Artificial Intelligence and Human-Computer Interaction
Y. Ye and H. Zhou (Eds.)

doi:10.3233/FAIA250108

Prediction of RMB Offshore Exchange Rate Driven by Investor Sentiment: Innovative Application of Deep Learning Model

Nan WANG[a], Bochuan ZHOU[b] and Po NING[b,1]
[a] *Changchun Sci-TechUniyersity*
[b] *Shanghai Normal University*

Abstract. The offshore Renminbi (CNH) exchange rate against the United States dollar (USD) better reflects the immediate changes in market supply and demand and investor sentiment due to its exemption from foreign exchange control in mainland China. This paper explores how to improve the forecasting accuracy of the offshore RMBUSD exchange rate by constructing an investor sentiment index based on online forums and news comments. This paper firstly collects and analyzes many financial news headlines on the English for Treasury website, and applies the BERT model in natural language processing technology to identify and quantify the sentiment tendencies in the news headlines, to construct a daily investor sentiment index. Subsequently, this sentiment index is combined with traditional financial market and macroeconomic indicators, and a variety of advanced machine learning and deep learning methods, including Random Forest, Support Vector Machines, Long Short-Term Memory Networks (LSTM), and Gated Recurrent Units (GRUs), are applied to forecast the offshore RMB exchange rate. It is found that the introduction of sentiment indices significantly improves the accuracy of the prediction models. Especially in LSTM and GRU models, the inclusion of sentiment index makes the models perform better in capturing the nonlinear features of exchange rate fluctuations.

Keywords. Investor sentiment index, offshore RMB exchange rate, machine learning model, deep learning model, Bert Model

1. Introduction

Against the backdrop of the current increasing integration of the global economy, the volatility of the offshore RMB-dollar exchange rate has attracted a great deal of attention from the global market. In particular, under the influence of the Federal Reserve's monetary policy adjustments, intensifying geopolitical conflicts, and uncertainties in the global trade environment, the RMB is facing increasing appreciation pressure. In this environment, accurate exchange rate forecasting is of great significance for multinational enterprises to carry out risk management and optimize international capital layout, as well as for governments and financial institutions to formulate relevant macroeconomic policies. Therefore, this study explores how investor sentiment affects the RMB exchange rate by integrating traditional financial market and macroeconomic data and tries to improve the accuracy of offshore RMB exchange rate forecasts by introducing

[1] Corresponding Author: Po NING, 1000528341@smail.shnu.edu.cn.

an investor sentiment index.

With the rapid development of data science and computational technology, research on exchange rate forecasting has gradually evolved towards integrating complex economic environments, machine learning techniques, and sentiment analysis. Recent studies have shown that deep learning models, especially Long Short-Term Memory Networks (LSTMs) and Transformer models (Transformers), have demonstrated the ability to effectively capture long-term dependencies and nonlinear patterns when dealing with financial time series data such as exchange rates. For example, Zhao and Yan [1] verified that the model can significantly improve the accuracy of forecasting after integrating complex economic indicators and market sentiment indices by using a transformer-based model to forecast multiple currency pairs. These studies not only extend traditional forecasting models but also provide a more comprehensive and accurate methodology for exchange rate forecasting by introducing machine learning and sentiment analysis.

On the other hand, the introduction of sentiment analysis, especially text data analysis based on news and social media, provides a new perspective for exchange rate forecasting. A study by Ding and Liu [2] shows that when there is a high level of uncertainty about monetary policy, market disagreement about future policy judgments intensifies, which in turn affects the RMB exchange rate's response to macroeconomic news. This response of the exchange rate to macroeconomic news is significantly weaker when uncertainty rises, showing that the impact of macroeconomic news on the exchange rate in an environment of monetary policy uncertainty is limited. A study by Sun [3] explored an exchange rate prediction model based on sentiment mining of online foreign exchange news, pointing out that analyzing news sentiment using machine learning methods can effectively predict exchange rate movements in the short term. This approach, especially after adding sentiment analysis indicators in the prediction model, can improve the accuracy and response speed of prediction, indicating that news sentiment is an important factor in exchange rate prediction. A study by Guo et al. [4] used the ChatGPT technique to identify sentiment tendencies in trade news and investigate how these sentiments affect short-term RMB exchange rate fluctuations. Their study finds that negative news, in particular, significantly affects the short-term fluctuations of the RMB exchange rate more than positive news.

In addition, the fusion of data from different sources, such as the combination of economic indicators with market sentiment indicators, is shown to enhance the predictive performance of the model. By analyzing data across markets and sources, Cao [5] proposes an integrated modeling framework that includes not only traditional economic and financial variables but also sentiment data collected from multiple perspectives to forecast the RMB exchange rate.

These studies show that the research trend in the field of exchange rate forecasting has shifted from relying on traditional macroeconomic models to adopting machine learning techniques, big data analytics, and cross-disciplinary data integration approaches. These advanced methods have significantly improved the ability of models to capture the nonlinear dynamics of the market by mining complex patterns and implicit correlations in large amounts of data, thereby enhancing forecasting accuracy and model robustness. As technology evolves, future research is expected to further explore more sophisticated computational models and richer data sources to continue to improve the performance of exchange rate forecasting and to meet the demand for high-precision forecasts in the global financial markets.

A central innovation of this study is the introduction of an investor sentiment index as a new variable for exchange rate forecasting. Unlike traditional methods, this study constructs a sentiment index by analyzing public comments on online forums and social media and applies natural language processing techniques to analyze textual data for sentiment tendencies. This approach effectively incorporates real-time changes in market sentiment into the exchange rate forecasting model, providing a new dynamic input variable to the model. The results of the study show that the model incorporating the sentiment index outperforms the traditional model under multiple time windows, especially in the presence of high market volatility. In addition, this innovation not only increases the sensitivity of the model to changes in market sentiment but also improves the timeliness and reliability of the forecasting results, enabling the model to more accurately reflect the latest market trends and possible turning points.

By integrating machine learning techniques and investor sentiment indices, this study provides a new methodological perspective in the field of exchange rate forecasting, demonstrates how technological innovations can be used to address the challenges faced by traditional models in modern financial markets, and also provides a rich experimental and theoretical foundation for future research.

2. Research design

First, based on user comments in online forums, we construct an investor sentiment index. Then, this sentiment index is used as one of the features and input into the exchange rate prediction model to predict the offshore exchange rate of RMB against USD using machine learning and deep learning techniques. Finally, the utility and performance of the sentiment index in exchange rate prediction are evaluated by comparing the prediction results of different models.

2.1 Data indicator description

The core data of this paper covers structured data such as the daily closing price of the offshore RMB exchange rate, financial market data and macroeconomic variables, and an investor sentiment index constructed based on the headlines of Investing.com. The timeframe is from April 2021 to August 2024, with exchange rate and market data sourced from Investing.com. The exchange rate and market data are sourced from the Investing.com website, while the macroeconomic data are sourced from iFinD, covering key economic indicators, as shown in Table 1. The investor sentiment index, on the other hand, is constructed by applying the BERT model to sentiment analysis of collected news headlines, which is calculated by scoring each headline (positive scores indicate positive sentiment and negative scores indicate negative sentiment). After obtaining the sentiment value of each financial news item for each trading day, a daily SI is constructed to quantify the sentiment contained in the daily news text, following Gu et al. [6], with the following equation:

$$SI = \ln\left(\frac{1 + News^{pos}}{1 + News^{neg}}\right) \tag{1}$$

where $News^{pos}$ and $News^{neg}$ represent the number of positive and negative news on a trading day, respectively. The combination of these data provides a multi-dimensional perspective for this study to analyze and forecast the volatility of the offshore RMB exchange rate, and to explore how macroeconomics, financial market dynamics, and market sentiment jointly affect exchange rate movements.

This study aims to analyze in-depth the volatility of the RMB offshore exchange rate by synthesizing these data and exploring the relationship between macroeconomic variables and the exchange rate in order to improve the accuracy and reliability of the forecasting model. This not only helps to understand the immediate impact of global economic and financial market dynamics on the RMB exchange rate but also provides deeper insights into the interaction between macroeconomic factors and the exchange rate.

Table 1. Variable description.

Class	Variable	Abbreviation
Financial Market Indexes	Offshore Chinese Yuan to US Dollar Opening Rate	USD_CNH_OPEN
	S&P 500 Opening Rate	US500_OPEN
	Dow Jones Industrial Average Opening Rate	US30_OPEN
	West Texas Intermediate Crude Opening Price	WTI_OPEN
	Shanghai Composite Index Opening Rate	SH_OPEN
	FTSE 100 Opening Rate	B25F_OPEN
	Natural Gas Opening Price	Natural_Gas_open
	FTSE China A50 Index Opening Rate	A50_Open
	Nikkei 225 Opening Rate	Nikkei_225_OPEN
	Silver Opening Price	Silver_OPEN
	US Dollar Index Opening Rate	Dollar_Index_OPEN
	Soybeans Opening Price	Soybeans_OPEN
	Copper Opening Price	Copper_OPEN
	Gold Opening Price	Gold_OPEN
	Hang Seng Index Opening Rate	Hang_Seng_OPEN
Macroeconomic Indicators	Euro to US Dollar Opening Rate	EUR_USD_OPEN
	Onshore Chinese Yuan to US Dollar Opening Rate	USD_CNY_OPEN
	Federal Funds Rate	FFR_RATE
	1-Month US Treasury Yield	US_Treasury_Yield_1M
	1-Year US Treasury Yield	US_Treasury_Yield_1Y
	1-Month US Dollar LIBOR Rate	LIBOR_USD_1M
	Overnight Shibor Rate	Shibor_Overnight

2.2 Normalization and rolling prediction

Since part of the structured data is non-daily data, this paper needs to expand this data into daily data. In this paper, we adopt the chained-equation multiple interpolation methods (nice forest) based on random forest in Shah et al. [7] and Sun et al. [8], which utilizes the chained-equation multiple estimation method (MICE) and RF to perform fast and efficient in-memory estimation for better handling of missing data situations. In addition, to avoid the problem of non-convergence or slow convergence of the model due to inconsistent feature scales, this paper adopts the min-max normalization module to normalize the soybean futures price series and scales the price data to the interval of [0,1] to eliminate the effect between different orders of magnitude. The formula is as follows.

$$x' = \frac{x - min(x)}{max(x) - min(x)}. \tag{2}$$

In this paper, the model is trained using an out-of-sample predictive power test with a rolling time window, i.e., keeping the training interval unchanged, and a continuous rolling forecast of the stock index price for the next day is made. A sliding window is set so that if the time window is s days, the data from day 1 to day s is used to predict day s+1; and the data from day 2 to day s+1 is used to predict day s+2. The two-dimensional input indices are converted into three-dimensional data in the format of (data length, time window, and number of features), which is used for the rolling prediction. To compare the effect of time window size on the prediction effect of the LSTM model, this paper chooses the time window lengths of 5, 10, 15, 20, 25, and 30 days to construct the training data, and the rolling prediction method is used for model training and evaluation.

2.3 Evaluation criteria

After the model gets the predicted value, it is compared with the actual value to get the magnitude of the error. In this paper, the root mean square error (MSE) is used as the loss function. To better measure the prediction effect, three indicators, MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), and SMAPE (Symmetric Mean Absolute Percentage Error) are added to evaluate the prediction results, and the smaller the value of all the evaluation indexes, the better the prediction effect. The equations for the four loss functions are as follows.:

$$MAPE = \frac{1}{N}\sum_{n=1}^{N}\frac{|\hat{y}_n - y_n|}{y_n}, \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (\hat{y}_n - y_n)^2}, \tag{4}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{n=1}^{N} |\hat{y}_n - y_n| \tag{5}$$

$$SMAPE = \frac{1}{N}\sum \frac{|\hat{y}_n - y_n|}{(|\hat{y}_n| + |y_n|)/2} \tag{6}$$

where y_n denotes the true value, $\hat{y}_n$ denotes the predicted value.

3. Forecasting results and analysis

In this paper, we use different lengths of historical data to predict the opening price of RMB offshore exchange rate by SVR, XGBoost, LSTM, and GRU models, and Table 2 and Figure 1 show their prediction accuracies as well as the trend of prediction accuracies with the increase of the number of days of historical data, respectively. In the comparison experiments, we find that the LSTM and GRU models are better compared to the SVR and XGBoost models. In addition, as the length of historical data increases, the SVR and XGBoost models become progressively worse, while the LSTM and GRU models are more stable, and the GRU model is slightly better than the LSTM model in general. The reason is that LSTM and GRU models can effectively capture long-term dependencies and complex nonlinear features when dealing with time series data, and thus the prediction effect is more stable when facing long historical data. In contrast, SVR and XGBoost perform moderately well in the short term, but their effects gradually weaken with the increase of the length of historical data due to the lack of the ability to model long-term dependencies. GRU, due to its simplified structure, high computational efficiency, and the ability to retain key information, slightly outperforms the LSTM in this experiment, and is especially suitable for exchange rate prediction of long time series.

Table 3 shows the comparison of the prediction accuracy of SVR, XGBoost, LSTM, and GRU models when predicting the opening price of the RMB offshore exchange rate without considering the media text sentiment factor. It can be seen that the addition of the media text sentiment factor brings some improvement to the prediction effect of all models, and the improvement rate of the deep learning model is higher than that of machine learning. In addition, for different models, as the time window increases from 5 to 20, the ability of the media text sentiment factor to improve the prediction effect of the SVR model gradually decreases, while the ability to improve the prediction effect of the XGBoost, LSTM, and GRU models gradually increases.

The reasons for this are, firstly, the SVR (Support Vector Regression) model is better able to capture short-term fluctuations in the data under a smaller time window, which makes the addition of the media text sentiment factor a more significant effect enhancement to the SVR model when the time window is shorter. However, as the time window expands, the ability of the SVR model to fit the data gradually stabilizes, and the marginal utility of the additional information provided by the media text sentiment factor decreases, so its ability to improve the effect gradually decreases. Second, XGBoost, LSTM, and GRU models are better able to capture long-term dependencies and complex patterns in time series data under larger time windows. The media text sentiment factor, as a kind of external information, can effectively enhance the ability of these models to perceive changes in market sentiment, thus further improving the predictive effectiveness of the models when the time window increases. Specifically, XGBoost models have advantages in handling nonlinear relationships and feature interactions, while LSTM and GRU models, as deep learning models, are good at capturing long- and short-term dependencies and complex temporal dynamics in time series. With the expansion of the time window, these models can more fully utilize the information provided by the media text sentiment factor, which gradually enhances their ability to predict the opening price of the offshore RMB exchange rate. Finally, there are lagged and cumulative effects of the media text sentiment factor, which can be better reflected under longer time windows. Under a long-time window, changes in market sentiment can gradually affect exchange rate fluctuations through media reports, and the model can therefore better capture the influence of the sentiment factor, thus improving the forecast accuracy of the opening price of the offshore RMB exchange rate.

Table 2. Side-by-side comparison of predictions by model for different historical data lengths.

Model	History length=5				History length=10				History length=15				History length=20			
	RMSE	MAE	MAPE	SMAPE	RMSE	MAE	MAPE	SMAPE	RMSE	MAE	MAPE	SMAPE	RMSE	MAE	MAPE	SMAPE
SVR	0.0633	0.0493	0.0068	0.0068	0.2103	0.1875	0.0259	0.0263	0.2874	0.2645	0.0365	0.0373	0.3242	0.2988	0.0413	0.0423
XGBoost	0.1005	0.0903	0.0125	0.0125	0.1047	0.0966	0.0133	0.0134	0.0991	0.0916	0.0127	0.0127	0.1076	0.0992	0.0137	0.0138
LSTM	0.0442	0.0357	0.0049	0.0049	0.0663	0.0549	0.0076	0.0076	0.0553	0.0436	0.0060	0.0060	0.0722	0.0592	0.0082	0.0082
GRU	0.0581	0.0508	0.0070	0.0070	0.0266	0.0210	0.0029	0.0029	0.0877	0.0810	0.0112	0.0111	0.0355	0.0304	0.0042	0.0042

Note: Bolded and underlined fonts are the optimal evaluation result values.

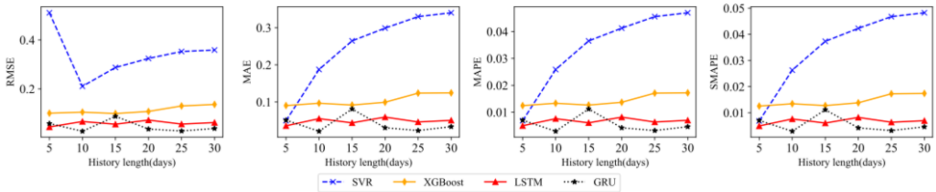

Figure 1. Side-by-side comparison of predictions by model for different historical data lengths.

Table 3. Enhancement of exchange rate news sentiment indicators on the effectiveness of exchange rate forecasting across models.

Model	Indicator	History length=5				History length=10				History length=15				History length=20			
		RMSE	MAE	MAPE	SMAPE	RMSE	MAE	MAPE	SMAPE	RMSE	MAE	MAPE	SMAPE	RMSE	MAE	MAPE	SMAPE
SVR	Include SI	0.0633	0.0493	0.0068	0.0068	0.2103	0.1875	0.0259	0.0263	0.2874	0.2645	0.0365	0.0373	0.3242	0.2988	0.0413	0.0423
	Excluding SI	0.1200	0.1032	0.0142	0.0144	0.2598	0.2358	0.0326	0.0332	0.2940	0.2713	0.0375	0.0383	0.3302	0.3079	0.0425	0.0436
	Improvement	47.20%	52.24%	52.11%	52.41%	19.06%	20.52%	20.54%	20.83%	2.25%	2.51%	2.51%	2.55%	1.82%	2.96%	2.97%	2.99%
XGBoost	Include SI	0.1005	0.0903	0.0125	0.0125	0.1047	0.0966	0.0133	0.0134	0.0991	0.0916	0.0127	0.0127	0.1076	0.0992	0.0137	0.0138
	Excluding SI	0.1072	0.0955	0.0132	0.0133	0.1165	0.1055	0.0146	0.0147	0.1070	0.0992	0.0137	0.0138	0.1242	0.1142	0.0158	0.0159
	Improvement	6.30%	5.49%	5.48%	5.54%	10.14%	8.49%	8.46%	8.56%	7.45%	7.61%	7.60%	7.66%	13.35%	13.18%	13.18%	13.29%
LSTM	Include SI	0.0442	0.0357	0.0049	0.0049	0.0663	0.0549	0.0076	0.0076	0.0553	0.0436	0.0060	0.0060	0.0722	0.0592	0.0082	0.0082
	Excluding SI	0.0481	0.0433	0.0060	0.0060	0.0736	0.0642	0.0089	0.0089	0.0894	0.0821	0.0113	0.0114	0.1111	0.1034	0.0143	0.0144
	Improvement	8.17%	17.59%	17.40%	17.57%	9.88%	14.44%	14.48%	14.43%	38.07%	46.93%	46.69%	47.14%	34.98%	42.74%	42.81%	42.95%
GRU	Include SI	0.0581	0.0508	0.0070	0.0070	0.0266	0.0210	0.0029	0.0029	0.0877	0.0810	0.0112	0.0111	0.0355	0.0304	0.0042	0.0042
	Excluding SI	0.0745	0.0655	0.0090	0.0091	0.0309	0.0246	0.0034	0.0034	0.1339	0.1226	0.0169	0.0171	0.0483	0.0424	0.0058	0.0059
	Improvement	22.00%	22.40%	22.38%	22.48%	13.96%	14.65%	14.73%	14.59%	34.50%	33.97%	33.85%	34.94%	26.44%	28.36%	28.30%	28.39%

Note: Bolded and underlined fonts are the optimal evaluation result values.

4. Conclusions and extensions

In this paper, the impact of sentiment factors on exchange rate fluctuations is explored by incorporating an investor sentiment index constructed based on online forum comments into a forecasting model of the offshore RMB exchange rate. Through the application and comparison of a variety of machine learning and deep learning models, the results show that the incorporation of the sentiment index significantly improves the forecasting accuracy of the model, especially in deep learning models (e.g., LSTM and GRU). In addition, this paper adopts the rolling window method to further enhance the effectiveness of time series forecasting and verifies the forward-looking guiding effect of sentiment factors on the offshore RMB exchange rate. The overall study shows that incorporating investor sentiment into exchange rate forecasting not only helps to improve enterprises' ability to prevent exchange rate risks but also provides valuable market signals for policymakers, thus enhancing the forward-looking guidance of exchange rate movements.

This paper suggests that future research can further optimize the construction of the sentiment index by introducing multi-dimensional data sources such as news media and social platforms to enhance the comprehensiveness of the sentiment index. In addition, increasing the analysis of macroeconomic variables such as inflation rate and international trade will also help to improve the accuracy of exchange rate forecasting. Meanwhile, the applicability and scalability of the model should be further explored by trying to combine the sentiment index with other methods such as econometrics or Bayesian modeling to verify its performance in different markets. Finally, in the future, the model can be considered to be applied to real-time forecasting and risk early warning systems to provide dynamic early warnings of exchange rate fluctuations for enterprises and investors, to help them avoid market risks.

References

[1] Zhao L, Yan Q. Prediction of Currency Exchange Rate Based on Transformers. Journal of Risk and Financial Management. 2024; 17(8):332.

[2] Ding J., Liu L. The RMB Exchange Rate Effect of Monetary Policy Uncertainty and Macroeconomic News in China. Finance and Trade Economics. 2020; 41(05):19-34.

[3] Sun S., Wei Y., Li J. Exchange Rate Forecasting with Online Forex News Sentiment Mining. China Journal of Econometrics. 2022; 2(02):441-464.

[4] Guo W., Han D., Wang B.. Trade News Identified Using the ChatGPT-based Emotion Perception Strategy and Its Short -term Volatility Effect on Renminbi Exchange Rates. Financial Economics Research. 2023; 38(05):160-174.

[5] Cao W., Liao C., Zhang F.. RMB Exchange Rate Forecasting Driven by Cross-Market and Cross-Source Sentiment Analysis. Data Analysis and Knowledge Discovery. 2023; 7(12):75-87.

[6] Gu W, Zhang L, Xi H, Zheng S.. Department of Statistics. Stock Prediction Based on News Text Analysis. Journal of Advanced Computational Intelligence and Intelligent Informatics. 2021; 25(5): 581-591.

[7] Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H.. Comparison of Random Forest and Parametric Imputation Models for Imputing Missing Data Using MICE: A CALIBER Study. American Journal of Epidemiology. 2014; 179(6), 764-774.

[8] Sun, Y., Li, J., Xu, Y., Zhang, T., & Wang, X.. Deep learning versus conventional methods for missing data imputation: A review and comparative study. Expert Systems with Applications. 2023; 227, 120201.