i^*\right)k\left(m_i, m\right) + b \tag{11}$$

After establishing the investment risk prediction function, it is necessary to set, in advance, the error test function and the investment project risk prediction accuracy standard value. The error test function and the standard value of investment project risk prediction accuracy are important tools for evaluating and verifying established investment risk prediction models. The error test function is a function used to measure the difference or error between the predicted results of a model and the actual observed values. It can help determine the predictive performance and accuracy of the model. This article uses Root Mean Square Error (RMSE) to detect errors. The standard value of investment project risk prediction accuracy is the prediction accuracy standard that you expect the model to achieve in practical application. This standard value is usually set based on specific business needs and risk tolerance. Once the standard values are set, you can use error testing functions to evaluate the performance of the model. Based on the test sample set, the final effect of the prediction function established by the algorithm model function in actual prediction is tested to verify whether the prediction function can meet the needs of practical application. If the final prediction accuracy does not meet the standard value, it is necessary to adjust parameters and conduct multiple tests to ensure that the prediction accuracy of the function meets the demand. The function obtained by meeting the standard value of prediction accuracy is the optimal prediction function, which is expressed as:

$$h^*(m) = \sum_{m_j \subset r} (u_i - u_i^*) k^*(m_i, m) + b^* \tag{12}$$

In the process of problem solving, the RMSE function is generally used to test the prediction accuracy of the SVM model. The calculation process of RMSE, which is used to evaluate the prediction effect of the prediction model on each test sample, is expressed as follows:

$$RMSE = Sqrt\left(\frac{1}{N}\sum_{n=1}^{N}\left(m(n, true) - m(n, pred)\right)^2\right) \tag{13}$$

In Formula (13), Sqrt represents square root calculation.

EMPIRICAL EVIDENCE ON PORTFOLIO SELECTION AND RISK PREDICTION IN FINANCIAL MARKETS

In order to verify the effectiveness of portfolio selection and risk prediction in financial markets based on the SVM algorithm, this paper analyzed the evidence from two perspectives: risk return and risk prediction. This article uses stock data from 500 listed companies in the United States as a sample to analyze the risk return and risk prediction performance of investment portfolio strategies based on the SVM algorithm. The basic information of this dataset is shown in Table 3. To ensure the objectivity of the experimental results, this article compared them with the investment strategies of index funds commonly used in the current financial market. From Table 3, this dataset covers the stock data of 500 listed companies in the United States from February 8, 2013 to February 7, 2018, with a sample size of 619,040. The information includes the observation date, opening price, maximum transaction price, closing price, and trading volume. This article used the dataset from January 12, 2016 to June 12, 2016 as the investment observation date, and used the data from that date to form a risk prediction sample dataset, with a total of 5490 sample data.

Risk Return

Risk return refers to the return that investors receive from taking risks in portfolio selection that exceeds the time value of capital. Generally speaking, the better the risk return of an investment strategy, the stronger the resistance of the strategy to risk. This article selects the traditional index

Table 3. Basic information of datasets

Sequence	Item	Information	Mathematical Types
1	Start date-	2013-02-08	Date Time
2	End date	2018-02-07	Date Time
3	Number of samples	619040	Integer
4	Data content	Observation date	Date Time
		Opening price	Floating point value
		Highest transaction price	Floating point value
		Closing price	Floating point value
		Trading volume	Floating point value

fund investment strategy as the baseline, because it is a widely accepted standard strategy in the financial field. Therefore, using it as a baseline can help researchers establish a universal reference point in experiments, to evaluate the performance of new strategies and help evaluate whether the new strategy can surpass or improve upon traditional methods. In addition, comparing the SVM algorithm strategies with traditional index fund investment strategies can be helpful for intuitively understanding the advantages or disadvantages of SVM algorithm strategies. If the SVM strategy performs better in terms of volatility, this may provide stronger evidence that it offers greater stability in the face of risk. This article has compared the risk returns of portfolio strategies based on the SVM algorithm with index fund investment strategies on the dataset. The risk returns of the two types of strategies used the Sharpe ratio as a standardized indicator for performance evaluation. The results are shown in Figure 1.

Figure 1(A) shows the Sharpe ratio results of the SVM algorithm strategy, and Figure 1(B) shows the Sharpe ratio results of the index fund investment strategy.

If the Sharpe ratio is positive, it indicates a high return on the portfolio strategy; if the ratio is negative, it indicates that the risk of the portfolio strategy is greater than the return. Therefore, the higher the Sharpe ratio, the greater the risk return that the portfolio can receive in the financial market. From Figure 1, it can be observed that under different time series, there are significant differences in the Sharpe ratio results between the two strategies. In Figure 1(A), the Sharpe ratio of the strategy based on the SVM algorithm under different time series was in the range of 0.1 to 0.3; in Figure 1(B), the Sharpe ratio of index fund investment strategies under different time series was in the range of - 0.1 to 0.2. From the comparison of results, it can be understood that the risk returns of index fund investment strategies were not ideal. In the time series, there were negative values, indicating that under this strategy, the unit risk assumed by the portfolio exceeded its return level. However, the portfolio strategy based on the SVM algorithm remained within a positive range, and its overall risk return was considerable.

In order to more intuitively reflect the stability of the two types of strategies in resisting risks, the authors of this article compared their volatility in the dataset. The volatility was calculated from the standard deviation of the portfolio price changes under both strategies during the investment observation date. The specific values are shown in Figure 2.

Figure 1. (A) Sharpe ratio result, (B) Sharpe ratio result

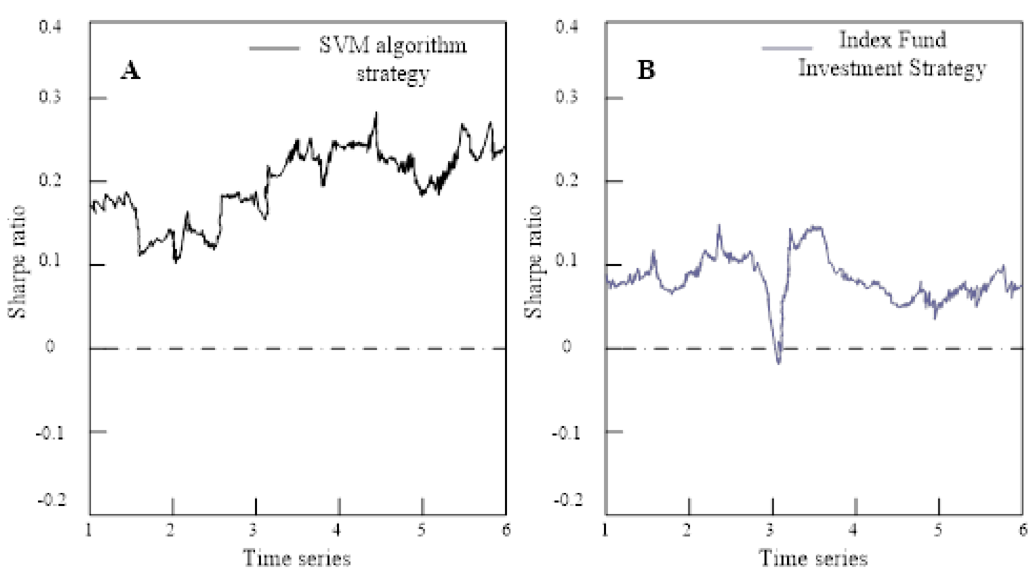

Figure 2. (A) Volatility results It is supposed there are m stocks in (B) volatility results

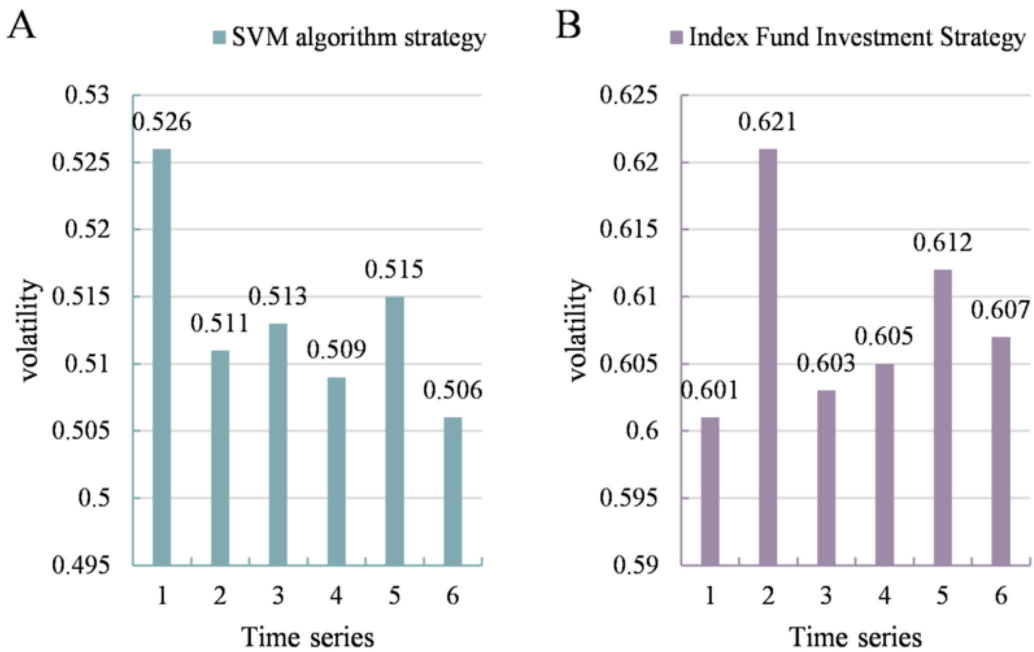

Figure 2(A) shows the volatility results of the SVM algorithm strategy, and Figure 2(B) shows the volatility results of the index fund investment strategy.

Volatility reflects the dispersion of portfolio prices in a specific time series. When the volatility result is large, it indicates that the greater the distance between portfolio yield and historical average, the weaker the risk resistance of the strategy. When the volatility result is small, it indicates that the return of the portfolio is relatively stable, and the stability of the strategy against risk is relatively strong. From the results in Figure 2, the volatility results of the algorithm strategy in this article were generally lower than the volatility of traditional index fund investment strategies. In Figure 2(A), the highest volatility generated by the strategy based on the SVM algorithm during the investment observation period was 0.526, and its average volatility was about 0.513. In Figure 2(B), the highest volatility generated by the investment strategy of an index fund during the investment observation period was 0.621, with an average volatility of about 0.608. In comparing results, it can be seen that the portfolio price dispersion under the algorithm strategy in this paper was relatively small, and the strategy had a strong ability to resist risks in the financial market. However, traditional index funds invest in portfolios with large price dispersion and weak strategic stability, making it difficult to effectively resist the risks that may exist in the financial market.

Risk Prediction Effect

In risk prediction, this research used the stock dataset from January 12, 2016 to June 12, 2016 to form a risk prediction sample dataset, with a total of 5490 sample data. The training set and test set were divided at a ratio of 7:3, with a total of 3843 pieces of training set data and 1647 pieces of test set data. This research then conducted portfolio risk prediction for the SVM algorithm in the training set and test set, and calculated the RMSE results of monthly data prediction, as shown in Figure 3.

Figure 3. (A) RMSE results of risk prediction, (B) RMSE results of risk prediction

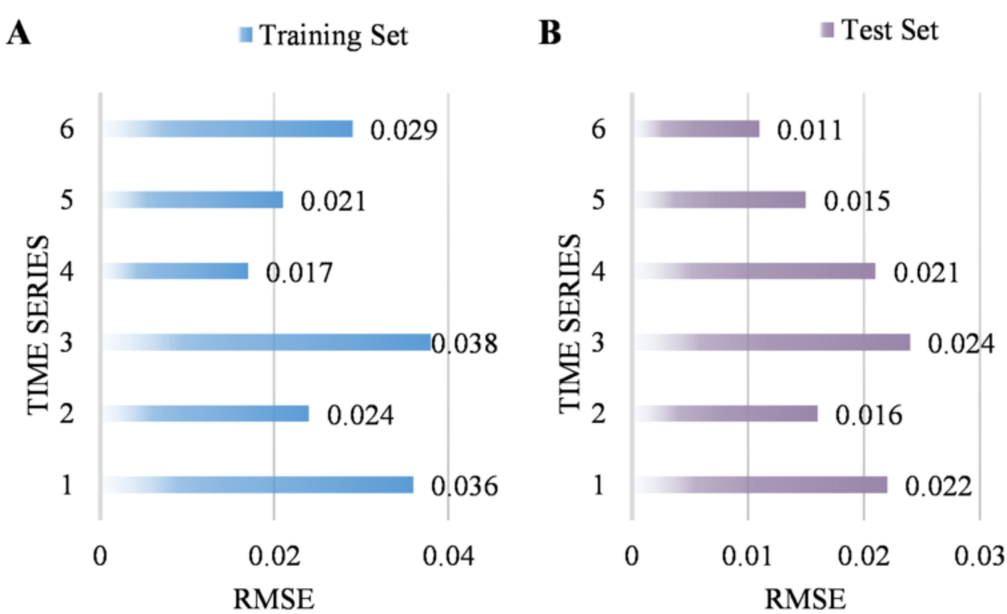

Figure 3(A) shows the RMSE results for the training set, and Figure 3(B) shows the RMSE results for the test set.

In Figure 3(A), the maximum value of RMSE results on the training set was 0.038, and the minimum value was 0.017. In Figure 3(B), the maximum RMSE result on the test set was 0.024 and the minimum value was 0.011. Based on the results, the SVM algorithm displays a high level of risk prediction accuracy, with a small difference between the actual value and the predicted value, and a low level of error. Moreover, the risk prediction results for the training set and the test set show little divergence. This indicates that the SVM algorithm has a strong fitting ability. For data under different time series, the SVM algorithm is able to accomplish effective prediction of risk, thus providing strong protection against objective and uncertain risk factors present in financial markets, and enabling effective decisions to achieve optimal allocation of investment portfolios.

CONCLUSION

Risk is the main factor that affects investment decisions and changes in investment returns. As an important locus for investment activities, financial markets are facing increasingly complex risk environments with the rapid development of the economy. In order to effectively avoid investment risks, assist investors in scientific investment, and achieve the optimal investment portfolio, these authors conducted in-depth research on financial market portfolio selection and risk prediction using the SVM algorithm. In terms of risk return and risk prediction, the SVM algorithm can not only formulate good portfolio strategies through flexible and effective learning methods, but also accurately predict potential risks in the market. As a result, an investment portfolio can achieve better risk returns on the basis of stable risk resistance. Although the research on portfolio selection and risk prediction in financial markets based on the SVM algorithm in this paper can promote the healthy development of financial markets to a certain extent, there are still some shortcomings and problems in the research process that call for improvement. For example, data quality issues, sample imbalance issues, risk management issues, and situations where the model is difficult to explain. In future research, these shortcomings and problems would be continuously addressed, and the depth of application of the algorithm would be enhanced, so as to promote the scientific development of investment activities in financial markets.

ACKNOWLEDGMENT

This work was not supported by any funds. The authors declare they have no conflict of interest.

The authors offer their sincere thanks to the designers of those techniques that contributed to this research.

REFERENCES

Abensur, E. O., & de Carvalho, W. P. (2022). Improving portfolio selection by balancing liquidity-risk-return: Evidence from stock markets. *Theoretical Economics Letters*, *12*(2), 479–497. doi:10.4236/tel.2022.122027

Aouni, B., Doumpos, M., Pérez-Gladish, B., & Steuer, R. E. (2018). On the increasing importance of multiple criteria decision aid methods for portfolio selection. *The Journal of the Operational Research Society*, *69*(10), 1525–1542. doi:10.1080/01605682.2018.1475118

Behera, J., Pasayat, A. K., Behera, H., & Kumar, P. (2023). Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. *Engineering Applications of Artificial Intelligence*, *120*, 105843. doi:10.1016/j.engappai.2023.105843

Breunig, C., Huck, S., Schmidt, T., & Weizsäcker, G. (2021). The standard portfolio choice problem in Germany. *Economic Journal (London)*, *131*(638), 2413–2446. doi:10.1093/ej/ueab006

Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean-variance portfolio optimization using machine learning-based stock price prediction. *Applied Soft Computing*, *100*, 106943. doi:10.1016/j.asoc.2020.106943

Endri, E., Kasmir, K., & Syarif, A. J. D. S. L. (2020). Delisting sharia stock prediction model based on financial information: Support Vector Machine. *Decision Science Letters*, *9*(2), 207–214. doi:10.5267/j.dsl.2019.11.001

Gadre-Patwardhan, S., Katdare, V. V., & Joshi, M. R. (2016). A review of artificially intelligent applications in the financial domain. In C. L. Dunis, P. W. Middleton, A. Karathanasopolous, & K. Theofilatos (Eds.), *Artificial Intelligence in Financial Markets: Cutting Edge Applications for Risk Management, Portfolio Optimization and Economics* (pp. 3–44). Springer. doi:10.1057/978-1-137-48880-0_1

Ghaddar, B., & Naoum-Sawaya, J. (2018). High dimensional data classification and feature selection using support vector machines. *European Journal of Operational Research*, *265*(3), 993–1004. doi:10.1016/j.ejor.2017.08.040

He, X. D., & Strub, M. S. (2022). How endogenization of the reference point affects loss aversion: A study of portfolio selection. *Operations Research*, *70*(6), 3035–3053. doi:10.1287/opre.2022.2309

Kouaissah, N., & Hocine, A. (2021). Forecasting systemic risk in portfolio selection: The role of technical trading rules. *Journal of Forecasting*, *40*(4), 708–729. doi:10.1002/for.2741

Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Annals of Data Science*, *10*(1), 183–208. doi:10.1007/s40745-021-00344-x

Li, B., Wang, J., Huang, D., & Hoi, S. C. (2018). Transaction cost optimization for online portfolio selection. *Quantitative Finance*, *18*(8), 1411–1424. doi:10.1080/14697688.2017.1357831

Liu, L., Chu, M., Gong, R., & Zhang, L. (2020). An improved nonparallel support vector machine. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(11), 5129–5143. doi:10.1109/TNNLS.2020.3027062 PMID:33055038

Miao, R. (2018). High technology investment risk prediction using partial linear regression model under inequality constraints. *Journal of Interdisciplinary Mathematics*, *21*(4), 869–881. doi:10.1080/09720502.2018.1475067

Nalepa, J., & Kawulok, M. (2019). Selecting training sets for support vector machines: A review. *Artificial Intelligence Review*, *52*(2), 857–900. doi:10.1007/s10462-017-9611-1

Padhi, D. K., Padhy, N., Bhoi, A. K., Shafi, J., & Yesuf, S. H. (2022). An intelligent fusion model with portfolio selection and machine learning for stock market prediction. *Computational Intelligence and Neuroscience*, *2022*, 2022. doi:10.1155/2022/7588303 PMID:35785077

Patalay, S., & Bandlamudi, M. R. (2020). Stock price prediction and portfolio selection using artificial intelligence. *Asia Pacific Journal of Information Systems*, *30*(1), 31–52. doi:10.14329/apjis.2020.30.1.31

Raj, J. S., & Ananthi, J. V. (2019). Recurrent neural networks and nonlinear prediction in support vector machines. *Journal of Soft Computing Paradigm*, *1*(1), 33–40. doi:10.36548/jscp.2019.1.004

Song, Q., Liu, A., & Yang, S. Y. (2017). Stock portfolio selection using learning-to-rank algorithms with news sentiment. *Neurocomputing*, *264*, 20–28. doi:10.1016/j.neucom.2017.02.097

Thakkar, A., & Chaudhari, K. (2021). Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions. *Information Fusion*, *65*, 95–107. doi:10.1016/j.inffus.2020.08.019 PMID:32868979

Thakur, G. S. M., Bhattacharyya, R., & Sarkar, S. (2018). Stock portfolio selection using Dempster-Shafer evidence theory. *Journal of King Saud University-Computer and Information Sciences*, *30*(2), 223–235. doi:10.1016/j.jksuci.2016.07.001

Yang, R., Yu, L., Zhao, Y., Yu, H., Xu, G., Wu, Y., & Liu, Z. (2020). Big data analytics for financial market volatility forecast based on support vector machine. *International Journal of Information Management*, *50*, 452–462. doi:10.1016/j.ijinfomgt.2019.05.027

Yu, L., Chen, H., Wang, S., & Lai, K. K. (2008). Evolving least squares support vector machines for stock market trend mining. *IEEE Transactions on Evolutionary Computation*, *13*(1), 87–102.

Yu, L., Wang, S., & Lai, K. K. (2005). Mining stock market tendency using GA-based support vector machines. In *Proceedings of the First International Conference on Internet and Network Economics*. Springer Berlin Heidelberg. doi:10.1007/11600930_33

Zhang, Y., Zhao, P., Wu, Q., Li, B., Huang, J., & Tan, M. (2020). Cost-sensitive portfolio selection via deep reinforcement learning. *IEEE Transactions on Knowledge and Data Engineering*, *34*(1), 236–248. doi:10.1109/TKDE.2020.2979700