

Research on Investment Strategy Based on Catboost Model and Bollinger Band Channel

Jiewen Zhong*

School of Economics, Guangdong University of Technology, Guangzhou, China, 510000

*Corresponding author: 18820829727@163.com

Abstract. As the economy continues to grow, more and more people tend to increase their income through asset investment. This paper devised a trading strategy and evaluated the value of an investment of \$1,000 after five years. The Catboost regression model was built to predict the price of gold and bitcoin with an accuracy rate of up to 99%. It showed that the model can effectively make predictions about gold and bitcoin prices. Trading signals for gold and bitcoin were established based on the Bollinger Bands channel theory. Simulated trading of assets began on September 11, 2016, and the total assets after five years were \$4,430,859.83. Then this paper conducted sensitivity analysis on the established trading model. Finally, the parameters of the established trading model are optimized to determine an optimal model parameter.

Keywords: Quantitative investing, Bollinger Band Channel, Catboost Model, Investment Strategy.

1. Introduction

1.1. Background

The emerging quantitative investment trading is a product of the mature development of computer science and capital markets. A quantitative strategy is a mechanized trading signal that does not require human judgment. It can replace people making judgments and plans. Therefore, quantitative investment has unique advantages in effectively controlling risks, improving decision-making efficiency, and achieving precise investment. In today's society, with the continuous development of the economy, more and more people tend to increase their income through asset investment. However, in the face of a large number of financial products in the market, how to accurately predict their price movements and maximize returns through the right buying and selling methods is what every market trader tends to know.

1.2. Literature Review

Al Khazali et al. [1] studied the impact of positive and negative macroeconomic news from large developed economies on the returns and volatility of gold and Bitcoin prices. Shahzad et al. [2] compared gold and bitcoin in the G7 stock markets and found that gold and bitcoin have obvious hedging and hedging characteristics. Zheng Zheng Li et al. [3] interpreted the volatility correlation between bitcoin, stocks, and gold in the context of uncertainty. Jingjing Li et al. [4] found that fluctuations in bitcoin prices and gold prices had a strong substitution effect.

The latest trend in asset prediction technology is the use of machine learning. Parmar I et al. [5] used regression and LSTM based machine learning to predict stock value. Mehtab S et al. [6] proposed a hybrid modeling method for stock price prediction, which involved establishing different models based on machine learning and deep learning. Chen W et al. [7] used a hybrid model based on machine learning for stock prediction.

In asset trading, trading strategies are often used to seek trading signals for more profitable transactions. Chen S et al. [8] used CSI to study 300 stock index futures and established a Bollinger band trading strategy for testing. Lehtoalho J [9] studied and optimized the standard deviation multiplier parameters of moving average length and Bollinger bands.

There was currently data on the trading prices of gold and bitcoin from September 11, 2016 to September 10, 2021. On each trading day, we will hold a portfolio consisting of cash, gold and bitcoin

represented by bitcoin. The heart of the problem was to build a quantitative trading decision model based on given data information, which we can use to determine whether we should buy, hold, or sell assets in their portfolio on a daily basis.

1.3. Main Work

First, we built the Catboost model to predict the next day's gold prices and bitcoin prices. We compared the predicted prices of these two assets with the real prices and evaluated the accuracy of the model results. It was determined whether the prediction model can be used to establish a trading model based on the accuracy of the model.

Then we set the trading signals for gold and Bitcoin by predicting prices based on the Bollinger band channel theory. We used trading signals to start simulating trading for gold and bitcoin on September 11, 2016. We got the cumulative amount of assets in gold and bitcoin five years later, and conducted sensitivity analysis on the commission cost of the model. Finally, the parameters of the established trading model were optimized to determine an optimal model parameter.

1.4. Arrangement

This paper was organized as follows. Section I introduced the research background, literature review, research problem and analysis of this paper. Section II presented the two theories used in this paper: Catboost algorithm, Bollinger Band channel theory, and proposed the basic assumptions of this paper. Section III built the Catboost model to predict the price of the asset and discussed the accuracy of price predictions. Section IV established trading signals for gold and bitcoin based on the Bollinger Bands channel, sensitivity analysis was conducted, and optimized the parameters of the model. Section V summarized the research questions, methods and results of this paper.

2. Preliminary

2.1. The Principle of the Catboost Algorithm

The Catboost model is a newer, iterative model with greater predictive power. As an ensemble learning algorithm of the Boosting family, it can use its excellent classification features and ordered boosting to greatly solve the problems existing in traditional statistical methods and individual machine learning algorithms[10]. The algorithm is based on symmetric decision trees, which has the characteristics of fewer parameters, support for categorical variables, and high accuracy. It handles categorical features efficiently and rationally[11].

Because its learning mechanism is through set learning and iterative learning, the effect of prediction is continuously improved, and the error of prediction is reduced, so that better prediction performance can be obtained.

2.2. Bollinger Band Channel Theory

Bollinger Bands are statistical charts used to indicate changes in the price of financial instruments over time[12]. Bollinger band is a technical analysis indicator that can be used to describe the volatility and range of stock prices. When analyzing stock price trends, the stock price average line is usually used as a reference line, and the Bollinger band adds two "stock price channel" lines to the average line[13].

The middle trajectory of the Bollinger band is the average of the stock price, and the upper and lower channels are the average of the stock price plus or minus a certain standard deviation multiple. The direction of the Bollinger Belt channel is mainly determined by the average line of the median. The bandwidth of the Bollinger band reflects the fluctuation range of stock prices.

2.3. Assumption

We made the following reasonable assumptions and conditional constraints based on the actual situation to construct a more accurate mathematical model.

Hypothesis 1: Assuming that the current prices of gold and bitcoin are strongly correlated with previous prices, we can use previous daily prices to predict future price trends.

Hypothesis 2: Do not considering systemic risk. Bitcoin and gold futures trading are subject to full market crashes, and we have constructed our investment strategy to not consider the possibility of such extreme scenarios occurring.

Hypothesis 3: Once we decide to trade, no matter how much volume we need to reach, we can make a deal instantly.

Hypothesis 4: We can trade any amount of gold or bitcoin on any day with the given price.

Hypothesis 5: The commission cost per transaction is $\alpha\%$ of the transaction amount, and there is no cost of holding the asset. ($\alpha_{gold} = 1\%$, $\alpha_{bitcoin} = 2\%$)

2.4. Notations

The symbol description of the paper was shown in Table 1.

Table 1. Notations

Symbol	Description
GP_n	The price of gold on the day
GFP_{n+1}	The forecast price of gold for the next day
BP_n	The price of bitcoin on the day
BFP_{n+1}	The forecast price of bitcoin for the next day
GR_{cost}	The cost of buying/selling gold
BR_{cost}	The cost of buying/selling bitcoin
μ_m	Bollinger band mid-band
$\overline{B_m}$	Bollinger band upper band
$\underline{B_m}$	Lower band of the Bollinger Band
σ_m	The standard deviation of an asset price
m	Average price days in the middle of the Bollinger band
a	A multiple of the standard deviation of the asset price
α_{gold}	Buy/sell rates for gold
$\alpha_{bitcoin}$	Buy/sell rates for bitcoin

3. Asset Prices Prediction Model Based on Catboost Algorithm

3.1. Modeling Steps

Step1: We used the Time Series Data Sliding Window transformation to convert the price of gold and the price of bitcoin into regression data.

Step2: The Catboost regression model was built from the training set data of gold price and bitcoin price.

Step3: The established Catboost regression model was applied to the training and testing data to obtain the prediction results of gold and bitcoin prices.

3.2. Result of the Model

We built the Catboost model to predict asset prices, and some of the prediction results were shown in Table 2 and Table 3.

Table 2. Prediction Results of Gold Price

Transaction date	The true price of gold	The predicted price of gold
2016/9/12	1324.6	1322.03478
2016/9/13	1323.65	1322.758262
2016/9/14	1321.75	1322.908007
...	...	...
2021/9/8	1786	1785.752867
2021/9/9	1788.25	1796.545304
2021/9/10	1794.6	1796.567144

Table 3. Prediction Results of Bitcoin Price

Transaction date	The true price of bitcoin	The predicted price of bitcoin
2016/9/12	609.67	624.1545642
2016/9/13	610.92	657.4203159
2016/9/14	608.82	624.1545642
...	...	...
2021/9/8	46809.17	46099.21816
2021/9/9	46078.38	45441.35032
2021/9/10	46368.69	45441.35032

The comparison between the real and predicted prices of gold and bitcoin were shown in Figure 1.

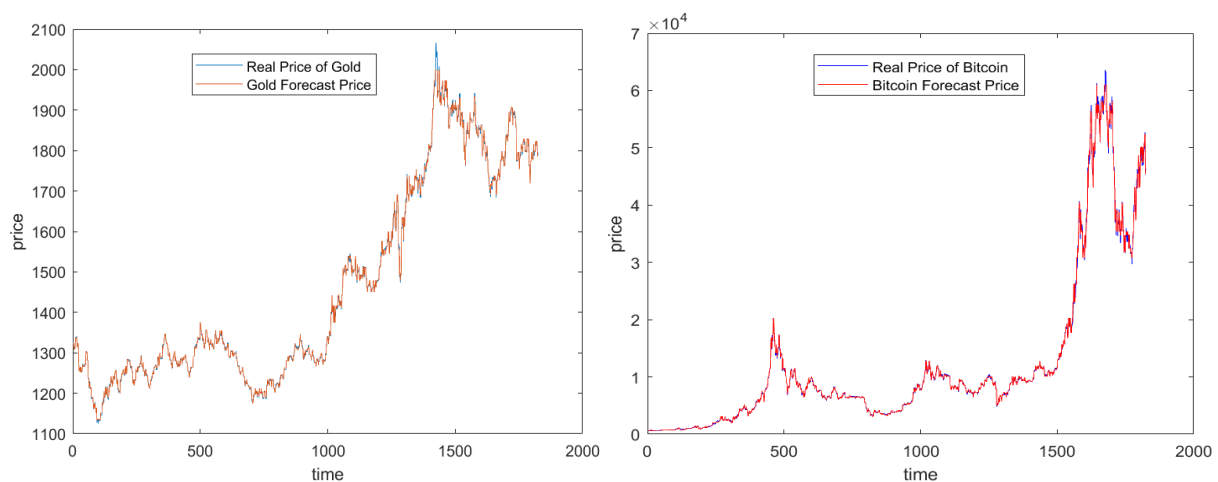


Figure 1. Comparison Chart of Real and Predicted Prices of Gold and Bitcoin

As shown in Figure 1, the true and predicted curves of gold and bitcoin prices were very close. This indicated that the Catboost model we had established can effectively predict test data.

3.3. Model Evaluation

After repeated iterative training of the Catboost model, we calculated the *MAPE* and *R²* evaluation indicators of the model.

Table 4. Model Evaluation Result

Asset		MAPE	<i>R²</i>
Gold	Training set	0.488	0.998
	Test set	0.516	0.998
Bitcoin	Training set	3.754	0.997
	Test set	3.517	0.996

The model evaluation results were shown in Table 4. *MAPE* (Mean Absolute Percentage Error) was a percentage value. The smaller the value, the more accurate the model was. Compared to the predicted value when only the mean was used, the closer the R^2 value was to 1, indicating that the model had a good fitting effect. The R^2 value of the training set for gold price and bitcoin price was 99%, and that of the test set was over 99%. This indicated that the model had a good fitting effect and high accuracy in predicting asset prices.

4. Investment Strategy Based on Bollinger Bands

The initial principal we had was \$1,000. We spent 1% of our initial stake on gold trading and the remaining 99% on bitcoin.

4.1. Calculation Method of Bollinger Band Channel

Let $p_1, p_2, \dots, p_m$ represent the asset price for the previous m days, the formula for calculating the middle band of the Bollinger band is

$$\mu_m = \frac{1}{m} \sum_{i=1}^m p_i \quad (1)$$

The upper Bollinger band is calculated as

$$\overline{B}_m = \mu_m + a\sigma_m \quad (2)$$

The lower Bollinger band is calculated as

$$\underline{B}_m = \mu_m - a\sigma_m \quad (3)$$

The standard deviation of an asset price is calculated as

$$\sigma_m = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - \mu_m)^2} \quad (4)$$

We initially chose $m=5$ and $a=1$.

4.2. Settings of Gold Trading Signals

The price of gold is relatively stable and less volatile. So when the forecast price of gold the next day is greater than the price of gold for that day, and the difference between the forecast price and the price of the day is greater than the forecast costs of buying, we buy gold. When the next day's forecast price minus the day's price is less than 0, the loss is greater than the sell forecast costs, and the predicted price breaks through the lower Bollinger band, we sell gold.

The buy signal is

$$GFP_{n+1} - GP_n > GR_{cost} \quad (5)$$

The set signal is

$$GFP_{n+1} - GP_n < GR_{cost} \quad (6)$$

$$GFP_{n+1} < \underline{B}_m \text{ and } GP_n > \underline{B}_m \quad (7)$$

where

$$GR_{cost} = \alpha_{gold} \times GFP_{n+1} \quad (8)$$

4.3. Settings of Bitcoin Trading Signals

The price of bitcoin fluctuates greatly. Therefore, buying and selling are set to be limited by Bollinger bands to prevent trading too often. When the predicted price of bitcoin the next day minus the price of the day is greater than 0, the profit is greater than the cost of buying, and the predicted price breaks through the upper Bollinger band, we buy bitcoin. We sell bitcoin when the next day's predicted price minus the day's price is less than 0, the loss is greater than the selling prediction cost, and the predicted price breaks through the lower Bollinger bands.

The buy signal is

$$BFP_{n+1} - BP_n > BR_{cost} \quad (9)$$

$$BFP_{n+1} > \underline{B_m} \text{ and } BP_n < \underline{B_m} \quad (10)$$

The set signal is

$$BFP_{n+1} - BP_n < BR_{cost} \quad (11)$$

$$BFP_{n+1} < \underline{B_m} \text{ and } BP_n > \underline{B_m} \quad (12)$$

where

$$BR_{cost} = \alpha_{bitcoin} \times BFP_{n+1} \quad (13)$$

4.4. Result of the Strategy Model

By means of our prediction and decision-making model, the asset accumulation curves of gold and bitcoin are shown in Figure 2.

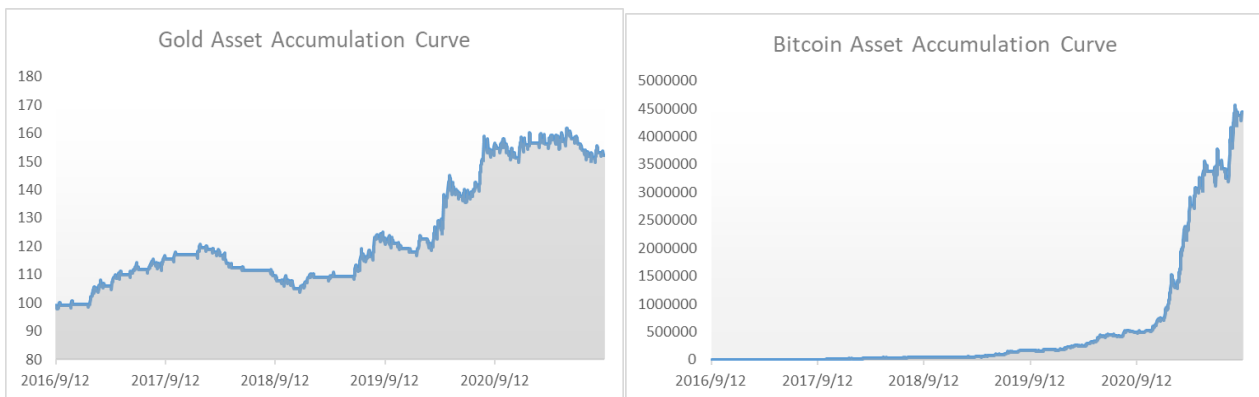


Figure 2. Gold Asset and Bitcoin Asset Accumulation Curve

We conclude that using 1% of the initial principal for gold trading and 99% of the initial principal for bitcoin trading, the final total assets will be \$4430859.83 by September 10, 2021. The resulting yield was 442985.98%. This gain was greater than the market gain. It showed that this trading model was suitable for trading gold and bitcoin, and the trading model worked well.

4.5. Sensitivity Analysis Based on Cost Changes

We analyzed the sensitivity of our established trading strategy to changes in gold and bitcoin trading commissions by changing the commission of one asset and keeping the commission of another asset alternating.

We kept the commission for one asset unchanged and set the buy/sell interest rates for the other asset at 1%, 1.5%, 2%, 2.5%, 3%, 3.5% and 4%, respectively. We incorporated them into our investment strategy model and obtained the final total values of the two assets as shown in Table 5.

Table 5. Sensitivity Analysis Table for Cost Changes

Buy/sell rates for gold	gold assets	Buy/sell rates for bitcoin	bitcoin assets
1%	152.1003	1%	30080969.2483
1.5%	145.2267	1.5%	10020210.9142
2%	134.1797	2%	4430707.7308
2.5%	130.4068	2.5%	1411139.8437
3%	126.5211	3%	729002.6626
3.5%	119.7875	3.5%	11310.0869
4%	113.1537	4%	1860.0128

The curve of gold cost change and the curve of bitcoin cost change are shown in Figure 3.

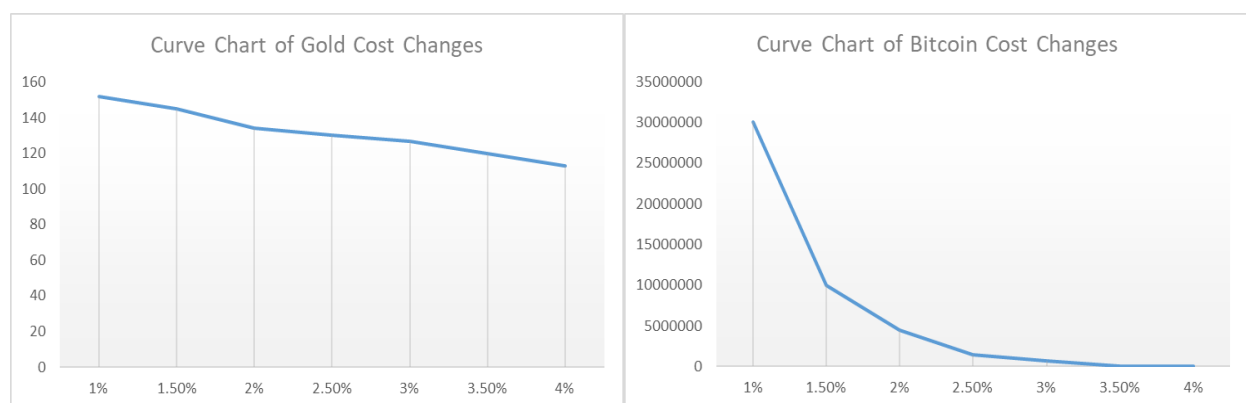


Figure 3. Curve Chart of Cost Changes in Gold and Bitcoin

From Table 5 and Figure 3, we can see that the cost of gold increased and the final value of gold assets showed a downward trend overall. However, in the end, the overall value of gold assets did not change much, indicating that this strategy was not sensitive to changes in gold costs. As the cost of bitcoin increased, the total assets of bitcoin showed a downward trend, and the decline was relatively significant. It indicated that the strategy was sensitive to changes in bitcoin costs. So when investing in bitcoin, we need to pay special attention to the changes in the commission cost of bitcoin.

4.6. Model Parameter Tuning

We tuned the model by adjusting the parameters m and a of the Bollinger Bands. We made m vary from 5 to 20 and a vary from 0.5 to 2. The resulting parameter optimization heatmap was shown in Figure 4.

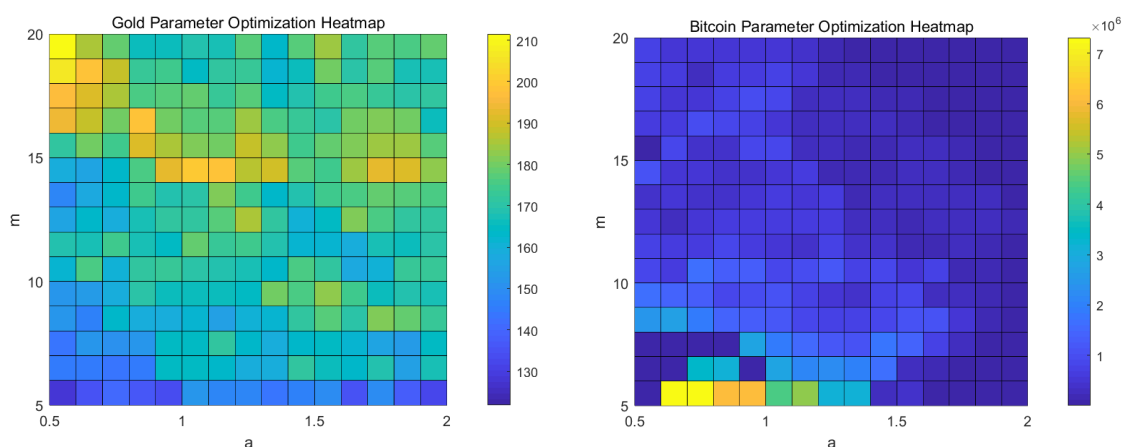


Figure 4. Gold and Bitcoin Optimization Heatmap

We found that in the parameter optimization of gold, $m=19$ and $a=0.5$ get the most returns. The final total gold assets were \$211.50. In the parameter optimization of bitcoin, $m=5$ and $a=0.6$ get the most returns. The final total bitcoin assets were \$7,310,266.48. So our maximum accumulated total equity can reach \$7,310,477.98 in this strategy.

5. Conclusion

This paper built a Catboost regression model to predict the price of gold and bitcoin, and the accuracy of the model was 99%. It showed that the model can effectively make predictions about gold and bitcoin prices. Based on the Bollinger Bands channel theory, this paper set up trading signals for gold and bitcoin based on the actual situation. Then we started demo trading on the asset on September 11, 2016, and got a total asset of \$4,430,859.83 five years later. This gain was greater than the market gain, which indicated that the trading mode was suitable for trading gold and bitcoin. Next, we conducted sensitivity analysis on the parameters of the established trading model. We found that this strategy was not sensitive to changes in the cost of gold, while it was relatively sensitive to changes in the cost of special currency. Finally, the parameters of the established trading model were optimized, and it can be found that the optimal parameters of the gold investment strategy were $m=19$ and $a=0.5$. The optimal parameters of the bitcoin investment strategy were $m=5$, $a=0.6$. The total equity after parameter tuning was \$7,310,477.98. Future studies could consider incorporating investors' investment risk into the investment strategy to optimize our model.

Reference

- [1] Al-Khazali, O. and Bouri, E. (2018) The Impact of Positive and Negative Macroeconomic News Surprises: Gold versus Bitcoin. *Economics Bulletin*, 38, 373-382.
- [2] Shahzad S J H, Bouri E, Roubaud D, et al. Safe haven, hedge and diversification for G7 stock markets: Gold versus bitcoin[J]. *Economic Modelling*, 2020, 87: 212-224.
- [3] Li Z Z, Su C W, Zhu M N. How does uncertainty affect volatility correlation between financial assets? Evidence from Bitcoin, stock and gold[J]. *Emerging Markets Finance and Trade*, 2022, 58(9): 2682-2694.
- [4] Li J, Rao X, Li X, et al. Gold and Bitcoin Optimal Portfolio Research and Analysis Based on Machine-Learning Methods[J]. *Sustainability*, 2022, 14(21): 14659.
- [5] Parmar I, Agarwal N, Saxena S, et al. Stock market prediction using machine learning[C]. 2018 first international conference on secure cyber computing and communication (ICSCCC). IEEE, 2018: 574-576.
- [6] Mehtab S, Sen J, Dutta A. Stock price prediction using machine learning and LSTM-based deep learning models[C]. *Machine Learning and Metaheuristics Algorithms, and Applications: Second Symposium, SoMMA 2020, Chennai, India, October 14–17, 2020, Revised Selected Papers 2*. Springer Singapore, 2021: 88-106.
- [7] Chen W, Zhang H, Mehlawat M K, et al. Mean–variance portfolio optimization using machine learning-based stock price prediction[J]. *Applied Soft Computing*, 2021, 100: 106943.
- [8] Chen S, Zhang B, Zhou G J, et al. Bollinger bands trading strategy based on wavelet analysis[J]. *Applied Economics and Finance*, 2018, 5(3): 49-58.
- [9] Lehtoaho J. Optimizing Bollinger band parameters: Individual stock and portfolio approach[J]. 2018.
- [10] Cai Ping A Study on the Arbitrage Strategy of Oil Futures Based on CatBoost Model [D]. Southwest University of Finance and Economics, 2021.
- [11] CHANG Wenfeng, WANG Xiao, YANG Jing, QIN Tao. Ecological suitability assessment model of blueberry based on CatBoost algorithm[J]. *Computer and Digital Engineering*, 2023, 51(01): 206-212.
- [12] Chen C H, Lai W H, Hung S T, et al. An advanced optimization approach for long-short pairs trading strategy based on correlation coefficients and bollinger bands[J]. *Applied Sciences*, 2022, 12(3): 1052.
- [13] Li Qiang, Chen Yanjiao. Research on the Quantitative Investment Strategy of Guiyang Bank's Stocks Based on the Bollinger Belt [J]. *Science and Industry*, 2022-22 (06): 189-193.