IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 March 2024, accepted 17 March 2024, date of publication 25 March 2024, date of current version 2 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3381500

RESEARCH ARTICLE

Long-Term Interbank Bond Rate Prediction Based on ICEEMDAN and Machine Learning

YUE YU, GUANGWU KUANG, JIANRUI ZHU, LEI SHEN, AND MENGJIA WANG
School of Economics, Wuhan University of Technology, Wuhan 430070, China
Corresponding author: Lei Shen (shenlei@whut.edu.cn)

ABSTRACT The application of time series forecasting utilizing historical data has become increasingly essential across a variety of industries including finance, healthcare, meteorology, and industrial sectors. The assessment of bond transaction rates in the interbank bond market serves as a crucial indicator for assessing bank risk. In this paper, we proposed a composite model to forecast the transaction interest rates of China's interbank bonds over a long period. Specifically, our model integrates an intrinsic complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) model along with various long-term prediction models including long short-term memory network, temporal convolutional network, transformer, and autoformer. Our findings reveal that: 1) predictive performance of different long-term prediction models varies across different frequencies of single time series data; 2) predictive efficacy of diverse model combinations differs across varying prediction time lengths; 3) best results can be realized by using different prediction model combinations for high-frequency, medium-frequency and low-frequency data under different time steps.

INDEX TERMS Long-term forecasting, ICEEMDAN, machine learning, interbank bond rate.

I. INTRODUCTION

Long-term time series forecasting plays a pivotal role in decision-making within the financial sector. It aids investors in formulating more rational expectations about future economic trends and enables regulatory authorities to extract potential risk factors from historical data for early warning purposes.

Initial time series forecasting models predominantly relied on linear predictive models such as autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving-average (SARIMA), generalized autoregressive conditional heteroskedasticity (GARCH), exponential generalized autoregressive conditional heteroskedasticity (EGARCH), and vector autoregression (VAR). However, these linear models often suffer from poor noise reduction capabilities and the inability to extract nonlinear signals, limiting their effectiveness in extracting comprehensive information from complex financial time series data, thereby hindering long-term and high-accuracy predictions [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Chun-Hao Chen.

Recent studies have commonly focused on nonlinear predictive models based on time series [2]. Uppala et al. [3] compared ARIMA, SARIMA, and long short-term memory (LSTM) networks in forecasting sales profits, finding that the LSTM model shows superior performance. Yang et al. [4] leveraged the LSTM model for global stock index predictions, finding that it preforms better accuracy than support vector regression (SVR), mobile location protocol (MLP), and ARIMA models in terms of precision and stability. Scholars have enhanced traditional LSTM models using techniques like whale algorithm, particle swarm optimization (PSO), least absolute shrinkage and selection operator (LASSO), and principal component analysis (PCA) for univariate forecasting of short-term financial data, improving various aspects of the traditional LSTM algorithm [1], [2], [5]. However, there still exist issues such as the lack of muti-factors and short forecasting steps. Bai et al. [6] proposed the temporal convolutional network (TCN), which effectively captures local dependencies in sequential data, making gradient propagation more stable compared to traditional convolutional neural networks (CNN) algorithms. Reference [7] proposed a short-term forecasting approach of multilayer perceptrons

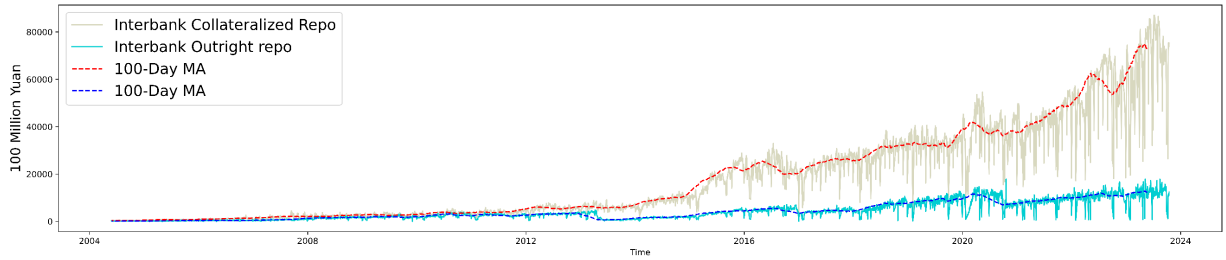


FIGURE 1. Scale of interbank bond trading.

FIGURE 2. Interbank AAA+ debenture transaction rates of China.

ensembles, combined dynamically with a long-term forecasting, achieving the continuous dynamical combination of forecasts. Reference [8] investigated LSTM and gated recurrent unit networks, demonstrating success in long-term load forecasting in hybrid distribution feeder systems.

With the development of the self-attention mechanism in the field of sequential data processing [9], deep learning frameworks based on transformer have demonstrated excellent performance in natural language processing [10], [11], audio processing [12], computer vision [13], [14], and other sequential data applications. Li et al. [15] introduced local convolution into the traditional Transformer model and proposed the LogSparse attention mechanism to select time steps, thereby reducing the time complexity to $O\left(L\left(\log L\right)^2\right)$, achieving the application of transformer-based models in time series prediction. Some literature [16], [17], and [18] applied the self-attention transfer mechanism to the traditional transformer model to reduce the computational time complexity of the model, while sacrificing information utilization. The autoformer model proposed in [19] replaced the point-to-point connection attention mechanism with the autocorrelation mechanism, achieving sequence-level connection and a time complexity of $O\left(L \log L\right)$. Sun et al. proposed the frequency decomposition transformer model (FD-Transformer), which improved the long-term prediction capability of the transformer model by introducing a frequency decomposition multi-head attention mechanism [20]. Specially, Yang et al. employed various machine learning methods and new generative pre-trained transformers for physical systems. Reference [21] proposed a new hybrid machine learning architecture named fast learning to understand and investigate dynamics with a generative pre-trained transformer, which shows excellent performance in predicting particle trajectories and erosion on an industrial-scale steam header geometry. Reference [22] harness CNN and LSTM machine learning methodologies to predict complex surface erosion profiles in steam distribution headers, achieving elevated precision and markedly accelerated computational efficiency compared to conventional models. Reference [23] present an efficient deep reinforcement learning approach to automatically construct time-dependent optimal control fields that enable desired transitions in dynamical chemical systems. OpenAI trained the large language model chatGPT based on the transformer framework. Reference [11] trained GPT-3 and tested its performance in a small sample setting, finding that GPT-3 achieved strong performance on many natural language processing datasets as well as some tasks that required immediate inference or domain adaptation.

As the research extends into the microstructure of financial markets and the psychology of trading behavior deepens, scholars increasingly recognize the inefficiency of single models in effectively mining and fitting the multidimensional rules of price and volume changes in complex financial markets. Hybrid or ensemble models, integrating methodologies from econometrics, signal processing, and machine learning, can identify different data patterns through their sub-modules, thereby aggregating and uncovering the complete change rules inherent in the data of high-precision financial time series forecasting. Colominas et al. [24] innovatively proposed the empirical mode decomposition (EMD) method. This approach hierarchically decomposes time series signals into intrinsic mode functions (IMFs) of different characteristic scales, theoretically enabling the decomposition of non-stationary and nonlinear time series

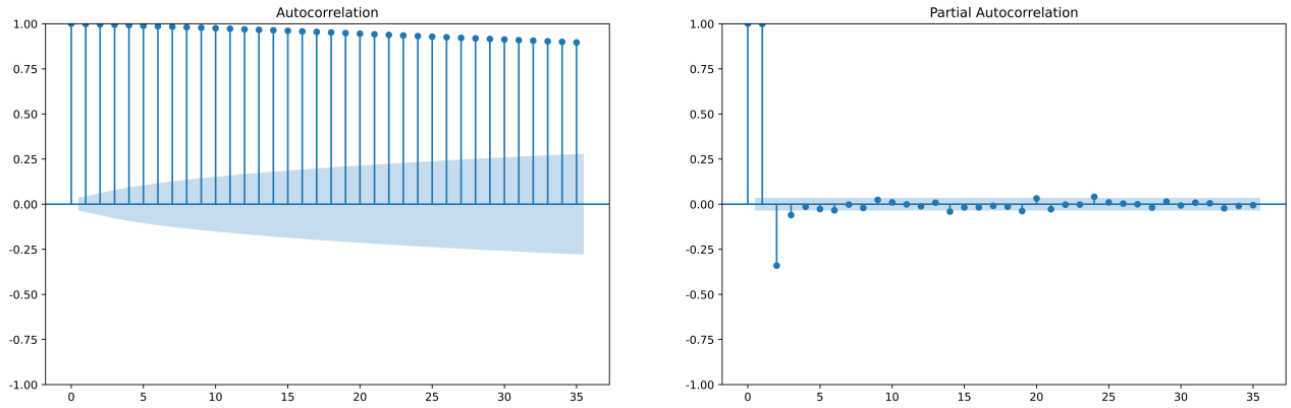


FIGURE 3. ACF and PACF of Interbank AAA+ debenture rates.

signals [25]. Addressing issues of incomplete decomposition and the emergence of false components in EMD, Wu et al. improved the EMD method by introducing uniformly distributed auxiliary noise, leading to the development of ensemble empirical mode decomposition (EEMD) which effectively resolves the problem of mode mixing. However, the Gaussian white noise added during the process is challenging to eliminate [26]. Torres et al. further enhanced EEMD by incorporating adaptive white noise, proposing the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) method that effectively overcomes the issues of incomplete decomposition and large reconstruction errors in EEMD, though it may still result in false components in the decomposed outcome. The intrinsic complete ensemble EMD with adaptive noise (ICEEMDAN) eliminates the problems of noise residue and false components in the derived intrinsic [27], [28]. The decomposed IMFs are relatively simple and independent, providing favorable conditions for fully extracting the volatility features of IMF sub-sequences, thus significantly reducing the difficulty of financial time series forecasting modeling [29]. EMD was mainly applied to signal denoising and meteorological science in the early days and has been introduced to economic and financial fields in recent years. Yang et al. improved the accuracy of exchange rate prediction by using the IMF input extreme learning machine obtained by EMD decomposition of exchange rate series [30]. Zhang et al. proposed to construct a mixed prediction model of EEMD and LSTM in the prediction of surface temperature, and the empirical results show that the prediction effect of this model is better than that of macshine learning prediction models such as Recurrent Neural Network (RNN), LSTM, and EMD-RNN. Reference [31] developed a two-stage hybrid model, including ICEEMDAN-recurrent deterministic policy gradient (RDPG) and variational mode decomposition-RDPG, to forecast the solar irradiance. Reference [32] proposed a hybrid influent forecasting model based on multimodal and ensemble-based deep learning, showing a good performance in predicting the long-term and short-term loads of circuits.

However, these studies use the same machine learning model to fit different frequency data, but the characteristics of different IMFs are significantly different, indicating that it may be better to use different models to fit them separately.

With the increasing marketization of bond trading in China and the expansion of the transaction scale, the fluctuation in interbank bond rates is increasingly influenced by external factors. Accurate interest rate forecasts, which are deeply considered by financial market participants, banks, firms, and policymakers, are essential for making informed investment decisions and formulating monetary policies. To ensure stability and growth in the financial market, scholars have made a series of predictions around bond interest rates. Avery et al. analyzed the sensitivity of the interest rate spread between bank related debt and comparable Treasury bills to bank risk measurement [33]. Lu et al. [34] found that when the zero bound is reached, the uncertainty of inflation and output gap forecasts increases, while the uncertainty of interest rate forecasts decreases. Özbekler, et al. [35] demonstrated that prior to the financial crisis, the information content of interest rate forecasts would affect market expectations of future short-term interest rates. Forecasting the long-term inter-bank bond transaction rate plays an important role in preventing financial risks and strengthening bank supervision. Most of the current research in the field of price forecasting takes the autoregressive approach to time series forecasting, using the time series of the price past to predict the future. The efficient market theory proposed by E. F. Foama pointed out that only in a perfectly efficient market can the price include all the information of itself. In reality, however, the price is often affected by factors such as securities index, market price index, exchange rate fluctuations, etc. Therefore, we take into account other influence factors as prediction indicators.

Previous research have made many useful attempts to forecast the bond market. Fan et al. [36] predicted the return volatility of major European government bond markets using

TABLE 1. Comparison of time series forecasting strategy.

Ref.	Mode decomposition	Multifactor prediction	Multi-step long time prediction	Hybrid model	Targeted combination of dynamic models
[3]			√		
[7]			√	√	√
[8]		√	√	√	
[15]		√	√		
[17]		√	√		
[21]		√		√	
[22]		√	√	√	
[31]	√	√		√	
[32]	√		√	√	
[47]	√			√	
This paper	√	√	√	√	√

the heterogeneous autoregressive (HAR) model. Cochrane and Piazzesi et al. [37] proposed a smoothed or projected principal component analysis (PCA) to forecast US bond risk premia. Huang et al. [38] found that a single tent-shaped linear combination of forward rates can be a good predictor of excess returns on one to five year bonds. Despite the achievements listed above, there is still room for improvement in the breadth and accuracy of their predictions.

In this paper, we use several factors such as SSE Index, Shibor rate, U.S. dollar index, gold price and so on to forecast interbank bond rates. A composite model is constructed based on ICEEMDAN with a variety of long-term series prediction models (LSTM, TCN, Transformer, Autoformer), which can adapt to different time steps. Based on different frequency and different forecasting steps of the time series, the long-term forecast of the bond transaction rate of Chinese banks is achieved. The main contributions of this paper are as follows.

1) A novel approach is introduced for employing different prediction models suited to various frequency components within different time steps in time series data. This strategy strategically harnesses the strengths of each prediction model, thereby enhancing overall forecasting accuracy. Simulation results substantiate the effectiveness of the proposed methodology, demonstrating superior performance compared to conventional time series prediction methods that rely on a single prediction model. The comprehensive use of diverse prediction models optimizes forecasting accuracy across different frequency domains within the time series data, presenting a notable advancement in predictive analytics methodologies.

2) Delves into the multi-factor prediction of interbank bond trading interest rates. Conventional single-factor time series forecasting encounters challenges in capturing intricate interdependencies among multiple factors, comprehensively assessing the collective influence of various factors on prediction outcomes, and forecasting long-term trends accurately. By employing multi-factor and prediction methodologies, these limitations can be mitigated, allowing for a more comprehensive consideration of diverse influencing factors. This approach enhances prediction accuracy and resilience by accommodating a broader spectrum of variables affecting prediction outcomes.

3) Introduce the ICEEMDAN model to decomposing interest rate data from interbank bond transactions. Conventional time series models encounter challenges such as excessive noise, inadequate stationarity, and diminished prediction accuracy in forecasting financial data. Through ICEEMDAN decomposition, we partition the original data into distinct intrinsic mode functions (IMFs), subsequently aggregating them into high, medium, and low-frequency components by evaluating the fuzzy entropy of each IMF. This approach effectively addresses the nonlinear and non-stationary characteristics of the data, thereby enhancing the performance of predictive models on the original dataset.

4)A long-term prediction of interest rates in the interbank bond market has been conducted. Compared to a large number of literature on short-term predictions of the financial market, this paper is able to achieve long-term prediction, which is relatively more practical. Moreover, the improvement of the prediction performance of the proposed model compared to other traditional models increases with the increase of steps. In addition, the interbank bond interest rate release reflects the internal risk of banks, and there is little literature predicting the interbank bond market. This article fills this gap by using ICEEMDAN and various machine learning algorithms to predict it.

The rest of the paper is structured as follows. A comprehensive description of the data is provided in Section II and

TABLE 2. ADF test for time series.

	X1	X2	X3	X4	X5	X6	X7	X8	XOT
Ori.seq.	−2.3975	−5.2667	−3.4122	−2.0896	−1.5732	−6.6574	−1.6489	−2.3850	−2.0422
p-value	(0.1424)	(0.0000)	(0.0106)	(0.2487)	(0.4971)	(0.0000)	(0.4576)	(0.1460)	(0.2684)
Diff1.seq.	−11.8168	−13.6147	−17.1067	−14.8967	−11.7653	−18.1040	−10.7299	−10.1358	−20.3501
p-value	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Section III describes the model. In Section IV, the framework of combination models to forecast the long-term data is proposed. Section V discusses the results. Finally, Section VI draws the conclusion.

II. DATA PROCESSING AND FACTORS SELECTING

A. TARGET DATA

In recent years, the share and influence of the interbank bond market in China has experienced significant growth. As shown in Fig. 1, the average daily trading volume of interbank bonds has notably increased from 8.852 billion yuan in late 2004 to 1322.521 billion yuan in September 2023. From the secondary market trading volume view, the average daily trading volume of interbank spot bonds has risen from 34.774 billion yuan in August 2004 to 7,385.375 billion yuan in 2023.

The interbank bond market has firmly established itself as the dominant player in China's bond market, and the transaction rate within this market has a considerable impact on the profitability and risk exposure of banks. This study focuses on examining the transaction rate of one-year AAA+ corporate bonds in China's interbank bond market (referred to as OT). As shown in Fig. 2, the time series data of 3111 trading days, spanning from December 15, 2009, to September 23, 2023, was obtained from iFinD. The dataset was divided into two subsets for analysis: 2811 sets of data from December 15, 2009, to June 17, 2022, were selected as the training set, and 300 sets of data from June 18, 2022, to September 23, 2023, were chosen as the test set.

This paper uses statsmodels in python3.11 to perform stationarity and autocorrelation tests on OT:

1. In this paper, we performed the augmented Dickey-Fuller (ADF) test on the original time series data of OT [36]. The *p*-value obtained from this test was 0.2684. The ADF test statistic after first-order differencing was –37.0732, which is lower than the critical value of –2.8624 at a 5% significance level, indicating that the null hypothesis of data non-stationarity can be rejected. Therefore, OT is considered to be stationary after first-order differencing.

2. Ljung-Box test is used to investigate autocorrelation within the OT time series [37]. The test was conducted with lag terms ranging from 1 to 10, and all resulting *p*-values were 0.00, as shown in Table 3. Therefore, the original hypothesis of a purely random sequence can be rejected, verifying that the time-series data OT is not a white-noise sequence but has strong autocorrelation.

TABLE 3. Ljung-box test.

Test	*p*-value	Hypothesis	conclusion
ADF	0.0	Reject	Stationarity
Ljung-Box Test	0.0	Reject	Autocorrelation
ACF	N/A	N/A	Autocorrelation
PACF	N/A	N/A	Autocorrelation

3. The autocorrelation function (ACF) and partial autocorrelation function (PACF) of the OT original time series were calculated, as depicted in Figure 3. The line height in the figure represents the autocorrelation coefficients of each order, while the blue area represents the 95% confidence interval for these coefficients. If the autocorrelation coefficient exceeds this confidence interval, it indicates statistical significance. The ACF plot reveals a significant tail, with all autocorrelation coefficients lying outside the blue range. Additionally, the PACF exhibits a truncated tail at the third order. These findings suggest the presence of strong autocorrelation and long-term dependence within the OT time series, implying that past historical data can provide valuable insights into future trends. The stationarity test results and autocorrelation test results of time series OT are summarized in Table 4:

TABLE 4. Statistical tests for OT.

Lag	lb_stat	*p*-value	lag	lb_stat	*p*-value
1	3105.8600	0.0	6	18436.4903	0.0
2	6200.4092	0.0	7	21456.2549	0.0
3	9281.7271	0.0	8	24459.4591	0.0
4	12348.6534	0.0	9	27446.1479	0.0
5	15400.4793	0.0	10	30416.5255	0.0

From the test results, it can be found that the first-order difference of the time series data of the transaction rate of one-year AAA+ corporate bonds in China's interbank bond market is stationary, and there is an obvious long-term dependence on the target data.

B. OTHER INFLUENCING FACTORS

The transaction rate of interbank corporate bonds is influenced by various factors. However, previous studies in the financial field mostly focused on forecasting the time series trend itself, which led to issues of incomplete information acquisition and inadequate credibility. In this

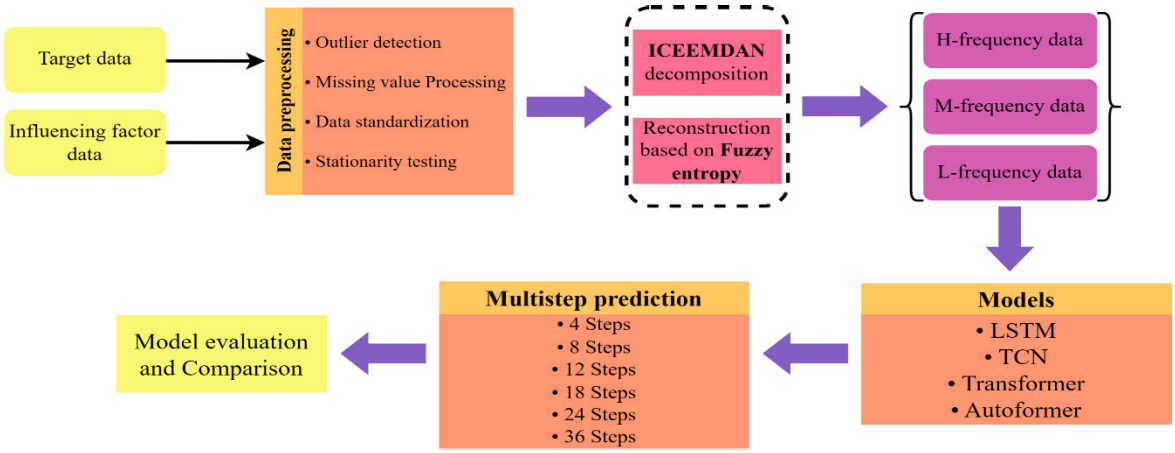


FIGURE 4. Flow chart of prediction model building.

paper, the Shanghai stock exchange (SSE) index, Shibor overnight rate, U.S. dollar index, gold price, commodity research bureau (CRB) index, fuel oil futures prices, treasury yield(10Year), interest rate swap(1Year), and other factors affecting the turnover interest rate of interbank corporate bonds are included in the input range, so that more information can be captured while studying the bond's closing rate.

1) SSE INDEXES

The SSE Indexes are vital indicators of stock price volatility in China as they reflect the price changes of stocks listed on the Shanghai Stock Exchange. The relationship between stock prices and the bond market is essential, as stock prices indicate changes in market confidence, economic cycles, and the market's perception of enterprise sustainability [39]. This, in turn, affects bond interest rate fluctuations in the interbank bond market. On the one hand, due to the stock and bond tilting effect, the rise of the SSE Index leads to the flow of funds to the stock market, and the attractiveness of corporate bonds decreases, increasing the transaction interest rate of corporate bonds. On the other hand, the higher the SSE index reflects the economic upswing, the stronger the overall profitability of the enterprise, the smaller the corresponding risk premium for corporate bonds, and the lower the transaction rates. Some scholars have found that there is a significant long-term correlation between the SSE Indexes and China's bonds, and there is a one-way Granger causality [40]. The relationship and the degree of influence of the SSE Indexs on China's bonds depend on the results of the above two power games.

2) SHIBOR OVERNIGHT RATE

The Shibor overnight rate directly affects the short-term financing costs and capital operation strategy of the bank. Changes in Shibor can reflect the tightness of capital supply and demand, which subsequently influences the bond trading market and the trading rate due to a certain degree of substitution relationship. Reference [41] found that the difference between the Shibor overnight rate and the interbank bond repo rate can be used as an effective indicator to measure the risk of the interbank money market.

3) US DOLLAR INDEX

The U.S. dollar index provides a comprehensive measure of the exchange rate of dollar against a basket of currencies in the international foreign exchange market. It affects Chinese bond yields by influencing inflation through the Fisher effect. According to [42], the U.S. dollar index can affect Chinese bond yields by influencing inflation and thus through the Fisher effect. In addition, some scholars found that exchange rate fluctuations caused by changes in the US dollar index would lead to changes in the investment desire of American investors in China, which would have a significant impact on the Chinese bond market [43].

4) GOLD PRICE

Gold prices serve as a haven asset during periods of market instability or uncertain economic outlooks. When the market is volatile, investors tend to purchase gold for value preservation. Reference [44] found that there is a significant negative correlation between the price of gold and bonds. In other words, gold was a hedge against bonds.

5) CRB INDEX

The CRB index is a measure of commodity price movements, which can reflect the public's inflation expectations to a certain extent, and thus affect bond prices. Reference [45] revealed a two-way spillover effect and interaction between the commodity market and the financial market.

6) FUEL OIL FUTURES PRICE

Fuel oil futures prices reflect macroeconomic risks, such as global supply chain conditions, geopolitical conflicts, and fluctuations in global demand. Reference [46] found a long-term equilibrium relationship between fuel oil futures

prices and macroeconomic growth, making it an indicator of a country's economic growth and development. Reference [47] studied the interaction between corporate bond yield and crude oil price, and found that there is a significant Granger causality between them.

7) TREASURY YIELD (10YEAR)

Treasury yield (10 years) represents the risk-free rate and serves as an important benchmark for evaluating bond yields. Reference [48] studied the risk transmission mechanism between treasury bonds and corporate bonds through the corporate bond spread and treasury yield VAR regression analysis and found that the market risk between treasury bonds and corporate bonds will be transmitted to each other, and there is a strong link between the two. In addition, [49] found that the yield to maturity of treasury bonds also contains complex macroeconomic information, reflecting the overall economic environment of business operations.

8) INTEREST RATE SWAP(1YEAR)

Interest rate swap(1 year) is a derivative trading tool in the Chinese interbank bond market. It reflects the credit risk and liquidity of the financial market. Some scholars studied the correlation between China's inter-bank bond market and interest rate swap market and found that there were significant two-way price guidance and long-term and short-term volatility spillover effects between the two markets. The negative correlation between these markets tends to gradually increase [50].

TABLE 5. Data source.

Variable	Impact Factor	Data Source
X1	SSE Index	iFind
X2	Shibor Overnight Rate	CFETS
X3	US Dollar Index	iFind
X4	Gold Price	World Gold Council
X5	CRB Index	Commodity Research Bureau
X6	Fuel oil futures prices	Shanghai Futures Exchange
X7	Treasury Yield (10Y)	ChinaBond
X8	Interest Rate Swap(1Y)	CFETS

To ensure the stationarity of the data, the augmented dickey-fuller (ADF) test was conducted to assess the stationarity of the time series of other influencing factors, and the results are shown in Table 2. The p-value in the table is a parameter used to determine the result of the hypothesis test. Generally speaking, the smaller the p-value, the stronger the rejection of the null hypothesis. The original time-series data for the Shibor overnight rate, the U.S. dollar index, and the fuel oil futures price were found to be smooth, the SSE index, the price of gold, the CBD Index, the treasury yield (10Year), and the interest rate swap (1Year) are integrated of order one I(1).

Johansen co-integration test is widely used in finance, economics and other fields to reveal the long-term relationship between different economic indicators, with the original hypothesis that there is no cointegration between the time series data. By applying the Johansen cointegration test, it has been determined that there are four cointegration relationships among the nine time series at the 95% confidence level, as shown in Table 6. i.e., there is at least one cointegration relationship, indicating that there is a long-term equilibrium relationship between variables, and no first-order difference processing is required for the data.

TABLE 6. Johansen test for time series.

	Johansen Trace Statistic	Critical Values	Hypothesis
1	1813.5905	197.3772	Reject
2	497.2874	159.529	Reject
3	171.5114	125.6185	Reject
4	104.3271	95.7542	Reject
5	54.7780	69.8189	Accept
6	21.3528	47.8545	Accept
7	9.9053	29.7961	Accept
8	3.4049	15.4943	Accept
9	0.0747	3.8415	Accept

III. METHODOLOGY

The utilization of machine learning and deep learning techniques for the prediction of time series data has become commonplace. Currently, research in this area typically involves the following steps: data acquisition, data decomposition, integration of sample mode and other data preprocessing steps, selection and establishment of prediction models, model evaluation, and result output. This paper adopts the conventional predictive modeling approach, and the specific modeling process is illustrated in Figure 4.

A. ICEEMDAN DECOMPOSITION AND FUZZY ENTROPY

CEEMDAN is an improvement on EMD and EEMD by adding multiple sets of independent and identically distributed white Gaussian noise to the time series. The averaging of results obtained after each EMD helps to effectively resolve the issue of mode aliasing. ICEEMDAN further enhances the early-stage problems of false components and mode aliasing in CEEMDAN, significantly reducing residual noise in the intrinsic mode function (IMF), and improving computational efficiency and algorithm performance [47].

Let $\boldsymbol{x}(t)$ be the original data, the main steps of ICEEMDAN are as follows:

1) Decompose the original data x (t) to construct a denoised composite signal of the original data:

$$\boldsymbol{x}^{(i)}(t) = \boldsymbol{x}(t) + \boldsymbol{\beta_0 E_1}\left(\boldsymbol{w}^{(i)}\right) \tag{1}$$

where $\boldsymbol{\beta_0}$ is the signal-to-noise ratio (SNR) obtained during the first decomposition, $\boldsymbol{w}^{(i)}$ is the $\boldsymbol{i}$th Gaussian white noise added, $\boldsymbol{E_1}(\cdot)$ is the first IMF component of EMD.

2) Repeat the previous step to obtain $\boldsymbol{x}^{(i)}(\mathbf{t})$, and then calculate the average of all local mean function values to

obtain the first residual component:

$$r_1 = \frac{1}{N}\sum_{i=1}^{N} M\left(x^{(i)}(t)\right) \tag{2}$$

where $M(\cdot)$ is the local mean function, derived from the original data minus the IMF value, r_1 is the first residual component.

3) Subtract the first residual component from the original data to obtain the first mode d_1:

$$d_1 = x(t) - r_1 \tag{3}$$

4) Calculate the local mean function value after the kth decomposition to obtain the kth residual component. Subtract the residual result from the previous residual component to obtain the kth modal d_m:

$$d_m = r_{k-1} - r_k \tag{4}$$

$$r_k = \frac{1}{N}\sum_{i=1}^{N} M\left(r_{k-1} + \beta_{k-1} E_k\left(w^{(i)}\right)\right) \tag{5}$$

where r_k is the kth residual component.

5) Repeatedly execute step (4) until the residual after decomposition does not exceed the two extreme points of EMD and the residual components cannot be decomposed.

The measurement of temporal disorder in signals is commonly done using entropy value. There exist three methods for calculating entropy, which are approximate entropy, sample entropy, and fuzzy entropy. Fuzzy entropy offers advantages of both approximate and sample entropy while enhancing the continuity and independence of the original signal in the extraction process, thereby increasing signal robustness. In this paper, fuzzy entropy is used to extract the features of the original data. First, the N-point sampling sequence is defined as $[u(1), u(2), \ldots, u(N)]$, and the similar capacitance limit r and phase space dimension are defined as m, thus the phase space is reconstructed as:

$$X(i) = [u(i), u(i+1), \ldots, u(i+m-1) - u_0(i)] \tag{6}$$

where $i = 1, 2, 3, \ldots, N - m + 1$, $u_0(i)$ is the mean of m consecutive $u(i)$, namely $u_0 = \frac{1}{m}\sum_{j=0}^{m-1} u(i+j)$. Then, the distance d_{ij}^m between any two sequences $X(i)$ and $X(j)$ is calculated by the formula:

$$\begin{aligned} d_{ij}^m = \max &\left(|u(i+p-1) - u_0(i)|\right. \\ &\left. - |u(j+p-1) - u_0(j)|\right) \end{aligned} \tag{7}$$

where $p = 1, 2, \ldots, m$. Fuzzy membership function is introduced to calculate the distance between two sequences:

$$A_{ij}^m = \begin{cases} 0, & x = 0 \\ exp\left[-ln(2)\left(\frac{d_{ij}^m}{r}\right)^2\right], & x > 0 \end{cases} \tag{8}$$

where ax is a fuzzy membership function, which can transform the uncertainty of a variable into a value that can be calculated. For each i, the average value $C_i^m(r)$ can be calculated to get:

$$C_i^m(r) = \frac{1}{N-m}\sum_{j=1, j\neq i}^{N-m+1} A_{ij}^m \tag{9}$$

Thus define:

$$\Phi^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} C_i^m(r) \tag{10}$$

At this point, the fuzzy entropy is obtained:

$$F(m, n, r, N) = \lim_{N\to\infty}\left[\ln\Phi^m(r) - \ln\Phi^{m+1}(r)\right] \tag{11}$$

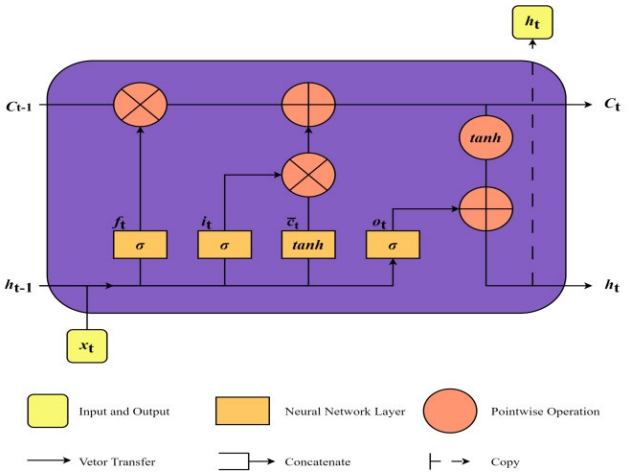


FIGURE 5. LSTM framework chart.

B. LSTM

LSTM is a specialized type of RNN capable of preserving long-term memory, achieved through a complex structure known as the LSTM Unit. This unit, the core of LSTM, comprises memory units designed for information storage and leverages a control gate mechanism for information updates, primarily through three distinct gates: the input gate, the forget gate, and the output gate [48]. The LSTM unit structure is shown in Figure 5.

Let x_t be the input at time t, h_{t-1} and h_t represents the output of the previous time and the output of time t, c_t is the memory unit value at time t. The main steps in the LSTM unit calculation process are as follows:

1) Calculate the candidate memory unit value $\tilde{c}_t$ at time t:

$$\tilde{c}_t = \tanh\left(w_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{12}$$

where w_c is the weight matrix, b_c is the corresponding offset.

2) Calculate the value of the input gate i_t, which is used to control the current input data to update the memory unit status value:

$$i_t = \sigma\left(w_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{13}$$

where $\sigma(\cdot)$ is the Sigmoid function.

3) Calculate the value of the forget gate f_t, which is used to control the update of historical data to memory unit status values:

$$f_t = \sigma\left(w_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{14}$$

4) Calculate the value of the memory unit at the current moment c_t:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{15}$$

where $\cdot$ represents the dot product. As can be seen from the above equation, the update of the memory unit depends on the values of the previous memory unit and candidate units, and is controlled by the input gate and forget gate.

5) Calculate the value of the output gate, which controls the output of the memory unit state value:

$$o_t = \sigma\left(w_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{16}$$

6) Calculate the output of the LSTM unit h_t:

$$h_t = o_t \cdot \tanh\left(c_t\right) \tag{17}$$

In the actual calculation process, some gradient descent optimization algorithms, such as the Adam algorithm, are often used to calculate the weight matrix. Furthermore, to mitigate overfitting, the incorporation of a Dropout mechanism is essential to enhance the network's robustness. In other words, traditional LSTM models perform poorly in predicting high-frequency sequences, thereby necessitating the exploration of alternative prediction techniques. Furthermore, the inherent gating mechanism in LSTM models tends to deteriorate over time, resulting in superior short-term memory but limited long-term memory capabilities. Consequently, the efficacy of processing long-term dependent information remains a challenge for LSTM models. Hence, alternative approaches are warranted for addressing this limitation.

C. TCN

TCN are variants of CNN that can be used to process time series. The basic idea is to obtain sufficient receptive fields(RFs) through a multi-layer network structure, making the CNN very deep [6]. Considering the benefits of large-scale parallel processing, TCN not only economizes time but also extracts more intricate data features.

1) CAUSAL CONVOLUTION

TCN needs to ensure that the output length of the network is the same as the input length and that future data will not leak, so 1D Full Convolutional Network and causal convolution are adopted. The following part is the definition of causal convolution, $X = (x_1, x_2, \ldots, x_t)$ is a time series, $F = \left(f_1, f_2, \ldots, f_t\right)$ is a wave filter, and the causal convolution at x_t is:

$$(FX)(x_t) = \sum_{k=1}^{K} f_k x_{t-K+k} \tag{18}$$

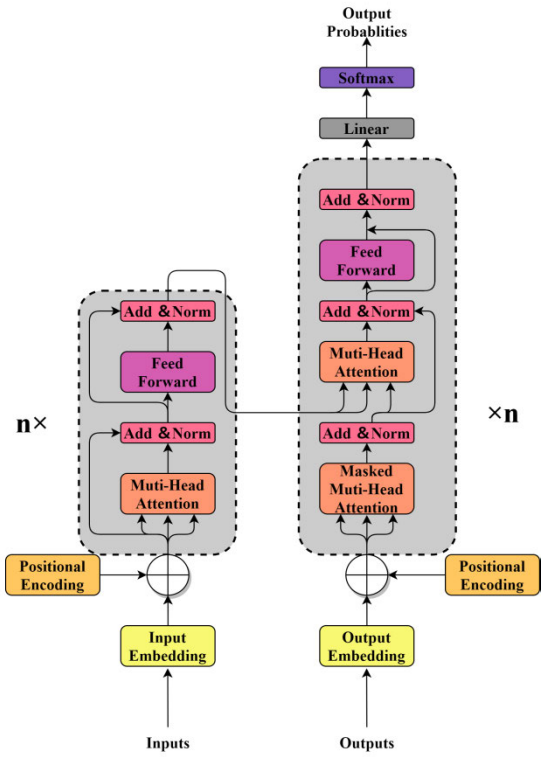


FIGURE 6. The transformer - model architecture.

2) EXPANSION CONVOLUTION

Since simple causal convolution cannot obtain longer time series, TCN adopts dilation convolution. The definition of expansion convolution with expansion factor d at x_t are as follows:

$$(F_d X)(x_t) = \sum_{k=1}^{K} f_k x_{t-(K-k)d} \tag{19}$$

3) RESIDUAL MODULE

To solve the problem of gradient vanishing or exploding caused by the introduction of causal convolution and dilation convolution, TCN introduces a residual module and weights the input x of the model into the output $F(x)$ of the model, ultimately obtaining the output y:

$$y = \text{Activation}\left(x + F(x)\right) \tag{20}$$

where Activation is the activation function.

In TCN, each layer uses the same filter, so convolution can be completed in parallel. Long input sequences can be processed as a whole in TCN, unlike sequential processing in RNN. Moreover, TCN can change the size of receptive fields in various ways, such as stacking more dilated convolutional layers, using a larger expansion factor, etc., to better control the memory size of the model and easily adapt to different domains. Additionally, TCN can accept input of any length through a one-dimensional convolution kernel, and the convolution window size increases over time, meaning that TCN

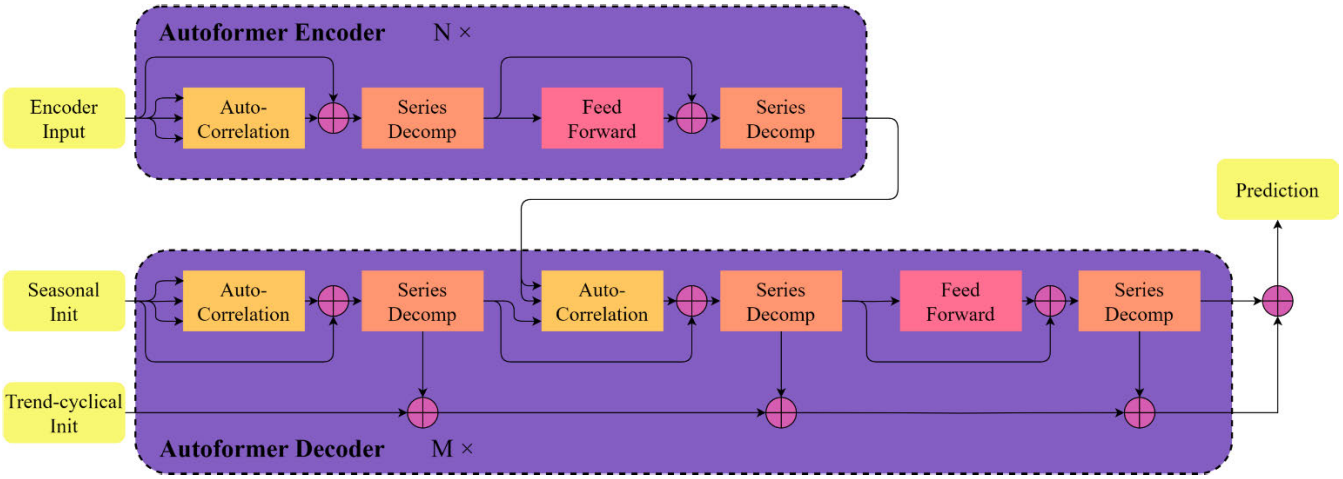


FIGURE 7. The Autoformer - model architecture.

can capture features of the long term. When shifting from smaller $\boldsymbol{k}$ and $\boldsymbol{d}$ domains to much larger $\boldsymbol{k}$ and $\boldsymbol{d}$ domains, TCN may fail to determine the overall trend of data changes over a long term due to insufficient receptive fields, indicating weak migration ability.

D. TRANSFORMER

The transformer model was introduced in 2017 as a solution to the sequence-to-sequence problem, utilizing a self-attention mechanism to replace the LSTM model [9]. This approach eschews the traditional decoder-encoder that must combine the inherent patterns of CNN or RNN and leverages parallel computing to enhance efficiency, improved results and reduced computational complexity. Its structure consists of encoder and decoder parts, as shown in Figure 6.

Among them, the encoder has two sub-layers, one is the multi-head attention layer, which uses self-attention to learn the internal relationships of time series, and the other is the feed forward layer, which is a simple fully connected network that performs the same operations on the vectors of each position, including two linear transformations and a ReLU activation function. The output is then passed to the decoder, and the entire process adopts parallel computing to improve efficiency. The decoder consists of three sub-layers, two of which are multi-head Attention layers. The attention layer utilizes self-attention to learn the internal relationships of the target, and the output obtained is input together with the results transmitted by the encoder into the attention layer above, which uses multi-head attention. Finally, after a negative feedback layer similar to the encoding part, the decoder output is obtained. However, due to the use of the self-attention function in both the encoder and decoder, the transformer has high time complexity and memory utilization increases twice within length L, resulting in memory bottlenecks when applied to time series prediction.

E. AUTOFORMER

Wu et al. [17] have introduced a novel transformer variant, termed the autoformer, which incorporates stochastic processes to enhance its predictive capabilities. Departing from conventional sequence decomposition approaches, the autoformer integrates an embedded deep decomposition architecture and auto-correlation mechanism to facilitate progressive sequence decomposition and connection, thereby overcoming limitations in information utilization for time series data. The autoformer architecture structure is shown in Figure 7.

1) DEEP DECOMPOSITION ARCHITECTURE

The deep decomposition architecture embeds sequence decomposition as an embedded module of autotransformer into the encoder and decoder. During the prediction process, the model achieves progressive decomposition, continuously separating periodic and trend terms from hidden variables. The Series decomposition block is based on the idea of moving average, smoothing the periodic term of $\mathcal{X}\in\mathbb{R}^{s\times 1}$ and stripping out the trend term. This process is defined as:

$$\begin{aligned} \mathcal{X}_t &= \text{AvgPool}\left(\text{Padding}\left(\mathcal{X}\right)\right) \\ \mathcal{X}_s &= \mathcal{X} - \mathcal{X}_t \end{aligned} \tag{21}$$

where $\mathcal{X}_s, \mathcal{X}_t \in R^{s\times 1}$ are trend terms and periodic terms. Sliding average is achieved through AvgPool$(\cdot)$, while Padding $(\cdot)$ maintains the sequence length. The above equation can also be written as $\mathcal{X}_t, \mathcal{X}_s = \text{SeriesDecomp}\left(\mathcal{X}\right)$.

The input of the Autotransformer encoder is the past $I = T$ time steps $\mathcal{X}_{en}\in\mathbb{R}^{I\times 1}$, decoder input includes periodic components $\mathcal{X}_{des}$ and trend component mathcal $\mathcal{X}_{det}$, both belong to $\mathbb{R}^{(I/2+O)\times 1}$.

In the encoder, the trend term is gradually eliminated to obtain the periodic term:

$$\mathcal{S}_{en}^{l,1}, \_ = \text{SeriesDecomp}\left(\text{Auto} - \text{Correlation}\left(\mathcal{X}_{en}^{l-1}\right) + \mathcal{X}_{en}^{l-1}\right)$$

$$\mathcal{S}_{en}^{l,2}, _{-} = \text{SeriesDecomp}\left(\text{FeedForward}\left(\mathcal{S}_{en}^{l-1}\right) + \mathcal{S}_{en}^{l-1}\right) \tag{22}$$

In the decoder, the autocorrelation mechanism aggregates subsequences with similar processes in different periods to obtain periodic terms, and gradually extracts and reconstructs trend terms from the predicted latent variables using a cumulative approach:

$$\begin{aligned}
\mathcal{S}_{de}^{l,1}, \mathcal{T}_{de}^{l,1} &= \text{SeriesDecomp}\left(\text{Auto} - \text{Correlation}\right. \\
&\quad \left. \times \left(\chi_{de}^{l-1}\right) + \chi_{en}^{l-1}\right) \\
\mathcal{S}_{de}^{l,2}, \mathcal{T}_{de}^{l,2} &= \text{SeriesDecomp}\left(\text{Auto} - \text{Correlation}\right. \\
&\quad \left. \times \left(\mathcal{S}_{de}^{l,1}, \chi_{en}^{N}\right) + \mathcal{S}_{en}^{l,1}\right) \\
\mathcal{S}_{de}^{l,3}, \mathcal{T}_{de}^{l,3} &= \text{SeriesDecomp}\left(\text{FeedForward}\left(\mathcal{S}_{de}^{l,2}\right) + \mathcal{S}_{en}^{l,2}\right) \\
\mathcal{T}_{\text{de}}^{l} &= \mathcal{T}_{\text{de}}^{l-1} + \mathcal{W}_{l,1} * \mathcal{T}_{\text{de}}^{l,1} + \mathcal{W}_{l,2} * \mathcal{T}_{\text{de}}^{l,2} + \mathcal{W}_{l,3} * \mathcal{T}_{\text{de}}^{l,3}
\end{aligned} \tag{23}$$

Progressive decomposition enables autotransformer to alternately perform sequence decomposition and prediction result optimization.

2) AUTOCORRELATION MECHANISM

autoformer achieves efficient sequence level connectivity through auto-correlation mechanisms, including periodic dependency discovery and delay information aggregation. The discovery of periodic dependence can be explained as: the similarity between a real discrete time process $\mathcal{X}_t$ and its τ-order delay $\mathcal{X}_{t-\tau}$ can be characterized by the autocorrelation coefficient $R_{\mathcal{X}\mathcal{X}}(\tau)$:

$$R_{\mathcal{X}\mathcal{X}}(\tau) = \lim_{L\to\infty} \frac{1}{L} \sum\nolimits_{t=0}^{L-1} \mathcal{X}_t \mathcal{X}_{t-\tau} \tag{24}$$

The similarity of this delay can be seen as the confidence level for period estimation. Secondly, sequence level connectivity is achieved through the step of time delay information aggregation, using Roll $(\cdot)$ to align information based on period estimation:

$$\begin{aligned}
\tau_1, \ldots, \tau_k &= \underset{\tau\in\{1,\ldots,L\}}{arg\text{Topk}}\left(R_{Q,K}(\tau)\right) \\
\hat{R}_{\text{Q,K}}(\tau_1), \cdots, \hat{R}_{\text{Q,K}}(\tau_k) &= SoftMax\left(R_{\text{Q,K}}(\tau_1), \right. \\
&\quad \left. \cdots, R_{\text{Q,K}}(\tau_k)\right) \\
\text{Auto} - \text{Correlation}(Q, K, V) &= \sum_{i=1}^{k} \text{Roll}(V, \tau_i) \hat{R}_{\text{Q,K}}(\tau_{\text{i}})
\end{aligned} \tag{25}$$

select $k = \lfloor c \times \log L \rfloor$ Top cycles with strong correlation, and use fast fourier transform (FFT) to ensure efficient calculation of auto-correlation coefficients:

$$\mathcal{S}_{\mathcal{X}\mathcal{X}}(f) = \mathcal{F}(\mathcal{X}_t)\mathcal{F}^*(\mathcal{X}_t) = \int_{-\infty}^{\infty} \mathcal{X}_t e^{-i2\pi tf}\,\mathrm{d}t$$
$$\overline{\int_{-\infty}^{\infty} \mathcal{X}_t e^{-i2\pi tf}\,\mathrm{d}t}$$

$$R_{\mathcal{X}\mathcal{X}}(\tau) = \mathcal{F}^{-1}\left(\mathcal{S}_{XX}(f)\right) = \int_{-\infty}^{\infty} \mathcal{S}_{XX}(f)\, e^{i2\pi f\tau}\mathrm{d}f \tag{26}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent FFT and its inverse transformation, respectively. The self phase shutdown system has a lightweight time complexity of $O(L\log L)$. Thanks to the use of autocorrelation mechanism, autoformer can improve both computational efficiency and information utilization compared with self-attention family.

F. MODEL EVALUATION

In terms of model evaluation, we use common indicators, including the mean squared error (MSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the mean squared percentage error (MSPE). The mathematical formulas are equations (27) — (30):

$$MSE\left(y, \hat{y}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{27}$$

$$MAE\left(y, \hat{y}\right) = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{28}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{29}$$

$$MSPE = \frac{100\%}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2 \tag{30}$$

where $\hat{y}_i$ is the predicted value, y_i is the original true value, n is the total number of predicted values, and $\overline{y}$ is the average of the original values. For MSE and MAE, their values represent the error between the original and predicted values. If MSE and MAE are closer to 0, it indicates that the model has a smaller error. MAPE and MSPE is similar to MSE and MAE. A MAPE or MSPE of 0% indicates that this is a perfect model.

Given that the aforementioned metrics only evaluate models based on their fitting performance, we introduce information criteria as indicators for model selection. Akaike's information criterion (AIC) was initially proposed by Akaike in 1973 [52]. This criterion operates on the principle that the quality of a fitted model should be determined not only by its fitting performance but also by the model's complexity, including the number of unknown parameters. Schwarz et al. [53] and Hannan and Quinn [54] respectively introduced the Schwarz Bayesian information criterion (SBIC) and the Hannan-Quinn information criterion (HQIC), which enhance the penalization of model complexity in different ways compared to AIC. However, in this study, the sample size is significantly smaller than the number of parameters in several machine learning models. Furthermore, considering that the model selection set in this study comprises TCN, LSTM, Transformer, and Autoformer, there are significant differences in the parameters between different models. TCN utilizes the CNN framework, LSTM operates on the RNN framework, while Transformer and Autoformer leverage attention mechanisms and their model complexities

TABLE 7. The total parameters of each model.

model name	TCN	LSTM	Autoformer	transformer
total params	60928	81049	11023872	15663105

often exceed those of TCN and LSTM. Employing the mentioned information criteria evaluation metrics could introduce a bias towards favoring simpler models. Akaike information criterion with a correction for finite sample sizes (AICc) [55] is a corrected version of AIC, specifically designed for situations with small sample sizes, aiming to alleviate the issue of AIC excessively penalizing model complexity under conditions of limited sample sizes. In situations with small sample sizes, AICc is more suitable for model selection as it mitigates the problem of overly favoring simpler models. Therefore, utilizing AICc as an evaluation metric aligns better with the models presented in this study. The formula for AICc is as follows:

$$AICc = 2k - 2ln\,(L) + \frac{2k\,(k+1)}{n-k-1} \tag{31}$$

where k represents the number of parameters in the model, L represents the likelihood function value of the model, and n represents the number of samples. It can be proven that as the sample size approaches infinity, AICc is equivalent to AIC.

G. PARAMETER SETTING OF EACH PREDICTION MODEL

1) LSTM

According to previous literature, data characteristics and continuous attempts, the LSTM structure parameters constructed in this paper are as follows: The number of nodes in the input layer is the number of lag periods, which is set as 36 here. Usually, the LSTM layer is set to be less than four layers and the number of cells is set to be less than 200, because too many layers or cells will lead to too long calculation time with only slightly improved prediction accuracy. Therefore, three LSTM layers are selected in this paper, with 128, 64 and 32 cells, respectively. At the same time, the structure also includes three Droupout layers with coefficients of 0.2 and a dense layer output, all of which are tanh functions. The optimizer uses an Adam optimizer with a mean square error (MSE) loss. In addition, the batch size for the fast calculation is 16, epochs is 5000, and the patience to adaptively reduce the learning rate is 100 runs.

2) TCN

Regarding existing literature and continuous attempts, the number of convolutional layers in TCN was set to 64, the kernel size of convolutional layers in TCN was set to 2, and the expansion rates of convolutional layers were set to 1, 2, 4, 8, respectively, the larger the expansion rate, the larger the receptive field. Meanwhile, the number of samples in each batch was set to 100 during training, with a total of 5000 iterations.

3) TRANSFORMER AND AUTOFORMER

We have the same parameter Settings for autoformer and transformer. The number of layers of the coding layer is set to 2, the number of layers of the decoding layer is set to 1, each layer contains 3 factors, the input dimension of both the coding layer and the decoding layer is 9, and the output dimension of the decoder is 1. The decoder we used is an Exponential decoder.

The total parameters of each model are shown as table:

It can be observed that the parameter size of the four selected models is significantly larger than the sample size to be tested, and the Transformer and Autoformer models boast a notably larger parameter count compared to the LSTM and TCN models. Thus, the incorporation of AICc of the information criterion as an evaluation metric serves to mitigate potential biases in model selection, ensuring a more balanced appraisal.

IV. FRAMEWORK OF COMBINATION MODEL

We first test the stationarity and autocorrelation test for the raw data of the forecasting target and other influencing factors. Then, the ICEEMDAN algorithm is used to divide the processed forecasting target data and influencing factors data into different IMFs. By calculating the fuzzy entropy of each IMF, the chaos degree in each IMF is measured, and based on this, all IMFs of the forecasting target and different factors are classified as data sequences of high, medium, and low frequencies. After that, multiple long-term sequence prediction models are used to predict the data of high frequency, medium frequency, and low frequency under different prediction time steps. The model with the best prediction performance at each step is selected as the sub-model. Combined with the long-term forecasting sub-models with the best prediction performance for the different frequency data of forecasting targets, a long-term forecasting combination model based on ICEEMDAN and machine learning is established.

A. MODE DECOMPOSITION

Use the ICEEMDAN algorithm mentioned above to perform mode decomposition on the forecasting target and other influencing factors. Obtain the intrinsic mode function i of the forecasting target and factor j, naming them as IMF_{0j} and IMF_{ij}. As shown in Figure 8, the smaller the number of i, the greater the degree of confusion in modifying the modality. In order to improve the calculation speed of the model and reduce the time complexity of the model, the fuzzy entropy value of each modality is calculated based on formula (7) to formula (11) in the previous text. According to the fuzzy entropy of IMF_{0j}, the $\boldsymbol{i}$-th IMF of data $\boldsymbol{j}$, multiple modes are divided into three frequencies: high-frequency, medium-frequency, and low-frequency. Among them, high-frequency time series data has the most severe fluctuations and the greatest chaos degree, mainly reflecting the short-term fluctuations and the impact of unexpected events in the original time series data; The fluctuation of intermediate frequency data

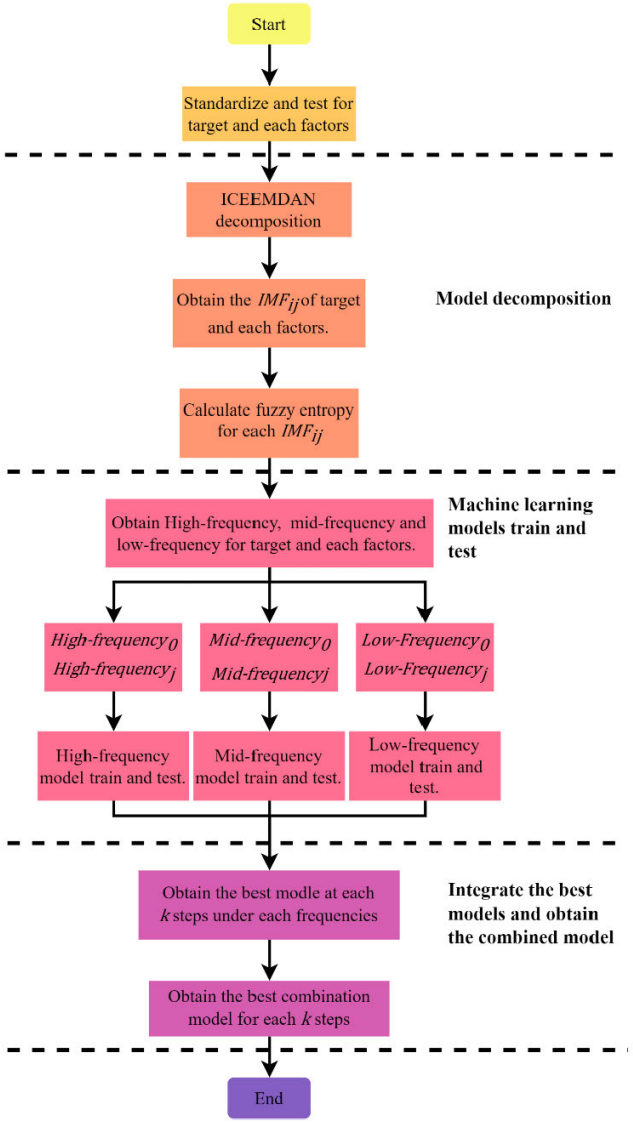


FIGURE 8. Framework of combination model.

is relatively slow compared to high-frequency data, mainly reflecting the periodic fluctuation of the original data in the time series; The low-frequency data fluctuates very slowly, reflecting the long-term trend term in the original data of the time series.

B. MACHINE LEARNING MODELS TRAIN AND TEST

Different Frequency data of forecasting target $High-Frequency_0$, $Mid-Frequency_0$, $Low-Frequency_0$ and factors $High-Frequency_j$, $Mid-Frequency_j$, $Low-Frequency_j$ are taken as the input of machine learning models. Test the predictive performance of each model at different frequencies and then select the model with the best prediction performance for different frequency data under k steps respectively.

C. INTEGRATE THE BEST MODEL AND OBTAIN THE COMBINED MODEL

Select the models with the best performance under different steps and combine the forecasting results of the three models with the best prediction performance for high-frequency, mid-frequency, and low-frequency data as sub-models. Finally, combine the optimal sub-models and obtain the best-integrated model under k steps.

In the next section, we take the interest rate of China's interbank bond as the prediction target and use the impact indicators selected in Section II as factor j to predict it using the combination model. The forecasting step k is 4, 8, 12, 18, 24, and 36, respectively. The LSTM, TCN, transformer, and autoformer are selected as the long-term forecasting models in the next section.

V. RESULTS AND COMPARATIVE ANALYSIS

When training the model, if the fluctuation of the original data is too large, the model may be unstable. We normalize the data and unify the range of data into a specific range to improve the stability and generalization ability of the model.

Then, the model introduced and designed in Chapters 3 and 4 was used to predict the target data through Python, and equations (26)-(29) were calculated to evaluate the results. The results of different prediction models under high-frequency data, mid-frequency data, and low-frequency data were compared respectively.

A. THE RESULT OF ICEEMDAN

The left part of Figure 9 shows the raw time series data and 9 IMF series of our forecasting target, where IMF9 is the residual. The horizontal coordinate represents the time series and the vertical coordinate represents the interest rate values. It can be found that from IMF1 to IMF9, the fluctuations of the time series become smaller and smaller. IMFs with large fluctuations tend to contain short-term information and noise, IMFs with slower fluctuations tend to contain cyclical information of the time series, and IMFs with flat fluctuations contain trend information. In this paper, we use fuzzy entropy to measure the complexity of the above IMF series. The fuzzy entropy value of each IMF calculated when the parameter n and r take different values is plotted as shown in the middle of Figure 9. Where the horizontal coordinates represent different IMF sequences and the vertical coordinates represent the values of fuzzy entropy.

It can be found that the fuzzy entropy values of IMF1 and IMF2 are much larger than that of other IMFs regardless of the parameter value variations, so IMF1 and IMF2 can be combined as a high-frequency sequence (H-IMF).Taking IMFs with fuzzy entropy significantly smaller than the other sequences (IMF6,IMF7,IMF8 and IMF9) as low-frequency sequence (L-IMF), and combine IMF3, IMF4, and IMF5 as medium-frequency sequence (M-IMF), the final decomposition results are shown in the right of Figure 9.

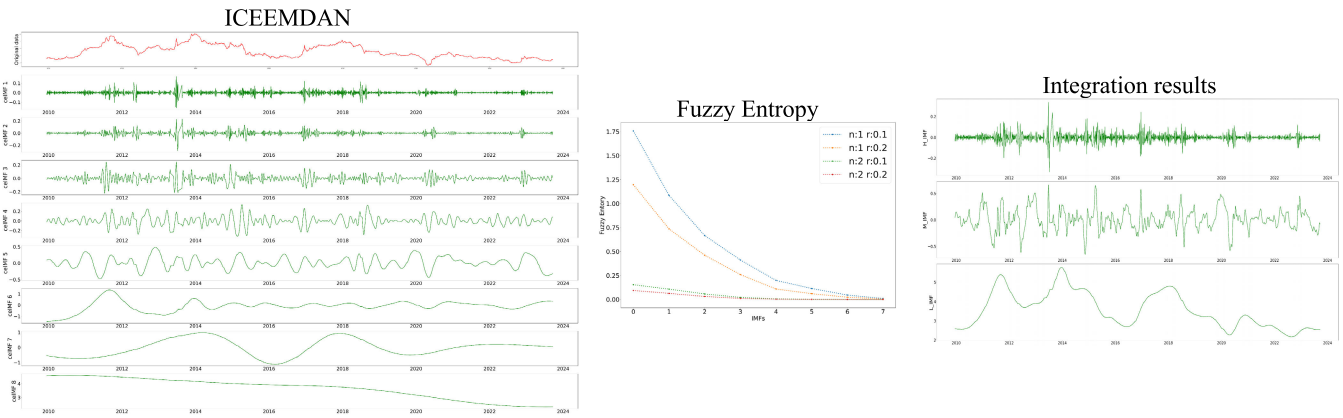


FIGURE 9. ICEEMDAN, Fuzzy Entropy, and integration results.

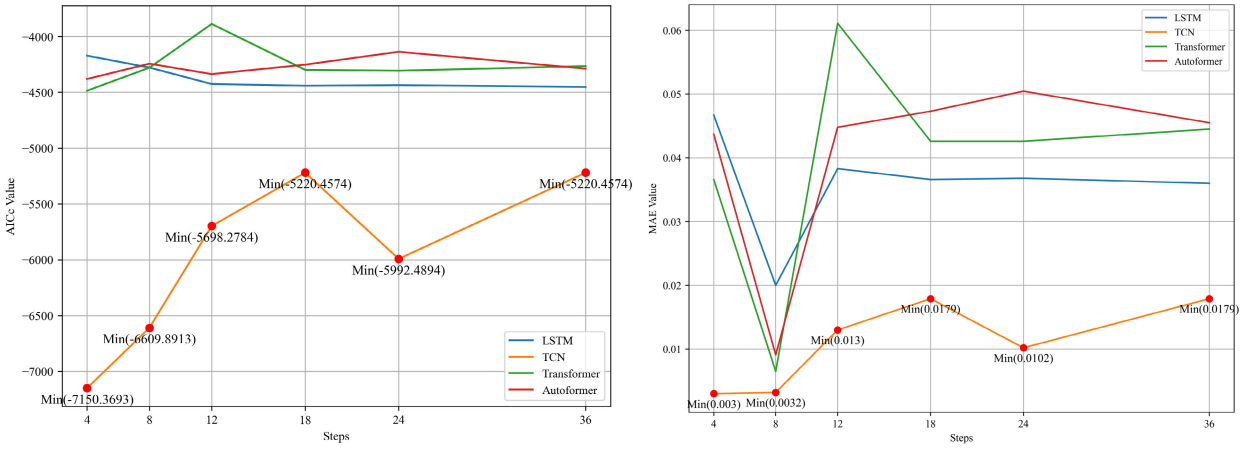


FIGURE 10. AICc and MAE of each prediction model at high frequency.

B. HIGH-FREQUENCY DATA

The calculation results of each evaluation index of different models for high-frequency data are shown in Table 8. Figure 10 shows the AICc and MAE of each model at high frequency, and it can be found that the two results are the same on the optimal model. It can be seen from the above table that under high-frequency data, TCN has the best performance among the four evaluation indicators under all step prediction. We believe that this is because TCN has good long-range dependency modeling ability. In high-frequency data, short term fluctuations between time series data are severe, and the distance between these time steps is really short, which make the short-distance dependency easier to be captured. and the characteristics of the data can be reflected more fully, thus amplifying the advantage of TCN. In order to more intuitively present the prediction effect, we perform the prediction results in the 8 steps situation as shown in Figure 11.

The MAE and AICc values corresponding to different models under different steps are shown in the following figure. The red markers mark the values of the evaluation indicators corresponding to the most effective model among all models at the corresponding number of steps. Due to the advantages of TCN in high-frequency data mentioned above, it can be observed that for high-frequency data, the TCN model exhibits the best fitting performance at all steps in case of both MAE and AICc.

C. MID-FREQUENCY DATA

The calculation results of each evaluation index of different models fors mid-frequency data are shown in Table 9. Figure 12 shows the AICc and MAE of each model at mid-frequency, and it can be found that the two results are the same on the optimal model.For the medium frequency data, LSTM performs optimally in the four evaluation indexes when the number of steps is low (4 steps). However, because LSTM controls the information flow of short-term memory and long-term memory through the gating mechanism of forgetting gate and input gate to realize the modeling of time series, it has short-term memory and forgetting property. As the step length increases, the information of past time step is

TABLE 8. Calculation results of each index under high-frequency data.

	Models	LSTM	**TCN**	Transformer	Autoformer
4 Steps	MAE	0.0467	**0.0030**	0.0366	0.0437
	MSE	0.0049	**0.0000**	0.0028	0.0034
	MAPE	0.1008	**0.0065**	0.0791	0.0900
	MSPE	0.1044	**0.0467**	0.0617	0.0752
	AICc	-4172.7992	**-7150.3693**	-4486.8295	-4381.7372
8 Steps	MAE	0.0200	**0.0032**	0.0065	0.0091
	MSE	0.4025	**0.1698**	0.2077	0.2039
	MAPE	1.2786	**0.2086**	0.3215	0.1015
	MSPE	0.0266	**0.0004**	0.0217	0.0236
	AICc	-4279.3245	**-6609.8913**	-4280.6631	-4244.5394
12 Steps	MAE	0.0383	**0.0130**	0.0611	0.0448
	MSE	0.0031	**0.0003**	0.0079	0.0036
	MAPE	0.0824	**0.0271**	0.1291	0.0944
	MSPE	0.0218	**0.0017**	0.0451	0.0198
	AICc	-4426.5887	**-5698.2784**	-3888.8941	-4337.7047
18 Steps	MAE	0.0366	**0.0179**	0.0426	0.0473
	MSE	0.0030	**0.0008**	0.0037	0.0042
	MAPE	0.0783	**0.0392**	0.0898	0.1005
	MSPE	0.0208	**0.0055**	0.0213	0.0259
	AICc	-4442.2553	**-5220.4574**	-4299.9967	-4253.3902
24 Steps	MAE	0.0368	**0.0102**	0.0426	0.0504
	MSE	0.0031	**0.0002**	0.0037	0.0051
	MAPE	0.0800	**0.0208**	0.0889	0.1068
	MSPE	0.0218	**0.0009**	0.0199	0.0294
	AICc	-4436.6744	**-5992.4894**	-4307.6490	-4137.9558
36 Steps	MAE	0.0360	**0.0179**	0.0445	0.0455
	MSE	0.0030	**0.0008**	0.0039	0.0039
	MAPE	0.0774	**0.0392**	0.0931	0.0958
	MSPE	0.0201	**0.0055**	0.0217	0.0215
	AICc	-4452.9402	**-5220.4574**	-4266.3748	-4289.3944

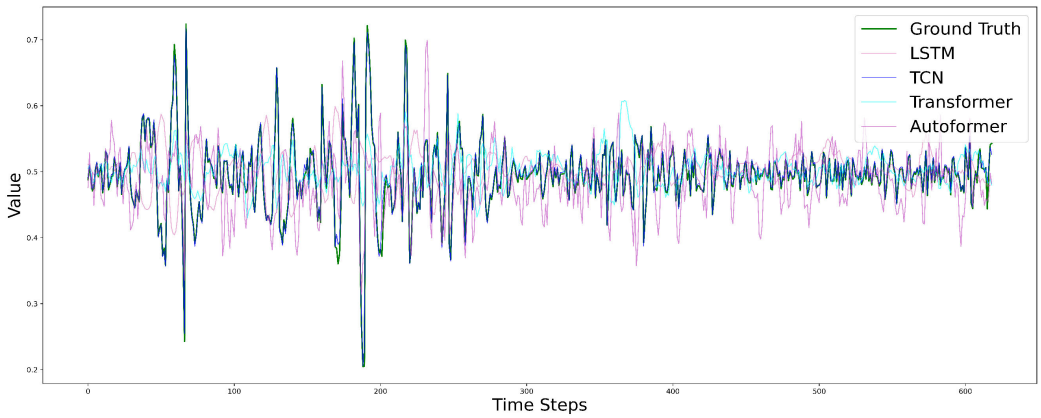


FIGURE 11. 8 steps prediction results for high-frequency data.

forgotten, and the prediction effect becomes worse, which is reflected in that the prediction result of LSTM with more than 4 steps is worse than that of TCN. The prediction results of 8 steps are shown in Figure 13.

D. LOW-FREQUENCY DATA

The calculation results of each evaluation index of different models for low-frequency data are shown in Table 10. Figure 14 shows the AICc and MAE of each model at low

TABLE 9. Calculation results of each index under mid-frequency data.

	Models	LSTM	TCN	Transformer	Autoformer
4 Steps	MAE	**0.0256**	0.0557	0.0546	0.0792
	MSE	**0.0011**	0.0048	0.0049	0.0103
	MAPE	**0.0782**	0.2214	0.1721	0.2173
	MSPE	**0.0308**	0.3803	0.1906	0.1357
	AICc	**-5018.0225**	-4186.43406	-4150.8933	-3741.6713
8 Steps	MAE	0.1044	**0.0467**	0.0617	0.0752
	MSE	0.0200	**0.0032**	0.0065	0.0091
	MAPE	0.4025	**0.1698**	0.2077	0.2039
	MSPE	1.2786	**0.2086**	0.3215	0.1015
	AICc	-3367.2875	**-4420.9775**	-3998.1838	-3822.5946
12 Steps	MAE	0.1550	**0.0662**	0.1250	0.0729
	MSE	0.0433	**0.0067**	0.0233	0.0087
	MAPE	0.5902	**0.2456**	0.4740	0.2270
	MSPE	3.4583	**0.2797**	1.4451	0.2107
	AICc	-2927.7451	**-3996.5869**	-3273.5934	-3850.1629
18 Steps	MAE	0.1906	**0.0520**	0.1346	0.0919
	MSE	0.0592	**0.0044**	0.0271	0.0128
	MAPE	0.6212	**0.2207**	0.5007	0.2249
	MSPE	2.6340	**0.3638**	1.6555	0.0949
	AICc	-2749.4327	**-4235.7678**	-3189.3243	-3623.5996
24 Steps	MAE	0.1920	**0.0586**	0.1641	0.0935
	MSE	0.0536	**0.0052**	0.0402	0.0122
	MAPE	0.6523	**0.2457**	0.6182	0.3389
	MSPE	2.7852	**0.4696**	2.2821	0.7150
	AICc	-2808.8988	**-4135.4152**	-2966.4915	-3642.6544
36 Steps	MAE	0.1746	**0.0237**	0.1690	0.0754
	MSE	0.0469	**0.0009**	0.0433	0.0084
	MAPE	0.0256	**0.0557**	0.6150	0.0792
	MSPE	0.0011	**0.0048**	2.1668	0.0103
	AICc	-2883.2443	**-5121.1951**	-2922.4784	-3858.3772

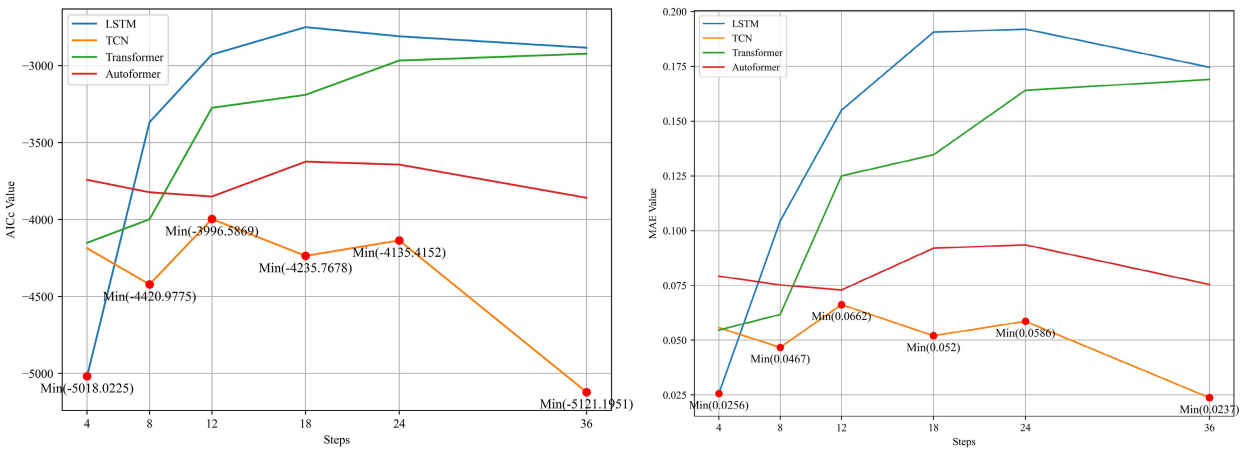


FIGURE 12. AICc and MAE of each prediction model at high frequency.

frequency, and it can be found that the two results are the same on the optimal model.

The LSTM prediction model performs best under low number of steps (4 steps and 8 steps) just as it does under medium frequency data, while the prediction results of autoformer stand out when the number of steps is increased. The prediction results of 8 steps are shown in Figure 15.

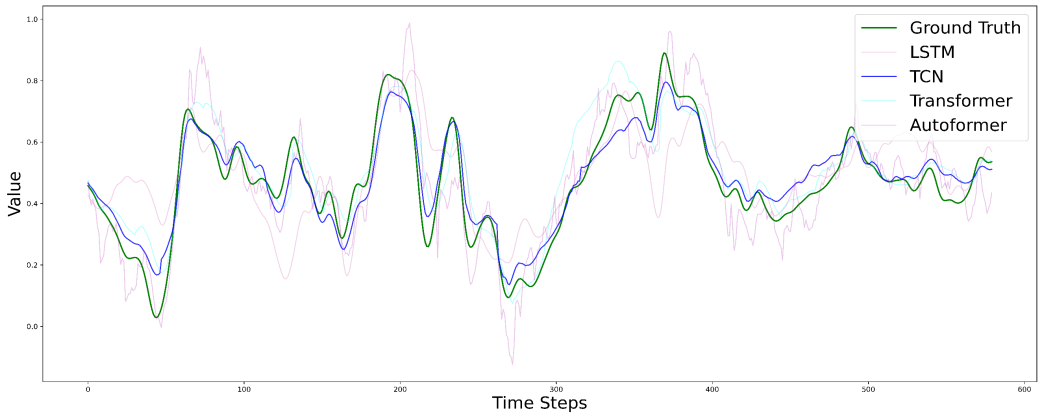


FIGURE 13. 8 steps prediction results for mid-frequency data.

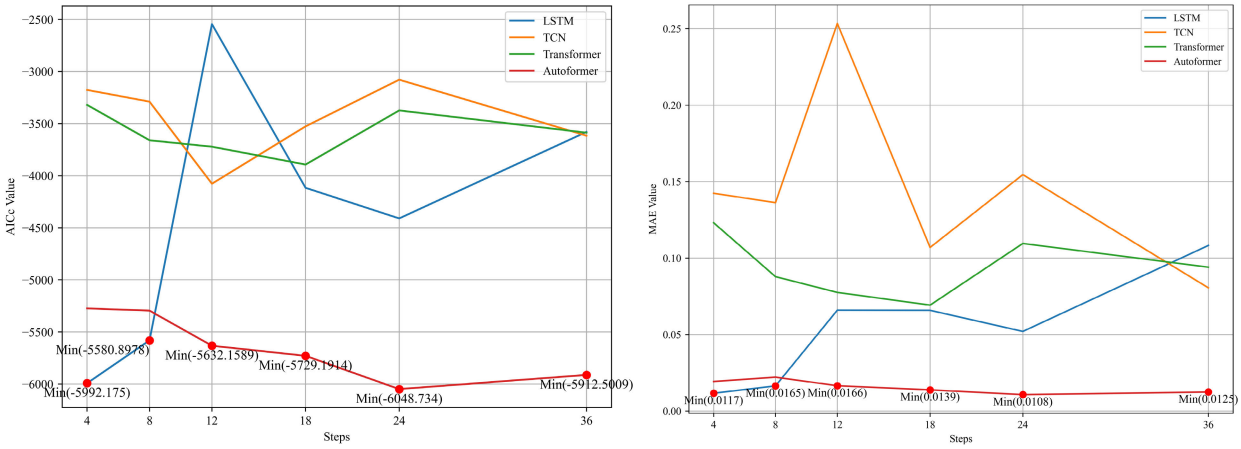


FIGURE 14. AICc and MAE of each prediction model at high frequency.

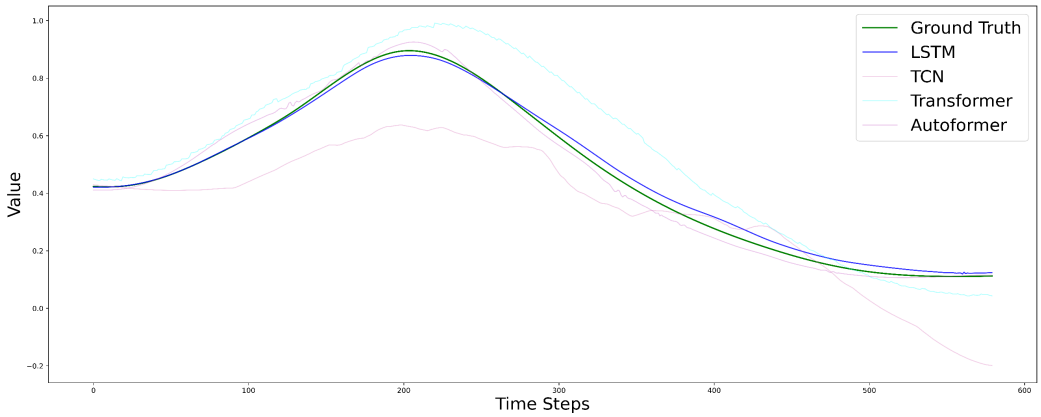


FIGURE 15. 8 steps prediction results for low-frequency data.

E. FINAL RESULT

We added and integrated the predicted values of the optimal prediction model under each component to obtain the optimal model, which is also the final model of this paper.

The process of obtaining the 8-step final prediction model is shown in Figure 16.

For other steps, the obtaining steps are the same as for 8 steps. The final calculation results of each evaluation index are shown in Table 11. In all tested steps, compared with each

TABLE 10. Calculation results of each index under low-frequency data.

	Models	LSTM	TCN	Transformer	Autoformer
4 Steps	MAE	**0.0117**	0.1424	0.1231	0.0194
	MSE	**0.0002**	0.0275	0.0208	0.0007
	MAPE	**0.0504**	0.4673	0.3425	0.0590
	MSPE	**0.0051**	0.3840	0.1826	0.0067
	AICc	**-5992.1750**	-3176.3341	-3320.0162	-5273.9119
8 Steps	MAE	**0.0165**	0.1363	0.0879	0.0222
	MSE	**0.0004**	0.0273	0.0116	0.0007
	MAPE	**0.0664**	0.5181	0.2486	0.0680
	MSPE	**0.0084**	0.8357	0.0928	0.0083
	AICc	**-5580.8978**	-3289.0874	-3659.7089	-5295.3554
12 Steps	MAE	0.0660	0.2534	0.0776	**0.0166**
	MSE	0.0058	0.1214	0.0104	**0.0004**
	MAPE	0.2004	1.4023	0.2309	**0.0609**
	MSPE	0.0724	7.5307	0.0878	**0.0070**
	AICc	-2544.7702	-4076.9155	-3721.4865	**-5632.1589**
18 Steps	MAE	0.0659	0.1070	0.0693	**0.0139**
	MSE	0.0054	0.0160	0.0079	**0.0003**
	MAPE	0.2600	0.3923	0.1868	**0.0506**
	MSPE	0.1244	0.3535	0.0548	**0.0069**
	AICc	-4116.2034	-3526.3039	-3893.0234	**-5729.1914**
24 Steps	MAE	0.0521	0.1546	0.1096	**0.0108**
	MSE	0.0036	0.0324	0.0193	**0.0002**
	MAPE	0.2578	0.5510	0.3638	**0.0378**
	MSPE	0.1563	0.6209	0.2284	**0.0036**
	AICc	-4410.1884	-3077.9246	-3373.2621	**-6048.7340**
36 Steps	MAE	0.1083	0.0806	0.0941	**0.0125**
	MSE	0.0144	0.0127	0.0136	**0.0002**
	MAPE	0.4717	0.1747	0.2583	**0.0402**
	MSPE	0.4496	0.0424	0.1055	**0.0028**
	AICc	-3579.4601	-3615.7120	-3586.9210	**-5912.5009**

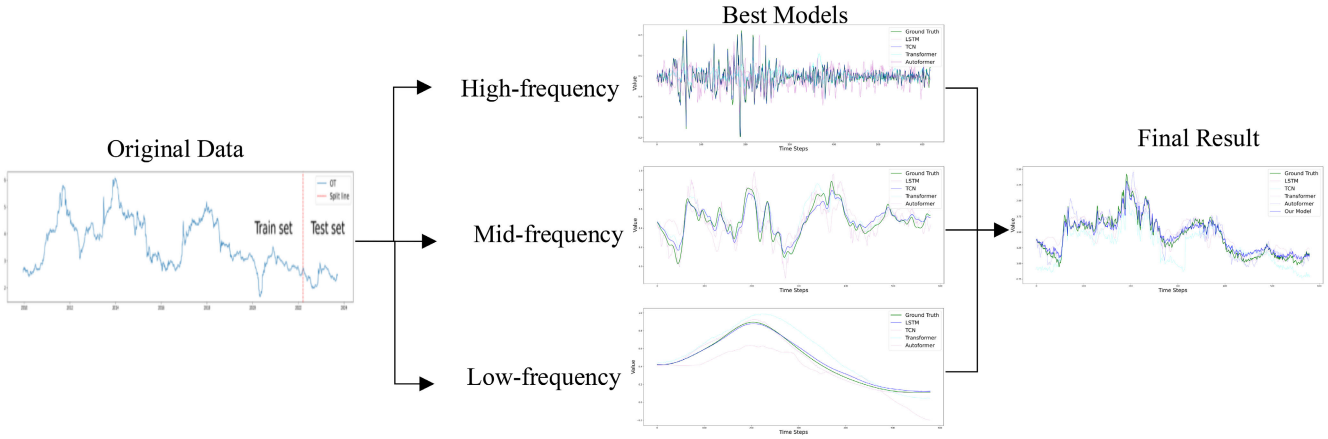


FIGURE 16. 8 steps prediction process diagram.

other long-term forecasting models, the composite model we proposed always shows better performance in terms of AICc and MAE, which indicates that our model has better prediction accuracy at different time steps.

TABLE 11. The results of each evaluation index under the final model.

	Models	LSTM	TCN	Transformer	Autoformer	Our method
4 Steps	MAE	0.0541	0.1359	0.1683	0.0860	**0.0334**
	MSE	0.0058	0.0282	0.0390	0.0128	**0.0020**
	MAPE	0.0371	0.0993	0.1203	0.0592	**0.0234**
	MSPE	0.0026	0.0157	0.0198	0.0056	**0.0009**
	AICc	-4061.4513	-3137.9139	-2952.6378	-3598.8381	**-4966.2049**
8 Steps	MAE	0.1243	0.1800	0.0733	0.0961	**0.0514**
	MSE	0.0255	0.0453	0.0096	0.0162	**0.0037**
	MAPE	0.0890	0.1238	0.0491	0.0654	**0.0382**
	MSPE	0.0133	0.0211	0.0041	0.0072	**0.0022**
	AICc	-3259.5139	-2910.3419	-3799.1523	-3578.8527	**-4323.8804**
12 Steps	MAE	0.1651	0.1877	0.1972	0.1063	**0.0720**
	MSE	0.0504	0.0633	0.0602	0.0185	**0.0077**
	MAPE	0.1158	0.1380	0.1423	0.0743	**0.0498**
	MSPE	0.0252	0.0407	0.0319	0.0089	**0.0037**
	AICc	-2865.7463	-2583.2588	-2759.3394	-3679.1324	**-3927.6299**
18 Steps	MAE	0.1991	0.1054	0.3396	0.0926	**0.0589**
	MSE	0.0607	0.0177	0.1447	0.0140	**0.0053**
	MAPE	0.1426	0.0744	0.2379	0.0660	**0.0434**
	MSPE	0.0323	0.0087	0.0708	0.0072	**0.0031**
	AICc	-2757.9802	-3448.8313	-2798.4209	-3408.5010	**-3992.3286**
24 Steps	MAE	0.2010	0.1890	0.2948	0.1063	**0.0573**
	MSE	0.0570	0.0496	0.1193	0.0176	**0.0049**
	MAPE	0.1450	0.1369	0.1976	0.0754	**0.0442**
	MSPE	0.0306	0.0273	0.0496	0.0089	**0.0033**
	AICc	-2787.0660	-2875.9410	-2483.9381	-3441.9267	**-4161.2919**
36 Steps	MAE	0.1961	0.0868	0.0885	0.0885	**0.0383**
	MSE	0.0567	0.0146	0.0127	0.0127	**0.0021**
	MAPE	0.1409	0.0544	0.0620	0.0620	**0.0278**
	MSPE	0.0541	0.1359	0.0062	0.0860	**0.0012**
	AICc	-2761.6188	-3581.1338	-2579.6160	-3615.0327	**-4644.0939**

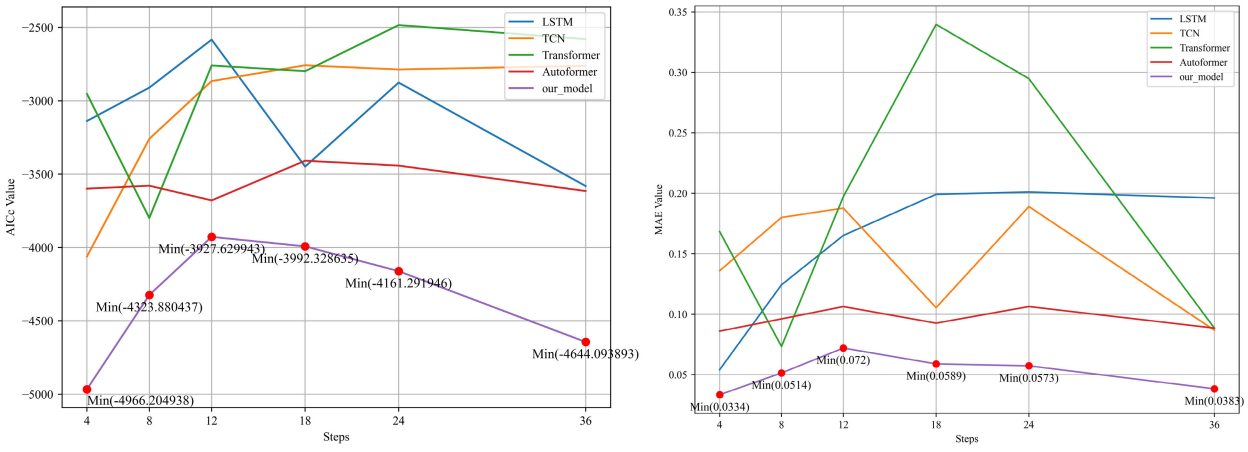


FIGURE 17. Comparison of AICc and MAE calculation results between our method and other single models under asynchronous number.

Figure 18 shows the optimal prediction model under each step number by our method is highly coincident with the original data, which can better reflect the trend of the original data, and is significantly better than the result obtained by using a single model.

As can be seen from Table 11, for the four evaluation indicators, the method we proposed is superior to other single models at different steps. Compared with only using LSTM, the accuracy of the model in terms of MAE from 4 steps to 36 steps was improved by 38.21%, 58.61%, 56.41%, 70.43%,

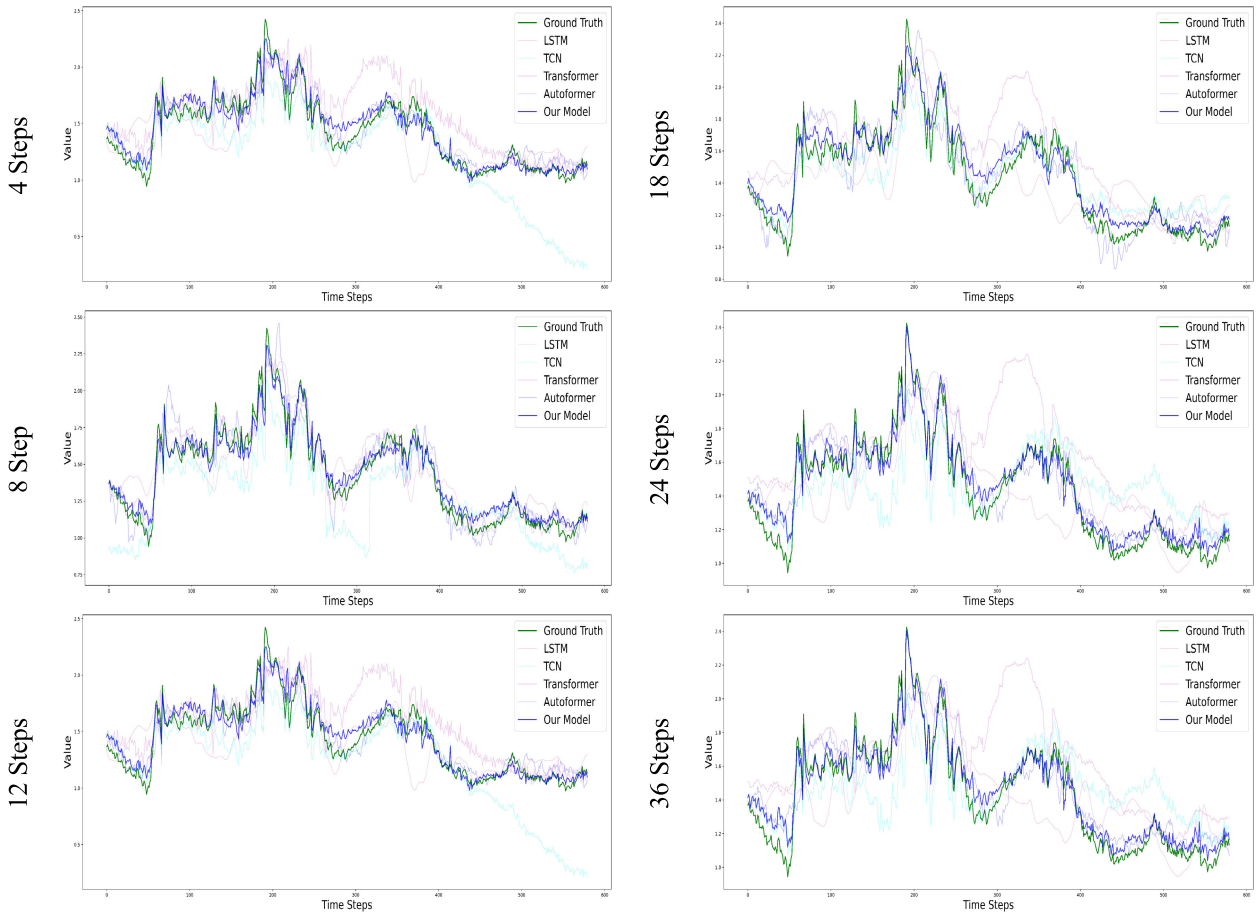


FIGURE 18. Final forecast result graph.

71.51%, and 80.49%, respectively. Compared with only using TCN, the accuracy of the model is improved by 68.30%, 71.43%, 61.66%, 44.15%, 69.71%, and 55.91%, respectively. Compared to only using transformer, model accuracy is improved by 80.15%, 29.82%, 63.51%, 82.66%, 80.58%, and 83.02%, respectively; Compared with only using autoformer, the model accuracy is improved by 61.16%, 46.46%, 32.20%, 36.46%, 46.12% and 56.78%, respectively. Then focus on the improvement of AICc, Compared with only using LSTM, the accuracy of the model in terms of AICc from 4 steps to 36 steps was improved by 22.28%, 32.65%, 37.05%, 44.76%, 49.31% and 68.17%, respectively. Compared with only using TCN, the accuracy of the model is improved by 58.26%, 48.57%, 52.04%, 15.76%, 44.69% and 29.68%, respectively. Compared to only using transformer, model accuracy is improved by 68.20%, 13.81%, 42.34%, 42.66%, 67.53% and 80.03%, respectively; Compared with only using autoformer, the model accuracy is improved by 37.99%, 20.82%, 6.75%, 17.13%, 20.90% and 28.47%, respectively. At the same time, from the prediction result chart, the curve predicted by our method is highly coincident with the original data, which can better reflect the trend of the original data, and is significantly better than the result obtained by using a single model.

VI. CONCLUSION

The interbank bond rate serves as a crucial indicator in the financial secondary market and offers insight into the liquidity and risk conditions of the banking industry. This study introduces a novel hybrid model based on ICEEMDAN, fuzzy entropy, and various long-term forecast models to predict the interbank bond rate. Compared to the traditional forecasting models such as LSTM, TCN, transformer, and autoformer, the proposed model demonstrates superior accuracy and long-term prediction capabilities. The results show that: 1) in the tested case, the forecasting accuracy of the proposed model in 4steps, 8steps, 12steps, 18steps, 24steps, 36steps is respectively improved by 38.21%, 29.82%, 32.20%, 36.46%, 46.12%, and 56.78% compared with the best performance among the contrastive models; 2) as the prediction step size increases, the advantages of the proposed model in improving prediction accuracy become more apparent compared with the contrastive models, indicating that the proposed model has better long-term prediction ability.

This paper contributes to the literature on interbank bond rate prediction in three ways. Firstly, the application of ICEEMDAN and fuzzy entropy facilitates the decomposition of raw data into high-frequency, mid-frequency, and

low-frequency components, thereby enhancing the model's capacity to capture the volatility characteristics of interbank bond market transaction rates. Secondly, the incorporation of multiple influencing factors into the model's input comprehensively captures financial market information in the time series. Thirdly, the utilization of different models for different frequencies significantly improves the accuracy of long-term interbank bond rate forecasts compared to a single model approach. Different model combinations have been proposed at different prediction time steps. Overall, we also present various model combinations at different prediction time steps, further advancing the understanding of long-term interbank bond rate prediction.

REFERENCES

[1] Y. Hu, "Stock forecast based on optimized LSTM model," *Comput. Sci.*, vol. 48, no. 1, pp. 151–157, Jun. 2021.

[2] X. Zhang, "Financial time series forecasting based on LSTM neural network optimized by wavelet denoising and whale optimization algorithm," *Academic J. Comput. Inf. Sci.*, vol. 7, no. 5, pp. 1–9, 2022.

[3] U. M. Sirisha, M. C. Belavagi, and G. Attigeri, "Profit prediction using ARIMA, SARIMA and LSTM models in time series forecasting: A comparison," *IEEE Access*, vol. 10, pp. 124715–124727, 2022.

[4] Q. Yang and C. Wang, "A study on forecast of global stock indices based on deep LSTM neural network," *Stat. Res.*, vol. 3, no. 36, pp. 65–77, Mar. 2019.

[5] T. Leng, "Stock price prediction research using attention-based LSTM neural network," M.S. thesis, Dept. Econ., Nanjing Univ. Sci. Technol., Nanjing, China, 2023.

[6] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[7] D. S. P. Salazar, P. J. L. Adeodato, and A. L. Arnaud, "Continuous dynamical combination of short and long-term forecasts for nonstationary time series," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 1, pp. 241–246, Jan. 2014.

[8] M. Dong and L. Grumbach, "A hybrid distribution feeder long-term load forecasting method based on sequence prediction," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 470–482, Jan. 2020.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017, *arXiv:1706.03762*.

[10] J. Devlin, M.-W. Chang, K. Lee, and L. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2019, *arXiv:1810.04805*.

[11] T. B. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 1877–1901.

[12] C. A. Huang et al., "Music transformer," 2018, *arXiv:1809.04281*.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[14] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[15] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y. X. Wang, and X. Yan, "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting," 2019, *arXiv:1907.00235*.

[16] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 12, 2021, pp. 11106–11115.

[17] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 22419–22430.

[18] E. Fu, Y. Zhang, F. Yang, and S. Wang, "Frequency decomposition transformer based model for long term time series prediction," *Manuf. Autom.*, vol. 44, no. 11, pp. 177–181, 2022.

[19] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[20] L. Sun, T. Wang, M. Xu, X. Li, and H. Piao, "Thin fouling ultrasonic detection signal denoising based on improved CEEMD," *Chin. J. Sci. Instrum.*, vol. 38, no. 12, pp. 2879–2887, Dec. 2017.

[21] S. D. Yang, Z. A. Ali, and B. M. Wong, "FLUID-GPT (fast learning to understand and investigate dynamics with a generative pre-trained transformer): Efficient predictions of particle trajectories and erosion," *Ind. Eng. Chem. Res.*, vol. 62, no. 37, pp. 15278–15289, Sep. 2023.

[22] S. D. Yang, Z. A. Ali, H. Kwon, and B. M. Wong, "Predicting complex erosion profiles in steam distribution headers with convolutional and recurrent neural networks," *Ind. Eng. Chem. Res.*, vol. 61, no. 24, pp. 8520–8529, Jun. 2022.

[23] Y. Gao, X. Wang, N. Yu, and B. M. Wong, "Harnessing deep reinforcement learning to construct time-dependent optimal fields for quantum control dynamics," *Phys. Chem. Chem. Phys.*, vol. 24, no. 39, pp. 24012–24020, 2022.

[24] M. A. Colominas, G. Schlotthauer, and M. E. Torres, "Improved complete ensemble EMD: A suitable tool for biomedical signal processing," *Biomed. Signal Process. Control*, vol. 14, pp. 19–29, Nov. 2014.

[25] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 4144–4147.

[26] H.-L. Yang and H.-C. Lin, "Applying the hybrid model of EMD, PSR, and ELM to exchange rates forecasting," *Comput. Econ.*, vol. 49, no. 1, pp. 99–116, Jan. 2017.

[27] R. B. Avery, T. M. Belton, and M. A. Goldberg, "Market discipline in regulating bank risk: New evidence from the capital markets," *J. Money, Credit Banking*, vol. 20, no. 4, p. 597, Nov. 1988.

[28] M. Bodenstein, J. Hebden, and R. Nunes, "Imperfect credibility and the zero lower bound," *J. Monetary Econ.*, vol. 59, no. 2, pp. 135–149, Mar. 2012.

[29] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *J. Amer. Stat. Assoc.*, vol. 74, no. 366, p. 427, Jun. 1979.

[30] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, p. 297, Aug. 1978.

[31] M. Sibtain, X. Li, S. Saleem, Qurat-Ul-Ain, M. S. Asad, T. Tahir, and H. Apaydin, "A multistage hybrid model ICEEMDAN-SE-VMD-RDPG for a multivariate solar irradiance forecasting," *IEEE Access*, vol. 9, pp. 37334–37363, 2021.

[32] S. Heo, K. Nam, J. Loy-Benitez, and C. Yoo, "Data-driven hybrid model for forecasting wastewater influent loads based on multimodal and ensemble deep learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 6925–6934, Oct. 2021.

[33] J. Gong, "Do macro factors affect credit spreads on credit bonds? Empirical analysis based on the interbank bond market," M.S. thesis, Dept. Finance, Southwestern Univ. Finance Econ., Chengdu, China, 2022.

[34] Z. Lu, "The equilibrium and causal analysis of the price fluctuation in China's bond market," *Securities Market Herald*, vol. 16, no. 2, pp. 74–77, 2006.

[35] A. G. Özbekler, A. Kontonikas, and A. Triantafyllou, "Volatility forecasting in European government bond markets," *Int. J. Forecasting*, vol. 37, no. 4, pp. 1691–1709, Oct. 2021.

[36] J. Fan, Y. Ke, and Y. Liao, "Augmented factor models with applications to validating market risk factors and forecasting bond risk premia," *J. Econometrics*, vol. 222, no. 1, pp. 269–294, May 2021.

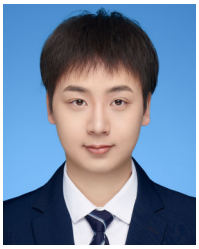
[37] J. H. Cochrane and M. Piazzesi, "Bond risk premia," *Amer. Econ. Rev.*, vol. 95, no. 1, pp. 138–160, Mar. 2005.

[38] M. Huang, "An empirical analysis of the difference between the interbank bond repo rate and Shibor," *J. East China Nommal Univ., Philosophy Social Sci.*, vol. 39, no. 6, pp. 114–116, Dec. 2007.

[39] F. Duan and L. Sun, "On mechanism of influence of the dollar index on China's bond market," *Shanghai Finance*, vol. 34, no. 6, pp. 60–65, 2013.

[40] Q. Wang, S. Gao, J. Li, and M. Leibrecht, "Do foreign investors affect the volatility of local currency bond prices? Empirical evidence from China," *Heliyon*, vol. 9, no. 6, Jun. 2023, Art. no. e16658.

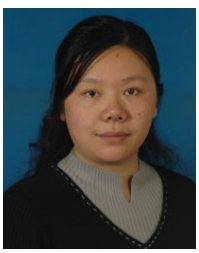
[41] D. G. Baur and B. M. Lucey, "Is gold a Hedge or a safe haven? An analysis of stocks, bonds and gold," *Financial Rev.*, vol. 45, no. 2, pp. 217–229, May 2010.
[42] X. Tan, J. Zhang, and X. Zheng, "The two-way spillover effects among international commodity markets and financial markets an empirical study based on BEKK-GARCH model and spillover index method," *China Soft Sci.*, vol. 8, pp. 31–48, Sep. 2018.
[43] X. Wen, Z. Liu, and Y. Xiong, "The impact of fluctuations in fuel oil futures prices on China's economic growth," *Statist. Decision*, vol. 29, no. 8, pp. 120–123, 2013.
[44] Z. Dai and J. Kang, "Bond yield and crude oil prices predictability," *Energy Econ.*, vol. 97, May 2021, Art. no. 105205.
[45] H. Qing, "Risk structure for enterprise bond and risk transmission between enterprise and treasury yield curves," M.S. thesis, Dept. Finance, Xiamen Univ., Xiamen, China, 2018.
[46] X. Liu, "Research on the information content of the yield curve in China's bond market," Ph.D. Dissertation, Dept. Finance, Univ. Int. Bus. Econ., Beijing, China, 2022.
[47] F. Zhou, Z. Huang, and C. Zhang, "Carbon price forecasting based on CEEMDAN and LSTM," *Appl. Energy*, vol. 311, Apr. 2022, Art. no. 118601.
[48] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[49] Y. Zhang, T. Du, and C. Yang, "The linkage between inter-bank bond market and interest rate swap market-based on DCC-MIDAS model," *China, Syst. Eng.*, vol. 36, no. 1, pp. 13–21, 2018.
[50] J. R. Booth and D. T. Officer, "Expectations, interest rates, and commercial bank stocks," *J. Financial Res.*, vol. 8, no. 1, pp. 51–58, Mar. 1985.
[51] H. Davies, "Averaging in a framework of zero reserve requirements: Implications for the operation of monetary policy," Monetary Econ., Bank England, Work. Paper 84, Oct. 1998.
[52] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, 1974.
[53] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.
[54] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression," *J. Roy. Stat. Soc., B, Methodol.*, vol. 41, no. 2, pp. 190–195, Jan. 1979.
[55] C. M. Hurvich, J. S. Simonoff, and C.-L. Tsai, "Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 60, no. 2, pp. 271–293, Jul. 1998.
[56] J. M. O'Brien, "Estimating the value and interest rate risk of interest-bearing transactions deposits," *SSRN Electron. J.*, vol. 2000, no. 53, pp. 1–46, 2000.
[57] E. T. Swanson, "Federal reserve transparency and financial market forecasts of short-term interest rates," *Finance Econ. Discuss. Ser.*, vol. 2004, no. 6, pp. 1–36, 2004.
[58] G. Song, Y. Zhang, F. Bao, and C. Qing, "Stock prediction model based on particle swarm optimization LSTM," *J. Beijing Univ. Aeronaut. Astronaut.*, vol. 45, no. 12, pp. 2533–2542, 2019.
[59] Y. He, P. Li, and J. Han, "Research on predictive modeling on stock market index based on CEEMDAN-LSTM," *Statist. Inf. Forum*, vol. 35, no. 6, pp. 34–45, 2020.
[60] H. Di, X. Zhao, and Z. Zhang, "Commodity price forecasting based on EEMD-LSTM-AdaBoost," *Stat. Decision*, vol. 34, no. 13, pp. 72–76, 2018.
[61] Y. Yao, Z. Zhang, Y. Zhao, Y. Li, F. Li, and L. Huang, "Heston options pricing model based on the principle of decomposition-reassembly-prediction-integration," in *Operations Research and Management Science*. Hefei, China: Hefei Univ. of Technology, pp. 1–8.
[62] J. Cui and H. Zou, "Carbon financial market price forecasting based on CEEMDAN-MR-PE-NLE multi-frequency optimization combined model," *Math. Pract. Theory*, vol. 50, no. 3, pp. 105–120, 2020.
[63] M. M. Andreasen, T. Engsted, S. V. Møller, and M. Sander, "The yield spread and bond return predictability in expansions and recessions," *Rev. Financial Stud.*, vol. 34, no. 6, pp. 2773–2812, May 2021.
[64] N. R. Sabri, "Stock return volatility and market crisis in emerging economies," *Rev. Accounting Finance*, vol. 3, no. 3, pp. 59–83, Mar. 2004.
[65] L. Tian and D. Tan, "Financialization and americanization of bulk commodities' pricing: A study of the relation between stock indices and spot commodities," *China Ind. Econ.*, vol. 32, no. 10, pp. 72–84, 2014.
[66] Y. Ma and L. Zhang, "The impact of international oil price fluctuations on sovereign bond markets—Empirical research based on network model," *J. Financial Develop. Res.*, vol. 41, no. 4, pp. 15–24, 2022.



YUE YU is currently pursuing the B.S. degree in financial technology with the School of Economics, Wuhan University of Technology, Wuhan, China. His research interests include financial technology, artificial intelligence, and operations optimization.

GUANGWU KUANG is currently pursuing the B.S. degree in mathematical finance with the School of Economics, Wuhan University of Technology, Wuhan, China. His research interests include mathematical finance and financial technology.

JIANRUI ZHU is currently pursuing the B.S. degree in mathematical finance with the School of Economics, Wuhan University of Technology, Wuhan, China. His research interests include energy and environmental economics, technical economics and management, and supply chain.



LEI SHEN received the B.S., M.S., and Ph.D. degrees degree from the Department of Economics, Wuhan University of Technology (WHUT), Wuhan, China, in 1997, 2002, and 2012, respectively. She is currently an Assistant Professor with the School of Economics, WHUT. She has more than 30 publications in national and international conference proceedings and journals. Her research interests include financial engineering, green finance, and financial risk management.

MENGJIA WANG is currently pursuing the B.S. degree in financial technology with the School of Economics, Wuhan University of Technology, China. Her research interests include financial technology and commercial banks.