ISSN 2616-5902 Vol. 5, Issue 14: 82-87, DOI: 10.25236/AJBM.2023.051415

Futures Trading Price Prediction Based on Big Data

Yunzhe Liu[1,a,*]

[1]*School of Economics and Management, Qinghai Minzu University, Xining, 810007, Qinghai, China*
[a]*liuyunzhe0815@163.com*
[*]*Corresponding author*

Abstract: *The futures market offers a secure, precise and quick exchange platform for traders, offering efficient transactions and no triangle bonds, which is beneficial for establishing and improving the market economy. However, for trading prices, it is generally difficult to predict. Therefore, this article uses big data analysis methods to predict futures trading prices. The research results show that after using big data for prediction analysis, the highest accuracy rate reaches 99.5%, and the highest recall rate also reaches 84.2%, which is much improved compared to traditional prediction methods. Therefore, the final conclusion can be drawn that big data can effectively predict futures trading prices.*

Keywords: *Futures Trading, Big Data, Price Forecasting, Feature Variables*

1. Introduction

The commodity futures exchange can conduct real-time trading of the varieties in trading, with two corresponding transactions per second. How to use big data analysis methods to predict prices and make market judgments in commodity futures trading, in order to obtain stable returns, has become increasingly important for traders.

Many financial professionals around the world have conducted research on futures trading. Reznik, N. P conducted a retrospective analysis on the basic risks of Ukrainian futures trading. The results showed that the futures price dynamics and main futures price patterns of wheat, corn, and barley are constrained by spot prices, which are the main risks of futures trading. Yang, Hao Chang used data from ten trading markets from 2011 to 2020 to empirically analyze the impact of foreign exchange futures volatility on macroeconomic variables [1-2]. The research results indicate that the volatility of foreign exchange futures has a significant impact on indicators such as government budget revenue and inflation. Luo and Jiawen predicted five agricultural product futures (corn, cotton, indica rice, palm oil, and soybeans) based on high-frequency data from the Chinese futures market, and evaluated the prediction results using statistical and economic evaluation methods. The results indicate that the heterogeneous autoregressive model has better accuracy compared to the benchmark heterogeneous autoregressive model based on high-frequency data, which has an infinite hidden Markov state transition structure. The constructed investment portfolio also achieves higher portfolio returns than the benchmark heterogeneous autoregressive model for short-term volatility prediction [3]. Jiang and Ying studied the impact of the introduction of night trading in Shanghai Futures Exchange on other markets. The results show that longer trading hours will lead to less market segmentation and more information flow [4]. Ogrand and Atle studied the impact of trading partner diversification on the futures trading market. The results indicate that the greater the degree of diversification of trading partners, the lower the basis risk of enterprises [5]. The above research results are helpful for the futures market, but they do not take into account the importance of big data in predicting futures trading prices.

In order to improve the accuracy of predicting futures trading prices, big data analysis methods were adopted. The final results indicate that the big data method has a significant effect on improving prediction accuracy and has great potential in the future field of futures trading price prediction.

2. Futures Trading

Futures trading are a buying and selling process. Its characteristics of "hedging", "preventing excessive market volatility", "reducing commodity circulation costs", and "promoting market fairness" have played a positive role in the healthy development of China's commodity circulation market.

ISSN 2616-5902 Vol. 5, Issue 14: 82-87, DOI: 10.25236/AJBM.2023.051415

China's futures market has made some progress, but due to the lack of relevant laws and regulations, various regions have acted independently, leading to the lawlessness of the futures market and the widespread occurrence of excessive speculation [6]. Therefore, it is necessary to establish a specialized futures trading system.

The buyer of a futures contract must purchase the corresponding subject matter of the contract when it is about to expire; The seller of a futures contract is obligated to sell the corresponding subject matter of the futures contract when the contract expires. Of course, traders in futures contracts can also offset this responsibility through reverse trading before the contract expires [7]. In the transaction, both parties have to pay a small amount of security, that is, performance security. The first contract to buy is called long, and the first contract to sell is called short. And the futures contracts in hand need to be delivered day by day, that is, traded day by day [8].

After establishing a buying and selling position (referred to as an open position), it is not necessary to wait until it expires. Instead, a reverse transaction can be conducted at any time before the expiration of the stock index futures contract to offset the original position, which is called a close position [9]. The basic characteristics of futures are:

(1) Small cap game

In the futures market, you only need to pay a margin of 5% to 10% to trade several or even dozens of prices. In futures trading, the "leverage effect" of the margin system determines the characteristics of its "small cap game", allowing traders to make larger trades with less capital, thereby saving more working capital [10].

(2) Bilateral trade

In the futures market, there are two forms of buying and selling, with great flexibility.

(3) No need to worry about compliance

All futures transactions are settled through futures exchanges, which are trading partners for both buyers and sellers, providing guarantees for each transaction. This way, traders don't have to worry about completing the transaction [11].

(4) Improve market transparency

The trading information is completely public and conducted in the form of public quotations, allowing all traders to compete openly and fairly [12].

(5) Well organized and efficient

Futures trading are a standardized transaction that has certain operating procedures and rules, with each link being linked to each other, and each link can be effectively operated. Generally, a transaction can be completed in just a few seconds [13].

At the same time, it also has five major trading characteristics: contract standardization, centralized trading, two-way trading, and hedging. Among them, the standardization of futures contracts provides great convenience for futures trading, as both parties no longer need to negotiate specific terms of the transaction, thereby saving trading time and reducing trading disputes [14-15]. Trading centralization refers to the fact that futures trading must be carried out within a futures exchange, which means that the futures market is a highly organized market with strict management systems implemented, and futures trading is ultimately concentrated within the futures exchange. The bidirectional trading and hedging mechanisms of futures exchanges have attracted many futures speculators to the futures market. In the futures market, speculators have two different profit opportunities. When futures prices rise, they can buy low and sell high to earn profits, and when futures prices fall, they can sell high and buy low to earn profits. Moreover, due to the existence of hedging mechanisms, speculators can avoid the troubles caused by physical delivery, thus their participation greatly improves liquidity in the futures market [16].

3. Big Data

Regarding "big data", some research institutes have defined it as requiring new processing methods to adapt to a large number, high growth rate, and diversified information assets, thereby possessing stronger decision-making, discovery, and process optimization capabilities.

ISSN 2616-5902 Vol. 5, Issue 14: 82-87, DOI: 10.25236/AJBM.2023.051415

The McKinsey Global Research Institute defines it as an enormous collection of data that goes beyond the capabilities of conventional databases in acquiring, storing, managing and analyzing. It has the features of big size, rapid data stream, multiple kinds of data, and low density of values [17].

Instead of handling huge quantities of data, the GDI has a significant effect on the data that includes a lot of semantic information. In other words, when you compare large data with an industry, you have to improve your "handling ability" so you can "process" it and add value to it [18].

From a technological point of view, Big Data and Cloud Computing are inseparable, just like two sides of the same coin. You can't handle that amount of data by yourself, so you'll need a distributed structure. The feature of this system is that it can deal with a great deal of distributed data mining. This, however, is dependent on Cloud Computing, Distributed Data Base, Cloud Storage and Virtualization [19].

Along with the coming of the Cloud Computing Age, Big Data is getting more and more concerned. Analysts say the word "Big Data" is commonly applied to the mass of unstructured or semi-structured data produced by a corporation, which takes time and money to be analyzed in relation to a relational database.

Large data needs to be handled efficiently with a large volume of data. Large data-handling techniques are: Large scale Parallel Database, Data Mining, Distribution File System, Distributed Database, Cloud Computing Platform and Scalable Memory System.

Large data consists of structural, semistructural and non-structural data, among which the bulk is non-structured. Big Data is only a kind of phenomenon or feature of the web's progress so far, it doesn't have any divinity or reverence. Along with the development of cloud computing and other techniques, the collection and utilization of data that used to be hard to gather and use is getting more and more convenient. Along with each profession's constant innovation, Big Data will be more and more valuable to us.

Second, for the sake of systemic knowledge, we need to make a thorough analysis on the basis of 3 aspects.

Firstly, it is the essential way for human beings to know something, and it is also a standard that is generally accepted and transmitted by humans. We analyses the value of Big Data from the perspective of Big Data, which can give us an insight into the development of Big Data, and Study Based on Large Data Privacy to investigate the Long Game of Human and Data.

The second layer is the technique, it is a way to achieve the value of large data, and it is also the basis of the large data exploitation. In this study, we can expect to extend the scope of collecting, handling, storing, analyzing, and other large data from the cloud, distribution, storage, and sensing fields.

Thirdly, it's practical, and that's what Big Data is all about. This is where we describe large amounts of information about the Internet, from governments, businesses, and individuals, and about the wonderful environment and future Big Data Map [20].

4. Experiment on Using Big Data for Price Prediction

The price fluctuations of futures trading are not only related to past prices, but also influenced by many other factors, including supply and demand relations, macroeconomic factors, related markets, and so on.

Research on price lag models and establish the relationship between price volatility, trading volume, and holdings. Obtain a model as shown in formula (1):

$$\delta = \lambda + \sum_{n=1}^{m} \eta_n \hat{\mu}_{i-n} + \sigma RW_i + \rho RU_i + \omega_i \tag{1}$$

Among them, δ represents the estimation of price fluctuations per unit time; RWi and RUi represent trading volume and holding volume, respectively. The price fluctuation is an estimate of German Klass fluctuation, calculated using formula (2):

ISSN 2616-5902 Vol. 5, Issue 14: 82-87, DOI: 10.25236/AJBM.2023.051415

$$\hat{\delta} = \left\{1.5 \times \left[\log\left(Q_{a,b}/Q_{a,c}\right)\right]^2 - \left[2\log(2)-1\right]\left[\log\left(Q_{a,u}/Q_{a,v}\right)\right]^2\right\}^{1/2} \quad (2)$$

Among them, Qab and Qac represent the maximum and minimum transaction prices per unit time, while Qau and Qav represent the opening and closing prices per unit time.

In order to refine the predicted results, the experiment decided to use agricultural products as the prediction object and conduct price prediction through a combination of long-term and short-term methods. Firstly, the short-term prediction effect was measured using 15 trading days, and then the long-term prediction effect was measured using 150 trading days. When training the prediction model, use the information data from the trading day to predict the settlement price for the next trading day. The predicted price on the following trading day is used as the output variable in the experiment. Firstly, the correlation coefficient between the characteristic variable and the output variable is calculated. The results are shown in Table 1. Due to significant differences in magnitude between the characteristic variables, standardization is performed before the prediction to eliminate the differences. When the correlation coefficient is below 0.5, it indicates a low linear correlation. From Table 1, it can be seen that except for the three variables of the Yangtze River average price, the previous period return index, and the previous period price index, which have a strong linear relationship with the settlement price on the next trading day, all other variables have a weak linear relationship with the target, which also indicates that linear models are not suitable for predicting futures settlement prices.

Table 1: Correlation coefficient between input variables and next day settlement price.

Classification	Variable name	Mean	Standard deviation	Correlation coefficient
Supply and demand relationship	Market volume	277334.89	188352.913	0.213
	Market open interest	216315.33	63759.641	0.413
Macroeconomics	A-share index	2836.940	573.635	-0.279
	Dow Jones index	19524.353	5483.566	-0.251
	Nasdaq index	5633.461	2596.318	-0.183
	S&P index	2230.495	688.647	-0.277
	Exchange rate	6.765	0.375	-0.176
Related markets	Yangtze River average price	50423.665	8366.379	0.991
	Previous income index	4885.804	674.398	0.985
	Previous price index	3310.523	531.379	0.983

According to the analysis of the correlation coefficient, three variables with a correlation coefficient exceeding 0.5 with the settlement price on the next trading day can be input variables into the model. Although the remaining variables have weak linear relationships, they also have a certain degree of correlation. When the number of input variables in the big data model is large, it is easy to cause overfitting problems. Therefore, principal component analysis is used to reduce the dimensions of the remaining variables, and some analysis results are given. According to the results of principal component analysis of variance, and in combination with the principle that the eigenvalue is greater than 1, five principal components are selected from the experiment, and their cumulative contribution rate reaches 74.652%, indicating that these five principal components can better reflect the information of all variables. The details are shown in Table 2.

Table 2: Principal Component Variance Contribution.

Composition	Eigenvalues	Variance contribution rate (%)	Cumulative contribution rate (%)
1	6.537	32.436	32.436
2	3.395	16.135	48.571
3	2.620	13.879	62.450
4	1.318	6.879	69.329
5	1.194	5.323	74.652

Next, the experiment introduced characteristic variables into the big data prediction model for 5 price predictions, with accuracy as the main evaluation indicator and recall as the secondary evaluation indicator. The results are shown in Figures 1 to 2.

ISSN 2616-5902 Vol. 5, Issue 14: 82-87, DOI: 10.25236/AJBM.2023.051415

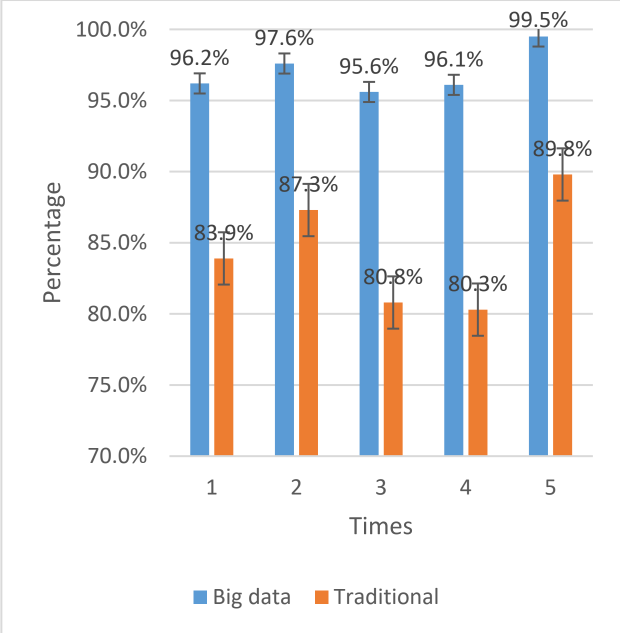

Figure 1: Comparison of Price Prediction Accuracy.

From Figure 1, it can be seen that the accuracy of futures trading price prediction based on big data reached the highest level at 99.5% in the fifth round, and the lowest level at 95.6% in the third round, with an average accuracy of 97% in the fifth round; The futures trading price prediction based on traditional prediction methods has the highest accuracy of 89.8% in the fifth round, and the lowest accuracy in the fourth round, with an accuracy of 80.3% and an average accuracy of 84.4%.

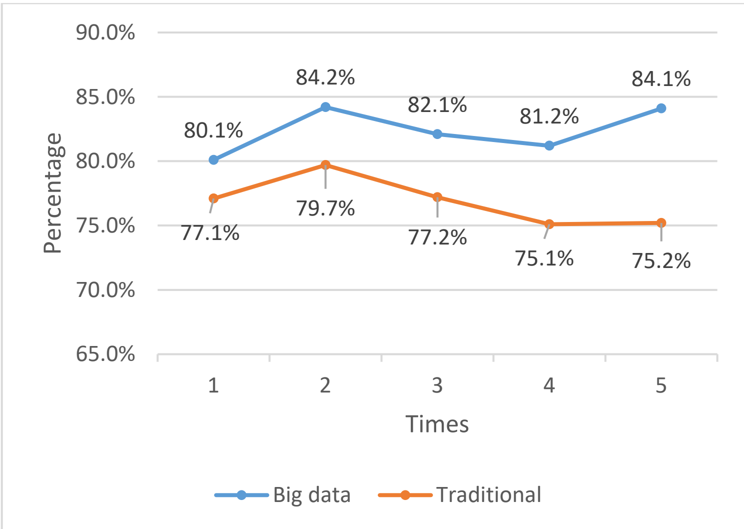

Figure 2: Comparison of Price Forecast Recall Rates.

From Figure 2, it can be seen that the recall rate of futures trading price prediction based on big data reached the highest in the second round, with a recall rate of 84.2%, and the lowest in the first round, with a recall rate of 80.1% and an average recall rate of 82.3; The futures trading price prediction based on traditional prediction methods reached its highest level in the second round, with a recall rate of 79.7%, and reached its lowest level in the fourth round, with a recall rate of 75.1% and an average accuracy of 76.9%.

Based on the above results, the final conclusion can be drawn that applying big data analysis methods to futures trading price prediction can not only improve the accuracy of prediction, but also increase the recall rate of prediction, which will have significant assistance for future price prediction.

ISSN 2616-5902 Vol. 5, Issue 14: 82-87, DOI: 10.25236/AJBM.2023.051415

5. Conclusions

Nowadays, big data has been applied in many fields, and futures trading are also growing day by day. Its low risk and fast trading characteristics are superior to stocks, so price prediction for futures trading is particularly important. This article uses big data analysis to evaluate the price prediction results of futures trading from two aspects: prediction accuracy and prediction recall. The final conclusion indicates that big data has a significant effect on improving the prediction accuracy and recall rate of futures trading. However, the article still has shortcomings, such as selecting fewer variables in the feature variable section, and future research will consider introducing more variables.

References

[1] Reznik, N. P., S. V. Dolynskyi, and N. Y. Voloshchuk. "Retrospective analysis of basic risk as a part futures trading in Ukraine." International Journal of Scientific and Technology Research 9.1 (2020): 3419-3423.

[2] Yang Haochang, et al. "The impact of exchange rate futures fluctuations on macroeconomy: Evidence from ten trading market." Emerging Markets Finance and Trade 58.8 (2022): 2300-2313.

[3] Luo, Jiawen, et al. "Forecasting realized volatility of agricultural commodity futures with infinite Hidden Markov HAR models." International Journal of Forecasting 38.1 (2022): 51-73.

[4] Jiang, Ying, Neil Kellard, and Xiaoquan Liu. "Night trading and market quality: Evidence from Chinese and US precious metal futures markets." Journal of Futures Markets 40.10 (2020): 1486-1507.

[5] Oglend, Atle, and Hans‐Martin Straume. "Futures market hedging efficiency in a new futures exchange: Effects of trade partner diversification." Journal of Futures Markets 40.4 (2020): 617-631.

[6] He, Feng. "Price discovery and spillover dynamics in the Chinese stock index futures market: a natural experiment on trading volume restriction." Quantitative Finance 20.12 (2020): 2067-2083.

[7] Avellaneda, Marco. "Trading signals in VIX futures." Applied Mathematical Finance 28.3 (2021): 275-298.

[8] Aysan, Ahmet Faruk, Behar Sadriu, and Humeyra Topuz. "Blockchain futures in cryptocurrencies, trade and finance: a preliminary assessment." Buletin Ekonomi Moneter Dan Perbankan 23.4 (2020): 525-542.

[9] Yang Jian, Zheng Li, and Tao Wang. "Price discovery in Chinese agricultural futures markets: A comprehensive look." Journal of Futures Markets 41.4 (2021): 536-555.

[10] Wellum, Caleb. "Energizing Finance: The Energy Crisis, Oil Futures, and Neoliberal Narratives." Enterprise & Society 21.1 (2020): 2-37.

[11] Liwang, Minghui, Zhibin Gao, and Xianbin Wang. "Let's trade in the future! A futures-enabled fast resource trading mechanism in edge computing-assisted UAV networks." IEEE Journal on Selected Areas in Communications 39.11 (2021): 3252-3270.

[12] Couleau, Anabelle, Teresa Serra, and Philip Garcia. "Are corn futures prices getting "jumpy"?." American Journal of Agricultural Economics 102.2 (2020): 569-588.

[13] Sharma, Arpita, and Shailesh Rastogi. "Spot volatility prediction BY futures and options: an INDIAN scenario." International Journal of Modern Agriculture 9.3 (2020): 263-268.

[14] Rastogi, Shailesh, Vikas Tripathi, and Sunaina Kuknor. "Informational role of futures volume for spot volatility." Pacific Accounting Review 34.1 (2022): 49-69.

[15] Shynkevich, Andrei. "Impact of bitcoin futures on the informational efficiency of bitcoin spot market." Journal of Futures Markets 41.1 (2021): 115-134.

[16] Chen Xinhua, Liu Jie. "Research and inspiration on Soybean meal futures price Prediction method based on ARIMA model." Rural South 36.2 (2020): 23-26.

[17] Yang, Jin. "Brief introduction of medical database and data mining technology in big data era." Journal of Evidence‐Based Medicine 13.1 (2020): 57-69.

[18] Wang, Jin. "Big data service architecture: a survey." Journal of Internet Technology 21.2 (2020): 393-405.

[19] Wang, C. Jason, Chun Y. Ng, and Robert H. Brook. "Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing." Jama 323.14 (2020): 1341-1342.

[20] Dai Hongning. "Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies." Enterprise Information Systems 14.9-10 (2020): 1279-1303.