

# *Research on Forecasting of China's Monetary Policy Based on Random Forest Algorithm*

*Qiu Chuanxin*

College of Management  
Shenzhen University  
Shenzhen, China  
e-mail: [670367646@qq.com](mailto:670367646@qq.com)

*Li Tong*\*

College of Management  
Shenzhen University  
Shenzhen, China  
e-mail: [litong@szu.edu.cn](mailto:litong@szu.edu.cn)

*Qiu Xuelin*

College of Management  
Shenzhen University  
Shenzhen, China  
e-mail: [2815342076@qq.com](mailto:2815342076@qq.com)

**Abstract**—This paper uses the random forest algorithm model to quantify and predict the monetary policy of the People's Bank of China under the input of 16 macroeconomic indicators. It is compared with three other machine learning algorithms (CART decision tree, support vector machine and neural network algorithm), discrete selection model and combined prediction model. The results show that the random forest algorithm shows better prediction accuracy in predicting the direction of the central bank's monetary policy.

**Keywords**—Change forecast; random forest algorithm; machine learning; monetary policy; quantitative analysis

## I INTRODUCTION

China is in a new normal economic period. At present, China's economy is developing healthily and economic growth remains resilient. This has a huge relationship with the rational use of the central bank's monetary policy. From the current policy orientation, China's central bank's use of monetary policy is more accurate than ever. The existing research results show that macroeconomic trends are closely related to the direction of the central bank's monetary policy [1]. The central bank's monetary policy trend will affect the behavior of market participants such as the public and financial institutions to a certain extent, thus helping the economy grows steadily. It is of theoretical and practical significance to predict the direction of the central bank's monetary policy.

The change of the central bank's monetary policy is a complex issue involving many aspects. At present, scholars generally adopt discrete selection models and combined forecasting models to predict the operation direction of the central bank's monetary policy. However, with the emergence of new algorithms and new ideas, more and more researches have been conducted on the interpretation of complex financial systems from the perspectives of machine learning and economic physics [2]. Compared with traditional prediction models, machine learning algorithms have higher predictions accuracy.

This paper collects monthly data that can reflect some of the current domestic and international economic and financial situation indicators, and attempts to use the relevant algorithms in machine learning to predict the direction of China's central bank monetary policy. We believe that using

machine learning algorithms to predict the central bank's monetary policy trend has certain innovation. Because machine learning algorithms are rare in the study of predicting the direction of monetary policy. This paper mainly applies random forest algorithm to predict the direction of China's central bank's monetary policy. The three machine learning algorithms of CART decision tree, support vector machine and neural network are compared, in order to inspire the formulation, implementation and innovation of China's central bank monetary policy.

## II LITERATURE REVIEW

In practice, the central bank's adjustment of the target benchmark interest rate is adjusted in a discrete manner both in time and in scale. Therefore, the discrete selection model is widely used in forecasting the central bank's monetary policy operation direction.

Dueker [3] made an early attempt to model the discreteness of the Fed's federal funds target interest rate, and proposed a method based on the ordered Probit model to estimate the extent of changes in the federal funds target interest rate. Hu and Phillips [4,5] proposed an orderly Probit model suitable for both stationary and non-stationary variables to predict the timing of changes in the federal funds target rate. Kim [6] conducted out-of-sample predictions on Hu and Philips models. Li Haitao [7] also applied the non-stationary discrete selection model to predict the operational status of China's monetary policy.

All of the above scholars use a single discrete selection model to predict the trend of the target benchmark interest rate. Because the combined model can not only improve the accuracy of model prediction, but also allow the prediction combination to consider more variables than the single multivariate model, Pauwels [8] presented data and models for applying prediction combinations to Hu and Philips. The prediction results are formed by assigning different weights to each single model. The empirical results show that the prediction combination improves the accuracy of model prediction [7]. Ruthira Naraidoo [9], Andrey Vasnev [10] and Pauwels L [11] used a combined prediction method based on the ordered Probit model to study the decision-making of monetary policy operations of the Reserve Bank of South Africa, the Reserve Bank of Australia and the Federal Reserve.

Compared with the existing research results, the innovation of this paper is mainly reflected in the attempt to use the random forest algorithm in the machine learning algorithm to predict the direction of the central bank's monetary policy. The predicted accuracy can not only be compared with other three kinds of machine learning. The algorithm (cart decision tree, support vector machine and neural network algorithm) is compared, and also can be compared with discrete selection model and combined prediction model. The structure of the rest of the paper is as follows: The third part introduces the random forest algorithm; the fourth part is The selection of indicators and data preprocessing; the fifth part is empirical analysis, the random forest classification algorithm is applied to predict the direction of the central bank's monetary policy; the sixth part is the conclusion of this paper.

### III RANDOM FOREST ALGORITHM

#### A. Introduction of random forest algorithm

Random forest is an integrated learning algorithm based on decision tree. It is also a supervised learning algorithm with good performance in classification and regression problems. The direction of the central bank's monetary policy can be summarized as the monetary policy tends to expand, the monetary policy remains unchanged, and the monetary policy tends to be tightened. Therefore, the trend of the central bank's monetary policy should be predicted using a random forest classification algorithm.

#### B. The basic knowledge of random forest algorithm

##### 1) Decision tree

The decision tree is a tree structure that classifies samples based on features. From a given training data set, according to the criteria of feature selection, the optimal partition features are recursively selected, and the training data is segmented according to the features, so that each child is divided. The data set has a process of best classification results [12]. The random forest algorithm uses the cart decision tree as the base learner, and the cart tree selects the features through the Gini index.

##### 2) Gini index

The Gini index indicates the probability that a randomly selected sample is split in the sample set. The smaller the Gini index, the higher the purity of the set [12]. Suppose a sample set  $D$  have  $k$  Categories, if used  $|C_k|$  Representation category Number of samples, Indicates the total number of samples, then the probability of each category is , then sample collection The Gini index is:

$$Gini(D) = 1 - \sum_{k=1}^K \left( \frac{|C_k|}{|D|} \right)^2 \quad (2)$$

#### C. Generation of random forests

The framework diagram of the random forest algorithm is shown in Figure 1. The generation steps are as follows:

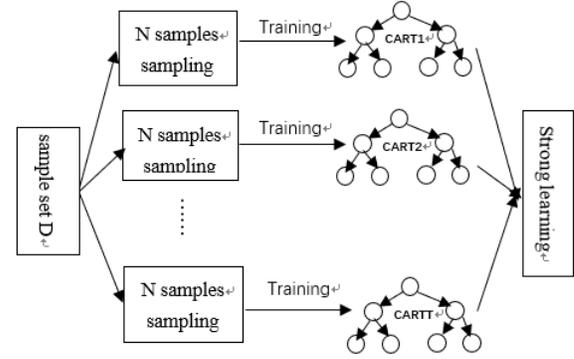


Figure 1 Random forest algorithm framework

#### D. Evaluation index of random forest algorithm performance

In this paper, the accuracy of the accuracy, precision, and recall are used to evaluate the prediction results of the random forest algorithm [12]. The calculation of these indicators is based on the confusion matrix of Table 1. Their calculation formula is as follows:

In this paper, the accuracy of the accuracy, precision, and recall are used to evaluate the prediction results of the random forest algorithm [12]. The calculation of these indicators is based on the confusion matrix of Table 1. Their calculation formula is as follows:

TABLE 1 CONFUSION MATRIX

	The predicted value is 1	The predicted value is 0
Actual value is 1	TP	FN
Actual value is 0	FP	TN

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$precision = \frac{TP}{TP+FP} \text{ or } \frac{TN}{TN+FN} \quad (4)$$

$$recall = \frac{TP}{TP+FN} \text{ or } \frac{TN}{TN+FP} \quad (5)$$

### IV INDICATOR SELECTION AND DATA PREPROCESSING

#### A. Selection of economic indicators

In the prediction study of the central bank's monetary policy trend, there is no uniform standard in the academic community regarding the selection of macroeconomic indicators. Hu and Philips selected 11 indicators [5] such as annual inflation rate, unemployment rate, number of initial claims for unemployment insurance, consumer confidence index, and industrial production index when forecasting the Fed's target fund rate. Andrey Vasnev [10] selected unemployment rate, consumer confidence index, commodity price, total export value, M3 annual growth rate, owner's housing loan, and monthly growth rate of S&P 500 index when forecasting the direction of the Australian Reserve

Bank's monetary policy. Indicator [11]. Li Haitao selected 40 indicators including CPI, GDP, commodity retail price index, PPI and other real economic activities, currency and capital markets, stock index, foreign exchange market, interest rate and price level when forecasting the operation status of China's monetary policy [7].

Since China's national conditions are different from those of the United States and Australia, this paper mainly refers to the indicators selected by domestic scholars in constructing the macro-control system when selecting macroeconomic indicators, such as Li Haitao [7] and Qiao muzi [14]. Shang Yuhuang [15]. In addition, this article also included some macroeconomic indicators mentioned in the "China Monetary Policy Implementation Report" [14] in the index system of this paper, such as the RMB exchange rate index and consumption. The confidence index, etc. Finally, this paper selects 16 indicators from macroeconomic entities, financial institutions, interest rate markets, real estate markets, external financial markets, and stock markets. This article obtains monthly data from January 2007 to June 2018. The data comes from the National Bureau of Statistics, the People's Bank of China, and the Wind database. Its information is shown in Table 2.

### B. Trend data of the central bank's monetary policy

The deposit reserve ratio and the regular deposit and loan interest rate are the most important operational tools of China's monetary policy [7]. Therefore, this paper only pays attention to the central bank's adjustment of the deposit reserve ratio and the regular deposit and loan interest rate when sorting out the data of the central bank's monetary policy, did not consider the central bank's open market operations, TMLF and other monetary policy tools.

Through the policy documents issued by the official website of the People's Bank of China on adjusting the deposit reserve ratio and the regular deposit and loan interest rate, the data on the trend of China's central bank monetary policy from January 2007 to June 2018 was compiled. In the data set of this paper, the central bank's monetary policy trend is organized into three categories: "1" means that the central bank's monetary policy tends to expand, "0" means that the central bank's monetary policy remains unchanged, and "-1" means that the central bank's monetary policy tends to tighten. The time series of China's central bank's monetary policy trend is shown in Figure 2.

As can be seen from Figure 2, between January 2007 and June 2018, the central bank took 21 months to expand the monetary policy, and 31 months the central bank adopted a tight monetary policy with 86 months of central bank. The monetary policy remains unchanged.

### C. Data preprocessing

This article uses Eviews 8.0 to preprocess the data. For indicators that have only monthly data missing from quarterly data (such as GDP), quadratic-match-average is

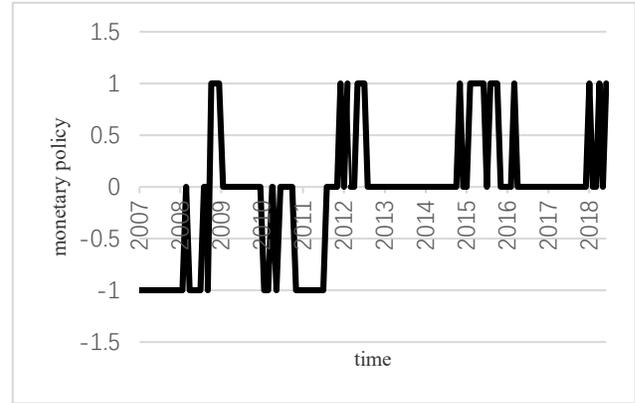


Figure 2 Timeline chart of central bank monetary policy

used to convert quarterly data into monthly data. For indicators that only have monthly data missing from the daily data (such as price-to-earnings ratio), use the quadratic-match-sum (constant matching sum) to convert the daily data into monthly data.

After the data of each indicator is preprocessed, its descriptive information is shown in Table 2.

TABLE 2 DESCRIPTIVE INFORMATION OF INDICATORS

index	mean	std	min	50%	max
Real interest rate	2.388	0.932	0.1	2.413	4.84
M1/M2	0.32	0.032	0.261	0.314	0.378
PMI	51.562	2.65	38.8	51.2	59.2
Deposit ratio	0.686	0.022	0.639	0.683	0.746
Short-term loan growth rate	0.01	0.009	0.033	0.01	0.047
Credit/GDP	18.093	3.744	12.267	17.229	27.577
P/E ratio	20.559	11.95	9.76	16.685	69.64
Total market value /GDP	6.912	1.867	3.785	6.759	12.655
Actual effective exchange rate index	110.315	12.897	87.07	109.31	131.03
Deposit spread	1.683	1.389	2.59	1.935	3.4
Industrial added value	10.578	5.02	2.93	9.55	29.2
Foreign currency loan / loan total	0.009	0.001	0.007	0.009	0.012
Solid vote / GDP	3.094	1.017	0.99	3.209	6.982
Interbank Offered Rate	0.009	0.001	0.007	0.009	0.012
Consumer information index	99.482	3.849	92.43	100.555	106.59
National Housing Climate Index	0.009	0.001	0.007	0.01	0.012

Source: People's Bank of China website, Wind

## V RANDOM FOREST ALGORITHM PREDICTS THE TREND OF CENTRAL BANK'S MONETARY POLICY

### A. Define abbreviatio the Division of the data set

This paper uses the Random Forest Classifier in the Scikit-learn library of Python3.7 to predict the direction of the central bank's monetary policy. Using the train\_test\_split function in the scikit-learn library, 70% of the data is used as the training set, and the remaining 30% is used as the test

set, which is 138 data from January 2007 to June 2018, of which 96 data are divided. For the training set, the remaining 42 pieces of data are divided into test sets.

### B. Random forest algorithm parameters

Because the sample size of the data set used to predict the direction of the central bank's monetary policy is small (138 sample data), and the characteristics of the data set are not many (16 features), `min_samples_split` (the minimum number of samples required for internal nodes to be subdivided) The parameters such as `min_samples_leaf` (the minimum number of leaves nodes) are taken as default values, and only the parameters of `n_estimators` (the largest number of decision trees in the random forest) and `max_depth` (the maximum depth of the decision tree) are adjusted to determine the best. The effect parameters, the best performance parameters obtained in this experiment are `n_estimators` (the maximum number of decision trees in the random forest) is 500, and `max_depth` (the maximum depth of the decision tree) is 25.

This paper uses the grid search method to determine the optimal parameters of the random forest algorithm. Auto-tuning can be achieved by calling `GridSearchCV` in the `scikit-learn` library. It can systematically traverse multiple parameter combinations and determine the best-performing parameters by setting cross-validation parameters. In this paper, the cross-validation parameter is set to 10, which uses ten-fold cross-validation to determine the best effect parameters.

### C. Random forest algorithm prediction accuracy

The data set that predicts the direction of the central bank's monetary policy is input into the Random Forest Classifier, and finally the confusion matrix as shown in Table 5 can be obtained. The accuracy, precision and recall rate of the random forest algorithm can be calculated by the confusion matrix. 42 of the 138 data from January 2007 to June 2018 are test sets, in Table 5.  $Y = 1$  means the central bank's monetary policy is expanding,  $Y = 0$  Said that the central bank's monetary policy remains unchanged.  $Y = -1$  It indicates that the central bank's monetary policy is tightening.

TABLE 3 CONFUSION MATRIX PREDICTING THE DIRECTION OF CENTRAL BANK MONETARY POLICY

	Prediction $Y = 1$	Prediction $Y = 0$	Prediction $Y = -1$	Recovery rate
Actual $Y = 1$	1	6	0	0.1429
Actual $Y = 0$	0	29	1	0.9667
Actual $Y = -1$	0	2	3	0.6
Precision rate	1	0.7838	0.75	0.7857

From the confusion matrix in Table 3, the accuracy rate of using the random forest algorithm to predict the direction of the central bank's monetary policy is 78.57%, and the

forecasting precision of each category has reached 75% and above.

### D. Comparison with other machine learning algorithms for predicting accuracy

In order to evaluate the prediction accuracy of the random forest algorithm more objectively, this paper also inputs the data set of the central bank's monetary policy trend into other machine learning algorithms (cart decision tree, support vector machine, neural network) for training, and obtains different machine learning. The prediction accuracy of the algorithm is shown in Table 4.

TABLE 4 PREDICTION ACCURACY OF DIFFERENT MACHINE LEARNING ALGORITHMS

algorithm	Predictive accuracy
Random forest algorithm	78.57%
Cart decision tree	71.43%
Support Vector Machines	66.67%
Neural Networks	71.88%

It can be concluded from Table 4 that the prediction accuracy of random forest algorithm is higher than that of cart decision tree, support vector machine and neural network, which indicates that the random forest algorithm used in this paper has better prediction accuracy.

### E. Comparison with other research results prediction accuracy

At present, the methods used to predict the central bank's monetary policy are generally adopted: non-stationary discrete selection model and combined prediction model. The application of random forest classification algorithm to predict the monetary policy trend of China's central bank is an innovation of prediction methods. Table 5 shows that Li Haitao uses the non-stationary discrete selection model to predict the direction of China's central bank monetary policy, Hu and Philps use the orderly Probit model to predict the Fed target fund interest rate decision, and Andrey Vasnev [9](2013) uses the combined forecasting model to predict the Australian Reserve Bank's currency. Policy operations and Pauwels L [11](2012) use a combined forecasting method to predict the forecast accuracy of the Fed's monetary policy operations.

TABLE 5: PREDICTIVE ACCURACY OF DISCRETE SELECTION MODEL AND COMBINED PREDICTION MODEL

Author	Forecasting subject	method of prediction	Predictive accuracy
Li Haitao	Central bank of china	Discrete selection model	46.45%
Hu and Philps	Federal Reserve	Discrete selection model	67%
Andrey Vasnev	Reserve Bank of Australia	Combined forecasting model	67.7%
Pauwels L	Federal Reserve	Combined forecasting model	71.9%

Note: The prediction accuracy in Table 5 corresponds to the prediction accuracy of the model with the highest accuracy of out-of-sample prediction in the literature.

It can be seen from Table 7 that the prediction accuracy of the random forest classification algorithm used in this paper (78.57%) is higher than that of the discrete selection model and the combined prediction model. It is verified once again that the random forest algorithm has higher prediction accuracy.

## VI CONCLUSION

This paper proposes four kinds of machine learning-based forecasting methods for the direction of the central bank's monetary policy. Sixteen indicators of macroeconomic entities, financial institutions, interest rate markets, real estate market, external financial market and stock market are selected to construct this paper. The index system collected data from January 2007 to June 2018. The empirical results show that the random forest classification algorithm is used to predict the monetary policy trend of China's central bank, and the accuracy rate is 78.57%. The prediction accuracy is not only higher than that. The three machine learning algorithms, cart decision tree, support vector machine and neural network, are also higher than the out-of-sample prediction accuracy of discrete selection models and combined prediction models used in other papers.

This shows that the random forest classification algorithm performs better in predicting the direction of the central bank's monetary policy. As the effectiveness of the predictions is further enhanced, the significance of the predictions not only stabilizes the emotions and behaviors of market participants, but also helps them form good expectations. It can also judge the demand for funds from the macroeconomic perspective and further the central bank's currency. Provide reliable advice on the formulation and use of policies.

The central bank's monetary policy decisions are often based on a wide range of statistical indicator systems. Therefore, the empirical results of this paper show that using representative macroeconomic indicators, the random forest classification algorithm can better predict the monetary policy trend of China's central bank.

## REFERENCES

- [1] Xu Wenbing. Liquidity overflows, austerity policy will continue – macroeconomic and monetary policy review and trend analysis [J]. *New finance*, 2006, (12):17-20. (in Chinese)
- [2] Rouhi Rad M, Manning D, Suter J F, et al. Spatial Spillovers from Conservation Policies in Common Property Resources[J]. 2019.
- [3] Dueker M. Measuring monetary policy inertia in target fed funds rate changes[J]. *Federal Reserve Bank of St Louis Review*, 1990, (81): 3–10
- [4] Hu L, Phillips PCB. Nonstationary discrete choice[J]. *Journal of Econometrics*, 2004a, (120): 103–138
- [5] Hu L, Phillips PCB. Dynamics of the federal funds target rate: a nonstationary discrete choice approach[J]. *Journal of Applied Econometrics*, 2004b, (19): 851–867
- [6] Kim H, Jackson J, Saba R. Forecasting the FOMC's interest rate setting behaviour: a further analysis[J]. *Journal of Forecasting*, 2009, (28): 145–165
- [7] Li Haitao, Fang Zhaoben, Wei Yu. The Analysis and Forecasting on Monetary Policy Stance of China-Based on the Comparison between Estimations of Nonstationary Discrete Model and SVM [J]. *Journal of Applied Statistics and Management*, 2013,32(06):1079-1089. (in Chinese)
- [8] Pauwels L, Vasnev AL. Forecast combination for discrete choice models: predicting FOMC monetary policy decisions[D]. University of Sydney, 2011
- [9] Ruthira Naraidoo, Ivan Paya. Forecasting monetary policy rules in South Africa[J]. *International Journal of Forecasting*, 2012, (28): 446–455
- [10] Andrey Vasnev, Margaret Skirtun, Laurent Pauwels. Forecasting Monetary Policy Decisions in Australia: A Forecast Combinations Approach[J]. *Journal of Forecasting*, J. Forecast, 2013, (32) :151–166
- [11] Pauwels L , Vasnev ALL . Forecast Combination for Discrete Choice Models: Predicting FOMC Monetary Policy Decisions[J]. *SSRN Electronic Journal*, 2012
- [12] Ren Cairong, Xie Gang. Prediction of PM2.5 Concentration Level Based on Random Forest and Meteorological Parameters [J]. *Computer Engineering and Applications*, 2019,55(02):213-220. (in Chinese)
- [13] Ren Pengzhou, Yue Cairong. Masking of Cloud and Cloud Shadow for Hyperion Image Based on Random Forest Algorithm [J]. *Forest Inventory and Planning*, 2018, (03):16-21. (in Chinese)
- [14] Qiao Muzi. Research of Dynamic Effect about Monetary Policy on China's Systemic Financial Risk [D]. Master's thesis of Jilin University, 2018. (in Chinese)
- [15] Shang Yuhuang, Zheng Guoting. The Mixed Frequency Measuring of Chinese FCI and Its Risk Warning Behavior [J]. *Financial Research* , 2018,(03):21-35. (in Chinese)