

Stock Prediction Analysis and Risk Conduction Path Research Based on EMD-LSTM Model

Jingwen Xu

School of Qianweichang, Shanghai University, Shanghai 200444, China.

xjw_math@163.com

Abstract. Stock market, government bond market and corporate bond market are important components of my country's financial market. It is of great significance to study stock market price fluctuations and their forecasting methods, and to reveal the risk transmission path between them. The first part of this paper mainly studies the construction of the index price prediction model, and makes a comprehensive comparison of the high-frequency fluctuation characteristics, long-term trend characteristics and average trend characteristics of the EMD decomposition results of the Shanghai Composite Index, the SSE Government Bond Index, and the SSE Corporate Bond Index. The LSTM prediction model and the EMD-LSTM prediction model were established for the three sequences respectively, and the prediction effects of the models were comprehensively compared by calculating the RMSE and MAPE of the predicted values of the test set. The results show that after the introduction of the EMD method, the prediction effects of the SSE Composite Index and the SSE Corporate Bond Index prediction model have been greatly improved, but the prediction effect of the SSE Government Bond Index prediction model is far inferior to that of the LSTM model. The second part of this paper mainly studies the financial market risk analysis and transmission path. First, the fluctuation risk analysis and stationarity analysis of the logarithmic return series of the three index price series are carried out. The yield series are stable. Then, the LSTM models based on different training data predict three index price series and compare the prediction results to study the similarity of the price fluctuation mechanism of the three financial markets. The results show that the price fluctuation mechanism of the stock market is similar to that of the corporate bond market, and the price fluctuation mechanism of the government bond market is more similar to the corporate bond market than the stock market. The paper conducts E-G cointegration test on three logarithmic return series that have been tested for stationarity to study the long-term equilibrium relationship between the three markets and finds that there is a cointegration relationship between the SSE Composite Index and the SSE Government Bond Index, as well as the SSE Government Bond Index and the SSE Corporate Bond Index. Finally, the Granger causality test is carried out between the sequences with cointegration relationship, and the result shows that the risk transmission mechanism between the national debt market and the corporate bond market is strong, while the risk transmission mechanism between the stock market and the bond market is weak.

Keywords: EMD decomposition; LSTM model; E-G cointegration test; Granger causality test

1. Introduction

1.1 Background

As an important part of my country's financial market, the stock market, government bond market and corporate bond market play a positive role in promoting my country's economic and social development. The stock market is a place for stock issuance and trading. Joint-stock companies raise capital for enterprises by issuing stocks. At the same time, investors in the society seek to increase their wealth by buying and selling stocks. A large amount of capital flows into the market and into joint-stock companies, which promotes capital concentration, improves the organic composition of the company, and promotes the development of the commodity economy. The stock market has the functions of accumulating capital, transferring capital, transforming capital and stock prices, regulating the macro-economy, being the "barometer" and "weather station" of the national economy, and playing a pivotal and core role in the market mechanism. As the saying goes, "everything is established in advance, if not foreseen, it will be abandoned". For stock investors, there is a direct

connection between the prediction of the changing trend of the stock market and the acquisition of profits. The more accurate the prediction is, the more certain the risk prevention will be. For companies, the stock index reflects the company's operating conditions and future development trends, and affects the benefits of the entire company. It is the main technical indicator for analyzing and studying the company. For the country's economic development, the stock forecast research also plays an important role, so the research on the intrinsic value and forecast of the stock market is of great importance. theoretical significance and application prospects.

Since the development of my country's financial market, many scholars and professionals have continuously used new technologies to study and predict the future development trend and overall trend of stocks and bonds. In recent years, with the development and rise of computer technology, the prediction technology of financial market has become more and more mature. Affected by multiple factors such as politics, economy and human psychology, the stock market has the characteristics of huge amount of data and noisy data. It is a complex nonlinear and non-stationary system. To conduct predictive research on financial data, the first step is to de-noise nonlinear noisy financial data. In 1998, Huang proposed the signal analysis method of EMD, which has been well applied in different engineering fields. Compared with other time-frequency domain feature extraction methods, the adaptability of EMD makes it very suitable for processing financial time series data. On the other hand, traditional statistical-based models also do not match the characteristics of financial market data, making it difficult to analyze and predict. On the other hand, traditional statistical-based models also do not match the characteristics of financial market data, making it difficult to analyze and predict. Stock and bond prices change over time and are typical time series data, so LSTM, which is widely used in the field of natural language processing, is also considered as a potential technology for financial time series forecasting.

1.2 Research significance

This paper takes the stock market and bond market, an important part of China's financial market, as an example to study the price fluctuation forecast and risk transmission path of China's stock market, government bond market and corporate bond market. Financial market stability has certain practical significance. First, this paper combines the machine learning model LSTM with the EMD decomposition method. Long Short-Term Memory (LSTM) is a special time series model. Compared with traditional GNN, LSTM can avoid many repeated calculations and accumulation of data, and retain important data while simplifying the calculation process. Using LSTM to predict the decomposed data can improve the accuracy and efficiency of model prediction. It is now widely used in processing vibration signals of mechanical equipment, natural language processing and other fields. The empirical mode decomposition (EMD) method can remove the noise of financial data and is mostly used in signal processing. As a processing method in the time-frequency domain, the advantage of EMD compared with other processing methods in the time-frequency domain is that it overcomes the problem of non-adaptive basis functions. In other words, for an unknown signal, the decomposition can be started directly without pre-analysis and research. In this paper, a lightweight and accurate operation is achieved by the combination of the above two methods.

Secondly, this paper studies the applicability of the stock market data forecasting model. Considering the complexity and diversity of financial data, we need to use EMD to decompose the original data to remove noise items before making formal predictions. In addition, deep learning algorithms can extract features from a large amount of raw time series data without relying on prior knowledge, so they are very suitable for financial time series forecasting, especially LSTM networks have long-term memory due to their recurrent structure. The stock market is an environment that changes with the external environment, with strong randomness and complex inherent nonlinear relationships between different phenomena. The current time series prediction model relies on a single method to directly identify sequences and cannot fully extract complex sequences. Therefore, it is very suitable for this paper to use the EMD-LSTM combined prediction method for prediction.

1.3 Content and structure

The research content of this paper mainly includes two parts:

Part 1: Stock prediction analysis based on EMD-LSTM model. This paper uses the LSTM prediction model and the EMD-LSTM prediction model for the data test sets of the Shanghai Composite Index, SSE Government Bonds, and SSE Corporate Bonds to make price predictions. The prediction accuracy is compared to explore whether the prediction effect of the stock price has been further improved after the introduction of the EMD decomposition method, and the applicability of the EMD-LSTM model is analyzed.

The second part: Research on the risk transmission path. First, the LSTM models based on the training of the SSE Composite Index, the SSE Government Bond Index, and the SSE Corporate Bond Index will predict the above price series, and analyze the stock market, the national debt market, and the corporate bond market. The similar mechanism of price fluctuations between bond markets, and further through the cointegration test to determine whether there is a long-term equilibrium relationship between the Shanghai Composite Index, the SSE Government Bonds, and the SSE Corporate Bonds. Finally, the risk transmission path between the stock market, government bond market and corporate bond market is studied through the Granger causality test.

1.4 Research methods

1.4.1 RNN and LSTM

A recurrent neural network (RNN) generally consists of three layers: input, hidden and output, as shown in Figure 1.1:

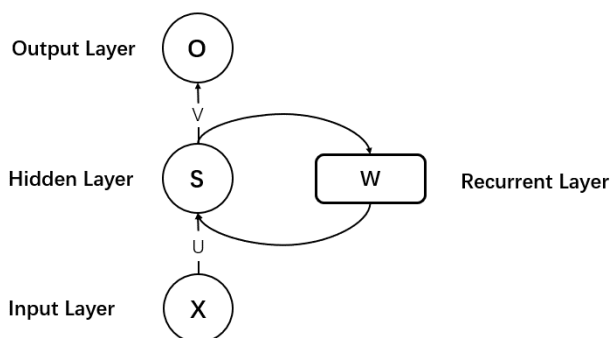


Figure 1. RNN

As shown in the figure, X_t, S_t, O_t are the input variables, hidden variables, and output variables of the neural network at time t , respectively. The reason why RNN can solve the sequence problem is that it can remember the information at each moment, and the hidden layer at each moment is not only determined by the input layer at this moment, but also determined by the hidden layer at the previous moment. The calculation formula of the hidden variable and output value at the current moment is as follows:

$$\begin{aligned} O_t &= g(V \cdot S_t) \\ S_t &= f(U \cdot X_t + W \cdot S_{t-1}) \end{aligned} \quad (1)$$

Where U is the weight matrix from the input layer to the hidden layer, V is the weight matrix from the hidden layer to the output layer, W is the weight matrix for each iteration of the hidden layer. It can be seen from the above formula that the calculation of the hidden variable at the current moment needs to use the value of the hidden variable at the previous moment, and the output at each moment retains historical information, so RNN is suitable for processing sequence data. It is worth noting that the same W is used at every moment throughout the training process.

In order to achieve accurate prediction of long-term sequences, the LSTM neural network improves the forgetting layer of the recurrent neural network from only a single layer to four interactive layers.

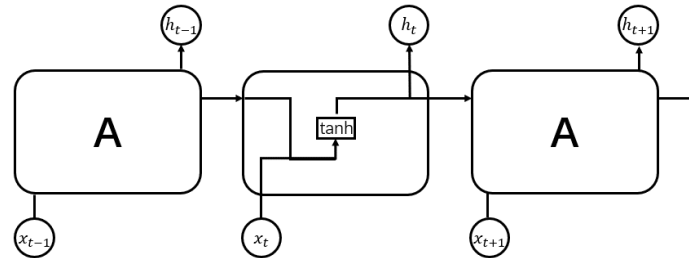


Figure 2. Simple RNN

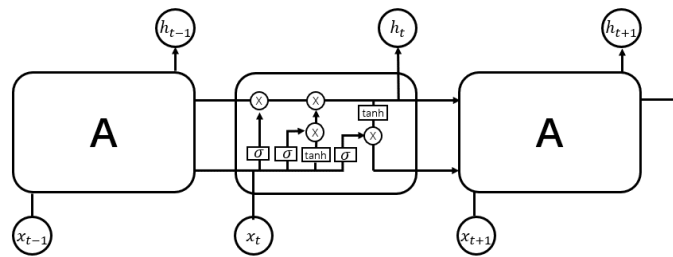


Figure 3. the structure of LSTM

LSTM introduces three gates based on simple RNN: input gate, forget gate and output gate, using memory cells to record additional information.

(1) Forget gate

The forget gate is used in the first step of LSTM to decide to discard information to determine whether the information in the memory cells of the previous time step is used for the calculation of the current time step. Splicing hidden layer output h_{t-1} with input x_t to form vector $[h_{t-1}, x_t]$. Get the vector f_t through the forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (2)$$

where σ is sigmoid function, W_f is the weight parameter of the forget gate, and b_f is the bias parameter of the forget gate.

(2) Input gate and cell state

The input gate calculates the cell state $\tilde{C}_t$ and vector i_t to be input according to the current input information.

$$\begin{aligned} \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \end{aligned} \quad (3)$$

where the vector i_t is called the input gate, which is used to control which features in $\tilde{C}_t$ are used to update C_t , i.e.

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t, \quad (4)$$

where $\times$ represents the multiplication of the corresponding elements of the vector.

(3) Output gate

First, the output gate calculates the output result o_t of the LSTM model, and then the cell state C_t determines which information in o_t is finally output as h_t .

$$\begin{aligned}o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o), \\h_t &= o_t \times \tanh(C_t),\end{aligned}\tag{5}$$

where o_t, W_o, b_o are the output vector, weight matrix and bias term of the output gate, respectively. The model output of the LSTM at time t is h_t .

1.4.2 EMD-LSTM algorithm

Wavelet analysis has good multi-scale analysis capability in both time domain and frequency domain, and can decompose time series into different levels according to different scales. The artificial neural network based on wavelet analysis combines the advantages of both and has been widely used in the analysis and research of the stock market. In recent years, EMD, as an alternative to traditional wavelet transform, has been shown by many studies to be effective in analyzing non-stationary time series. Huang et al. mainly established the basic framework of Hilbert-Huang transform: analyzed the basic basis of HHT, introduced the concept of intrinsic mode (IMF), proposed empirical mode decomposition (EMD) and continuous mean sieve method (SMS), defined the concept of Hilbert spectrum and marginal spectrum, discussed the completeness and orthogonality of HHT; compared the difference between HHT and wavelet transform and other signal analysis methods. Using the characteristic wave method of boundary processing, the application of HHT in nonlinear system analysis, water wave analysis, wind speed analysis, etc. is studied. Aiming at the modal aliasing problem in the EMD process, a solution based on the periodic scale is given. Since it was proposed, it has been widely used in the field of engineering. For example, Zhu, W. Huang, K. Huang, and Tabor have studied the application of HHT in gravitational waves, biomedicine, bridge health monitoring, and the environment, respectively.

Accordingly, this paper adopts a combined forecasting method based on Empirical Mode Decomposition (EMD) and Long Short-Term Memory (LSTM) model to forecast the price of stock data. The model consists of two parts. The first part is the EMD algorithm, which aims to decompose the prices of stocks and bonds into IMF sequences and residual sequences. Then, the IMF sequence and the residual sequence are respectively constructed as LSTM models, and finally added and summed.

2. Research Status

Stock forecast analysis and risk transmission path model are mainly used to establish a prediction model for the specific market to study the stock price or future trend, which can have an important impact on quantitative decision-making and investment strategy. It is also an important research topic for quantitative investment theory and practice.

2.1 Research status of EMD

Empirical mode decomposition (EMD) was first proposed by Huang, a Chinese American from NASA. The original purpose was to solve the noise problem in engineering signal analysis. Because the intrinsic mode function (IMF) decomposed by EMD method can well describe the characteristics of instantaneous frequency, it adapts to the characteristics of stock index time series data. Huang (2010) applied EMD method to the analysis of financial time series. Li (2012) used EMD method to decompose the data of China's stock market and analyze the influencing factors of the stock market. It is found that the low-frequency component of the closing price of the Shanghai Composite Index

is significantly positively correlated with the low-frequency component of the repurchase rate, while the exchange rate market has no significant impact on the stock market. Ji (2013) decomposed the relevant financial data of Shanghai and Shenzhen stock markets through EMD, and then analyzed the new series using cointegration test and Granger causality test. It was found that the growth rate of money and industrial added value had a significant positive impact on the SSE Index; The interest rate and exchange rate have a significant negative impact on the SSE. Fang (2015) proposed a new exchange rate forecasting method by combining EMD method with support vector regression method and achieved certain results. Li (2015) combined EMD with PCA dimension reduction method to solve the tracking portfolio problem. He studied the sample error of the portfolio and found that the tracking error was smaller after using EMD noise reduction. It can be seen that EMD has better adaptability and wider application than traditional noise reduction methods.

2.2 Research status of LSTM

LSTM model has been used in handwritten font recognition, speech recognition system, and natural language processing. Compared with the traditional recurrent neural network, LSTM has the characteristics of selective memory and interaction in time series, so it is suitable for solving the problem of stock price time series with nonstationary and randomness. In the study of combined prediction model based on LSTM model, Xiong (2015) used LSTM to investigate the impact of quantitative public sentiment factors and macroeconomic factors obtained from Google from October 19, 2004 to July 24, 2015 on the volatility of S & P 500. The results showed that the average absolute percentage error (MAPE) of the prediction results of LSTM model was 24.2%, which was at least 31% higher than the linear ridge / lasso regression and autoregressive GARCH benchmark. This shows that the results of deep learning neural network can better predict stock behavior. Borovkov and Tsiamas (2019) compared lasso regression and ridge logistic regression with LSTM's combined model to predict the daily data of multiple US stocks. The results showed that LSTM model can handle unstable phenomena and has better performance. Bukhari et al. (2020) used the LSTM model optimized by Afrima to predict the stock price, and obtained smaller model error than Arima, generalized regression radial basis network and standard LSTM model. In conclusion, LSTM and LSTM hybrid methods have made some achievements in financial time series prediction.

3. Construction of Prediction Model Based on EMD-LSTM

3.1 Data acquisition and processing

3.1.1 Data selection

This paper selects the closing price data of Shanghai Composite Index (000001) from January 1, 2016 to March 18, 2021 to construct EMD-LSTM model. Shanghai Composite Index was released on July 15, 1991. It is the first flagship index in Shanghai that only reflects the overall trend of the market. It is also the most influential index in China's capital market. It includes all listed stocks of SSE, such as A-share and B-share. It is weighted by the total capital stock. It represents the 20-year development history of China's capital market and is the symbol of China's capital market (Li, Z.M.,2013). The index sample of Shanghai Composite Index is qualified stocks and depository receipts listed on SSE, which reflects the overall performance of listed companies on SSE. The specific conditions are as follows: the 10 securities with the highest daily average total market value since listing are included in the index calculation three months after listing, and other securities are included in the index calculation one year after listing. If risk warning measures are applied, the sample will be removed from the sample range from the next trading day, that is, the second Friday of the next month. The same is true for the samples after the risk warning measures are cancelled. The base date of the index is December 19, 1990, and the base point is 100.

3.1.2 Data acquisition

This paper obtains the data of Shanghai Composite Index, SSE Government Bond Index and SSE Corporate Bond Index from CEIC economic database. The time series diagrams of the data are shown in Fig. 3.1, FIG. 3.2, and Fig. 3.3 respectively. It can be seen from Fig. 3.1 that the overall trend of the SSE Index in 2016-2017 is a slow upward trend, while in 2018, it shows a downward trend. After recovering to a certain level in the first quarter of 2019, it shows a stable trend. Around July 2020, the closing price of the SSE Index rose rapidly, and it maintained a gentle upward trend in the second half of the year. It can be seen from Fig. 3.2 and Fig. 3.3 that the SSE Government Bond Index and the SSE Corporate Bond Index showed a steady upward trend from 2016 to 2021. The training set of this article is from January 1, 2016 to December 31, 2019, and the test set is from January 1, 2020 to March 18, 2021.

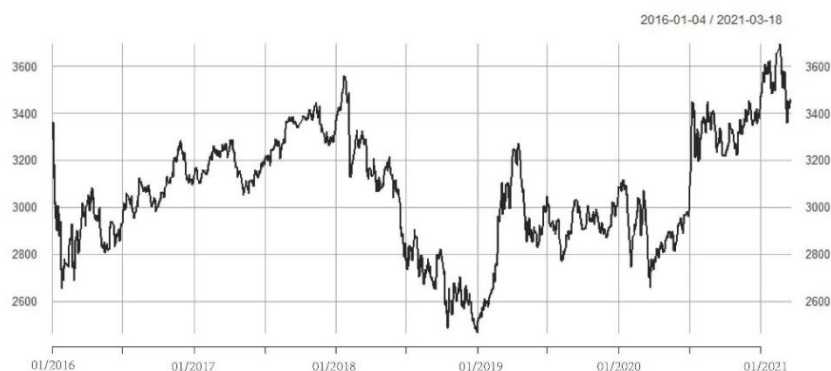


Figure 4. Time series graph of Shanghai Composite Index from 2016 to 2021

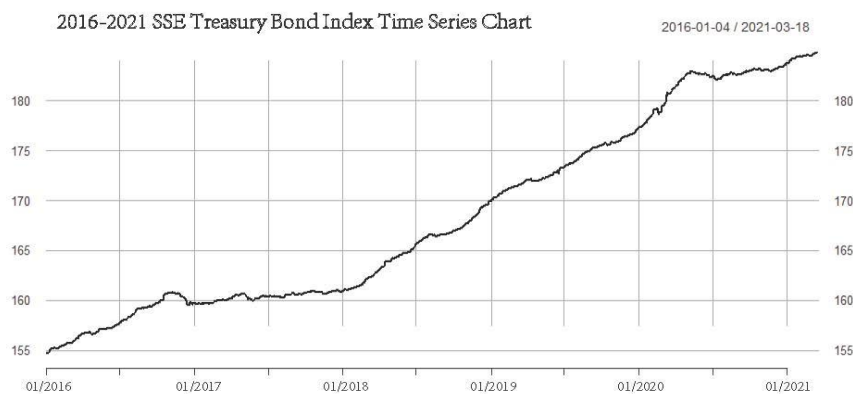


Figure 5. 2016-2021 SSE Government Bond Index time series graph

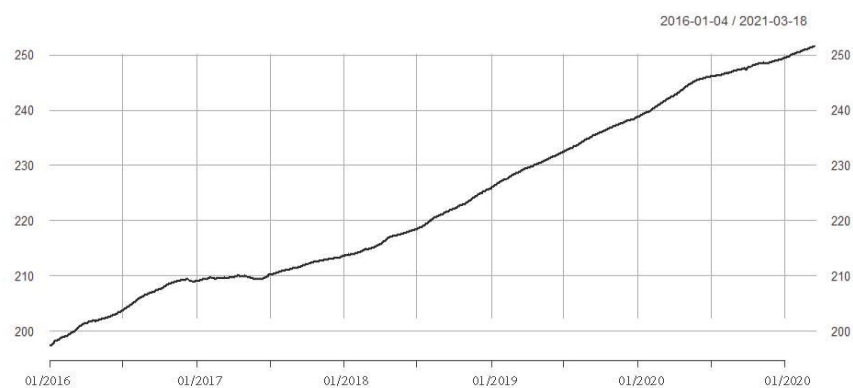


Figure 6. 2016-2021 SSE Corporate Bond Index time series graph

3.1.3 Data processing and EMD decomposition

This paper uses EMD function in EMD package of R language to decompose. After EMD decomposition of data, six IMF values and one residual value are obtained. See Figure 3.4 for EMD decomposition results. Each IMF decomposed is independent and orthogonal to each other, but the frequency range is different. The frequency of IMF1 is the highest, reflecting the characteristics of high-frequency fluctuations. The frequency of the last IMF value is the lowest, reflecting the long-term trend of the series. Residual is the remaining part of the original sequence excluding all IMF components, which is actually the average trend part of the original sequence. By comprehensively comparing the EMD decomposition results of Shanghai Composite Index, SSE Government Bond Index and SSE Corporate Bond Index, from the perspective of high-frequency fluctuation characteristics, the extreme risk of "peak" of Shanghai Composite Index is large, and the extreme risk of SSE Government Bond Index and SSE Corporate Bond Index is small. From the long-term trend, the SSE Index shows the characteristics of periodicity and gradual increase in fluctuation range, while the Shanghai Stock Exchange Government Bond Index has a long fluctuation period, and the fluctuation range has hardly changed. The fluctuation range of the SSE Corporate Bond Index gradually decreases. From the perspective of the average trend, the Shanghai Composite Index shows the characteristics of periodic fluctuation, while the SSE Government Bond Index and the SSE Corporate Bond Index both show the characteristics of stable rise. It can be seen that there is certain extreme risk in the short-term fluctuation of the stock market price, the long-term trend also tends to fluctuate cyclically and the average trend will fluctuate cyclically; The extreme risk of the short-term fluctuation of the government bond market price is small, and the long-term trend is periodic fluctuation, but the fluctuation range remains unchanged, and the average trend rises steadily; The extreme risk of short-term fluctuation of corporate bond market price is small, and the long-term trend tends to be periodic and small fluctuation, and the average trend also rises steadily.

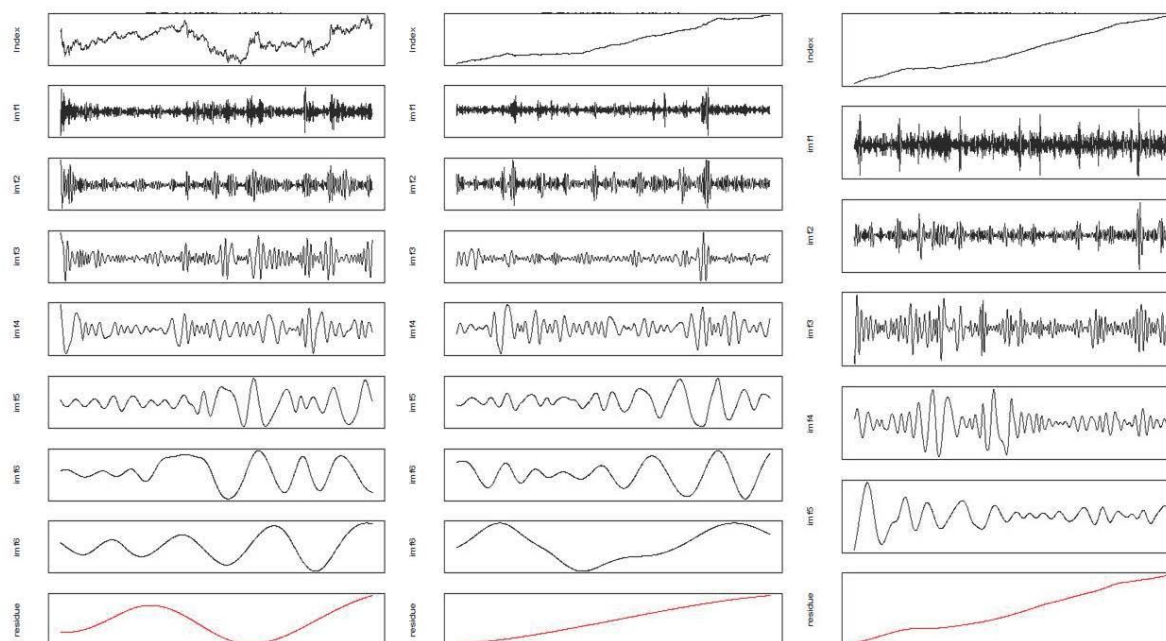


Figure 7. EMD decomposition results

3.2 Construction of LSTM and EMD-LSTM prediction models

3.2.1 Construction of prediction models of Shanghai Composite Index

This paper first uses the LSTM prediction model on the test set of the Shanghai Composite Index data, and the fitting results are shown in the figure below. We can see that LSTM can reflect the overall trend of the Shanghai Composite Index to a certain extent, but in some special cases (such as

February 2020 and March 2021), the forecast results will deviate from the actual situation to a certain extent.

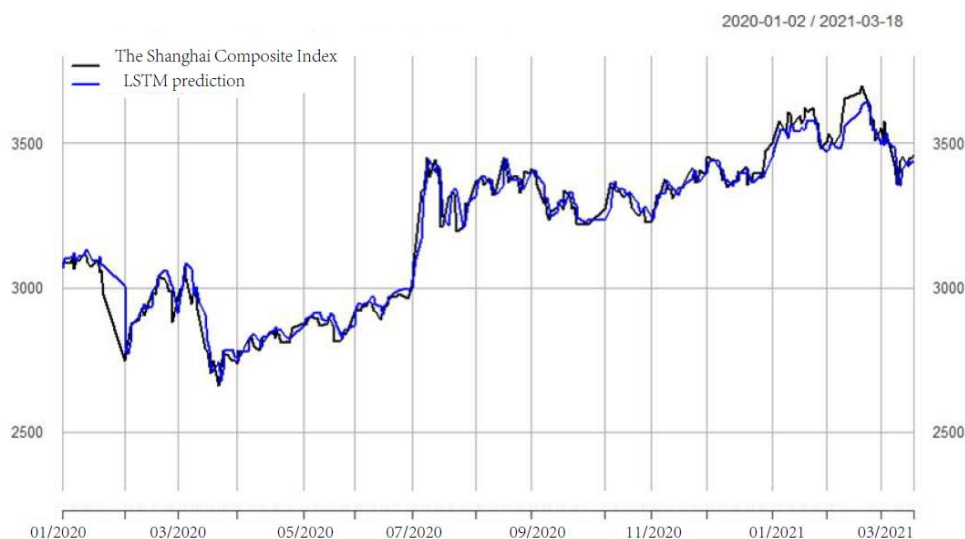


Figure 8. LSTM model (based on SSE Index training) SSE Index test set prediction results time series graph

The fitting results of LSTM model of each decomposition value of Shanghai Composite Index after EMD decomposition are shown in the following figure. The predicted values of other decomposition values are highly fitted to the original values except that IMF1 is a constant, which indicates that the LSTM model of the seven decomposition values from IMF2 to residual has a good prediction effect.

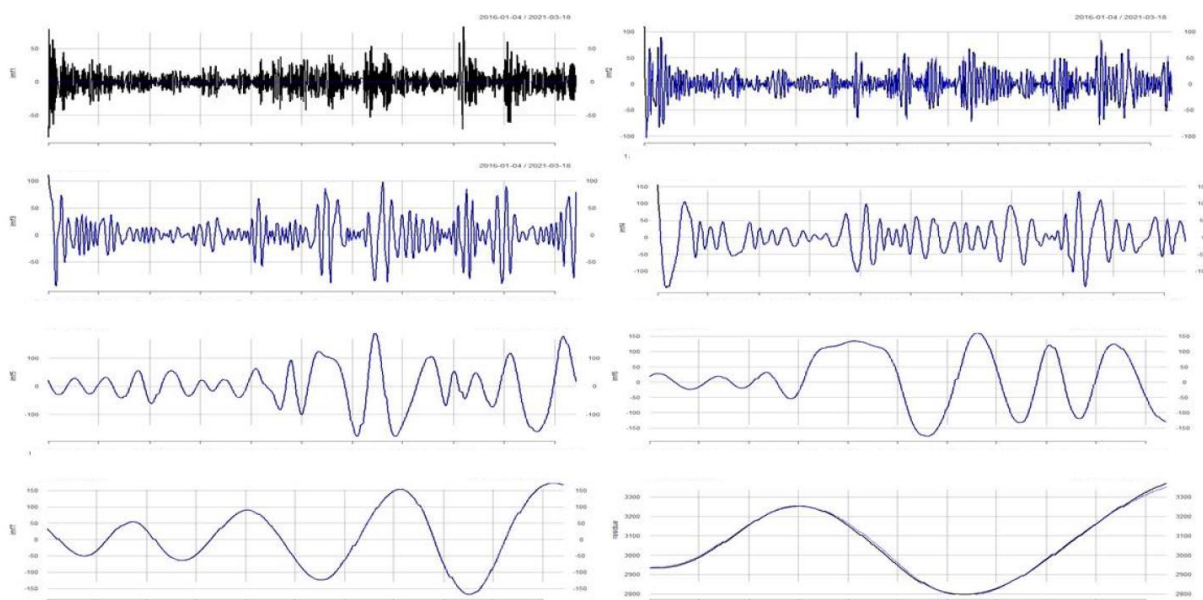


Figure 9. Time series graph of EMD decomposition value prediction results of Shanghai Composite Index

The fitting results of Shanghai Composite Index EMD-LSTM model test set are shown in the following figure. It can be seen that EMD-LSTM has a good prediction result on the future trend of SSE Index, and the EMD decomposition method decomposes the data into more stable results than using LSTM alone, which makes up for the shortcomings of LSTM.

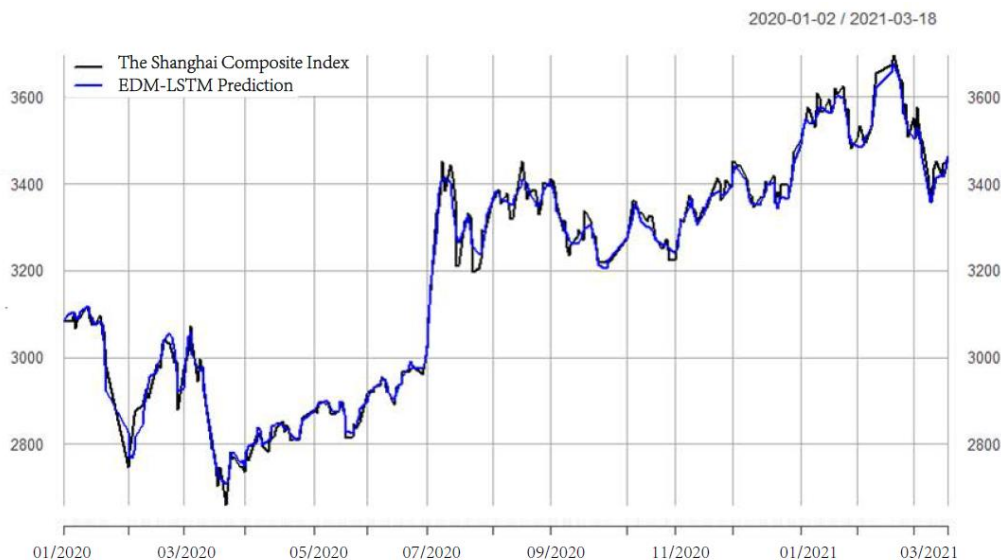


Figure 10. Prediction results time series graph of EMD-LSTM model based on SSE test set

Comprehensively analyze the prediction effect of Shanghai Composite Index LSTM model and EMD-LSTM model. The RMSE and MAPE of the test set prediction results of the two models are shown in the following table. The fitting results of the two test sets are shown in the following figure. The root mean square error and average absolute percentage error of EMD-LSTM model are smaller than that of LSTM model, that is, EMD-LSTM model has better prediction effect on Shanghai Composite Index data than LSTM model.

Table 1. Prediction accuracy table of Shanghai Composite Index Prediction Model

Prediction model of Shanghai Composite Index	RMSE	MAPE
LSTM model	43.1083	0.9731
EMD-LSTM model	23.4046	0.5566

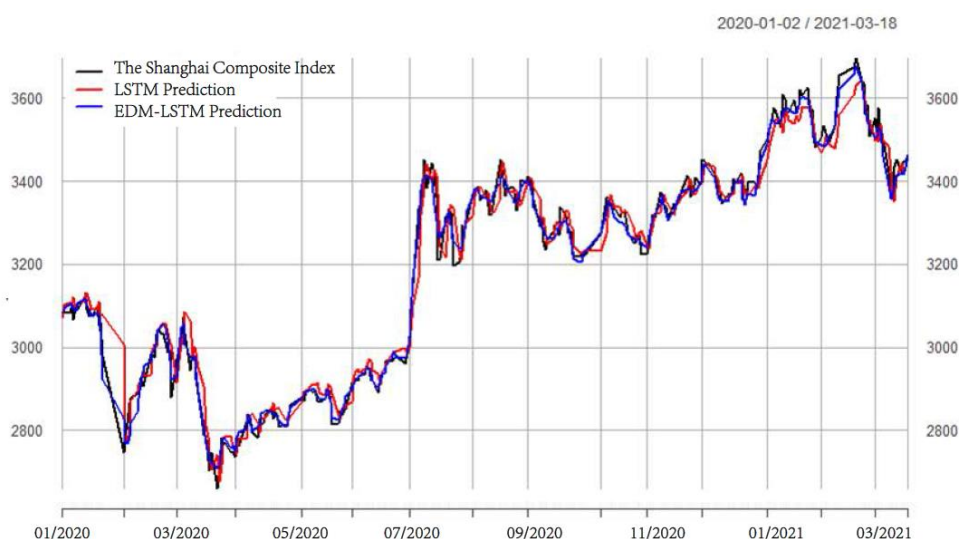


Figure 11. Comparison time series graph of prediction results of SSE test set

3.2.2 Construction of prediction models of SSE Government Bond Index

In this section, the LSTM prediction model is used for the test of the data of SSE Government Bond Index. The fitting results are shown in the following figure. It can be seen that the time series

curve of the predicted value basically coincides with the original time series curve, but there is a large difference between January 2020 and March 2020 and between January 2021 and March 2021.

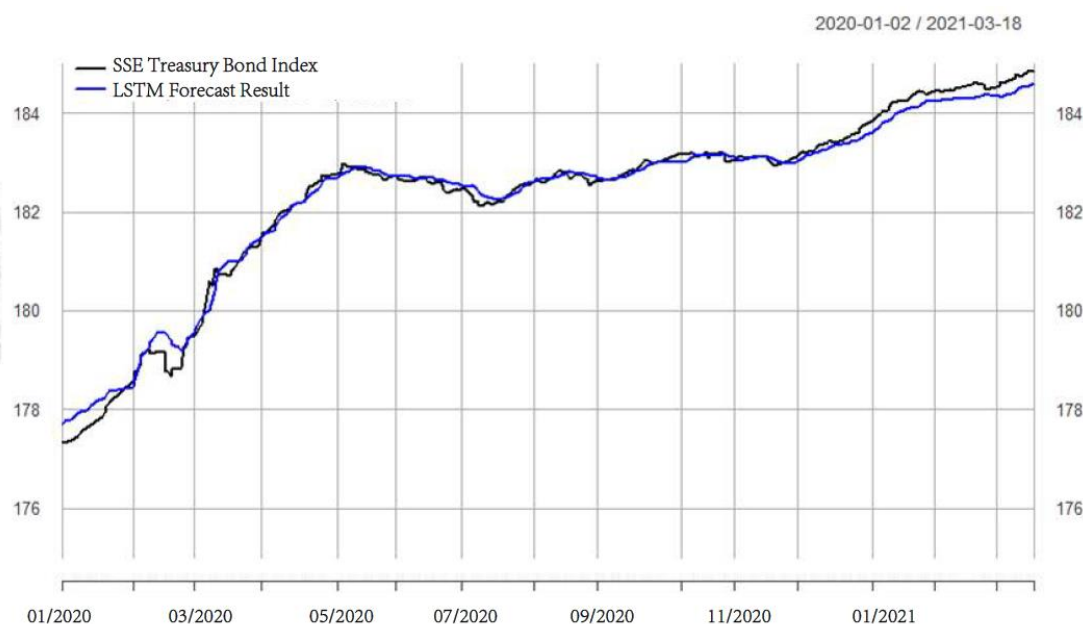


Figure 12. Time series graph of prediction results of LSTM model (based on the training of national debt index) test set of SSE Government Bond Index

The LSTM model fitting results of the EMD decomposition values of the SSE Government Bond Index are shown in the following figure. It can be seen that the predicted values of the LSTM models of IMF1 to IMF6 are all constants, indicating that these decomposition values are closer to the white noise sequence. The predicted values of IMF7 and residual highly fit the original values, and the prediction effect is good.

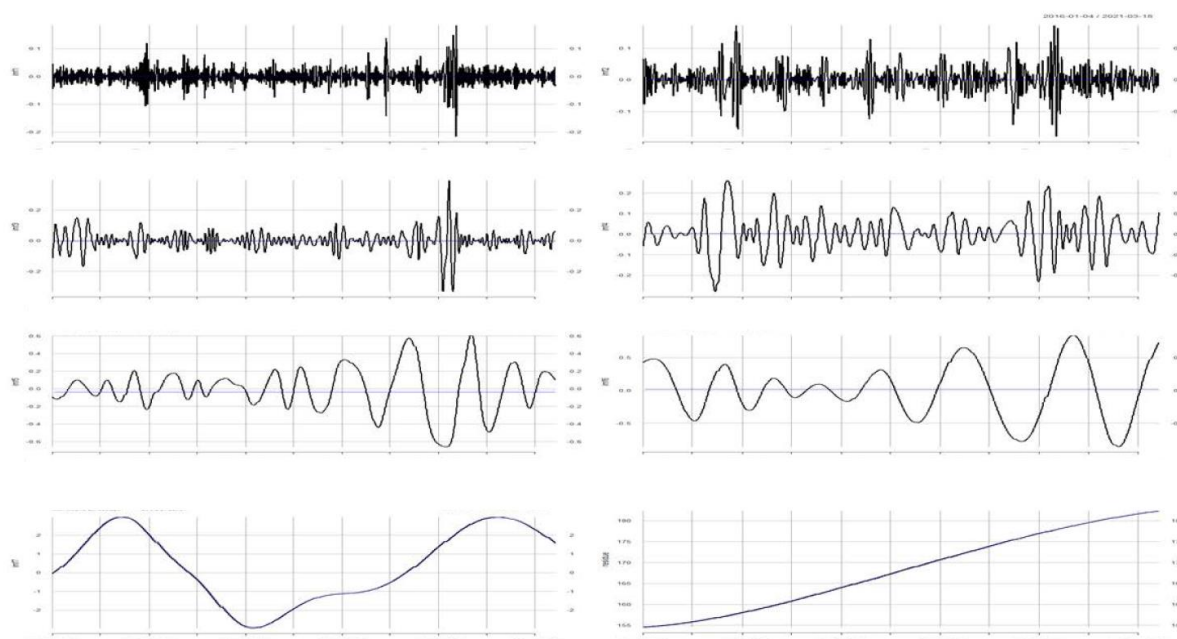


Figure 13. Time series graph of forecast results of EMD decomposition value of SSE Government Bond Index

The RMSE and MAPE of the EMD-LSTM model and the test set prediction results of LSTM model of Shanghai Stock Exchange government bond index are shown in the following table. The fitting results of the test sets of the two models are shown in the following figure. It can be seen that the RMSE and MAPE values of LSTM model are far less than those of EMD-LSTM model for the SSE Government Bond Index, and the prediction results of EMD-LSTM model are relatively smooth curves, which are far from the original sequence curves, that is, the prediction effect of LSTM model is better than that of EMD-LSTM model.

Table 2. Prediction accuracy table of prediction model of SSE Government Bond Index

Prediction model of SSE Government Bond Index	RMSE	MAPE
LSTM model	0.1915	0.0779
EMD-LSTM model	0.7108	0.3393

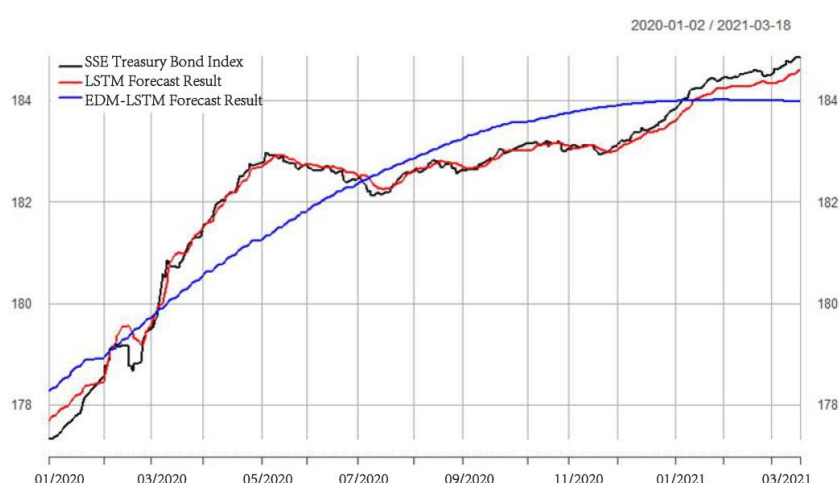


Figure 14. Comparison time series graph of prediction results of SSE Government Bond Index test set

3.2.3 Construction of prediction models of SSE Corporate Bond Index

In this section, the LSTM prediction model is used for the test of SSE Corporate Bond Index data. The fitting results are shown in the following figure. It can be seen that the prediction effect is similar to that of the SSE Government Bond Index. The time series curve of the predicted value basically coincides with the original time series curve, but there is a big difference between January 2020 and March 2020, and between January 2021 and March 2021.

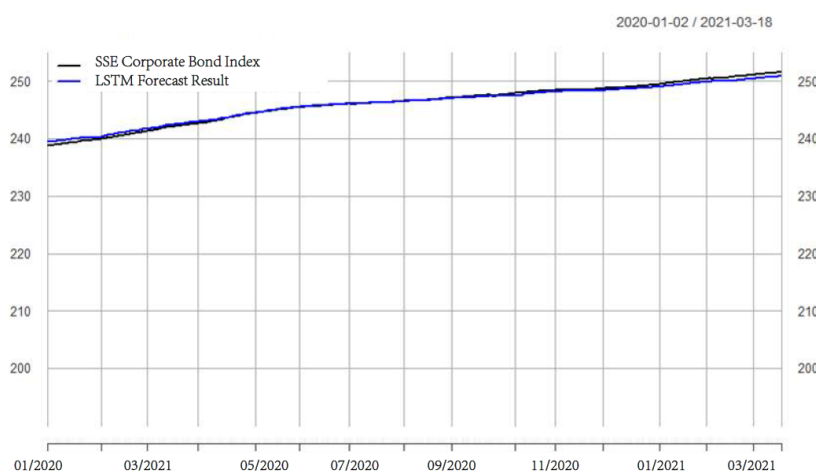


Figure 15. LSTM model (trained based on corporate debt index) time series graph of prediction results of SSE Corporate Bond Index test set

The LSTM model fitting results of EMD decomposition values of Shanghai corporate bond index are shown in the following figure. The LSTM model prediction values of IMF1 to IMF5 are all constant, indicating that these decomposition values are closer to the white noise sequence. The residual prediction value is highly fitted to the original value, and the prediction effect is good.

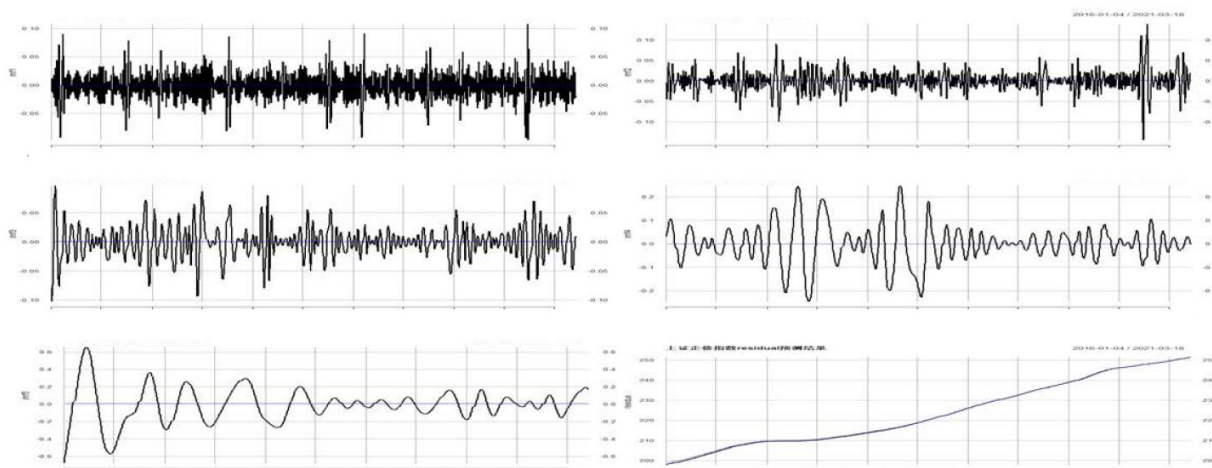


Figure 16. Time series graph of EMD decomposition value prediction results of SSE Corporate Bond Index

The RMSE and MAPE of the predicted results of the EMD-LSTM model and the test set of LSTM model of SSE Corporate Bond Index are shown in the following table. The fitting results of the test sets of the two models are shown in the following figure. It can be seen that both the root mean square error and the average absolute percentage error of EMD-LSTM model are smaller than that of LSTM model, and the EMD-LSTM prediction sequence curve between January 2020 and March 2020 and between January 2021 and March 2021 is obviously more consistent with the original sequence curve, that is, the EMD-LSTM model has better prediction effect on the data of Shanghai corporate debt index than LSTM model.

Table 3. Prediction accuracy table of SSE Corporate Bond Index prediction model

Prediction model of SSE Corporate Bond Index	RMSE	MAPE
LSTM model	0.3527	0.1146
EMD-LSTM model	0.1329	0.0449

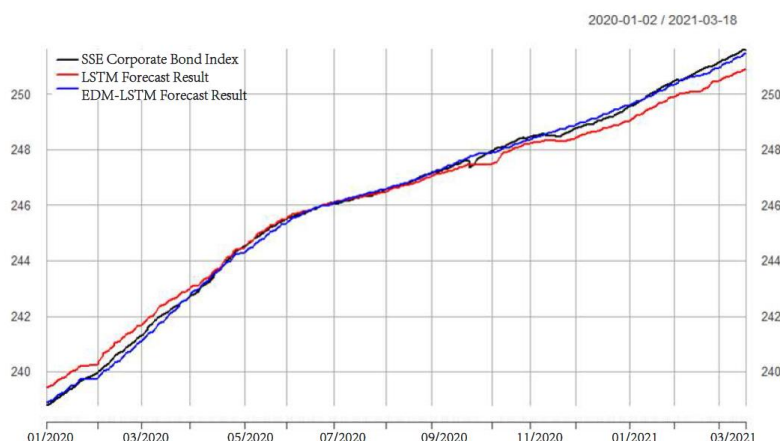


Figure 17. Time series graph for comparison of prediction results of SSE Corporate Bond Index test set

3.3 Result Analysis

Through comprehensive analysis of the prediction results of Shanghai Composite Index, SSE Government Bond Index and SSE Corporate Bond Index, it can be seen from the table below that the prediction results of the prediction models of Shanghai Composite Index and SSE Corporate Bond Index have been greatly improved after the introduction of EMD method, but the prediction results of SSE Government Bond Index models are far less than those of LSTM model. According to the analysis of EMD decomposition diagram, the introduction of EMD decomposition method will improve the prediction accuracy of LSTM model for the time series with short fluctuation period of long-term trend; For the time series with long fluctuation period of long-term trend, the introduction of EMD decomposition method will reduce the prediction accuracy of LSTM model.

Table 4. Prediction accuracy table of prediction models

Prediction models	Modeling method	RMSE	MAPE
Prediction models of Shanghai Composite Index	LSTM	43.1083	0.9731
	EMD-LSTM	23.4046	0.5566
Prediction models of SSE Government Bond Index	LSTM	0.1915	0.0779
	EMD-LSTM	0.7108	0.3393
Prediction models of SSE Corporate Bond Index	LSTM	0.3527	0.1146
	EMD-LSTM	0.1329	0.0449

4. Research on Risk Analysis and Transmission Path of Financial Market

4.1 Risk analysis of volatility of index price and yield

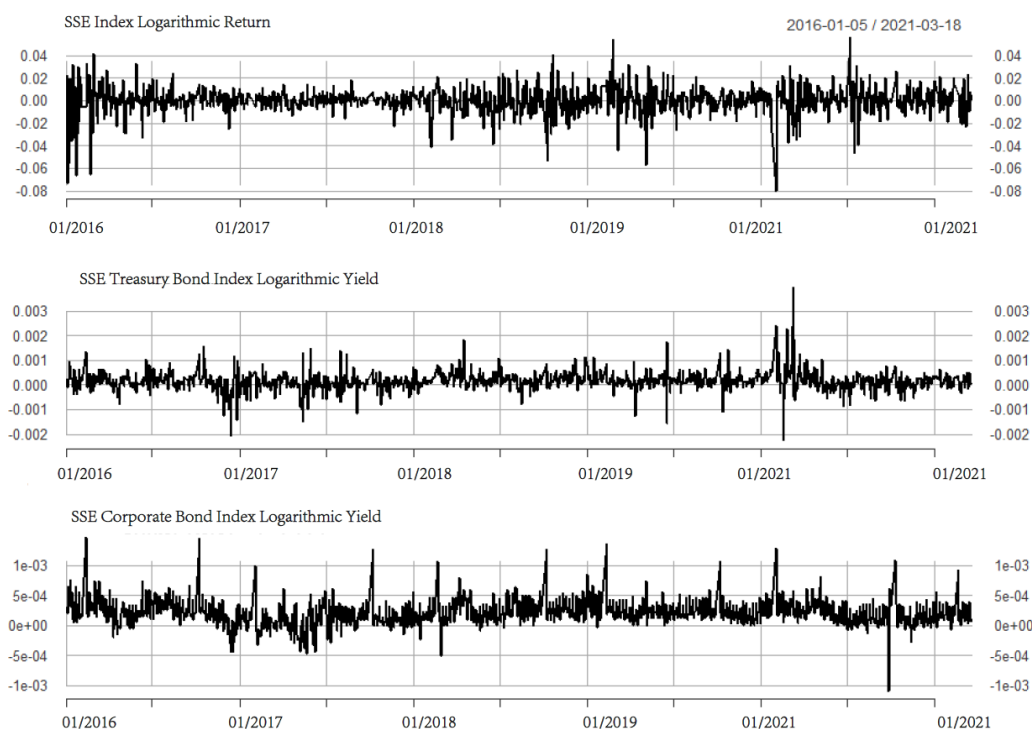


Figure 18. Time series graph of logarithmic yield of Shanghai Composite Index, SSE Government Bond Index and SSE Corporate Bond Index

Using function *ur.df* in R software to test the unit root of the logarithm return rate of the three sequences, the results show that the absolute value of t statistic is greater than the absolute value of the critical value at the 1% significance level, indicating that the sequence is stable without unit root.

Table 5. Unit root test result of logarithmic rate of return

Time series	T-Statistic	1pct	5pct	10pct
Logarithmic yield of Shanghai Composite Index	-24.2706	-2.58	-1.95	-1.62
Logarithmic yield of SSE Government Bond Index	-17.7908	-2.58	-1.95	-1.62
Logarithmic yield of SSE Corporate Bond Index	-11.7628	-2.58	-1.95	-1.62

4.2 Similarity analysis of price fluctuation mechanism

The LSTM model based on the training of Shanghai Composite Index, the training of SSE Government Bond Index and the training of SSE Corporate Bond Index is used to predict the three price series. The time sequence diagram of the prediction results is shown in Figure 4.2 below, and the RMSE value and MAPE value of the prediction results are shown in the following table.

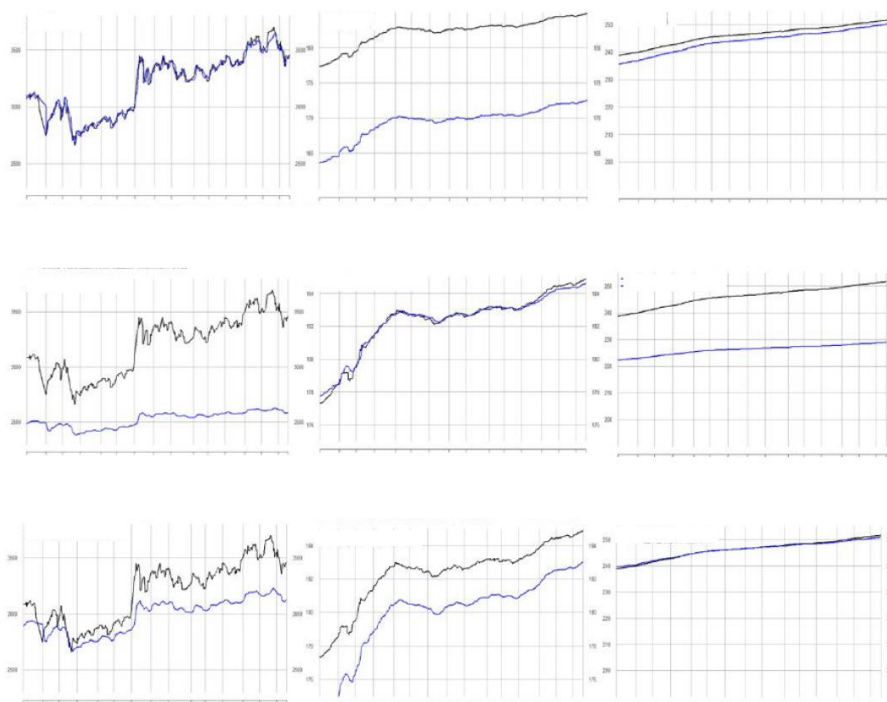


Figure 19. Time series graph of prediction results of three LSTM models based on different training data

Table 6. Prediction accuracy table of three LSTM models based on different training data

Model	Prediction series	RMSE	MAPE
Prediction model of Shanghai Composite Index	Shanghai Composite Index	43.1083	0.9731
	SSE Government Bond Index	12.8396	7.0409
	SSE Corporate Bond Index	2.22817	0.8867
Prediction model of SSE Government Bond Index	Shanghai Composite Index	702.6217	20.7161
	SSE Government Bond Index	0.1915	0.0779
	SSE Corporate Bond Index	20.0778	8.1202
Prediction model of SSE Corporate Bond Index	Shanghai Composite Index	246.3909	6.6064
	SSE Government Bond Index	2.3885	1.2912
	SSE Corporate Bond Index	0.1329	0.0449

For the Shanghai Composite Index, the prediction result of LSTM model (trained based on corporate bond data) is significantly better than that of LSTM model (trained based on national bond data), which to some extent indicates that the price fluctuation mechanism of the stock market may be more similar to that of the corporate bond market. For the SSE Government Bond Index, the prediction results of LSTM model (trained based on corporate bond data) are significantly better than those of LSTM model (trained based on comprehensive index data), which to some extent indicates that the price fluctuation mechanism of the bond market may be more similar to that of the corporate bond market. For the SSE Corporate Bond Index, the prediction result of LSTM model (trained based on the comprehensive index data) is obviously better than that of LSTM model (trained based on the national debt data), which to some extent indicates that the price fluctuation mechanism of the corporate bond market may be more similar to that of the stock market. Therefore, for the similarity of the price fluctuation mechanisms of the three markets, we can draw a relationship chart (see Figure 4.5).

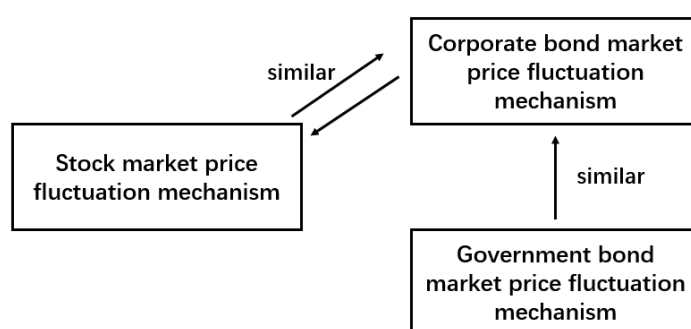


Figure 20. Results of price fluctuation mechanism similarity analysis

4.3 Research on equilibrium relationship of financial market based on cointegration test

According to the economic theory, there is indeed a long-term equilibrium relationship between some economic variables. This equilibrium relationship means that there is no internal mechanism to destroy the equilibrium in the economic system. If the variable deviates from its long-term equilibrium point after being disturbed in a certain period, the equilibrium mechanism will be adjusted in the next period to make it return to the equilibrium state (Luo, 2011). Cointegration test can test the equilibrium relationship between variables and determine whether the linear combination equilibrium relationship composed of a group of non-stationary time series is stable. The commonly used test methods include Engle Granger test and Johansen test. This paper uses E-G cointegration test to test the cointegration relationship among stock market, government bond market and corporate bond market. The basic steps are as follows: firstly, the stationarity test is carried out on the variables to determine that the variables are single integration processes with the same order; Then the cointegration regression equation is established; Finally, the cointegration relationship is judged according to whether the residual sequence is stable.

The inspection results are shown in table 4.5 below. The test results show that the P value of ADF test is less than 0.5 at the significance level of 5%, rejecting the original hypothesis, indicating that the sequence is stable, that is, there is a cointegration relationship between the Shanghai Composite Index and the SSE Government Bond Index, and between the SSE Government Bond Index and the SSE Corporate Bond Index.

Table 7. Cointegration Test Results Table

Cointegration data	Pseudo regression			P value of residual ADF test		
	coefficient	t-value	P-value	Lag=1	Lag=2	Lag=3
Shanghai Composite Index & SSE Government Bond Index	-2.0138	-2.369	0.018	0.01	0.01	0.01
Shanghai Composite Index & SSE Corporate Bond Index	1.0776	0.682	0.496	0.01	0.01	0.01
SSE Government Bond Index & SSE Corporate Bond Index	0.6787	13.968	<2e-16	0.01	0.01	0.01

4.4 Research on risk transmission path based on Granger causality test

If there is a cointegration relationship between variables, Granger causality test can be performed. If the prediction effect of y is better under the condition of using the past data of two variables x and y than using only y data, then variable x is the Granger cause of variable y .

Granger test requires estimation of the regression model $y_t = \beta_0 + \sum_{i=1}^p \beta_i y_{t-i} + \sum_{i=1}^q \alpha_i x_{t-i} + \mu_t$,

where p and q are Hysteresis coefficients. *grangertest* function in *lmtest* package of R language can be used to perform *grangertest* test.

In Section 4.3, we have proved that there is a cointegration relationship between the SSE Index and the SSE Government Bond Index, and between the SSE Government Bond Index and the SSE Corporate Bond Index. In Section 4.1, we have tested that the logarithmic yield series is a stable series. In this section, we can directly perform Granger causality test analysis on the log return series. Based on AIC and SC criteria, the lag order is selected as 2. The inspection results are shown in table 4.6 below.

Table 8. Granger causality test results

Original hypothesis	F-Statistic	P-value
SSE Government Bond Index accounts for the change of Shanghai Composite Index	2.339	0.09685
Shanghai Composite Index accounts for the change of SSE Government Bond Index	1.8052	0.1649
SSE Corporate Bond Index accounts for the change of SSE Government Bond Index	3.9149	0.02019
SSE Government Bond Index accounts for the change of SSE Corporate Bond Index	7.4664	0.0005977

It can be seen from the test results that, at a significant level of 5%, we reject the hypothesis that the SSE Government Bond Index is the reason for the change of the Shanghai Composite Index and the Shanghai Composite Index is the reason for the change of the SSE Government Bond Index, and accept the hypothesis that the SSE Government Bond Index is the reason for the change of the Shanghai Corporate Bond Index and the Shanghai Corporate Bond Index is the reason for the change of the SSE Government Bond Index. In other words, in the long run, there is a two-way causal relationship between the SSE Index and the Shanghai Corporate Bond Index. The risk transmission mechanism between the government bond market and the corporate bond market is strong, while the risk transmission mechanism between the stock market and the bond market is weak.

5. Conclusion and Policy Recommendations

In this paper, EMD-LSTM model is applied to predict the index price based on the data of Shanghai Composite Index, SSE Government Bond Index and SSE Corporate Bond Index, aiming at the three important components of China's financial market: stock market, government bond market and corporate bond market. Compared with the results of using LSTM model only, it can be seen that EMD-LSTM model does not have good results in any case, That is, not all financial time series data are suitable for decomposition into different levels according to different scales. For the time series with short fluctuation period of long-term trend, the introduction of EMD decomposition method will improve the prediction accuracy of LSTM model. However, for the time series with long fluctuation period of long-term trend, the introduction of EMD decomposition method will reduce the prediction accuracy of LSTM model.

The LSTM models trained based on Shanghai Composite Index, SSE Government Bond Index and SSE Corporate Bond Index is used to predict the three price series respectively, and the similarity of price fluctuation among the stock market, corporate bond market and government bond market is analyzed. It is found that the price fluctuation mechanism of corporate bond market is more similar to that of the stock market to a certain extent. Further, the traditional statistical measurement method is used for step-by-step analysis. The long-term equilibrium of the stable logarithmic yield series is analyzed by cointegration, and then the risk transmission path between China's financial sub markets is obtained according to the causality test. We found that the risk transmission mechanism between the Treasury bond market and the corporate bond market is strong, while the risk transmission mechanism between the stock market and the bond market is weak. Using the EMD-LSTM portfolio model to predict the index price, analyze the complex causal relationship between the stock market, the government debt market, and the corporate bond market, and study the risk transmission path of the financial market will help investors adjust their investment portfolio in time to reduce investment risk, and better grasp the operation law of China's financial market, providing a certain basis for the monitoring and supervision of cross market risks.

EMD method has a wide range of practical application value, which is helpful to promote the in-depth research of prediction and decision-making issues in the financial field of China, and also can greatly promote the further development of prediction theory. While analyzing the future trend or fluctuation of the stock price in the whole market can stabilize the future state of the national macro-economy at large, and stabilize the economic situation of small and medium-sized investors, especially retail investors, In short, the healthy and steady development of the stock market is a good thing for the country and the people.

Based on the above empirical results, the main recommendations of this paper are as follows:

First, in view of China's stock market, the government should strengthen supervision, carry out more regulatory measures, give full play to the role of the securities regulatory authorities in maintaining the dynamic balance of the stock market, rationally use monetary and fiscal policies, and improve stock credit trading and stock index futures trading.

Second, it is suggested to improve the linkage mechanism between the stock, government bond and corporate bond markets. For example, improve the information exchange and sharing mechanism to ensure the effective transmission of price, transaction, order and other information, and reduce the cross-market risk transmission caused by information asymmetry, market manipulation and other behaviors.

Third, in view of the risk transmission path of the financial market, it is necessary to establish and improve risk monitoring and early warning, adopt more targeted risk control measures, use the policy tool of "monetary policy + capital control" rationally, optimize the investor structure, reduce the harm of risk transmission to the market, and constantly promote the healthy development of the capital market.

Fourth, it is suggested that investors should combine risk transmission analysis based on causality test with portfolio management, grasp the dynamic nature of risk transmission structure among

financial markets, and adjust their portfolios in time according to the risk transmission path, so as to effectively reduce risks.

Reference

- [1] Bian, L., & Lili, Q. (2015). Study on the frequency forecast method of hydrogen maser based on EMD. 2015 12th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), 01, 164-168.
- [2] Borovkova, S., & Tsiamas, I. (2018). An Ensemble of LSTM Neural Networks for High-Frequency Stock Market Classification. *Capital Markets: Asset Pricing & Valuation eJournal*.
- [3] Bukhari, A.H., Raja, M.A., Sulaiman, M., Islam, S., Shoaib, M., & Kumam, P. (2020). Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting. *IEEE Access*, 8, 71326-71338.
- [4] Ding, Y. (2018). Research on the relationship between China's stock market volatility and economic growth cycle -- based on linear and nonlinear Granger causality test. *Jiangsu Social Sciences*. 03, 175-182.
- [5] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1007.
- [6] Fama, E.F. (1970). EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK. *Journal of Finance*, 25, 383-417.
- [7] Fan, Z.H., & Han, Y. (2021). Research on the Relationship between Economic Development and Resident Savings in Yunnan Province——Based on Cointegration Analysis and Granger Causality Test. *China Economic & Trade Herald*. 46-47.
- [8] Fang, Z.B., & Li, Y. (2012). Denoising and prediction of exchange rate data based on HHT-SVR Model. *Journal of Applied Statistics and Management*, 05, 900-909.
- [9] Hochreiter, S., & Schmidhuber, J. (1997). Long Short -- Term Memory. *Neural Computation*, 9, 1735-1780.
- [10] Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N., Tung, C.C., & Liu, H.H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454, 903 - 995.
- [11] Huang, N.E., Wu, M.C., Qu, W., Long, S.R., & Shen, S.S. (2003). Applications of Hilbert–Huang transform to non-stationary financial time series analysis. *Applied Stochastic Models in Business and Industry*, 19, 245-268.
- [12] Ji, C.L. (2013). Macroeconomics factors of stock market fluctuates and prediction of stock market price. (MA thesis, Dongbei University of Finance and Economics)
- [13] Ji, Y., Liew, A., & Yang, L. (2021). A Novel Improved Particle Swarm Optimization With Long-Short Term Memory Hybrid Model for Stock Indices Forecast. *IEEE Access*, 9, 23660-23671.
- [14] Lev, G., Sadeh, G., Klein, B., & Wolf, L. (2016). RNN Fisher Vectors for Action Recognition and Image Annotation. *ECCV*.
- [15] Li, Y. (2013). Multiresolution analysis of financial time series-Basing on Hilbert-Huang Transform (PhD dissertation, University of Science and Technology of China)
- [16] Li, Z.M. (2013). Study on co-movement between Shanghai and world stock markets. (MA thesis, Nanjing University of Finance & Economics)
- [17] Luo, J.H. (2011). A study of the influence of money supply upon China's agricultural prices. (MA thesis, Huazhong Agricultural University)
- [18] Xiong, R., Nichols, E.P., & Shen, Y. (2015). Deep Learning Stock Volatility with Google Domestic Trends. *arXiv: Computational Finance*.
- [19] Yang, H., & Lin, H. (2017). Applying the Hybrid Model of EMD, PSR, and ELM to Exchange Rates Forecasting. *Computational Economics*, 49, 99-116.
- [20] Zhang, Y.F. (2020). Prediction and analysis of Shenzhen Composite Index Based on LSTM neural network. *Business News*, 36, 81-82.