The Research of Stock Index Futures Price Forecasting Using DFA-LSTM Model Based on Panic Index

Shuo Bian

20120494@bjtu.edu.cn

Beijing Jiaotong University, No.3 Shangyuan Village, Haidian District, Beijing, China

Abstract. With the rapid development of artificial intelligence, it is possible to use machine learning models to predict financial market prices more accurately. Based on the analysis of the efficiency and fractal characteristics of the Chinese stock index futures market, combined with the "panic index", this paper established an LSTM neural network model to predict the price of the CSI 300 stock index futures. The results show that: (1) Stock index futures market has long memory, and the strength of memory has time-varying characteristics; (2) Combining LSTM neural network with fractal method and panic index, the prediction accuracy is greatly improved; (3) Compared with the GARCH (1,1) model, the model constructed in this paper has better prediction effect.

Keywords: Stock index futures; Fractal market; The VIX index; LSTM neural network

1 Introduction

With the advent of the post-pandemic era, along with a series of unstable factors such as the Russia-Ukraine conflict, the global energy crisis, and the global spread of inflation, finding a more effective way to hedge financial assets has become a more concerned issue for all kinds of investors. As a professional risk management market, futures market plays an important role in the continuous development and improvement of China's financial market. Investors can use stock index futures to avoid systematic risks in the stock market. At the same time, stock index futures trading can make the price of the stock market more reasonable. Therefore, it is of certain practical significance to study the price characteristics of stock index futures and find a more effective price prediction method of stock index futures.

Fama (1965)[1] put forward the efficient market hypothesis (EMH) and divided the market into three types: weak efficient, semi-strong efficient and strong efficient according to the different transmission efficiency of market prices to inside information, historical information and fundamental information. If the market is weak-form efficient, it will not be possible to use historical information to forecast prices. Fractal phenomenon was first discovered by Mandelbrot (1967)[2]. Peters (1994)[3] applied the methods of fractal geometry to the study of financial markets and proposed the Fractal Market Hypothesis (FMH). Many scholars have studied the fractal phenomenon in the futures market. Some scholars have found that market fractal characteristics can be used to predict prices. Mishra and Bhanumurthy (2011)[4] believed that the chaotic structure in the fractal stock market made it possible to predict short-term profits. Wang and Zhang (2011)[5] used the fractal interpolation model to predict the short-term trend of Shanghai Composite Index. In addition, different types of financial markets have different

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334279

speeds of response to information, showing a long-term lead-lag relationship in asset prices. Back (1993)[6] believed that options can often provide additional information for the price prediction of the underlying asset. Zheng et al. (2021)[7] proved that the net buying pressure of real-valued option trading can predict volatility. The VIX index, calculated using option prices, reflects market expectations of volatility over the next 30 days. In turn, the market's expectation of volatility will affect people's trading behavior, and due to the trading restrictions in the spot market, the demand for speculative trading or hedging trading using futures will change, and then affect the price of futures.

Traditional price prediction methods in financial markets are mainly divided into three categories: fundamental analysis, technical analysis and time series prediction, among which ARCH model proposed by Engle (1982)[8] and GARCH model proposed by Bollerslev (1986) [9]are widely used. Recently, more and more researchers have tried to apply machine learning models to the prediction of financial market prices. Yang Qing (2019)[10] predicted 30 major stock indexes in the world and found that LSTM neural network showed the highest prediction accuracy and optimal prediction stability. Matsuoka et al. (2021)[11] found that RNN model had higher prediction accuracy than AR-GARCH or VAR model.

In the existing researches on the prediction of stock index futures price, traditional econometric models are used to forecast, or basic volume-price information is combined with shallow machine learning models such as BP neural network. This paper analyzes the price characteristics of stock index futures, combines the fractal method with the LSTM neural network in machine learning to establish a deep learning model, and combines the price discovery function of financial derivatives to establish a DFA-LSTM neural network model based on the panic index, so as to achieve a more accurate prediction of the price of stock index futures.

2 Test of stock index futures market characteristics

When analyzing the characteristics of stock index futures market, this paper chooses CSI 300 stock index futures as the research object. In this paper, the contract with the largest daily turnover is selected as the main futures contract. The sample time range is from April 16, 2015 to January 19, 2023, with a total of 1,893 closing price data. The data were obtained from Flush iFind financial data terminal.

2.1 Statistical characteristics of stock index futures prices

Table 1 shows the results of the descriptive statistics of the return series. Figure 1. (left) shows the trend of closing prices and returns of CSI 300 stock index futures, and Figure 1. (right) shows the distribution of returns. The rate of return is calculated as equation (1).

$$r_t = ln(P_t/P_{t-1}) \tag{1}$$

Table 1. Descriptive statistics of returns

	n	Mean	Median	Max	Min	Std.	Skewness	Kurtosis	J-B statistic
R_HS300	1892	-4.89E-05	0.0003	0.0974	-0.1064	0.0165	-0.5130	11.0602	5204.5610

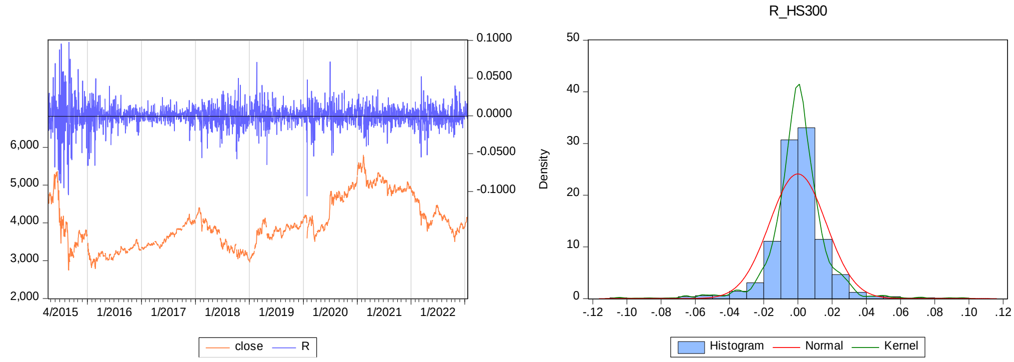

Fig. 1. Trend of closing prices and returns(left) & Distribution of returns(right)

The skewness of stock index futures return is less than 0 and the kurtosis is much greater than 3. The distribution of the return series has the characteristics of sharp peaks and long tails on the left, which can also be confirmed in the kernel density plot. The p-value of the J-B statistic is much less than 0.05, indicating that the stock index futures return series does not follow a normal distribution. According to the efficient market hypothesis, if the stock index futures market is an efficient market, the price series should be a random walk process, and the return rate conforms to the characteristics of normal distribution. Therefore, the stock index futures market does not conform to the price distribution characteristics in the efficient market, and the future price trend can be predicted by historical information.

2.2 Hurst exponent

The Hurst Exponent is an index used to measure whether a time series has Long-Term Memory, reflecting the long-term trend implied in the series. When Hurst exponent = 0.5, the time series is a Random Walk process. When the Hurst exponent is less than 0.5, the time series has the feature of Mean Reversion. When the Hurst exponent > 0.5, the time series has long-term memory and persistence. In this paper, the R/S method is used to calculate Hurst exponent. Kantelhardt (2002)[12] gives a detailed description of this approach[1].

Calculate the overall Hurst index of CSI 300 stock index futures yield series in the sample period:

$$H = 0.6024$$

H > 0.5, which proves that the stock index futures price series does not obey the random walk, the market has fractal characteristics, and has persistence and long memory on the whole. The Hurst index studies the volatility change of the return series after de-equalization, which can eliminate the influence of the drift rate of the time series. Therefore, the Hurst exponent describes the autocorrelation of log returns after removing the drift rate term. A Hurst exponent greater than 0.5 means that if the price is rising in the early stage, the price may continue to rise in the next moment.

---

[1]The main calculation formula of R/S method is:$(\mathrm{R/S})_{\mathrm{n}} = \mathrm{Kn}^{\mathrm{H}}$. Where, n is the length of the subinterval of the time series x_t, and $(\mathrm{R/S})_{\mathrm{n}}$ is the mean of the rescale range of the subinterval of x_t.

The overall Hurst index calculated using the full sample data can describe the overall situation of the market in the sample period, but as time goes by, the characteristics of the price in the market may also change. Figure 2. shows the moving Hurst exponent calculated using the moving window method, and the length of the moving window is 120 trading days. During the sample period, the moving Hurst index of CSI 300 stock index futures is all greater than 0.5, and most of the time is between 0.6 and 0.7, showing strong persistence.

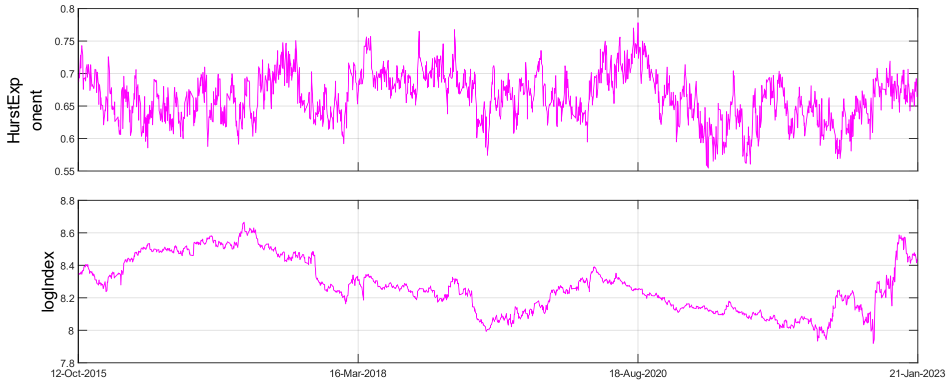

Fig. 2. Moving Hurst index of CSI 300 stock index futures

2.3 Fractal dimension

The fractal dimension is another index to analyze the fractal. The fractal dimension can be calculated in various forms, such as Hausdorff Dimension, Similarity Dimension, Box Dimension, etc. Mandeldrot proposed an equivalent definition of Box Dimension D(A) in 1975, which can be used to calculate the set A of fractional dimensions. Calculate the box dimension of the log price series of CSI 300 stock index futures, and the results are as follows:

$$D = 1.3469$$

Calculate the box dimension of logarithmic price series of CSI 300 stock index futures under the moving window, and the window length is 120 trading days (see Figure 3.). From the beginning of September 2016 to the end of February 2017, the fractal dimension of CSI 300 stock index futures is large. Since the sliding window length is about half a year, the fractal dimension of a certain trading day measures the roughness of the price series in the past half year. It can be seen that in the corresponding period from March 2016 to the end of February 2017, the price series continues to fluctuate, and does not show an obvious trend. In the interval with a small fractal dimension, such as the beginning of September 2020 to the end of December 2020, the corresponding price period from the beginning of June 2020 to the end of December 2020 shows a clear downward trend, and the curve is very smooth.

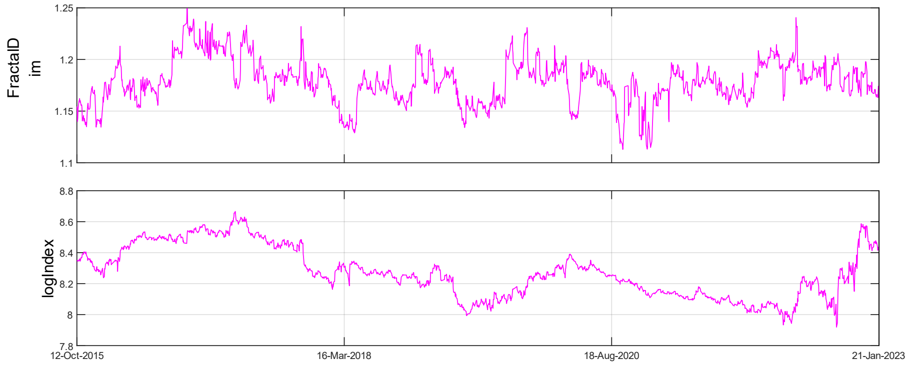

Fig. 3. Moving fractal dimension of CSI 300 stock index futures

2.4 The China VIX Index (IVIX)

From June 26, 2015 to February 14, 2018, the Shanghai Stock Exchange published the China VIX Index (iVX), which is China's first volatility index calculated based on real trading data in the options market. It is used to measure the expected volatility of the SSE 50ETF in the next 30 days and can be used as a representative of the actual implied volatility of the market. Based on the *Shanghai 50ETF Volatility Index Preparation Plan*, combined with the original VIX index preparation plan published by Chicago Board Options Exchange (CBOE), this paper calculates the China VIX Index (IVIX) between February 9, 2015 and February 23, 2023. Among them, the risk-free interest rate uses the *Shanghai Interbank Offered Rate (Shibor)*, and the interpolation method is used to calculate the risk-free interest rate for the remaining duration of the option. The options data contain the basic daily information of all listed and delisted options during the sample period, and exclude the option contracts with 0 trading volume or position volume on that day. The data are obtained from the Flush iFinD options database. This paper uses Python language to process and calculate option data when calculating IVIX index. Table 2 shows the results of the descriptive statistics of the IVIX index. The IVIX index shows the characteristics of sharp peaks and right-skewed.

Table 2. Descriptive statistics of IVIX index

n	Mean	Std.	Min	Median	Max	Skewness	Kurtosis
1934	21.7814	7.9246	2.6633	20.3865	91.4552	1.7313	6.4257

Figure 4. is a sequence diagram of the calculated IVIX index, which has the following characteristics: (1) As a representative of market implied volatility, IVIX index has volatility clustering effect. (2) The IVIX index is negatively correlated with stock index futures prices and index prices to a certain extent. This shows that when the futures and spot markets continue to fall, the panic in the market will continue to accumulate, and the expectation of future market volatility will gradually increase. (3) The IVIX index in the bear market is larger, while the IVIX index in the bull market is smaller. This indicates that the market is more sensitive to a bear market. Because IVIX index is closely related to market sentiment, investor sentiment will directly affect investors' trading behavior, and then have an impact on the price of index and

futures. Therefore, the IVIX index is used as a variable to predict stock index futures prices in this paper.

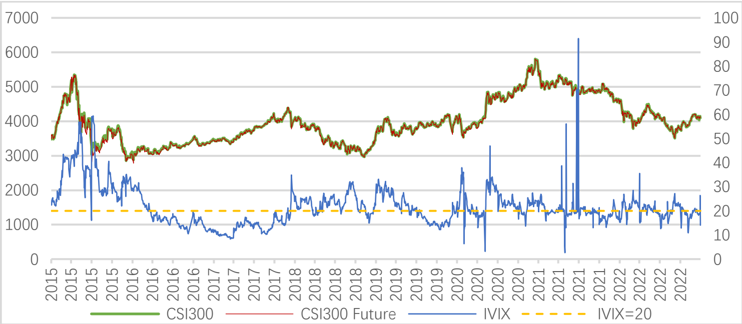

Fig. 4. Trend chart of CSI 300 index, stock index futures and IVIX index

3 DFA-LSTM stock index futures price prediction model based on panic index

3.1 Long and Short Term Memory (LSTM) Neural Network

The Long Short-Term Memory (LSTM) Neural Network is a variant of the traditional Recurrent Neural Network (RNN) model. The input and output of the traditional neural network are independent each time, while in the RNN model, the output at a moment depends not only on the current input, but also on the output at the previous moment, which allows the RNN network to "remember" historical information and learn the relationship between the sequence.

When RNN neural network learns long-term time series, it is easy to produce the problem of gradient "disappearance" or gradient "explosion". To solve this problem, LSTM networks add "Cell State", "Forget Gate", "Input Gate" and "Output Gate" structures to neurons specifically to process the input data, which is able to process information with long-term dependencies[2].

3.2 LSTM stock index futures price prediction model

In this section, the LSTM neural network will be used to forecast the price of CSI 300 stock index futures. The selection of the sample interval in this section is from February 1, 2016 to January 19, 2023. The total length of the sample data is 1696 trading days, and the ratio of the training set data to the test set data is 8:2. In the training and prediction of the model, the sliding window is used, the window time length is 20 trading days, and the prediction step is 1 trading day. Before training the model, the data were normalized. When making the forecast, the actual output data are de-normalized to obtain the predicted value[3]. This article uses the Pytorch module in the Python language for modeling.

---

[2] A detailed introduction to the LSTM neural network model is available at http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[3] The formula for normalizing $\{X_i\}$: $X_{norm} = (X_i - X_{min})/(X_{max} - X_{min})$. The formula for the inverse normalization: $Y_i = Y_{min} + Y_{norm}(Y_{max} - Y_{min})$.

When evaluating the model, this paper selects Mean Square Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), R Squared (R^2) and adjusted R^2 to evaluate the prediction effects of different models.

Table 3. Input and output variables of the model[4]

Input	Open price (OPEN), High price (HIGH), Low price (LOW), Close price (CLOSE), Settlement price (STMT), Trading volume (TDVOL), Position volumn (PSVOL)
Output	Close price (CLOSE)

The input and output variables of the model are shown in Table 3. In this section, the input layer dimension of the neural network is established as 8, the output layer dimension is 1, and three hidden layers are set, including two LSTM layers and a fully connected layer, and the number of neurons in each layer is 256. When training the model, the batch_size (that is, the batch size required to update a weight) of the input model is set to 32, the number of iterations (epochs) is set to 50, the learning rate in the gradient descent algorithm is set to 0.0003, and the AdaW optimizer is used as the optimizer. In order to prevent the phenomenon of "overfitting", the Dropout method is adopted to randomly deactivate the neurons during each training, and the Dropout probability is set to 0.5. Figure 5. shows the prediction effect of the model on the training set (left) and the test set (right).

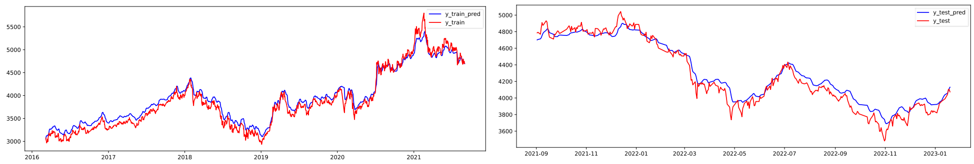

Fig. 5. Prediction effect of LSTM neural network model

It can be seen that although the neural network model can predict the trend of the stock index futures price relatively accurately, there is a certain deviation between the predicted value and the true value on the test set. After the current price has experienced a sharp rise, the predicted value of the subsequent time period is generally higher than the actual value. In particular, in real markets, long-term increases are often followed by declines in the next period, and neural networks are unable to learn this change in time. Therefore, after a generally continuous bull market from early 2016 to early 2018, the future price prediction using only historical volume price information will remain higher than the actual price.

3.3 DFA-LSTM stock index futures price prediction model based on panic index

The moving Hurst index and moving fractal dimension of stock index futures contain the historical volatility information of stock index futures market. The IVIX index is the expectation of future volatility, and the time range of information contained in fractal index and IVIX index is different. In order to improve the prediction effect, this paper combines the fractal index with the IVIX index, and establishes the IVIX-DFA-LSTM model to predict the price of CSI 300

[4] Among the variables, the price index is in units of index points, and the unit of trading volume and position volume is the number of lots calculated by side.

stock index futures. Table 4 shows the input and output variables of the model. The dimension of the input layer is 10 and the dimension of the output layer is 1.The model prediction errors under different hyperparameter settings are shown in Table 5.

Table 4. Input and output variables of the model

Input	Open price (OPEN), High price (HIGH), Low price (LOW), Close price (CLOSE), Settlement price (STMT), Trading volume (TDVOL), Position volumn (PSVOL), Panic index (IVIX), Moving Hurst index (Ht), Moving fractal dimension (Dt)
Output	Close price (CLOSE)

Table 5. Model optimization

	Epoches	Batch_size	Cells	Dropout		MSE	R2	MAE	MAPE
Model 1	50	64	128	0.1	train	5098.2880	0.9871	51.2682	0.01346
					test	4280.9985	0.9754	50.9206	0.01213
Model 2	50	32	128	0.1	train	11204.8970	0.9717	89.8985	0.02312
					test	10596.4570	0.9391	93.7547	0.02163
Model 3	50	64	256	0.1	train	3629.4200	0.9908	43.1466	0.01131
					test	2962.2600	0.9830	42.7179	0.01020
Model 4	200	64	256	0.1	train	1002.6995	0.9975	22.9523	0.00604
					test	928.6445	0.9947	23.4867	0.00546

Model 4 is the optimal model, and the optimal model is used to predict the price of CSI 300 stock index futures. The loss function of the model during training is shown in Figure 6.. Figure 7. shows the prediction effect of the model on the training set (top) and the test set (bottom).

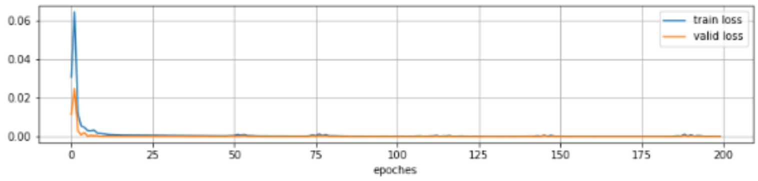

Fig. 6. Loss function

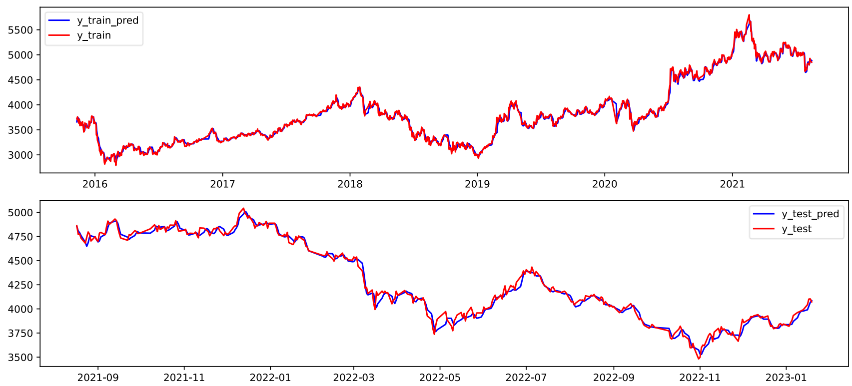

Fig. 7. Prediction effect on training set (top) and test set (bottom)

The R^2 of the model on the test set is 0.9947, which can well fit the actual trend of CSI 300 stock index futures and perform well. According to the trend of the predicted value and the actual value, it can be found that the model can not only timely capture the trend of price change in general, but also react more sensitive to the signal of "falling", which is related to the characteristics of the IVIX index. According to the analysis in the previous section of this paper, in a bear market, the market is more prone to panic, and the uncertainty of the expected future increases. When investors are bearish on the future market, such panic will be reflected in the changes of the IVIX index, and then have an impact on the prediction of the stock index futures price.

3.4 Comparison of prediction effects of different models

In addition to the neural network model, as a comparison, this paper also uses the GARCH (1,1)[5] model to predict the price of stock index futures. When GARCH(1,1)-t model is used to predict the closing price of stock index futures, the rolling window prediction method is still used, the window length is 120 trading days, and the prediction step is 1.The forecast errors of the different models are shown in Table 6 below.

Table 6. Error comparison of stock index futures price prediction model

Evaluation index	MSE	MAE	MAPE	R2	Adj R2
LSTM	28105.402	160.6396	0.03829	0.8347	0.8317
IVIX-DFA-LSTM	928.6445	23.4867	0.00546	0.9947	0.9946
GARCH(1,1)-t	2932.5146	37.6137	0.00955	0.9928	0.9928

The IVIX-DFA-LSTM neural network model constructed in this paper has a better prediction effect than the GARCH model. From the perspective of the prediction process, GARCH model has the best one-step prediction effect, so it is necessary to re-model according to historical data in each prediction to obtain good prediction effect, which requires a large amount of computation. However, in the prediction of neural network model, the trained model can be used to make multi-step prediction forward, and then the value of input variable can be updated each time of prediction, without the need to re-train the model, which improves the prediction efficiency.

4 Conclusions

This paper takes CSI 300 stock index futures as a representative, analyzes the characteristics of China stock index futures market, and uses LSTM neural network model to predict the price. The main conclusions of this paper are as follows:

(1) The yield distribution of China's stock index futures market has an obvious "peak and thick tail" feature, which is not in line with the efficient market hypothesis, and its price can be predicted through historical information.(2) China's stock index futures market has long-term memory, and the price series of stock index futures have different roughness in different time

[5] In order to better describe the non-normality of returns in the market, this paper sets the conditional variance to follow the T-distribution in the GARCH (1,1) model.

periods, and the strength of its memory changes with time.(3) By combining machine learning with fractal analysis method, and using the panic index of the Chinese market as the input of the model, the DFA-LSTM stock index futures price prediction model based on panic index established shows good prediction effect. The fitting degree is 99.47%. Compared with the traditional financial econometric model, the prediction effect has been greatly improved.

The existence of long memory in China's stock index futures market, on the one hand, is due to the large proportion of small and medium investors in A-share investors, and the prevalence of irrational investor behaviors such as herd behavior; On the other hand, it is also due to the short sale restrictions in China's stock market and the high cost of margin lending and short selling. Therefore, compared with the mature capital market, China's stock index futures market still has a large space for development, and more diversified stock index futures are needed. In addition, the research in this paper also finds that the information in the option market can be used to predict the price of stock index futures, which also means that the risk is easy to spread in different markets. Therefore, only by strengthening the risk capture and prevention by regulators can the healthy and stable development of China's capital market be better promoted.

In the future research, higher order information can be considered for the prediction of stock index futures prices, in order to further improve the prediction accuracy.

References

[1] Fama, E. F. (1965). The Behavior of Stock-Market Prices. The Journal of Business, 38(1), 34–105. http://www.jstor.org/stable/2350752

[2] Mandelbrot, B. (1967). How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. Science, 156(3775), 636–638. https://doi.org/10.1126/science.156.3775.636

[3] Peters E E. (1994). Fractal Market Analysis—Applying Chaos Theory to Investment & Economics. John Wiley & Sons.

[4] Mishra, R. K., Sehgal, S., Bhanumurthy, N. R. (2011). A Search for Long-range Dependence and Chaotic Structure in Indian Stock Market. Review of Financial Economics, 20(2), 96–104. https://doi.org/10.1016/j.rfe.2011.04.002

[5] Wang, H.Y., Zhang, Q.G. (2011). Fractal Interpolation Model and R/S Analysis of Stock Index Time Series. Statistics and Information Forum, 26(8), 23–27.

[6] Back, K. (1993). Asymmetric Information and Options. The Review of Financial Studies, 6(3), 435–472. https://doi.org/10.1093/rfs/5.3.435

[7] Zheng, Z.L., Xu, J, Chen, R. (2021). Implicit Information on the "Net Buying Pressure" of Options. Journal of Management Sciences, 24(6), 42-56+126.

[8] Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), 987–1007. https://doi.org/10.2307/1912773

[9] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31(3), 307–327. https://doi.org/https://doi.org/10.1016/0304-4076(86)90063-1

[10] Yang, Q, Wang, C.W. (2019). Research on Global Stock Index Prediction Based on Deep Learning LSTM Neural Network. Statistical Research, 36(3), 65–77. https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C44YLTlOAiTRKgchrJ08w1e7tvjWANqNvp8aduMdl4nu9wzJnv9y3jbX-yPW9TbVg41p7va8RIKtC_S8k7QW0P31&uniplatform=NZKPT

[11] Matsuoka, K., Hamori, S. (2021). Forecasting WTI Futures Prices Using Recurrent Neural Networks. Review of Integrative Business and Economics Research, 10(1), 34–50.
[12] Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., et al. (2002). Multifractal Detrended Fluctuation Analysis of Nonstationary Time Series. Physica A: Statistical Mechanics and Its Applications, 316(1), 87–114. https://doi.org/https://doi.org/10.1016/S0378-4371(02)01383-3