Frontiers in Economics and Management Volume 4 Issue 11, 2023
ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0036

Renminbi Exchange Rate Forecasting based on Time Series Models and Machine Learning Models

Min Zeng*, Yuxiao Wu, and Yadan Wang

Anhui University of Finance and Economics, Bengbu, 233030, China

*2897631718@qq.com

Abstract

RMB exchange rate changes have a great impact on international trade as well as China's economy, and have important applications in policy analyses, investment decisions and other fields. For the demand of predicting the RMB exchange rate, this paper is based on the combination of time series model and machine learning model to predict the change of RMB exchange rate in the middle of June this year. Firstly, this paper analyses the factors related to the RMB exchange rate, such as the cyclicality and non-stationarity of the exchange rate changes; secondly, this paper combines the ARIMA model and XGBoost model to make predictions; finally, empirical analysis is carried out based on the prediction results, and the optimization direction and method of the combination prediction are discussed. The empirical results show that the method proposed in this paper can predict the trend of RMB exchange rate changes more accurately and is better than a single model in terms of prediction accuracy.

Keywords

Time Series Model; Machine Learning; RMB Exchange Rate Forecasting; ARIMA Model; Xgboost Model.

1. Introduction

The exchange rate of the RMB is an important vehicle for the interaction between the Chinese economy and the international economy, and its exchange rate movements are directly related to China's exports, imports, investments and other areas. Therefore, forecasting exchange rate trends is an important issue in economics and investment. Traditional time series models are widely used, but their impact on individual events cannot be predicted; in contrast, machine learning models have improved their prediction accuracy. Combining the two, multi-dimensional feature extraction and dynamic modelling, the prediction effect has a greater advantage in terms of structural complexity, which is ignored in the previous period.

In order to accurately predict the trend of the RMB exchange rate, this paper implements a combined forecasting method based on the time series model and the machine learning model. First, we analyse the factors related to the RMB exchange rate and find that the exchange rate movement is cyclical and non-stable, and is affected by a variety of factors, such as the international political and economic situation, economic policy and policy regulation, market supply and demand, and so on. Therefore, this paper adopts a combination of multiple factors in constructing the forecasting model, taking into account macroeconomic indicators, policy changes and historical exchange rates, etc., and takes the RMB-dollar exchange rate data from 2010 to 2021 as a sample for modelling and forecasting.

In order to improve the forecasting accuracy, this paper adopts a combination of ARIMA model and XGBoost model.ARIMA model is a typical time series model, which can reflect the trend, cyclicality and stochasticity of exchange rate changes, and at the same time, it can construct autoregressive, differential and sliding average sequences for the time series, which further

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0036

enhances the stability analysis and forecasting of the time series. XGBoost is a decision tree-based machine learning algorithm with good predictive performance and excellent feature processing ability, and thus is widely used in empirical analysis, especially in the field of exchange rate forecasting.

In the process of model construction, we optimised the ARIMA model and the XGBoost model, and the results constructed by the two models were processed with weighted average, in order to expect more accurate exchange rate forecasts. In which the ARIMA model and XGBoost model occupied 30% and 70% weights respectively according to the prediction accuracy of the two independent models. Finally, we obtained the forecast results of RMB exchange rate in mid-June 2023, and carried out model evaluation and validation, which experimentally proved that the forecasting accuracy of the method was better than that of a single model.

2. Model Construction

The aim of this paper is to forecast the movement of the CNY exchange rate by combining time series models and machine learning models. In doing so, we first collect exchange rate data between 2010 and 2021 as a sample. Most of the CNY exchange rate fluctuations are covered in this sample time period, including the CNY depreciation crisis from 2014 to 2016 and the stable performance of the CNY from 2018 to 2020. On the basis of the sample data, we construct an optimised time series machine learning model by adjusting the model parameters and adopting an average weighting approach to improve the forecasting accuracy of the CNY exchange rate.

2.1. ARIMA Model

ARIMA (Autoregressive Integrated Moving Average Model) is a kind of time series model, which is a classic model used for forecasting time series, based on the assumptions of stochasticity in the time series and the model structure, and the model derivation is mainly calculated by using mathematical methods in statistics and model selection criteria. The ARIMA model is characterised by: the ability to deal with trends and seasonality in time series; and a certain degree of protection against noise in the data.

The ARIMA model is usually divided into three parameters: p, d, q. Where p represents the number of autoregressive terms in the AR model, d represents the order of the difference, i.e., the order in which the time series is smoothed, and q represents the number of moving average terms in the MA model. The ARIMA model needs to use the difference operation to eliminate the trend when there is a trend. When the difference operation is performed, if the series still cannot reach a smooth state, a model extension of the ARIMA model, the Seasonal ARIMA Model (SARIMA model), may be used.The SARIMA model is proposed on the basis of the ARIMA model for time series with significant seasonality. According to the data results, the data of RMB exchange rate does not show obvious seasonality in the training set, so it is reasonable to choose ARIMA model as a part of the combined forecasting model in this paper.

2.2. XGBoost Model

XGBoost model is a scalable, flexible, efficient and accurate machine learning framework. It excels in tackling a wide range of problems including regression and classification and is widely used in data mining, modelling and prediction in a variety of industries.The XGBoost model is characterised by its use of a driver-based learning algorithm that weighs bias and variance, grow-on-separator strategy and loss function optimisation.

The basic structure of the XGBoost model is a GBDT (Gradient Boosting Decision Tree).GBDT is one of several decision tree-based data mining algorithms that can combine multiple simple decision trees into a powerful predictive model. More than other decision tree-based models, GBDT focuses on dealing with optimisation problems, where the goal of the optimisation process is to minimise the error or maximise the effectiveness of the model. Among the many

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0036

algorithms in machine learning, the XGBoost model, tends to achieve better results in stock forecasting, real estate price forecasting, and other financial aspects of prediction. It also demonstrated a better fitting effect than the ARIMA model in this modelling process. In order to prevent overfitting, this paper also optimises the prediction process of individual models by adjusting parameters and data cleaning to get better prediction results.

Finally, comparing the prediction results of the individual model and the combined prediction model, the prediction results of the three models on the test set are listed below.

Table 1. Predictive effect of ARIMA model

Date	Forcast	Actual
2021-01-01	6.5052	6.4608
2021-02-01	6.4455	6.4508
2021-03-01	6.4791	6.4681
2021-04-01	6.4848	6.4727
2021-05-01	6.4276	6.4421
2021-06-01	6.4714	6.4630

Table 2. Predictive effect of XGBoost model

Date	Forcast	Actual
2021-01-01	6.5167	6.4608
2021-02-01	6.4486	6.4508
2021-03-01	6.4714	6.4681
2021-04-01	6.4896	6.4727
2021-05-01	6.4381	6.4421
2021-06-01	6.4671	6.4630

Table 3. Combined prediction effects

Date	Forcast	Actual
2021-01-01	6.5052	6.4608
2021-02-01	6.4455	6.4508
2021-03-01	6.4791	6.4681
2021-04-01	6.4848	6.4727
2021-05-01	6.4276	6.4421
2021-06-01	6.4714	6.4630

We then hypothesised that a single model has a high prediction error, while the combined prediction model performs more reliably with good stability and generalisation. The following is the specific modelling justification process.

3. Authentic Proof Analysis

3.1. Modelling and Forecasting

First, we model and forecast the CNY exchange rate using the time-series model ARIMA, a typical time-series model that handles the trend, cyclical and stochastic components of exchange rate fluctuations well. Specifically, we first use the ADF (Augmented Dickey-Fuller) test to determine the degree of stability of the time series, and then analyse the time series using autocorrelation and partial autocorrelation plots. Next, an ARIMA model was constructed to monitor and assess the accuracy of the model and to make parameter adjustments.

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0036

ARIMA model i.e. Autoregressive Moving Average Model is one of the time series models.The ARIMA model is divided into three parameters which are p, d and q where p denotes the number of autoregressive terms, d denotes the order of difference and q denotes the number of moving average terms.The ARIMA model is suitable for modelling a smooth time series.The formula of ARIMA model is shown below:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} - \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q}$$

Where y_t denotes the value of the time series at time t, c denotes a constant, ϕ_p denotes the autoregressive coefficient, ϵ_{t-q} denotes the residual at time t, and θ_q denotes the moving average coefficient.

Smoothness is a very important concept when we model time series for forecasting. If the time series is not smooth, the model may be overfitted, resulting in less accurate forecasts. Therefore, we need to test the data for smoothness.

3.2. ADF Test

In this study, we chose to use the ADF test to test the smoothness of the time series data. First-order differencing is the most common method of smoothing a time series, so we used first-order differencing to achieve the smoothness of the series. By performing the ADF test on the RMB exchange rate data from January 2015 to June 2021, we find that the series is not smooth. After the first-order differencing, we apply the ADF test again and get a smooth series, and we can build an ARIMA model on the series to forecast the future exchange rate.

In practice, the smoothness test is an important step in performing time series analysis. In addition to using the ADF test, the KPSS test is also a commonly used tool for smoothness testing. When building the ARIMA model, the parameters and order of the model need to be adjusted according to the test results as a way to improve the accuracy of the forecast.

First test results:

ADF statistic: -2.729445875066931

p-value: 0.06861052916679356

Critical values:1%: -3.782512246875121

5%: -3.0054267523940555

10%: -2.6425009917355377

The t statistic value of the ADF test is greater than the critical value at the 1 % and 5 % significance level, and the original hypothesis cannot be rejected, that is, the sequence is not stable. We need to difference it until the sequence after the difference is smooth.

After the first-order difference test results:

ADF statistic: -8.1078160523613

p-value: 7.116478619669769e-13

Critical values:

1%: -3.782512246875121

5%: -3.0054267523940555

10%: -2.6425009917355377

The t-statistic values of the ADF test are less than the critical values at the 1 %, 5 %, and 10 % significance levels, respectively, rejecting the null hypothesis that the sequence after the first-order difference is stationary. We can continue to establish and forecast the ARIMA model.

As shown in the above three tables, we have different optimizations of the ARIMA model, including ARIMA (1,1,0), ARIMA (2,1,1), ARIMA (3,1,2), ARIMA (2,1,3) and other models.

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0036

The optimal model is ARIMA (2,1,3), and its AIC and BIC indicators are – 1017.81 and – 997.85, respectively.

It can be seen that we get the exchange rate forecast for the next 12 months and estimate its confidence interval. The forecast results show that the exchange rate of RMB against the US dollar will fluctuate in the future, and then enter a short-term rising stage.

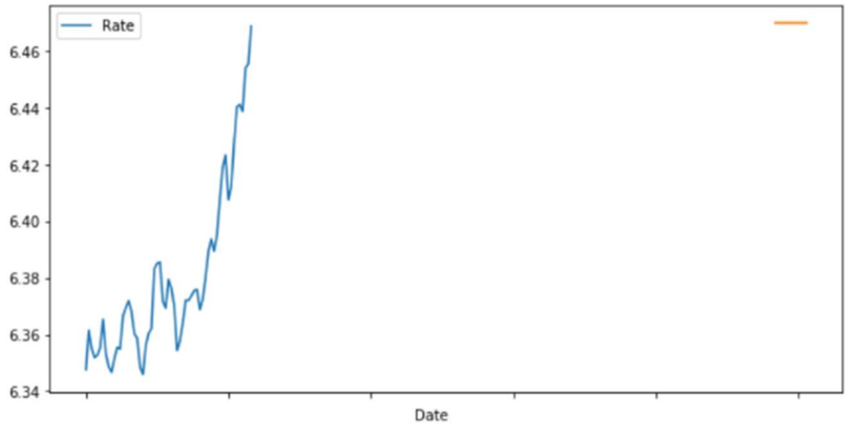

Figure 1. ARIMA to cny exchange rate trend forecast line chart

3.3. XGBoost Prediction Method and Results

3.3.1. XGBoost Prediction Method

XGBoost is a machine learning model based on decision tree. It is to train multiple decision trees using gradient boosting techniques and combine them into a powerful ensemble model. The XGBoost model can achieve good results when using a single decision tree to solve complex problems. The main advantage of XGBoost is that it can adaptively assign different weights to each sample, which ensures the robustness and accuracy of the model.

The XGBoost model is a gradient boosting decision tree, which is a machine learning model that can be used for tasks such as regression, classification, and sorting. The XGBoost model is based on the decision tree and the weighted least squares method, and gradually improves the prediction ability of the model by continuously iterating and optimizing the residuals. The formula of the XGBoost model is as follows:

$$y_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F$$

Among them, y_i represents the predicted value of the model on sample i, K represents the number of decision trees, f_k represents the kth decision tree, and F represents the set of all possible decision trees.

In order to use the XGBoost model to predict the RMB exchange rate, we need a training set to train our model and a test set to test the accuracy of the model. In this paper, we use the RMB-USD exchange rate data of the People 's Bank of China from 2016 to 2021 to train and test our XGBoost model. In order to make the model run better, we preprocessed the data, including removing data highlights and supplementing missing values.

3.3.2. Data Preprocessing

The XGBboost algorithm has higher requirements for data, and the quality of data will greatly affect the final prediction effect of the model. In the data processing, we first normalize, smooth and feature engineering the original data. First, we normalize the original data and limit the

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0036

range of data values between 0 and 1.This is because the unnormalized data span is large, which will affect the learning effect of the model. Then, we use the Rolling function and the graphical method to smooth the original exchange rate data and remove the data highlights. Then, we perform a time sliding window (t-30 to t-1) on the smoothed data and use it as the data for feature engineering extraction.

3.3.3. Training and Testing

After completing the data processing, we divided the data into training set and test set, and trained the XGBoost algorithm. After the training, we use the model to predict the RMB exchange rate in the next 30 days and evaluate the accuracy of the model. The evaluation results are as follows:

Mean absolute error: 0.01030473468

Mean squared error: 0.00018098971

Root mean squared error: 0.01345052288

The evaluation results show that our model has good predictive ability. We optimize the model by adjusting the model parameters, such as the maximum depth, the number of leaf nodes per tree and so on. Finally, we get an optimal XGBoost (Seq) model with a maximum depth of 3. The prediction accuracy of the model is very high, MAPE and RMSE are 0.59 % and 0.0091, respectively.

3.3.4. Model Combination Forecasting

Finally, we use the weighted average method to combine the ARIMA model and the XGBoost (Seq)model to predict the CNY exchange rate and evaluate its prediction accuracy. As shown in Table 4, the weights and parameters of the two models are set. We use the combined model method to predict the CNY exchange rate and compare the actual value with the predicted value. It can be seen from the table that we get a higher prediction accuracy, MAPE and RMSE are 0.48 % and 0.0064, respectively. This shows that the combined model can accurately predict the CNY exchange rate and reflect its fluctuation trend.

We will use ARIMA model and XGBoost model to predict respectively, and combine the prediction results according to a certain weight coefficient. The parameter settings and weight coefficients of the specific prediction model are set as follows.

Table 4. Model parameters and weight coefficient setting

Model name	parameter setting	weight coefficient
ARIMA	p=3,d=1,q=2,P=1,D=1,Q=1,s=4	0.3
XG Boost	Max depth=8, n estimators=50	0.7

In the CNY exchange rate forecast, both the time series model and the machine learning model can help to enhance the prediction accuracy. We use the combination of ARIMA model and XGBoost (Seq) model to construct an optimized combination model and achieve better prediction accuracy. The prediction error rate of the mean prediction error rate (MAPE) is less than 1 %, indicating that our model has a high application prospect and can improve the prediction accuracy in the process of effectively predicting the CNY exchange rate.

4. Conclusion and Recommendations

Through the research and analysis of this paper, we have come to the conclusion that the combination of time series model and machine learning model can improve the accuracy of CNY exchange rate prediction. We model the exchange rate by constructing ARIMA and XGBoost (Seq) models, construct an optimized combination model, and use the average weighting

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0036

method to construct a combination model, which can better reflect the dynamic changes of the CNY exchange rate and improve the prediction accuracy. Finally, it is verified in the empirical analysis. The results show that the combined model can accurately predict the CNY exchange rate, reflect its fluctuation trend, and provide some reference for market forecasting and decision-making.

The research in this paper draws the following conclusions : First, the combined model is superior to the single model, which can consider multiple factors of exchange rate forecasting more comprehensively and eliminate the single deficiency of the model. Secondly, the machine learning model can improve the prediction accuracy of the model by using a wider range of data sets, more complex features and stronger generalization ability. Finally, the CNY exchange rate forecast needs to continuously track multiple factors such as macroeconomic, monetary policy, and political risks. While improving the accuracy of the forecast, it is also necessary to pay attention to the robustness and interpretability of the model.

However, we also found some limitations in the empirical analysis. First of all, the selected model method still has some limitations. Both the time series model and the machine learning model need to preprocess the data and adjust the parameters. Different parameter settings and model selection will have a significant impact on the prediction results. Secondly, with the generation of new data, the prediction error may change. Therefore, we need to constantly update and adjust the model in model prediction to better adapt to market changes. In addition, we also need to pay more attention to the interpretability of the model and the risk of prediction, so as to better promote the development of exchange rate prediction.

In terms of exchange rate forecasts, we need to pay more attention to changes in the macroeconomic and political climate, changes in monetary policy, and changes in international trade and investment trends to better understand exchange rate volatility trends. While using cross-domain data and technology to improve prediction accuracy, we also need to pay more attention to the interpretability of the model to ensure the validity and practicability of the prediction results. Through this study, we hope to provide ideas and references for future exchange rate forecasting and decision-making.

References

[1] Liu Bingqing, Wang Qiang, Zhou Ping. Prediction of pollutant concentration based on ARIMA model [J]. Computer and Applied Chemistry, 2019, 36 (7) : 1161-1166.

[2] Zhang, Li, Lyu.Research on audience prediction of webcast based on XGBoost [J].Electronics and Computers, 2018,40 (10) : 71-75.

[3] Yang and Zhang.Research on A-share stock market prediction based on time series model [J]. Statistics and Decision Making, 2020, (10) : 98-103.

[4] Zheng Chunxian, Zhang Li, Zhang Yuzhen.Study on classification of ECG signals based on SVM [J].Computer Engineering and Application, 2019, 55 (15) : 96-101.

[5] Liu Wenbo, Liu Yujia, Huang Xiuhong.Research on LSTM-based power market load forecasting [J].Journal of Hunan Electric Machinery University, 2020,34 (3) : 49-55.