International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-11

https://doi.org/10.70917/ijcisim-2026-0122
Article

Research on Support Vector Regression-based Profit Forecasting Model for Enterprise Economic Analysis

Yuan Wang *

Faculty of Economics and Management, Shanxi Normal University, Taiyuan 030031, Shanxi, China; wy1446339240@163.com

Abstract: Profit forecasting based on corporate economics serves as an important indicator of market returns and plays a pivotal role in market investment. This paper randomly selects financial statement data from 3,248 A-share listed companies in the H database for the years 2006–2022 as the research data source, determines the model indicator variables, and performs descriptive statistics on the sample. In constructing the prediction model, to address the limitations of traditional Support Vector Regression (SVR) models in handling large-scale data, the Whale Optimization Algorithm (WOA) was employed to optimize the penalty factor and kernel function parameters within SVR, thereby proposing an SVR-WOA-based profit prediction model. After 50 iterations, the model demonstrated a mean squared error as low as 0.075 on the validation set, exhibiting superior predictive capabilities.

Keywords: profit prediction model; support vector regression model; whale optimization algorithm; corporate economics

1. Introduction

Following the pandemic, China's economic growth began to slow down, the stock market experienced several fluctuations, and the debt of A-share listed companies continued to rise, leading to various financial risks [1]. To sustain development in intense market competition, enterprises must establish a sound financial foundation. Once financial difficulties arise, they can hinder corporate development and potentially lead to bankruptcy [2-4]. Corporate bankruptcy not only harms the interests of stakeholders but also obstructs sustained economic growth and may even cause social unrest [5].

Additionally, the development of economic and trade globalization in recent years has made business cooperation between countries and industries more closely intertwined, making corporate profitability more susceptible to crises and exposing companies to greater risks [6]. However, financial difficulties for companies typically do not occur suddenly; they follow a developmental cycle, progressing from normal financial conditions to gradual deterioration, and ultimately to financial distress [7-8]. On one hand, if companies can predict their profitability based on data and identify the risks they may face, managers can decide to make strategic adjustments in advance to overcome the challenges they currently face, thereby avoiding financial difficulties [9-11]. On the other hand, if companies can establish a profitability risk firewall, it will help them adjust their development strategies according to business needs [12]. Therefore, corporate profitability forecasting holds significant importance for corporate survival.

Over the years, traditional methods for predicting corporate profitability have primarily included time series models, regression models, and grey prediction models [13]. For example, Evans, M. E., et al. [14] utilized a regression model to construct a cross-sectional corporate profitability prediction model and empirically tested the model's predictive accuracy and economic viability. Apergis, N. [15] utilized statistical models to construct a profitability prediction model for U.S. small and medium-sized enterprises using panel data sets. Finding that imperfections in financial markets significantly impact the profitability of SMEs. Nguyen, H., and Nguyen, T. [16] utilized past accounting profits and cash flow information, employing time and spatial series techniques to predict future capital cash flow trends, thereby forecasting future corporate cash flows, i.e., profitability status. While these prediction methods each have their advantages, they primarily focus on analyzing causal regression models and time series models, failing to comprehensively and fundamentally reflect the intrinsic structure and complex characteristics of prediction data, thereby losing certain information content [17-18].

Machine learning algorithms possess exceptional nonlinear mapping capabilities and the ability to learn from data, thereby gaining significant attention and application in the financial field. To date, the most widely used model for profit forecasting is the BP network model; however, this network suffers from drawbacks such as slow convergence speed and susceptibility to local minima during prediction [19-21]. For example, Kayakus, M et al. [22] noted in their research that return on equity (ROE) and return on assets (ROA) are the most important indicators in corporate profit forecasting and can to some extent reflect the sustainability of a company's profitability. They constructed prediction models using artificial neural networks (ANN), multiple linear regression (MLR), and support vector regression (SVR), achieving good accuracy rates respectively. In the study by Zoričák, M et al. [23], the SVM model was incorporated into the financial indicator prediction model, and it was concluded that the SVM model has higher financial indicator prediction capabilities than the previously used hierarchical analysis method, logistic model, and neural network model (NN).

This paper first briefly outlines the sources of the indicator variables used, proposes methods for processing the sample data, and defines the calculation methods for multiple key variables. Based on the selected variables, descriptive statistical analysis of the research sample and correlation analysis between variables are conducted. Subsequently, the objective function definition and mathematical operation methods of the traditional Support Vector Regression (SVR) model are explained, along with the basic principles and optimization process of the WOA optimization algorithm, to establish a profit prediction algorithm based on SVR-WOA. Subsequently, a comparison experiment is conducted between the proposed model and traditional profit prediction models to evaluate their mean squared error performance on the validation set and training set, assessing the predictive performance of the proposed model. Finally, the practical application capability of the proposed model is verified through quarterly and annual profit prediction comparison experiments.

2. Research Preparation

2.1. Data Sources and Data Processing

The indicator variables used in this paper are sourced from the H database. Many financial indicators are derived indicators that require separate calculation, while the basic indicators are sourced from the H database. Existing policy documents state that accounting information related to a company's financial condition should be provided to users to assist them in making economic decisions. This paper uses data from the H database for A-share listed companies from January 31, 2006, to December 31, 2022, excluding the financial industry. This paper uses all observations from the previous 10 years (i.e., t, $t-1$, $t-2$, ..., $t-9$) as training data to estimate the model, then applies the model to the prediction variables for year t to generate profit forecasts for year $t+1$. To maintain consistency, all existing models also use the same data from the first 10 years for estimation, and the resulting linear models are applied to the respective predictive variables for year t to generate profit forecasts for year $t+1$ Since variables such as AC (involving the HVZ model and SO model), AG (involving the SO model), and TACC involve three years of data, the first complete input variable is available as of December 31, 2007, enabling profit forecasts for 2008. The first 10-year training interval is from December 31, 2007, to December 31, 2016, generating the company's profit forecast for 2017. The second five-year training interval is from December 31, 2008, to December 31, 2017, generating the company's profit forecast for 2018. Similarly, the profit forecast set spans from 2018 to 2022.

The objective of this paper is to compare the out-of-sample profit forecast accuracy of profit forecast models. To ensure comparability of data across companies of different market capitalizations, this paper uses earnings per share (EPS) adjusted for non-recurring items as the profit forecast value. Therefore, model-related variables must be converted to per-share values. To exclude the potential adverse effects of extreme data on empirical results, this paper excludes data from the financial industry.

2.2. Calculation of Key Variables in the Model

The following are the main variable calculations for the profit forecast model.

(1) HVZ model variables

E: Earnings per share excluding non-recurring gains and losses, calculated by dividing the net profit attributable to shareholders of listed companies excluding non-recurring gains and losses by the total number of shares.

AT: Total assets, calculated using the total assets in the H database.

D: Dividend distribution, calculated using the distribution type, ex-dividend date, and distribution ratio from the H database.

AC: Operating accruals, calculated by subtracting cash and cash equivalents from total current assets, then subtracting total current liabilities, adding short-term borrowings, adding trading financial liabilities, adding non-current liabilities due within one year, adding taxes payable, subtracting EBITDA, adding net profit, adding income tax expenses, and adding financial expenses.

(2) SO model variables

E and AC variables refer to HVZ model variables.

AG: Percentage change in total assets, calculated as the difference between year-end total assets and the previous year's beginning total assets divided by the beginning total assets.

BTM: Book-to-market ratio, calculated as the inverse of the price-to-book ratio.

Price: Defined as the closing price of the stock at the end of the fourth month following the end of the year, calculated using monthly closing prices.

(3) EP and RI Model Variables

E and NegE variables refer to HVZ model variables.

BVE: Book value of equity, calculated as total shareholders' equity minus minority interest minus other equity instruments.

TACC: Total operating accruals, which, in addition to operating accruals, also consider non-operating accruals and accruals from financial activities. Defined as the annual change in working capital plus the annual change in non-operating accruals plus the annual change in accruals from financial activities. The annual change in working capital is calculated by subtracting cash and cash equivalents, net short-term investments, trading financial assets, accounts receivable financing, and non-current assets due within one year from total current assets, then adding short-term borrowings, trading financial liabilities, and non-current liabilities due within one year. The annual change in non-operating accruals is calculated by total assets minus total current assets minus debt investments minus net available-for-sale financial assets minus other debt investments minus net held-to-maturity investments minus net long-term receivables minus investments in other equity instruments minus other non-current financial assets minus net long-term debt investments minus net long-term investments minus net investment in investment real estate minus total liabilities plus long-term borrowings plus bonds payable plus long-term payables plus long-term payables plus lease liabilities. The annual change in accrued financial activities is calculated by adding the net amount of short-term investments, trading financial assets, accounts receivable financing, non-current assets due within one year, debt investments, net available-for-sale financial assets, other debt investments, net held-to-maturity investments, net long-term accounts receivable, other equity instruments investments, other non-current financial assets, net long-term debt investments, net long-term investments, and net investment property, then subtracting long-term borrowings and accounts payable. bonds, less long-term payables, less lease liabilities, less short-term borrowings, less trading financial liabilities, and less non-current liabilities due within one year.

2.3. Descriptive Statistics of the Sample

The descriptive statistics results for 10 variables of 3,248 listed companies from 2006 to 2022 are shown in Table 1. It can be seen that analysts' forecast value for the earnings per share (E) of the sample companies was 0.946 yuan, while the actual earnings per share of the sample companies during this period was 0.653 yuan, indicating that analysts had an optimistic bias in their forecasts.

Table 1. Variable descriptive statistic.

Variable	Average	Sd.	Minimum	Maximum	Number
E	0.653	1.231	-10.598	49.931	3248
Forest	0.946	1.449	-8.946	51.164	3248
AT	0.151	1.582	-4.601	53.002	3248
D	0.467	0.926	0.527	49.162	3248
AC	0.544	110.234	-7924.429	10964.457	3248
AG	0.964	54.672	-3222.322	4658.634	3248
BTM	0.766	8.064	-2.564	603.708	3248
Price	90.397	2807.652	1.237	392618.364	3248
NegE	12.794	42.464	13.264	35.074	3248
BVE	1.851	1.489	1.255	2.793	3248
TACC	23.671	44.036	1431.234	2051.234	3248

2.4. Correlation Analysis

Figure 1 shows the correlations between the variables in the proposed model, with Pearson correlations above the diagonal. It can be seen that the correlations between variables (<0.3) are economically and statistically significant, indicating the predictive usefulness of the proposed model variable framework.

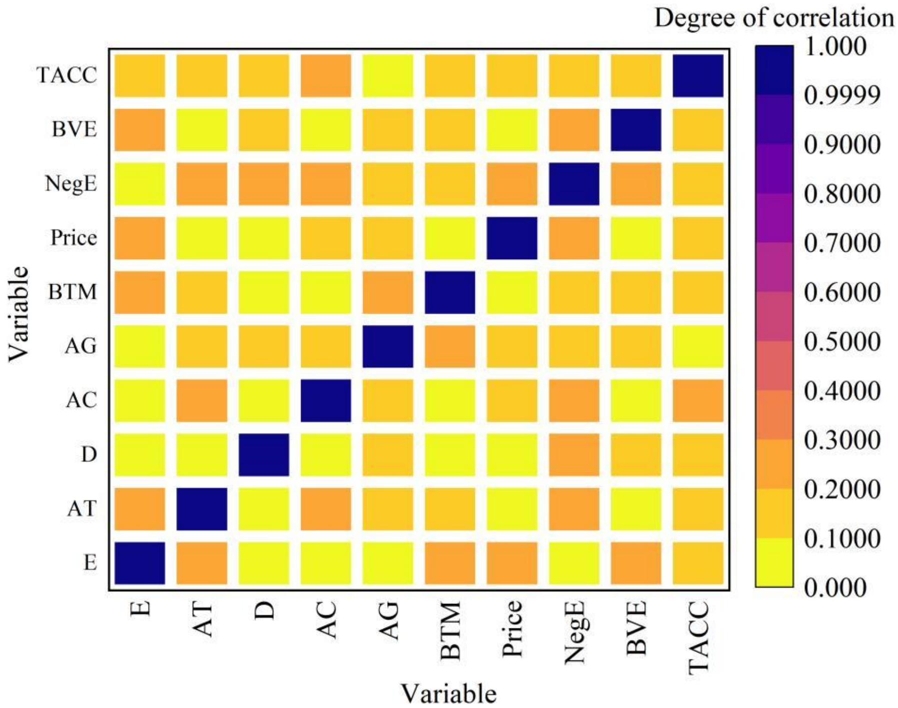

Figure 1. Correlation analysis of model variables.

3. Profit Prediction Model Based on SVR-WOA

3.1. Support Vector Regression Model

Suppose we are given a sample set $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$ in a feature space, where $x_i \in \mathbb{R}^p$ is the explanatory variable of the sample set and y_i is the response variable of the sample set. In the support vector machine regression model, the most critical issue is to find an optimization function such as (1):

$$f(x) = w \cdot \phi(x) + b \tag{1}$$

Minimize the convex quadratic programming problem, where w is the weight vector and b is the bias term. The objective function of the traditional support vector regression model is defined as equations (2)-(3):

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \rho_\varepsilon(y_i - w \cdot \phi(x_i) - b) \tag{2}$$

$$\rho_\varepsilon(y_i, f(x_i)) = \begin{cases} 0, & |y - f(x_i)| \le \varepsilon \\ |y - f(x_i)| - \varepsilon, & |y - f(x_i)| > \varepsilon \end{cases} \tag{3}$$

Among them is equation (4):

$$f(x_i) = w \cdot \phi(x_i) + b \tag{4}$$

$\phi(\cdot)$ is a mapping function from the input space to the feature space, $\|w\|^2$ is the structural risk, C

is the penalty factor, and ξ_i is the introduced slack variable.

When the support vector machine method is extended to regression problems, an appropriate loss function must be selected to maintain its sparsity. In this paper, we will use the ε -insensitive loss function, where ε represents the insensitivity loss degree. Let $\varepsilon = 0.1$, then the learning problem of the SVR model becomes the convex quadratic programming problem shown in equation (5):

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i^{\wedge} + \xi_i^{\vee})$$
$$s.t.\begin{cases} y_i - w\cdot\phi(x_i) - b \le \varepsilon + \xi_i^{\wedge} \\ w\cdot\phi(x_i) - b - y_i \le \varepsilon + \xi_i^{\vee} \\ \xi_i^{\wedge}, \xi_i^{\vee} \ge 0 \end{cases} \tag{5}$$

Introducing Lagrange multipliers α_i and α_i^*, equation (5) is transformed into equation (6):

$$\begin{aligned} &L\left(w,b,\xi,\xi^{\wedge},\xi^{\vee},\alpha,\alpha^*,u^{\wedge},u^{\vee}\right) \\ &= \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i^{\wedge} + \xi_i^{\vee}) - \sum_{i=1}^{n}\alpha_i\left(\varepsilon + \xi_i^{\wedge} - y_i + w\cdot\phi(x_i) + b\right) \\ &-\sum_{i=1}^{n}\alpha_i^*\left(\varepsilon + \xi_i^{\vee} + y_i - w\cdot\phi(x_i) - b\right) - \sum_{i=1}^{n}u_i^{\wedge}\xi_i^{\wedge} - \sum_{i=1}^{n}u_i^{\vee}\xi_i^{\vee} \end{aligned} \tag{6}$$

where $\alpha_i, \alpha_i^* \ge 0$, $u_i^{\wedge}, u_i^{\vee} \ge 0$. Solving the constrained optimization problem using the KKT conditions yields equation (7):

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi(x_i) \\ \dfrac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \\ \dfrac{\partial L}{\partial \xi_i^{\wedge}} = 0 \Rightarrow C - \alpha_i - u_i^{\wedge} = 0 \\ \dfrac{\partial L}{\partial \xi_i^{\vee}} = 0 \Rightarrow C - \alpha_i^* - u_i^{\vee} = 0 \end{cases} \tag{7}$$

Substituting equation (5) into equation (4) yields equation (8):

$$\begin{aligned} &\min \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\phi(x_i)\phi(x_j) \\ &+\varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) - \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i \end{aligned} \tag{8}$$

Since the kernel function is as in (9):

$$K(x_i, x_j) = \phi(x_i)\phi(x_j) \tag{9}$$

Equation (8) can be converted to equation (10):

$$\begin{aligned} &\min \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x_j)(\alpha_j - \alpha_j^*) \\ &+\varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) - \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i \end{aligned} \tag{10}$$

When solving linear regression problems using a support vector regression model, $\phi(\cdot)$ is a linear mapping function, and equation (11) applies:

$$K(x_i, x_j) = x_i x_j^T \tag{11}$$

When the support vector regression model is used to solve nonlinear regression problems, $\phi(\cdot)$ is a nonlinear mapping function. The kernel function used in this paper is the radial basis kernel function, i.e., equation (12):

$$K(x_i, x_j) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right) \tag{12}$$

Equations (13)-(15):

$$U = \alpha - \alpha^* \tag{13}$$

$$V = \alpha + \alpha^* \tag{14}$$

$$K = K(x_i, x_j) \tag{15}$$

U and V can be obtained through optimization equation (16):

$$\left(\widehat{U}, \widehat{V}\right) = \arg\min_{U,V} \frac{1}{2} U^T K U + \varepsilon V - U^T y \tag{16}$$

Support vector regression models are an important application branch of support vector machine models. They perform regression predictions by finding a hyperplane in the feature space and are characterized by their strong nonlinear capabilities. The kernel function introduced maps the input data from the original feature space to a high-dimensional feature space, transforming the nonlinear regression problem in the original space into a linear regression problem in the high-dimensional space. Therefore, support vector regression models can handle complex nonlinear regression problems.

3.2. WOA Optimization Algorithm

In SVR, both the penalty factor c and the kernel function parameter g affect the estimation results. The penalty factor c influences the algorithm's tolerance level; a larger c allows for more errors but increases the risk of underfitting, while a smaller c increases the likelihood of overfitting. The kernel function parameter g determines the distribution of data after mapping to a new feature space; a larger g results in fewer support vectors. Both c and g being too large or too small will affect the speed and accuracy of estimation. In this paper, the whale optimization algorithm (WOA) is used to optimize c and g.

WOA is an algorithm inspired by the behavior of whales hunting prey, using random or optimal search agents to simulate hunting behavior. The WOA algorithm has a simple mechanism, few parameters, and strong optimization capabilities. The algorithm flow is shown in Figure 2.

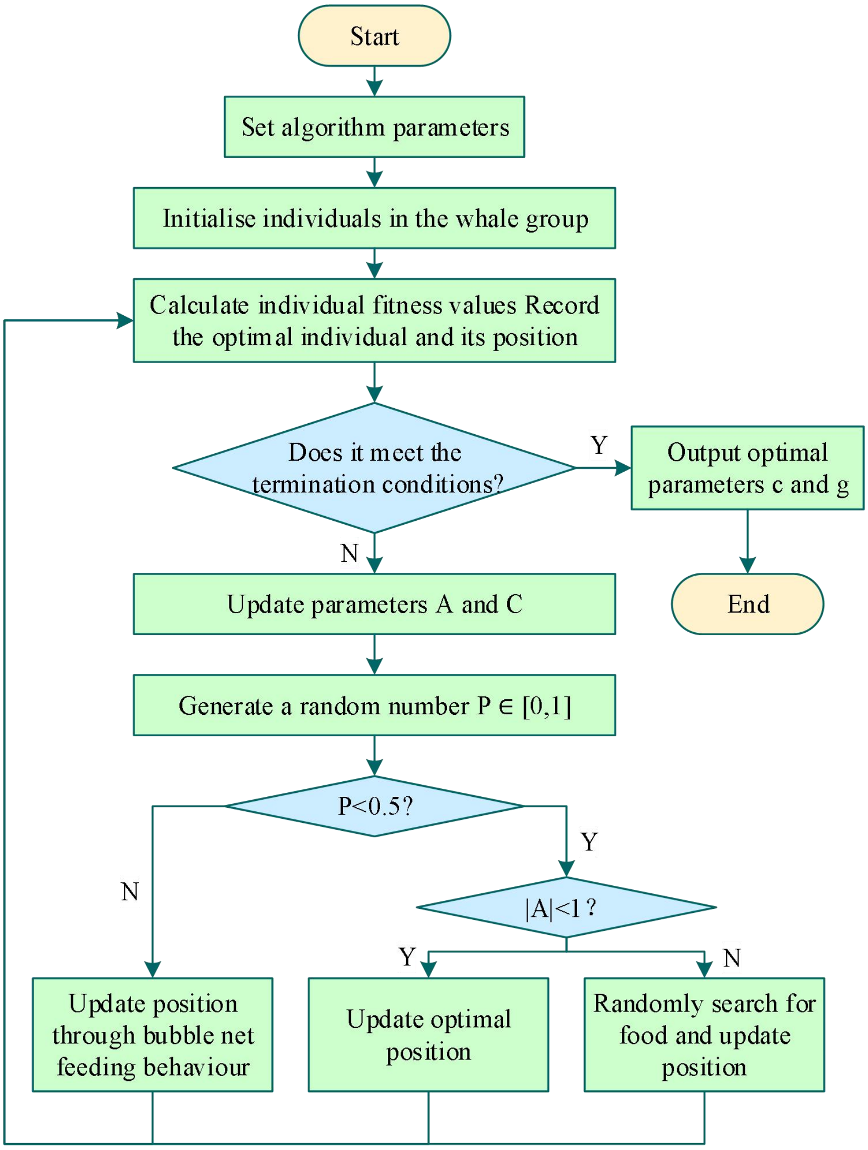

Figure 2. Flow chart of the WOA algorithm.

In the WOA algorithm, the distance and position vectors between individuals are given by Equations (17):

$$\left.\begin{aligned} D &= CX^*(t) - X(t) \\ X(t+1) &= X^*(t) - AD \end{aligned}\right\} \tag{17}$$

where t is the current iteration number. D is the distance vector between the current whale and the random whale. $X(t)$ is the position vector. $X^*(t)$ is the position vector of the currently obtained optimal solution. A and C are coefficients, with calculation formulas as shown in Equations (18)-(19);

$$A = 2a(t)r_1(t) \tag{18}$$

$$C = 2r_2(t) \tag{19}$$

Throughout the iteration process, the convergence factor a is linearly reduced from 2 to 0. r_1 and r_2 are random vectors in [0, 1].

In the WOA algorithm, whale predation primarily involves two mechanisms: encirclement predation and bubble net predation. Therefore, based on the probability P , either bubble net predation or encirclement contraction is selected, with the position update formula given by Equation (20):

$$X(t+1)=\begin{cases}X^{*}(t)-AD & P\le 0.5\\ D'\exp(BL)\cos(2\pi L)+X^{*}(t) & P>0.5\end{cases} \tag{20}$$

In the equation, D' is the distance vector between the current search individual and the current optimal solution. B is the spiral shape parameter. L is a random number uniformly distributed in $[-1,1]$. P is the predation mechanism probability, which is a random number in the range of $\left[0,1\right]$.

As the iteration count t increases, the parameters A and convergence factor a gradually decrease. If $|A|<1$, the whale swarm gradually surrounds the current optimal solution, entering the local optimization phase. To ensure that all whales can search the solution space thoroughly, WOA updates the positions based on the distances between whales to achieve random search. Therefore, when $|A|\ge 1$, the search individuals swim toward random whales, thereby obtaining the optimal solution as shown in Equation (21):

$$\left.\begin{aligned}D''&=CX_{rand}(t)-X(t)\\ X(t+1)&=X_{rand}(t)-AD\end{aligned}\right\} \tag{21}$$

In the equation, D'' is the distance vector between the current search individual and the random individual. $X_{rand}(t)$ is the position of the current random individual.

The root mean square error (RMSE) of the SVR training results is used as the objective function, as shown in Equation (22):

$$X_{RMSE}=\sqrt{\frac{\sum_{k=1}^{N}(X_{o,k}-X_{m,k})^2}{N}} \tag{22}$$

In the equation, X_{RMSE} is the root mean square error value. N is the number of estimates. $X_{o,k}$ is the true value. $X_{m,k}$ is the estimated value.

The optimization ranges for the penalty factor c and the kernel function parameter g are $[10^{-2},10^{3}]$ and $[2^{-5},2^{8}]$, respectively.

4. Model Performance Verification and Application Analysis

4.1. Predictive Performance of the Model

Figure 3 shows the comparison of the mean squared error values of the basic profit prediction model (M1) using only monthly revenue features and the profit prediction model (M2) proposed in this paper on the training set and validation set when the number of epochs is set to 50, i.e., after 50 epochs. Where: (VM1) is the basic profit prediction model (validation set), (VM2) is the basic profit prediction model (training set), (TM1) is the profit prediction model in this paper (validation set), and (TM2) is the profit prediction model in this paper (training set).

It can be observed that the mean squared error of both models on the validation set decreases oscillatingly with the increase in the number of epochs. However, the profit prediction model in this paper (M2) consistently maintains a lower mean squared error than the basic profit prediction model (M1) on the validation set, reaching the lowest value (0.075) after 50 iterations. This indicates that the profit prediction model in this paper (M2) has a high accuracy in profit prediction.

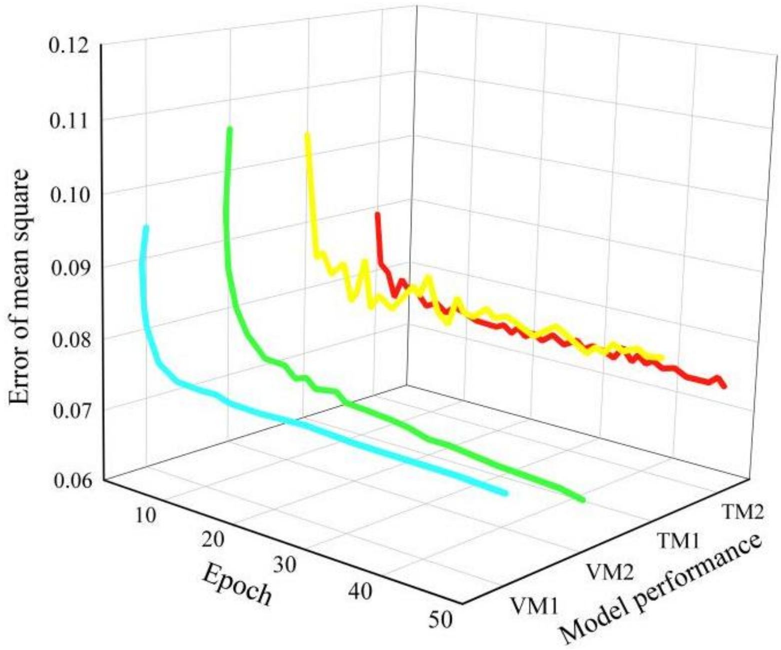

Figure 3. Comparison of mean square errors of the model.

4.2. Practical Application Evaluation

This section selects (M1) a basic profit forecasting model that only uses monthly revenue characteristics and (M3) a profit forecasting model that only uses quarterly revenue characteristics as comparison models, and compares them with the proposed profit forecasting model in terms of forecasting quarterly profits for the research sample from 2020 to 2022 and annual profit forecasts.

4.2.1. Quarterly earnings

Table 2 analyzes the quarterly profit forecasting capabilities of three quarterly profit models. In the table, (a) indicates that the former is more accurate than the latter, i.e., the median absolute value of the prediction error for the former is smaller than that for the latter, while (b) indicates that the latter is more accurate than the former, i.e., the median absolute value of the prediction error for the latter is smaller than that for the former. *, **, and *** indicate significant differences at the 0.1, 0.05, and 0.01 levels, respectively. It can be seen that the profit prediction model (M2) in this paper not only significantly outperforms the other two prediction models in all quarters but also shows strong significant differences (p<0.01) in profit predictions for 2020 and 2022.

Table 2. Quarterly profit forecast.

Forecasting methods	Year	Quarter			
		1	2	3	4
Friedman	2020	-	17.796***	30.079***	18.762***
M1-M3		-	-11.311(a)*	-12.459(a)*	-12.841(b)**
M2-M1		-	-17.395(a)***	-13.897(a)***	-13.448(a)***
M2-M1		-	-0.245(a)***	-2.79(a)***	-7.303(a)***
M1-M3	2021	312.877***	15.5***	27.221***	9.891***
M2-M1		-9.673(b)**	-7.533(a)**	-11.373(a)*	-10.482(b)
M2-M1		-8.958(a)**	-15.942(a)**	-4.465(a)**	-2.443(a)***
M1-M3		-14.756(a)*	-14.029(a)**	-9.019(a)**	-2.16(a)**
M2-M1	2022	273.639***	6.559***	30.965***	10.148***
M2-M1		-9.677(a)	-5.741(b)	-10.559(b)*	-9.677(b)*
M1-M3		-9.072(a)***	-5.914(a)***	-14.223(a)***	-9.072(a)***
M2-M1		-13.855(a)***	-3.046(a)***	-0.753(a)***	-13.855(a)***

4.2.2. Annual Profitability

Table 3 analyzes the annual profit forecasting capabilities of the three models. It can be seen that although the profit forecasting model (M2) in this paper showed statistically significant differences ($p<0.05$) from the other two models in terms of annual profit forecasts for 2020-2022, overall, the annual profit forecasting performance of the three models was slightly inferior to their quarterly profit forecasting performance. As the forecasting interval gradually shortens, the quarterly profit model incorporates more comprehensive and timely information. The information advantage gradually increases, thereby significantly improving the model's predictive performance and making it superior to annual profit forecasts.

Table 3. Annual profit forecast.

Forecasting methods	Year	Advance quarterly quantity			
		4	3	2	1
Friedman	2020	980.173*	24.095**	60.56**	102.146**
M1-M3		-26.492(b)	-1.396(b)	-5.66(b)	-0.487(b)
M2-M1		-18.729(a)*	-26.27(a)*	-22.37(a)**	-12.224(a)**
M2-M1		-19.282(a)*	-17.339(a)**	-11.456(a)*	-25.445(a)*
M1-M3	2021	287.786*	25.971**	23.034*	73.163*
M2-M1		-27.878(b)	-5.421(b)**	-19.454(b)*	-29.165*
M2-M1		-27.096(a)**	-27.951(a)	-8.054(a)*	-0.2(a)**
M1-M3		-4.597(a)*	-22.072(a)	-12.032(a)	-8.187(a)*
M2-M1	2022	296.025*	22.435*	52.662*	119.419**
M2-M1		-26.583(b)*	-27.755(b)*	-12.101(b)	-11.006(b)
M1-M3		-15.524(a)*	-5.66(a)*	-9.414(a)**	-2.197(a)*
M2-M1		-4.857(a)*	-15.367(a)**	-15.831(a)**	-17.965(a)*

5. Conclusion

This paper uses financial statement data from 3,248 A-share listed companies in the H database from 2006 to 2022 as the research data source, identifying 10 indicator variables with correlations all below 0.3: earnings per share (EPS) adjusted for non-recurring gains and losses, total assets, dividend distribution, operating accruals, percentage change in total assets, book-to-market ratio, stock closing price at the end of the fourth month following the end of the fiscal year, E as a negative indicator variable, book value of equity, and total operating accruals. Combining the traditional support vector regression model with the whale optimization algorithm, a profit prediction model based on SVR-WOA was established.

The SVR-WOA-based profit prediction model not only achieves the lowest mean squared error of 0.075 in overall prediction performance but also demonstrates statistically significant differences ($p < 0.05$) compared to similar model algorithms in practical applications for quarterly and annual profits.

References

1. Sun, C., Abbas, H. S. M., Xu, X., Gillani, S., Ullah, S., & Raza, M. A. A. (2023). Role of capital investment, investment risks, and globalization in economic growth. International Journal of Finance & Economics, 28(2), 1883-1898.
2. Laksmana, I., & Yang, Y. W. (2015). Product market competition and corporate investment decisions. Review of Accounting and Finance, 14(2), 128-148.
3. Miao, L., Zhuo, Y., Wang, H., & Lyu, B. (2022). Non-financial enterprise financialization, product market competition, and total factor productivity of enterprises. Sage Open, 12(2), 21582440221089956.
4. Ma, C., Ruan, Y., & Zheng, K. (2018). Product market competition, asset strategy and enterprise operating liabilities. Journal of Intelligent & Fuzzy Systems, 34(2), 1033-1043.
5. Jace, K., Koumanakos, D., & Tsagkanos, A. (2022). Bankruptcy prediction in social enterprises. Journal of Social Entrepreneurship, 13(2), 205-220.
6. Bogatska, N. (2019). Enterprise profit: its role, formation and reserves of growth in the market conditions. Efektyvna ekonomika,[Online], 9.
7. Dang, H. N., Vu, V. T. T., Ngo, X. T., & Hoang, H. T. V. (2019). Study the impact of growth, firm size, capital structure, and profitability on enterprise value: Evidence of enterprises in Vietnam. Journal of Corporate Accounting & Finance, 30(1), 144-160.
8. Liang, Y., Zhang, W. H., Lu, Y., & Wang, Z. S. (2020). Optimal control and Simulation for enterprise financial risk in Industry environment. Mathematical Problems in Engineering, 2020(1), 6040597.
9. Ai, J., Brockett, P. L., Cooper, W. W., & Golden, L. L. (2012). Enterprise risk management through strategic allocation of capital. Journal of Risk and Insurance, 79(1), 29-56.

10. Sievidova, I., Pakhucha, E., Chumak, V., Shovkun, L., Novikov, A., Zakharov, I., & Skrypnyk, O. (2021). Assessment of economic risks in the context of strategic enterprise management. European Journal of Sustainable Development, 10(3), 231-231.
11. Guryanova, L., Klebanova, T., & Trunova, T. (2017). Modeling the financial strategy of the enterprise in an unstable environment. Economic Studies, 26(3).
12. Pankratov, E. L. (2024). Assessing Profit Risk Prognosis in Industrial Enterprises. International Journal of Economics & Business Administration (IJEBA), 12(2), 132-140.
13. Belloni, M., Jarmuzek, M., & Mylonas, D. (2022). From modelling to forecasting bank profitability: Evidence from euro area banks. Journal of Risk Management in Financial Institutions, 15(4), 314-337.
14. Evans, M. E., Njoroge, K., & Yong, K. O. (2017). An examination of the statistical significance and economic relevance of profitability and earnings forecasts from models and analysts. Contemporary Accounting Research, 34(3), 1453-1488.
15. Apergis, N. (2020). Financial market imperfections and profitability: New evidence from a large panel of US SME firms. Journal of Forecasting, 39(2), 220-241.
16. Nguyen, H., & Nguyen, T. (2020). The prediction of future operating cash flows using accrual-based and cash-based accounting information: Empirical evidence from Vietnam. Management science letters, 10(3), 683-694.
17. Esplin, A., Hewitt, M., Plumlee, M., & Yohn, T. L. (2014). Disaggregating operating and financial activities: Implications for forecasts of profitability. Review of Accounting Studies, 19, 328-362.
18. P. Bauman, M. (2014). Forecasting operating profitability with DuPont analysis: Further evidence. Review of Accounting and Finance, 13(2), 191-205.
19. Yang, R. (2020). Enterprise Network Marketing Prediction Using the Optimized GA-BP Neural Network. Complexity, 2020(1), 6682296.
20. Lomakin, N., Kharlamova, E., & Polyanskaya, A. (2018, May). Sustainable development of regional enterprises based on the neural network profit forecasting model. In International Scientific Conference" Competitive, Sustainable and Secure Development of the Regional Economy: Response to Global Challenges"(CSSDRE 2018) (pp. 113-116). Atlantis Press.
21. Ding, X. I. A. O. W. E. I., Ye, L., Jin, J., & Wang, Y. (2023). Financial Profitability Evaluation and Forecasting Using the Deep Learning: A Case of the Chinese Petroleum Industry. Montenegrin Journal of Economics, 19(4), 55-64.
22. Kayakus, M., Tutcu, B., Terzioglu, M., Talaş, H., & Ünal Uyar, G. F. (2023). ROA and ROE forecasting in iron and steel industry using machine learning techniques for sustainable profitability. Sustainability, 15(9), 7389.
23. Zoričák, M., Gnip, P., Drotár, P., & Gazda, V. (2020). Bankruptcy prediction for small-and medium-sized companies using severely imbalanced datasets. Economic Modelling, 84, 165-176.