

Risk prediction of interest rate futures based on machine learning scenarios

Rui Chen¹, Xinrui Gu², Peiling Li³, and Kexin Tan^{4,*}

¹Information Engineering School, Shanghai Maritime University, 200135 Shanghai, China

²France Sino-European School of Technology, Shanghai University, 200444 Shanghai, China

³College of Mathematics and Computer Science, Zhejiang Agriculture and Forestry University, 311300 Hangzhou, China

⁴Electrical and Computer Engineering, Shenzhen MSU-BIT University, 518100 Shenzhen, China

Abstract. The interest rate futures market is a significant part of the financial market. It has a crucial impact on forecast the interest rate risk in global financial markets, which due to the complexity of financial markets and the volatility of interest rate futures. Based on machine learning scenarios to analyse and compare different algorithms, this paper analyses and forecast 2-year Treasury futures for the period 2022.6-2023.6 through regressions and other methods. Meanwhile, it is applied to construct charts and graphs to better compare and analyse models that are more suitable for forecasting future risk in interest rate futures. National policies, the volatility of the general market environment and its smoothness are utilized as the main factors to forecast its risk fluctuations. The main algorithms this paper use are: random forest regression, ARIMA model, BP Neural Network regression model, ARCH model (model validity test), GARCH model. In conclusion, though the predicted results of the random forest and ARIMA models exhibit a close to 0 and have strong stability, the predicted results of the GARCH are relatively better, none of them achieve the desired prediction performance.

1 Introduction

The interest rate futures market is a vital part of the financial market. It has a crucial impact on forecast the interest rate risk in global financial markets, which due to the complexity of financial markets and the volatility of interest rate futures. Interest rate futures are contracts for the purchase and sale of a currency as the underlying for delivery within a specified period of time. While its price fluctuations have a significant impact on both macroeconomic and financial markets [1]. Therefore, accurate forecasting the risk of the interest rate futures market has significant economic importance and practical value. Predicting the risk of the interest rate futures market can provide some assistance to different clients: firstly, it assists investors in developing better investment strategies to reduce investment risks and improve investment returns; then it helps money managers better develop risk management strategies suitable for financial practitioners such as banks and financial institutions to ensure the safety of their clients' assets. Meanwhile, it assists exchanges in setting effective margin levels in order to protect against and control, to a certain extent, the risks of price fluctuations faced by the market as a whole.

Current research shows that the use of machine learning methods can effectively predict the risk of interest rate futures markets but there are some shortcomings. For example, Zhang used LSTM neural networks to forecast interest rate futures prices,

international oil prices, and exchange rates as indicators, which were able to discover the long-range correlation in the time sequence data and improve the accuracy of risk prediction, but did not consider the impact of policy changes on risk [2]. Li and Xie used convolutional neural networks and LSTM models in combination with interest rate futures prices, gold prices, and stock prices as indicators, which were able to extract more time series features and further improved the accuracy of risk prediction, but did not consider the impact of market sentiment on risk [3]. Wang and Xing selected interest rate futures prices, CPI, and GDP growth rate as indicators and used a combination of grey correlation and neural networks to improve the accuracy of risk prediction, but also did not consider the impact of market sentiment on risk [4]. Ren used Logistic Regression, Lasso, Ridge Regression (Ridge), Elastic Net, Support Vector Machines are used for classification (SVC), Random Forest (RF), XGBoost model and Deep Feed-forward Neural Networks (DNN) to enrich the literature on the application of machine learning models for bond default prediction and the literature on the factors influencing bond defaults in China by quantitatively revealing the extent to which non-linear model settings contribute to bond default prediction, using very standard machine learning models, without developing corresponding machine learning algorithms independently for the characteristics of the dataset, in order to achieve better results in bond default prediction research [5]. The

* Corresponding author: 1120200259@smbu.edu.cn

identification of factor importance is applied with reference to the regression class task machine learning model, and no targeted factor importance analysis method is developed for the classification task problem, so further research can also be considered for the factor importance profiling method. Yang used the Rolling Window technique proposed in this paper to extract window data features from time series data as sample features and extracted three types of features, including features for technical analysis, fundamental features and marketing sentiment features [6]. Linear regression models, KNN models, support vector machine models (SVM) and multilayer perceptron neural network models (MLPNN) were used to normalise the data set and to evaluate each prediction model for correctness, cross-check accuracy, MSE and RMSE. The machine learning model with the best prediction performance is selected as the final prediction model. Zhang weighted the input feature parameters through an attention mechanism so as to extract deeper features, used the Fourier transform-based signal decomposition algorithm EMD and its improved algorithm CEEMDAN to perform adaptive processing on the original non-linear, non-smooth sequences, and improved accuracy by applying clustering algorithms and principal component analysis for feature dimensionality reduction [7].

Nelson Area used three machine learning skills: SVR, Artificial Neural Network and Random Forest [8]. Xin also used Random Forest, Artificial Neural Network and Support Vector Regression [9]. Meanwhile, Willem used Artificial Neural Network, SVR and Random Forest (RF) [10]. The paper written by Li used Convolutional Neural Networks, Support Vector Regression (SVR) and Random Forests [11]. These studies use different machine learning techniques and indicators, but most of them use some basic technical indicators such as moving averages and relative strength indicators [12, 13]. The indicators may not fully reflect the characteristics of the interest rate futures market, and there may be correlations between different indicators that may affect the predictive accuracy of the models. Judging from the domestic situation in China, most of the relevant international literatures adopt the same machine learning algorithm, but the algorithm adopted has certain national characteristics. It is mainly manifested in the differences in the selection of indexes and samples.

2 Data and methods

2.1 Data collection and preprocessing

This paper utilized a random forest model to preprocess and select relevant data from the "TS2303 two-year treasury bond futures" of China Financial Futures Exchange during the period from June 2022 to May 2023. The necessary data was extracted for analysis. Subsequently, a historical return rate graph was plotted using Python, the result is shown in Fig 1. This paper adopts an integrated approach that uses a combination of traditional ARIMA and GARCH models as well as

machine learning methods such as Random Forest and BP artificial neural networks to forecast the volatility of interest rate futures. This integrated approach takes full advantage of the respective strengths of the traditional models and machine learning methods, and further improves the accuracy and stability of the forecasts by complementing and enhancing each other.

ARIMA models and GARCH models, as traditional time series models, are able to capture historical patterns and trends in volatility and perform better for long-term and medium-term forecasting. In contrast, random forests and BP neural networks, as machine learning methods with strong non-linear fitting capabilities and adaptability to complex relationships, are able to extract valid information from a large number of features and have an advantage for short-term volatility forecasting. By integrating these models, one can combine their strengths to further improve predictions. While traditional models can provide more stable and interpretable forecasts, machine learning approaches can capture more complex correlations by learning patterns and regularities in large sample data. Integrated models can be flexible and adaptable across different time periods and market environments, and have high predictive accuracy and robustness.

In summary, this paper's integrated approach combines traditional models with machine learning methods to improve the accuracy and stability of interest rate futures volatility forecasts.

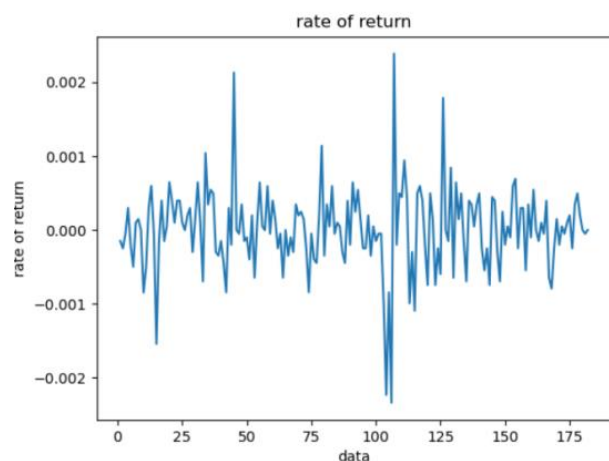


Fig. 1. Daily return (Photo/Picture credit: Original).

2.2 Machine learning algorithms

Machine learning aims to improve the performance metrics of computer programs by using data and past experience. It involves computers, big data and statistical thinking and is an intelligent discipline. The three types of machine learning are supervised learning, unsupervised learning and augmented learning for different types of problems respectively. In machine learning, we can perform model building and analysis by importing data sets, data pre-processing, selecting and training models, parameter selection and optimisation and building final

models to predict and analyse future data trends. The basic idea is shown in Fig 2.

Random Forest is an integrated learning method based on decision trees that aims to solve the problem of overfitting and noise interference in decision tree algorithms by introducing randomness. The algorithm mainly embodies randomness in two aspects: sample selection and feature selection. In terms of sample selection, the training set of each decision tree is obtained by randomly sampling and regrouping samples from the original dataset. In terms of feature selection, instead of considering all feature variables, the random forest selects some features at random from a certain range, and then selects the best features for node segmentation.

The steps of the Random Forest algorithm include extracting the training set from the original sample set, obtaining multiple base learners to vote or calculate the mean to obtain the final prediction. In addition, the Random Forest algorithm can also efficiently handle missing values in the dataset and be used for feature importance analysis. The process of random woods model prediction is shown in Fig 3.

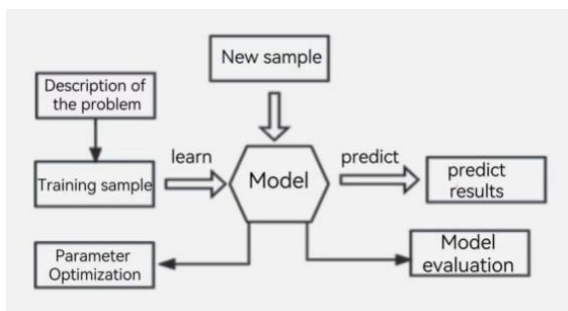


Fig. 2. Machine learning system architecture (Photo/Picture credit: Original).

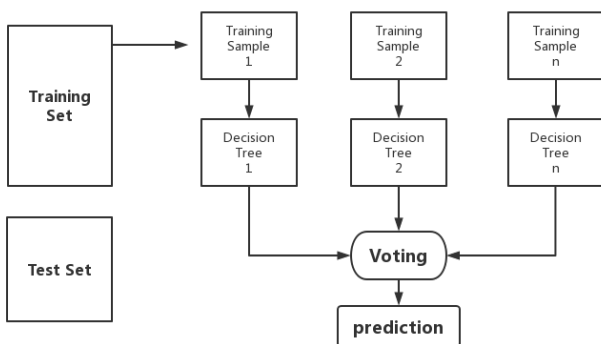


Fig. 3. Schematic diagram of the random forest process (Photo/Picture credit: Original).

A BP neural network, also known as backpropagation neural network, is a type of multi-layer feedforward neural network that is distinguished by the utilization of forward propagation to transmit signals and backward propagation to calculate errors. Specifically, for a neural network model containing only one hidden layer as follows. The general expressions are

$$Y = \sum_{i=1}^n W_{i=1} X_i + b \tag{1}$$

Here, $X_i (i = 1, 2, \dots, n)$ denotes the input value; $W_i (i = 1, 2, \dots, n)$ denotes their weights; b denotes the threshold value; Y denotes the neuron output. A backpropagation (BP) neural network is employed for regression analysis, which involves two distinct stages. The initial stage is the forward propagation, commencing from the input layer, traversing through the hidden layers, and ultimately reaching the output layer. Throughout this process, the neural network transforms input data into corresponding output data. The subsequent stage is the backward propagation of error, commencing from the output layer, then progressing through the hidden layers, and concluding at the input layer. During this process, the neural network automatically adjusts the weights and biases based on the disparity between the predicted values and the actual values, thereby enhancing the prediction accuracy. Following these two stages, the BP neural network is capable of performing regression analysis and forecasting future data. The diagram illustrating the prediction process of the BP neural network model is presented in Fig. 4, where

$$x_j^i = \sigma \left[\sum w_{ij}^1 x_i + b_j \right] \quad (j = 1, 2, \dots, n) \tag{2}$$

Here, w_{ij}^1 are the input layer weights. The output layer output functions are

$$y_j^i = \sum w_{ij}^2 x_j + b_k \quad (k = 1, 2, \dots, n) \tag{3}$$

Here, w_{ij}^2 are the output layer weights. The excitation function uses a S Type function:

$$\sigma(x) = \frac{1}{1 + e^{-\frac{x}{Q}}} \tag{4}$$

where Q is the tuned excitation function of the S type function.

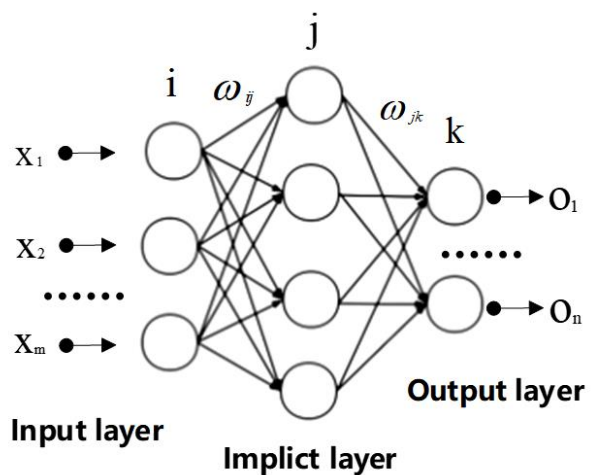


Fig. 4. Principle of BP Neural Network (Photo/Picture credit: Original).

2.3 ARIMA and GARCH

The ARIMA model is a statistical model commonly used for time series analysis and forecasting. ARIMA models combine the features of autoregressive (AR) models, difference (I) operations and moving average (MA) models and can be adapted to many different types of time series data. The ARIMA model consists of three essential components: the autoregressive (AR) component, the differencing (I) component, and the moving average (MA) component. Autoregression refers to the linear relationship that exists between the current observations and past observations. the AR term takes into account the effect of values from past moments of the time series on the current values. in the AR(p) model, p denotes the number of observations from past moments used. The difference operation is used to smooth a time series. A smoothed time series is characterised by its mean and variance being constant in time. If the time series does not satisfy the smoothness condition, it can be transformed into a smooth time series by the difference operation, denotes the order of the difference, i.e., the number of difference operations. A moving average is a linear combination of the errors between the current observations and the past observations. the MA term takes into account the effect of errors in past moments of the time series on the current values. in the MA(q) model, q denotes the number of error terms from past moments used:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1-L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \quad (5)$$

The ARIMA model is commonly represented as ARIMA(p, d, q), where p, d, and q symbolize the order of the autoregressive, differencing, and moving average components, respectively. The selection of appropriate values for these parameters is contingent upon the inherent characteristics of the time series data, which can be discerned by analysing the autocorrelation plot (ACF) and the partial autocorrelation plot (PACF).

GARCH models are statistical models used to model and predict volatility in financial time series. GARCH models can more accurately describe volatility in financial markets by capturing the heteroskedasticity characteristics of time series volatility. The fundamental concept underlying the GARCH model is to represent volatility as a function of previous volatility, taking into account the autoregressive and moving average properties of volatility. The model is based on a basic ARCH (Autoregressive Conditional Heteroskedasticity) structure which captures the autocorrelation of volatility by way of autoregression and a moving average component to capture the heteroskedasticity of volatility. The general representation of the GARCH model is GARCH(p, q), where p denotes the order of the ARCH and q denotes the order of the moving average component. By estimating the parameters in the model, forecasts of future volatility can be obtained. Commonly used GARCH models

include GARCH(1,1), GARCH(2,1), etc. The residual term (observable) is

$$\varepsilon_t = \sigma_t u_t \quad (6)$$

The autoregressive terms:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (7)$$

2.4 Model comparison criteria

Generally speaking, the prediction effect of the model is affected by many factors, the same set of data can simulate different models, but different models predict different effects, the prediction effect in this paper uses the absolute mean error MAE, mean absolute percentage error MAPE, the R² coefficient of determination, and MSE mean squared error, with the following equations:

$$MAE = \frac{1}{N} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

$$MSA = \frac{1}{N} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (9)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (11)$$

In the given context, y_i represents the observed value of the dependent variable, while \hat{y} symbolizes the estimated value of the dependent variable. In the context being considered, a Mean Absolute Percentage Error (MAPE) value of 0 signifies an optimal model, while a MAPE exceeding 100% is indicative of an inadequate model. For subsequent modelling, the MAE and MAPE will be used to evaluate the strengths and weaknesses of the prediction models, so that the optimal model can be selected.

Table 1. ADF Test Results Table.

ADF statistic	-7.062	
P-value	5.199e-10	
Lag order	5	
Number of Observations	176	
Critical value	1%	-3.468
	5%	-2.878
	10%	-2.576

3 Results and discussion

3.1 ADF and PADF test

For empirical research on time series data, a series of tests are conducted before establishing a regression equation. The testing approach is as follows. The ADF test is used to determine the stationarity of each variable. If a variable is stationary, the initial values are used to estimate the equation directly. If a variable is non-stationary and has a unit root, the ADF test is used to determine whether the residuals after regression exhibit cointegration. For variables with non-cointegrated unit roots, they are first transformed into first differences before estimation. When there are unit roots and cointegration, the equation system is estimated using the initial values of the variables. By applying the ADF test method, Table. 1 shows that the p-value is less than 0.05, which indicates stationarity, rejecting the null hypothesis. Therefore, an ARIMA model can be used for subsequent analysis.

3.2 Machine learning scenarios

Random Forest is a predictive modeling technique that utilizes an ensemble of decision trees. In the prediction process, Random Forest applies the test sample to each individual decision tree and then combines the predictions of all trees to generate the final prediction outcome. In the case of regression tasks, the final prediction outcome is commonly calculated as the mean of all the predictions made by the individual trees. Since Random Forest cannot obtain a deterministic equation like traditional models, the model is typically evaluated based on the prediction accuracy on the test data. By analysing the variation of errors over the range of trees from 1 to 10000, the optimal parameters are determined by finding the minimum error.

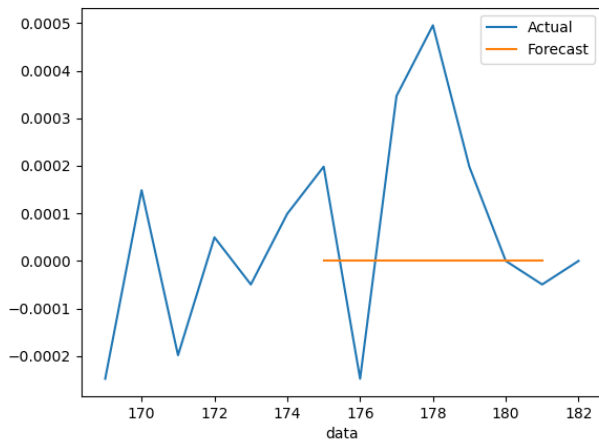


Fig. 5. Random Forest Prediction Performance (Photo/Picture credit: Original).

Fig 5 depicts the comparison between the predicted model and the actual volatility. The random forest regression model exhibits significant deviation from the displayed data across the entire sample set. Furthermore, the line appears as a smooth straight line without any fluctuations, indicating that the model's predictive performance is poor.

The BP neural network, which is a multilayer feedforward neural network model, utilizes the error backpropagation algorithm to address issues including classification, regression, and pattern recognition. The

training algorithm is as follows. First, network initialization, determining the relevant parameters of the network. By specifying the number of nodes in each hidden layer, the output of the hidden layers is calculated using the following formula:

$$l = \sqrt{m + n + \alpha} \tag{12}$$

$$H_j = f \left(\sum_{i=1}^n w_{ij} - \alpha_j \right) \quad j = 1, 2, \dots, l \tag{13}$$

Here, m is the number of neurons in the input layer, n is the number of neurons in the output layer, and l represents the number of neurons in the hidden layer. α is a constant between 1 and 10, and f is the activation function of the hidden layer. Then, compute the output layer:

$$P_k = \sum_{j=1}^l H_j W_{jk} - b_k \quad k = 1, 2, \dots, m \tag{14}$$

Subsequently, one can calculate the prediction error between the predicted output and the expected output:

$$e_k = Y_k - P_k \tag{15}$$

Then, we can update the weights, where η represents the learning rate:

$$w_{jk} = w_{jk} + \eta H_j e_k \quad j = 1, 2, \dots, l; k = 1, 2, \dots, m \tag{16}$$

Then, we update the thresholds:

$$\alpha_i = \alpha_j + \eta H_j (1 - H_j) \sum_{k=1}^m w_{jk} e_k \quad j = 1, 2, \dots, l \tag{17}$$

$$b_k = b_k + e_k \quad k = 1, 2, \dots, m \tag{18}$$

If the error meets the desired accuracy requirement, the process is finished. Otherwise, continue training until the error reaches the expected requirement. The simulation of the BP neural network will only end when the error criterion is satisfied.

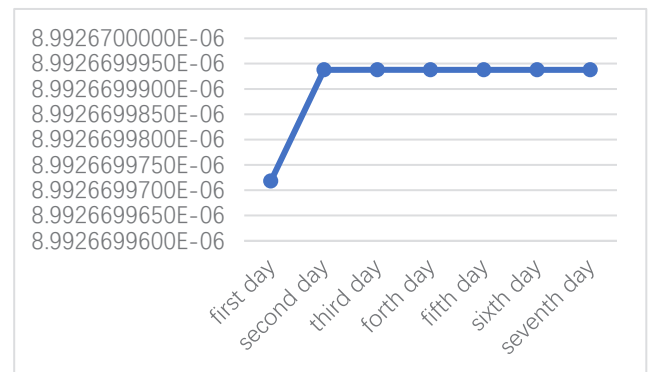


Fig. 6. Prediction of the Volatility for the last seven days in the dataset (Photo/Picture credit: Original).

3.3 ARIMA and GARCH

Fig 6 displays the prediction of the ARIMA model with parameters (1,0,1), which were selected based on the results of the ADF test. The forecast focused on the volatility of the dataset's last seven days. After conducting

the PADF test, the study opted for training the GARCH model with parameters (2, 1). The model was utilized to forecast the volatility of the dataset for the most recent sixty days, yielding the outcomes in Fig. 7.

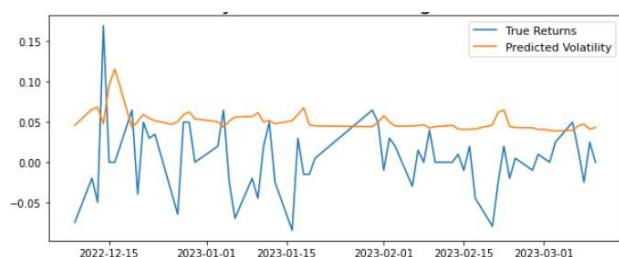


Fig. 7. GARCH Model Prediction Results (Photo/Picture credit: Original).

3.4 Empirical analysis

According to the empirical results, when using the random forest and ARIMA models for volatility prediction, the predicted results exhibit a close-to-zero and highly stable characteristic. This may indicate that these two models have a certain bias in the prediction process, resulting in an inability to accurately capture the true fluctuations in volatility. While such trendiness and stability may sometimes be considered good features, in the context of volatility prediction, it may suggest that the models fail to effectively capture the actual changes in volatility.

On the other hand, the prediction results of the GARCH model are relatively better compared to the random forest and ARIMA models, but still exhibit moderate performance. This may indicate that the GARCH model is relatively more accurate in considering the dynamic changes in volatility, but some errors still exist.

It is important to note that suboptimal prediction results can have multiple reasons. Firstly, the selected models may not be suitable for the dataset under study, or the parameter settings of the models may be inappropriate. Secondly, factors such as data quality, feature engineering processing, and the choices and adjustments made during the model training process can also affect the prediction results. Additionally, the complexity and uncertainty of financial markets also increase the difficulty of volatility prediction.

In conclusion, although the predicted results of the random forest and ARIMA models exhibit a close-to-zero and highly stable pattern, and the prediction results of the GARCH model are relatively better, none of them achieve the desired prediction performance. Further research can explore alternative models or improve existing models while refining the steps of data preprocessing and feature engineering to enhance the accuracy of volatility prediction for interest rate futures.

4 Limitations and prospects

There are still some challenges and limitations in this study, including insufficient indicator selection and feature extraction, failure to consider the impact of policy changes and market sentiment, as well as failure to implement the empirical part of the BP neural network and the joint prediction and benefit evaluation of the model. These shortcomings limit our exploration of the accuracy and comprehensiveness of interest rate futures risk prediction.

However, we are confident in future research and have developed a series of improvement measures. Firstly, we will strengthen the selection of indicators and feature extraction, further exploring key factors that affect volatility, in order to improve the accuracy of the model's prediction. Secondly, we will introduce factors such as policy changes and market sentiment to better understand the changes in volatility and improve the risk prediction model for interest rate futures. Thirdly, we will conduct in-depth research and empirical research on the application of BP neural networks, incorporate them into the empirical part, and combine various models for prediction, ultimately evaluating the effectiveness and performance of the models. In the next step of research, we will strive to achieve these improvements, optimize the performance of the model, and propose a more accurate and reliable interest rate futures volatility prediction model.

5 Conclusion

To sum up, given the complexity of financial markets and the volatility of interest rate futures, interest rate futures have become increasingly important as a tool for managing global interest rate risks. Accurately predicting the risk of the interest rate futures market is of great economic and practical value, given its substantial influence on the global macroeconomy and financial markets. Models based on machine learning techniques have shown great potential in predicting interest rate futures risk. These models can provide valuable information for investors and risk managers, helping them make more informed decisions to reduce risks and achieve better returns. However, it is important to note that any prediction model has its limitations, and therefore, the accuracy and applicability of the model should be carefully evaluated in practical applications.

Author contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

1. X. Li, *Fut. Mark.*, **6**, 20-23 (2019).
2. X. Zhang, *Data Ana. and Knowle. Dis.*, **3(2)**, 70-79 (2019).
3. M. Li, Y. Xie, *Mod. Fin.*, **2**, 89-95 (2020).

4. L. Wang, X. Xing, *Statistics and Decision Making*, 2018, **34(20)**, 86-89 (2018).
5. G. Ren, *Research on bond default risk prediction based on machine learning*, Dissertation of Southwest University of Finance and Economics, 2021.
6. X. Yang, *Sof. Engi.*, **25(12)**, 1-8 (2022).
7. Q. Zhang, *Research and application of neural network in stock price prediction model*, Dissertation of Tianjin University of Finance and Economics, 2021.
8. N. Areal, *J. of Fut. Mar.*, **38(12)**, 1566- 1585 (2018).
9. X. Huang, *J. of Trad.*, **14(4)**, 51-60 (2019).
10. J. Willem van den End, *J. of Fore.*, **36(6)**, 623-644 (2017).
11. C. Li, *Int. J. of Fin. Studies*, **8(4)**, 63-78 (2020).
12. F. Waldow, S. Matthias, K. Christopher, and Thomas G. Fischer, *J. of Risk and Fin. Mana.* **14(3)**, 119 (2021).
13. W. Liu, Y. Gui, G. Qiao, *Res. Int. Bus. Fin.*, **61**, 101669 (2022).