ISSN: 2688-9323

# A Stock Prediction Method based on Futures Data

Tingzhen Chang*

School of Economics and Management, Tongji University, Shanghai, 201800, China

*Corresponding Author

## Abstract

**Stock prediction has always been the focus of attention and difficulty in the field of finance, this thesis selects the stock data of the precious metals sector in the past ten years, and at the same time obtains the futures data related to the precious metals sector, and makes a prediction on it. We used LSTM and various machine learning models. In the regression prediction, basic stock indicators, market sentiment scores, stock technical indicators, and futures indicators are considered, and it is found that the inclusion of futures indicators reduces the MSE by more than 0.0001 compared to the other indicators, and reduces the MSE by more than 0.0003 compared to the other models, and the confidence is significant. This thesis illustrates that the inclusion of futures data in the consideration of stock forecasting can significantly improve the accuracy of stock forecasting.**

## Keywords

**LSTM; Stocks; Futures; Time Series Prediction.**

## 1. Introduction

To a certain extent, the financial market reflects the economic situation and level of the whole society, so the importance of paying attention to the trends and movements of the financial market can help to grasp the trend of the country's economic development, and to find the problems, laws and characteristics of the development, etc. The importance of this is self-evident. Yet stock trends are affected by many factors. For example, national policies can cause some sectors and fields to become hot spots, making stocks soar, and social opinions can put some companies and their stocks in great fluctuations. Nowadays, the international situation has become more complicated, especially the frequent occurrence of inter-regional conflicts, which leads to the rise of risk aversion and huge fluctuations in the demand for precious metals such as gold, therefore, the prediction of the stock market has become extremely important, but at the same time, it is also very difficult.

There are many products in the financial market, including stocks and futures. Stocks and futures are two different markets, but there is a strong link between them, for example, futures indices may move ahead of stock indices. Currently, most of the research focuses on stocks, even though there is also research on futures, but there is little literature that combines the two to improve the prediction accuracy.

Therefore, in this paper, we will use machine learning and deep learning algorithms, combined with real-time Internet public opinion and technical indicators, to find the connection between stocks and futures, innovatively link futures data with stock forecasts, and predict the closing price of stocks using the precious metals sector as an example, so as to provide new ideas and methods for the forecasting problems in the financial market.

ISSN: 2688-9323

## 2. Literature Review

### 2.1. Relationship between Futures and Stocks

Futures and stocks are both financial products with fluctuating prices, and the relationship between futures indices and stocks is also very close. According to the existing studies, for example, Yuanzheng Wang et al. studied the correlation between Chinese stock index futures IF1006 and HS300 stock index, and found that there is a univariate causality by Granger causality test, i.e., futures are the cause of stock index movement [1].

On the other hand, much research has been done on the prediction of futures prices, for example, Pan, HJ et al. modeled gold futures by constructing GARCH and LSTM models to improve the accuracy [2]. et al. improved the accuracy of crude oil futures price prediction based on empirical modal decomposition (EEMD), convolutional neural network (CNN), and improved long short-term memory (ILSTM) [3].

### 2.2. Technical and Sentiment Indicators

In the development of the financial market over a long period of time, through people's experience, many artificially designed indicators with reference value and significance have appeared, for example, Yaohu Lin et al. found that the incorporation of technical indicators, especially the momentum indicator, will have a positive effect on the prediction of the rise and fall of the stock [4]; Nagaraj Naik et al. through the incorporation of 33 kinds of technical indicators, including: Moving Average, Momentum Indicator, Relative Strength Indicator, Williams Index, etc., to regression prediction of stock prices in India and reduced the error rate of the model [5].

In addition to this, the volatility of the financial market is largely affected by external social opinion, for example, Pooja Mehta et al. believed that financial news coverage would have an impact on stock prices, so they incorporated the sentiment of the news and the public opinion sentiment into the indicators of stock prediction, which was successful [6]; Xiaodong Li et al. took the Hong Kong stock market as an example, and took the sentiment factor into consideration and proposed four sentiment dictionaries, which predicted better results than price, technical indicators, and news sentiment indicators alone, both on individual stocks and on the board [7].

All the above studies suggest that technical and sentiment indicators may help in predicting the movement of stocks in various regions of the world and improve the accuracy of the predictions.

### 2.3. Time Series Forecast

In terms of research methods, from machine learning long ago to deep learning models nowadays, the methods about stock prediction are constantly progressing and developing. Jibing Gong predicts the next month's stock data based on the current month's stock data through Logistic regression-based methods, which gives higher prediction accuracy with lower complexity [8]; Chuiyong Zheng et al. used gray correlation analysis and ARMAX model to regressively predict the stock price of PetroChina [9].

With the development of deep learning models, more deep models are applied to time series prediction, especially stock prediction; Wei Shen et al. obtained the conclusion that “the more indicators are considered, the more accurate the prediction will be” by combining BP neural network and genetic algorithm; Widodo Budiharto et al. used the LSTM model to predict the stock price of PetroChina by using the short-term forecasting model [9]. LSTM model, using short-term data, improved the accuracy of predicting the price of stocks in Indonesia during the New Crown epidemic [10].

Therefore, we observe that for the time series prediction of stocks, no scholars have yet taken into consideration the various indicators of the relevant futures as variables for stock

ISSN: 2688-9323

prediction; therefore, in this paper, we specifically take the futures characteristics into consideration and utilize the models, e.g., the LSTM model, etc., to make predictions about the rise and fall of stocks. We have selected the stock prices of the precious metals sector for forecasting.

# 3. Model Principle

## 3.1. Bert Model

For the sentiment classification problem, we have chosen the BERT model. BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model proposed by Google in 2018, and its structure is shown in Figure 1. The token embedding is obtained by encoding this paper with segment embedding and position embedding, which are used as inputs to the model.

The internal structure of Bert is the encoder part of the transformer, and its most central part, the attention mechanism, is shown in Figure 2. By multiplying the input by different weight matrices, the Key, Query, and Value matrices are obtained, and then Equation 2 is applied to get the final attention score.

$$Input = Token\ Embeddings + Segment\ Embeddings + Position\ Embeddings \quad (1)$$

$$output = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (2)$$

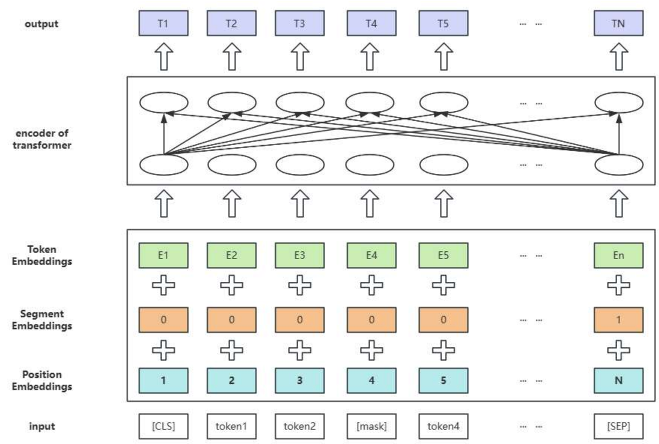


**Figure 1.** The BERT model structure

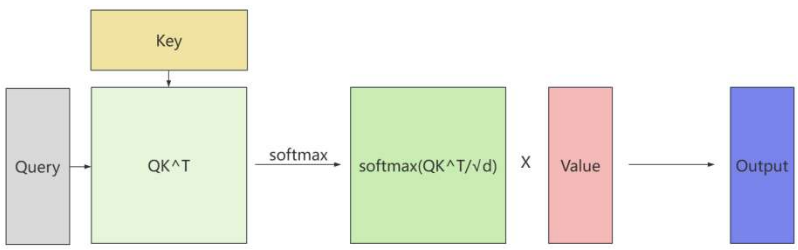


**Figure 2.** Principles of Attention Mechanisms

The way BERT dynamically adapts the vector representation of tokens to the context is by pre-training with a large corpus.

ISSN: 2688-9323

## 3.2. LSTM Model

LSTM named Long Short-Term Memory Network is a special kind of Recurrent Neural Network (RNN). Traditional recurrent neural networks will produce the phenomenon of gradient vanishing or gradient explosion when processing long time sequence data again, in order to solve this problem, LSTM controls the information flow by introducing the memory unit and three gating mechanisms (input gate, forgetting gate, and output gate, as shown in Figure 3), which can effectively mitigate the gradient vanishing and explosion, and is more conducive to its learning of long term dependency.

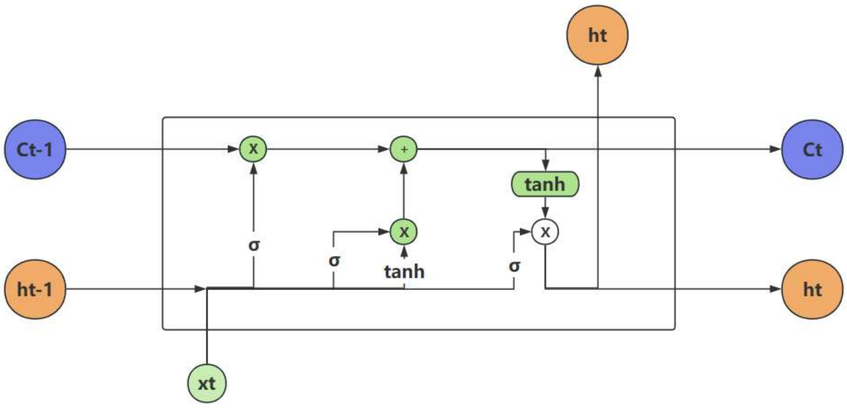


**Figure 3.** LSTM model structure

There are some special mechanisms among the structure of the LSTM model.

(1)Forget Gate

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{4}$$

(2)Input Gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{5}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{6}$$

$$\widetilde{C_t} = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{7}$$

(3)Cell State

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \tag{8}$$

(4)Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

ISSN: 2688-9323

$$h_t = o_t * \tanh(C_t) \quad (10)$$

# 4. Data Acquisition and Processing

## 4.1. Data Sources

### 4.1.1. Stock Futures Correspondence Table

Futures and stock list from the Oriental wealth enterprise choice financial terminal software, we get: enterprise individual stock list, plate stock list, commodity index list; which: each plate contains many stocks below, that is, the plate stock and individual stocks are subordinate to each other, and is a pair of many. The "commodity index" is a series of related futures contracts in accordance with certain rules of calculation of the price obtained by the synthesis of a number of futures together, the collective term, which is essentially a futures index of these commodities, so in the rest of this paper, we will "commodity index" as "futures index". Therefore, in the remainder of this thesis, we will refer to the "commodity index" as "futures index", as futures data. For example, the futures index "Shanghai copper" is a combination of futures data from "copper, copper, lead and zinc, copper and silver".

Some of the futures indices, futures and stock lists are shown in Table 1-3. The list of stocks is categorized according to the "Shenwan secondary sectors" (a classification rule).

**Table 1.** Futures index and its code (part)

| Index of futures | Code of futures |
|---|---|
| Shanghai-silver Index | AGFI.WI |
| Shanghai Aluminum index | ALFI.WI |
| Shanghai Gold index | AUFI.WI |

**Table 2.** Sector stock names and codes (part)

| Sector stock Name (Shenwan Level 2) | Sector code |
|---|---|
| General steel | 801044 |
| Textile manufacturing | 801131 |
| Precious metals | 801053 |

**Table 3.** Stock names and their codes (part)

| Name of individual stock | Individual stock code |
|---|---|
| Baosteel Corporation | 600019 |
| Shandong Iron and Steel | 600022 |
| Shougang Shares | 000959 |

In this thesis, we innovatively discover the connection between stocks and commodity indices, namely: "futures index - futures - main products - individual stocks - sector stocks", and these data can be obtained in the choice. Therefore, this thesis explores and organizes the correspondence table of "futures index - futures" and "futures - sector stock - individual stock", and some data are shown in Tables 4 and 5.

The important basis for this thesis to establish such a correspondence is that the futures index corresponds to futures corresponding to products that happen to account for a relatively large share of revenues, costs, profits, and gross margins of certain companies, which can be obtained from the CHOICE software.

ISSN: 2688-9323

**Table 4.** Futures index and corresponding futures name (part)

| Code | Name | Component 1 | Component 2 | Component 3 |
|---|---|---|---|---|
| AGFI.WI | Shanghai-silver Index | Copper and Silver | Silver | |
| AUFI.WI | Shanghai Gold index | Gold | | |
| CUFI.WI | Shanghai Copper Index | Copper | Copper lead zinc | Copper and Silver |

**Table 5.** Futures and proportion of financial indicators of products in the company (part)

| Sector Name | Sector Code | Futures Name | Company Name | Company code |
|---|---|---|---|---|
| Textile manufacturing | 801131 | Cotton yarn | Blum Oriental | 601339 |
| Textile manufacturing | 801131 | Cotton yarn | Huamao Shares | 000850 |
| The planting industry | 801016 | corn | Wanxiang Denong | 600371 |
| Product Name | Percentage of revenue | Percentage of cost | Percentage of profit | Gross profit ratio |
| Yarn yarn | 93.39% | 94.55% | 86.72% | 13.85% |
| Yarn yarn | 59.49% | 57.32% | 76.00% | 14.83% |
| Seed of corn | 96.83% | 96.37% | 97.34% | 48.01% |

The link we have established from the futures index to the sector stocks can be felt more intuitively in Figure 4.

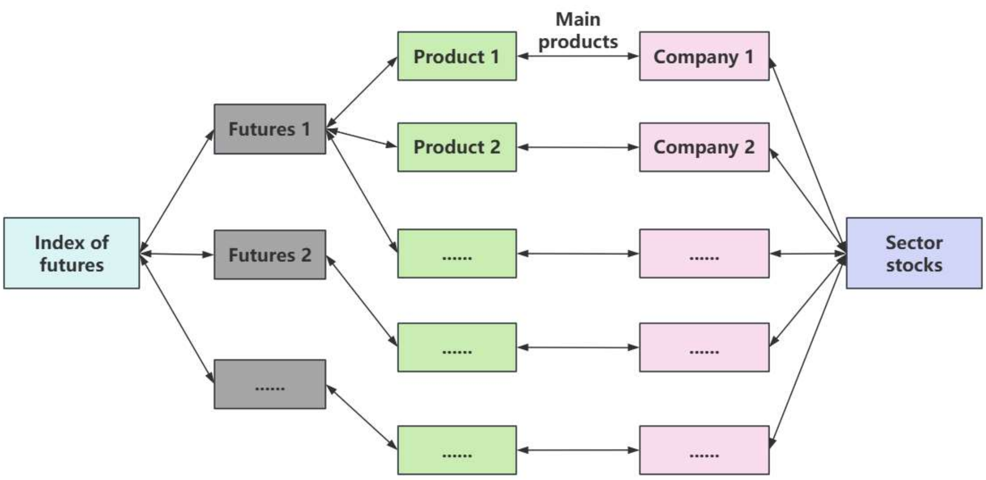


**Figure 4.** Schematic diagram of the link from futures index to sector stocks

### 4.1.2. Sector Selection

In this thesis, we have taken into consideration and selected the precious metals sector to predict its stock movement.

### 4.1.3. Futures Index and Sector Stock Indicators

The data of futures index and plate stock are all from the choice financial terminal of Oriental wealth, from the beginning of January 1, 2014 to June 30, 2024 period of weekdays. The indicators of the board stocks include: trading date, minimum price, maximum price, opening price, closing price, number of gainers, number of losers, number of gainers, number of losers, turnover, turnover, turnover, turnover, up and down; the indicators of the futures index include: opening point, maximum point, minimum point, closing price, settlement price, up and down, up and down (%), cumulative increase and decrease on the starting day, cumulative increase and decrease on the starting day, Volume (lot), turnover (million), position. Some of the data are shown in Tables 6 and 7. (Take precious metals as an example). And labeled, if the day's opening price is lower than the closing price, it is positive, labeled 1, the day's opening price is higher than the closing price, it is negative, labeled 0.

**Table 6.** Sector Stock Indicator (Precious Metals)

| Date of transaction | Lowest price | Maximum price | Opening price | Closing price | Number of limit up companies | Number of limit down companies |
|---|---|---|---|---|---|---|
| 20140102 | 4321.65 | 4482.49 | 4348.93 | 4419.9 | 0 | 0 |
| 20140103 | 4349.65 | 4443.05 | 4425.7 | 4383.25 | 0 | 0 |
| 20140106 | 4235 | 4404.62 | 4380.89 | 4263.64 | 0 | 0 |
| Number of increased companies | Number of falling companies | Trading volume | Total transaction amount | Rate of turnover | Up and down | Label |
| 9 | 0 | 1.25E+08 | 9.38E+08 | 0.54436 | 1.957066 | 1 |
| 3 | 6 | 76740374 | 5.51E+08 | 0.334776 | -0.8292 | 0 |
| 1 | 8 | 78241886 | 5.16E+08 | 0.341326 | -2.7288 | 0 |

**Table 7.** Futures Index Indicators

| Date of transaction | Closing price |
|---|---|
| 20140102 | 244.43 |
| 20140103 | 246.38 |
| 20140106 | 246.89 |

### 4.1.4. Text Data

All the text data come from the web version of Oriental Wealth's financial forums, in this thesis, we chose the precious metals as a sector of stocks, so we also selected some popular, corpus-rich topic forums related to precious metals, and crawled some of the stock bar forums with a crawler to extract the comments of the stockholders from January 1, 2014 onwards. These theme forums include: China Gold, Hunan Gold, Shandong Gold, Hengbang Stock, Chifeng Gold five boards, the number of comments are: 153375, 141279, 186826, 128187, 249298, respectively, the labels are artificially labeled with sentiment polarity, “2” means positive, “1” means positive, “1” means positive, “1” means positive, “1” means positive, “1” means positive, and “1” means positive. “1” means neutral, ‘0’ means negative.

## 4.2. Technical Indicators

(1) Daily mean $\bar{x}$

$$\overline{x_{date}} = \frac{\sum_{i=1}^{n} x_{date-i}}{n} \tag{11}$$

Where n is the length of the mean interval and x refers to the price.

(2) William's index WR

$$WR = \frac{H_n - C}{H_n - L_n} * 100 \tag{12}$$

Where C is the closing price on the calculation day, $H_n$ is the highest price in N periods, and $L_n$ is the lowest price in N periods, where N we take 10.

(3) KDJ index

$$\mathrm{RSV} = \frac{\mathrm{C_n - L_n}}{\mathrm{H_n - L_n}} * 100 \tag{13}$$

K, D, and J in the KDJ index are shown in Equations 14-16, respectively.

$$\mathrm{K_n} = \frac{2}{3} * \mathrm{K_{n-1}} + \frac{1}{3} * \mathrm{RSV_n} \tag{14}$$

$$\mathrm{D_n} = \frac{2}{3} * \mathrm{D_{n-1}} + \frac{1}{3} * \mathrm{K_n} \tag{15}$$

$$\mathrm{J_n} = 3 * \mathrm{D_n} - 2 * \mathrm{K_n} \tag{16}$$

(4) Relative strength index RSI

$$\mathrm{RS} = \frac{\mathrm{N_{up}}}{\mathrm{N_{down}}} \tag{17}$$

$$\mathrm{RSI} = 100 - \frac{100}{1+\mathrm{RS}} \tag{18}$$

Where $N_{up}$ represents the average amount of closing price increase within N days; $N_{down}$ represents the average amount of closing price decline on N days.

(5) Deviation rate BIAS

$$\mathrm{BIAS} = \frac{\mathrm{C} - \overline{\mathrm{x_n}}}{\overline{\mathrm{x_n}}} * 100\% \tag{19}$$

Where C refers to the closing price of the day, and $\overline{x_n}$ represents the n-day average.

(6) Homeopathy index CCI

$$\mathrm{TP} = \frac{\mathrm{H+L+C}}{3} \tag{20}$$

$$\mathrm{MA} = \frac{\sum_{i=1}^{n} \mathrm{C_i}}{\mathrm{N}} \tag{21}$$

$$\mathrm{MD} = \frac{\sum_{i=1}^{n} (\mathrm{MA_i - C_i})}{\mathrm{N}} \tag{22}$$

$$\mathrm{CCI} = \frac{\mathrm{TP - MA}}{\mathrm{MD} * 0.015} \tag{23}$$

Where N is the calculated period, H is the highest price, L is the lowest price, and C is the closing price.

(7) BOLL indicator

ISSN: 2688-9323

$$MB = \frac{\sum_{i=1}^{n} \overline{x_i}}{N} \tag{24}$$

$$UP = MB + 2 \times MD \tag{25}$$

$$DN = MB - 2 \times MD \tag{26}$$

Where MD is the standard deviation of N days and MB is the average of the average moving lines of the previous N days.

## 5. Empirical Analysis

### 5.1. Data Sources

#### 5.1.1. Stock Futures Correspondence Table

In order to verify the correlation between the sector stock indexes and futures indexes, we firstly selected some data of the sector indexes and their related futures indexes, and then counted all the data of working days from January 1, 2014 to June 30, 2024, and calculated the Pearson's correlation coefficients based on the closing prices of the sector indexes and the futures indexes, as well as statistically obtained the proportion of negative and positive days of the sector indexes and the futures indexes on the same day; part of the situation is shown in Table 8, in order to show the trend of the data more clearly. The proportion of negative and positive lines on the same day; part of the situation is shown in Table 8, in order to more clearly show the trend of the data, we deliberately selected the pair of data “precious metals - Shanghai gold” to plot the trend of the two (in order to compare the two obvious, we will be divided by 20, which does not affect the trend and the Pearson correlation coefficient calculation). Pearson correlation coefficient calculation) shown in Figure 5.

**Table 8.** Sector stock indices -futures

| Sector Indexes - Futures | Precious metal - Shanghai gold | Precious metal - Shanghai silver | Precious metal - Shanghai copper |
|---|---|---|---|
| Pearson's correlation coefficient | 0.8740 | 0.8478 | 0.8072 |
| The same rate of negative and positive | 53.16% | 38.71% | 39.30% |
| Sector Indexes - Futures | Precious metal - Shanghai gold | Precious metal - Shanghai silver | Precious metal - Shanghai copper |

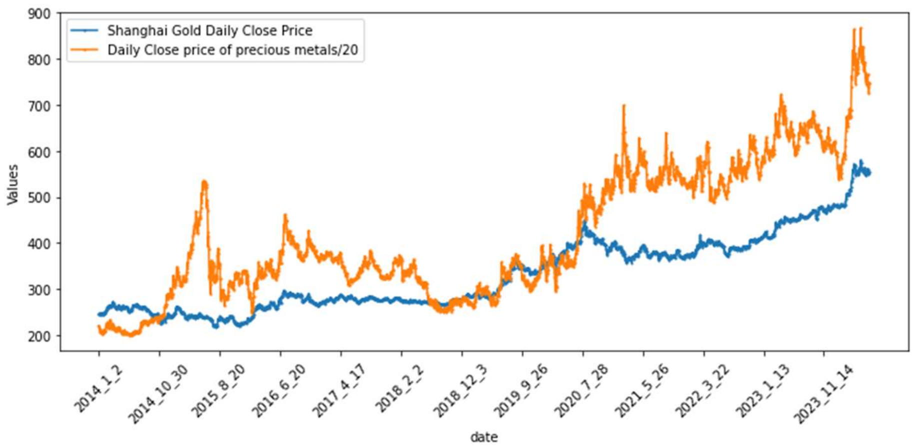


**Figure 5.** Precious Metals - Shanghai Gold Data Trend Chart

ISSN: 2688-9323

Through the above analysis, we get some conclusions:

(1) The absolute values of the Pearson correlation coefficients between the sector indices and their related futures are almost always above 0.8, which is a clear correlation.

(2) A significant portion of the probability that the negative and positive lines are the same is more than 53% or less than 40%, which indicates that there is also a correlation on the K-plot of the Sector Index and the related futures, which is more than 53%. This indicates that most of the time, the plate index within a day of the rise and fall with the futures, less than 40% means that most of the time, the plate index within a day of the rise and fall with the futures of the opposite, in either case, are more obvious correlation relationship.

Therefore, based on the above analysis, it can be tentatively concluded that the selection of futures closing price data related to a specific sector as an auxiliary variable is explanatory and persuasive.

## 5.2. Sentiment Analysis Results

We selected some forums related to precious metals in stock bar financial forums and crawled all the stock comment corpus and news corpus from January 1, 2014 to June 30, 2024, with a total of 966,547 data, as well as a total of 10,000 positive, neutral, and negative corpus, with 2 positive labels, 1 neutral label, and 0 negative labels, and trained 1 epoch in the test set with a bert-base-chinese model downloaded from hugging face bert-base-chinese model downloaded from hugging face, using AdamW optimizer with a learning rate of 5e-5, batch_size=32, and training for 1 epoch, and achieved an accuracy of 97.34% in the test set.

Eventually we obtained the mean value of the sentiment scores for each day and after normalizing the precious metal index, we obtained the trend graph shown in Figure 6.

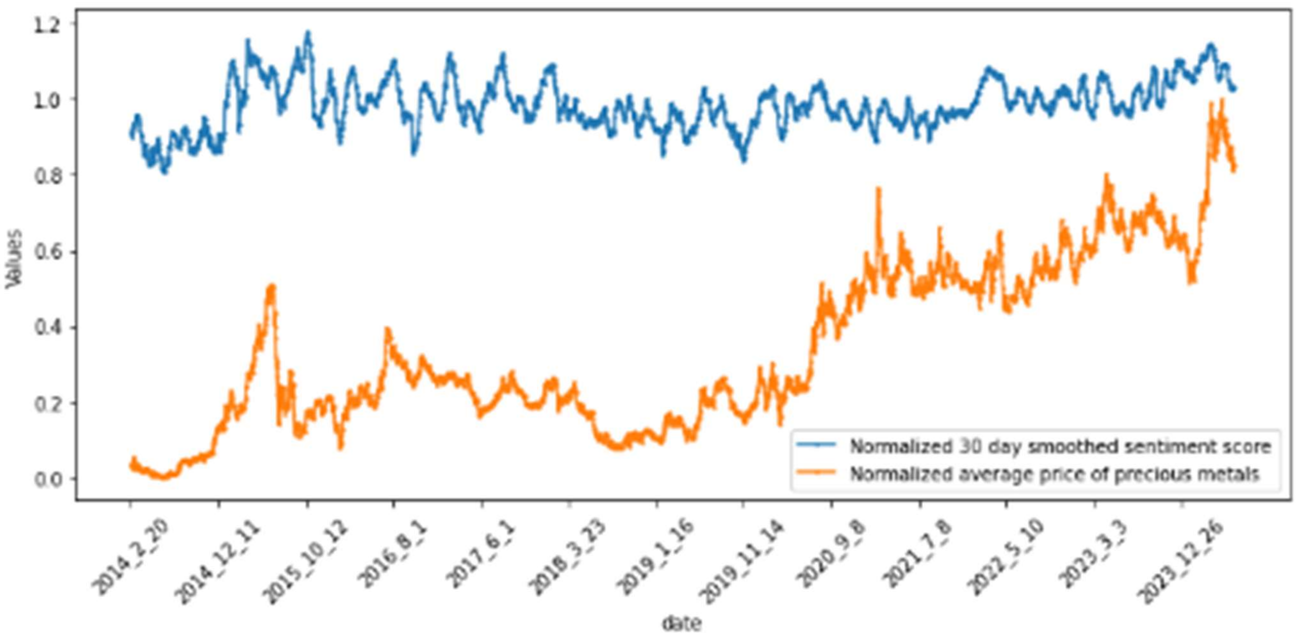


**Figure 6.** Sentiment Scores and Precious Metals Stock Trends

Therefore, we will choose the 30-day smoothed sentiment score as the sentiment profile in the subsequent forecasts.

Eventually we get all the metrics as shown in Table 9.

**Table 9.** Indicator statistics

| Indicators | In detail |
|---|---|
| Base | Lowest price, Highest price, Opening price, Closing price |
| Transactions | Number of limit up companies, Number of limit down companies, Up number<br>Down number, Transaction quantity, Transaction amount, Turnover rate<br>Up or down, negative and positive |
| Sentiment | 30-day smoothed sentiment score |
| Technology | 30-day mean, Rsv, William indicator, KDJ, RSI<br>Deviation rate, CCI, Boll line up, Boll line down |
| Futures | Closing price |

## 5.3. Regression Prediction

### 5.3.1. Results

For categorical prediction, we conducted the following experiment: the difference between the regression prediction accuracy of the base metrics with and without other metrics.

In this thesis, we mainly used three metrics to measure the prediction accuracy: mean square error MSE, root mean square error RMSE, mean absolute error MAE, and mean absolute percentage error MAPE.

The comparison of regression prediction for the LSTM model is shown in Table 10. It can be seen that the LSTM model has higher prediction accuracy after adding the closing price indicator in futures. And the overall MAPE of the LSTM model is much lower than the machine learning model. We chose the prediction images with only the base indicator and the closing price added to the futures, as shown in Figures 7 and 8.

**Table 10.** Regression prediction results

| Models | Indicators | MSE | RMSE | MAE | MAPE |
|---|---|---|---|---|---|
| LSTM | Base | 0.000874 | 0.029571 | 0.021011 | 0.025133 |
| LSTM | Transactions | 0.000813 | 0.028512 | 0.022069 | 0.027407 |
| LSTM | Sentiment | 0.000773 | 0.027796 | 0.019726 | 0.023850 |
| LSTM | Technology | 0.000829 | 0.028789 | 0.020418 | 0.024470 |
| **LSTM** | **Futures** | **0.000672** | **0.025929** | **0.019014** | **0.023430** |
| Random forest | Futures | 0.001040 | 0.032248 | 0.024350 | 0.317800 |
| XGBoost | Futures | 0.001090 | 0.033021 | 0.027286 | 0.327758 |
| Lightgbm | Futures | 0.001028 | 0.032057 | 0.026585 | 0.306663 |

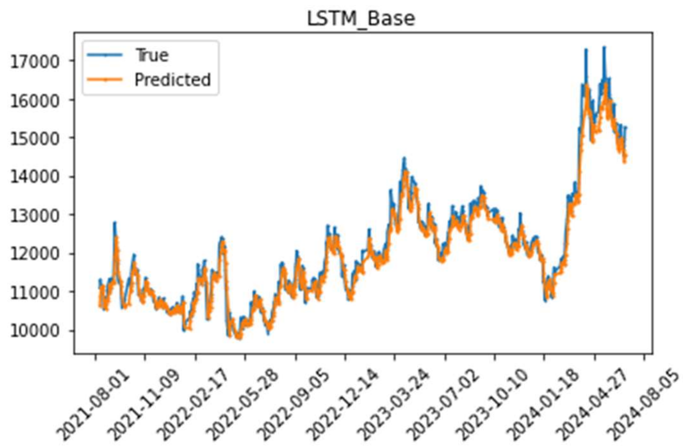


**Figure 7.** Projected line graphs for underlying indicators

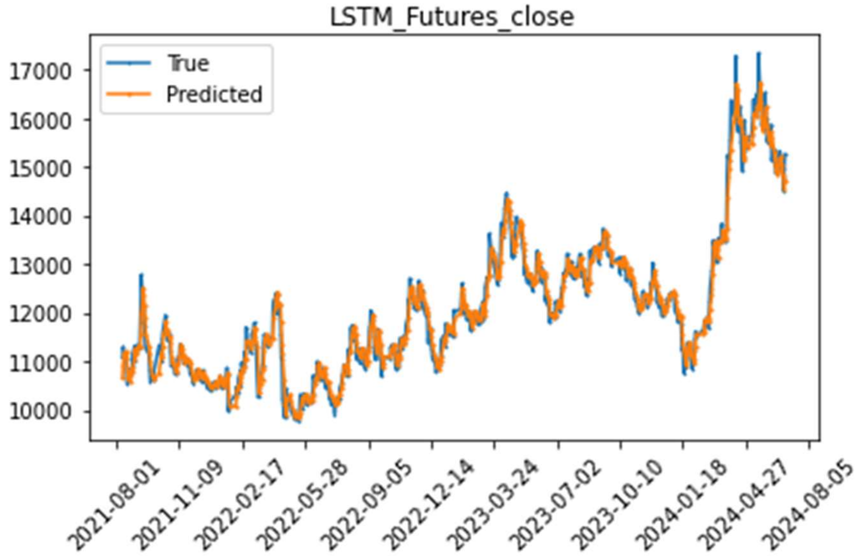


**Figure 8.** Add a predictive line chart of futures closing prices

### 5.3.2. Confidence Test

In order to verify the prediction difference between different models in regression prediction, we also need to perform Diebold-Mariano test. Since in the previous section, we found that the prediction accuracy of the integrated learning model is overall inferior to that of the LSTM model, we only perform an internal comparison of the LSTM model in the DM test. In addition, this is in line with the innovation of this thesis: to explore the value of this feature of futures rather than the differences between deep and machine learning models. The results of the DM test are shown in Table 11.

**Table 11.** Results of Diebold-Mariano test

| | Transaction | Sentiment | Futures | Technology | Transactions and Futures | Sentiment and Futures | Technology and Futures |
|---|---|---|---|---|---|---|---|
| Base | (1.2819) | (7.8964) | **(5.5569)** | (4.5379) | (-3.7537) | (6.9012) | (4.4627) |
| | 0.2003 | 0.0000*** | **0.0000***** | 0.0000*** | 0.0002*** | 0.0000*** | 0.0000*** |
| Transaction | | (1.0841) | **(6.1361)** | (-0.3498) | (-10.3860) | (2.1430) | (2.8451) |
| | | 0.2787 | **0.0000***** | 0.7266 | 0.0000*** | 0.0325** | 0.0046*** |
| Sentiment | | | **(3.9792)** | (-4.7918) | (-5.9099) | (2.5578) | (1.7457) |
| | | | **0.0001***** | 0.0000*** | 0.0000*** | 0.0107** | 0.0813* |
| Futures | | | | **(-4.6169)** | **(-10.7092)** | **(-3.0300)** | **(-4.2540)** |
| | | | | **0.0000***** | **0.0000***** | **0.0025***** | **0.0000***** |
| Technology | | | | | (-4.5231) | (4.6185) | (3.4360) |
| | | | | | 0.0000*** | 0.0000*** | 0.0006*** |
| Transactions and Futures | | | | | | (6.7544) | (8.0225) |
| | | | | | | 0.0000*** | 0.0000*** |
| Sentiment and Futures | | | | | | | (-0.1950) |
| | | | | | | | 0.8454 |

Note: ***p<0.01, **p<0.05, *p<0.1

We can see that in terms of prediction accuracy, the base indicator has worse accuracy compared to most indicators; however, most importantly, the inclusion of the closing price in futures improves all the prediction accuracies and the p-values are all below 1%. Therefore, we

ISSN: 2688-9323

can conclude that the inclusion of the closing price as a futures attribute in the regression prediction significantly improves the prediction accuracy.

## 6. Conclusion

### 6.1. Main Conclusion

This thesis focuses on the problem of predicting stocks in this sector of precious metals in the Chinese financial market, and the conclusions obtained are:In terms of closing price prediction, through Diebold-Mariano test, the prediction effect of LSTM model is generally better than that of machine learning model; moreover, for LSTM model, the prediction accuracy is the highest after adding the indicator of closing price of related futures, and the p-value of the related Diebold-Mariano test reaches within 0.01, with the confidence level exceeding 99%. 99%.

These conclusions show that the closing price of the futures related to the predicted stocks has a great reference value for the prediction of the K-plots of the stocks in the related sectors.

### 6.2. Contributions

This thesis provides a new perspective of stock forecasting, establishes a complete linkage of "sector stock - individual stock/company - main product - futures - commodity index/futures index", and statistically analyzes the data between several futures indices and the related sector stocks, and verifies that there is a high correlation between them.

Most importantly, this thesis verifies that the addition of futures indexes is very effective in improving the prediction accuracy.

### 6.3. Limitations

First, this thesis only verifies the effectiveness of the futures indicators for the prediction of the precious metals sector, and does not completely verify the other sectors; second, the LSTM model used in this thesis only represents a deep learning model, and this thesis does not explore other deep models; third, except for the futures indicators, the indicators, or features, selected in this thesis are relatively conventional and limited; it does not explore whether adding other indicators will improve the accuracy of the forecast. that do not explore whether the inclusion of other metrics would have an impact on the prediction.

## References


[1] Wang, Y., Xu, Y., Liu, P: The Co-integration and Causality Relationship Research of Stock Index Futures IF1006 and HS300 Stock Index[J]. Education and Management,2022,vol.210: 541–545.

[2] Pan, H, Tang, Y, Wang, G: A Stock Index Futures Price Prediction Approach Based on the MULTI-GARCH-LSTM Mixed Model[J]. Mathematics 2024,12(11):1677.

[3] Wang, J, Zhang, T, Lu, T, Xue, Z: A Hybrid Forecast Model of EEMD-CNN-ILSTM for Crude Oil Futures Price[J]. Electronics 2023; 12(11):2521.

[4] Lin, Y, Liu, S, Yang, H, Wu, H: Stock Trend Prediction Using Candlestick Charting and Ensemble Machine Learning Techniques With a Novelty Feature Engineering Scheme[J]. IEEE Access 2021; 9(10):1433-1446.

[5] Naik: BOptimal Feature Selection of Technical Indicator and Stock Prediction Using Machine Learning Technique. In Emerging Technologies in Computer Engineering: Microservices in Big Data Analytics 2019 (Jaipur, India, February 1–2, 2019). Vol. 985, p. 261–268.

[6] Pooja Mehta, Sharnil Pandya, Ketan Kotecha: Harvesting social media sentiment analysis to enhance stock market prediction using deep learning[J]. PeerJ Computer Science 2021; 7:1–21.

[7] Xiaodong Li, Pangjing Wu, Wenpeng Wang: Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. Information Processing & Management 2020; 57(5):102212.

ISSN: 2688-9323

[8] Gong, J, Sun: SA New Approach of Stock Price Prediction Based on Logistic Regression Model. In 2009 International Conference on New Trends in Information and Service Science 2009. (Beijing, China, 30 June–2 July, 2009). p. 1366-1371.

[9] Zheng, C, Zhu, JResearch on stock price forecast based on gray relational analysis and ARMAX model. In 2017 International Conference on Grey Systems and Intelligent Services (GSIS) 2017 .(Stockholm, Sweden, 8-11 Aug, 2017). p. 145-148.

[10] Budiharto W: Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM)[J]. Journal of big data, 2021; 8(1): 47.