

Analysis of Capital Flow in Commodity Futures Market Based on SVM

Zi-ang Lin¹, Shaozhen Chen², Hongtao Liang¹ & Hong Zhang¹

¹ Financial Management of International Business School, Jinan University, Zhuhai, Guangdong Province, China

² Finance Department of International Business School, Jinan University, Zhuhai, Guangdong Province, China

Correspondence: Shaozhen Chen, Finance Department of International Business School, Jinan University, Qianshan Road 206#, Zhuhai City, Guangdong Province, Post No. 519070, China. Tel: 86-131-1235-5361. E-mail: 1813012994@qq.com

Received: May 17, 2018

Accepted: June 5, 2018

Online Published: June 28, 2018

doi:10.5539/ijef.v10n8p28

URL: <https://doi.org/10.5539/ijef.v10n8p28>

Abstract

Commodity futures are futures contracts based on the physical commodities. Unlike commodity stocks, which must be “bought first and then sold”, commodity futures can also be “sold first and then bought”. Therefore, it is not possible to directly use the formula of capital flow in the stock market to characterize the capital flow in futures contracts. In this paper, the principal component analysis method is used to construct the principal component factors based on the K-line basic market data and one based on the K-line index data. Then the factors mentioned above are cross-validated using the Holdout verification form to generate the training set and test of the support vector machine. Then, this paper applies genetic algorithm to optimize the penalty parameters and kernel functions of SVM, and obtains the parameters with the highest accuracy of classification and prediction of capital flow. Finally, this paper uses the traversal algorithm to find the time window with the highest accuracy of the SVM classification to predict the capital flow. The research results of this paper show that the SVM-based classification of capital flow in commodity futures market is highly accurate.

Keywords: commodity futures, capital flow, principal component analysis, support vector machine, genetic algorithm, traversing algorithm

1. Introduction

Commodity futures are the futures contracts based on the physical commodities. Commodity futures have a long history and a wide range of types. As early as the ancient Greek period in ancient Greece, there had been a few central trading venues. China's commodity futures market was born in October 1990. With more the development of more than 20 years, the role of commodity futures has continued to increase in the national economy, and its trading system and related laws have also been constantly improved. At present, there are three futures exchanges in China: Shanghai Futures Exchange, Dalian Commodity Exchange and Zhengzhou Commodity Exchange.

At present, China's futures market is facing a golden period of development, and commodity futures investment has attracted more and more investors' attention. As of the end of November 2016, the cumulative volume of the national futures market was 3.85 billion lots throughout the year, and the cumulative turnover was 178.16 trillion yuan, which increased 32.16% and 31.67% over the same period of 2015. The variety of futures trade has also evolved from the first few varieties to about 20 species now, including agricultural products, metal products, chemical products, and forestry products.

The futures market has played two extremely important functions, namely price discovery and risk avoidance. It goes without saying that the correct grasp of commodity futures and the forecast of the price rise and fall of commodity futures will help us to take advantage of these two functions to seize the opportunities and avoid disadvantages. However, studies at home and abroad mainly focus on the prediction of stock index futures prices, and there are still not enough studies on commodity futures. In this context, this paper selects Shanghai Silver as an example, adopts a quantitative approach, analyzes the law of capital flow of commodity futures, and establishes a capital flow model of commodity futures market based on SVM method. This will provide investors with certain references and suggestions. It will also help the country to formulate relevant economic

policies.

1.2 Relevant Scholarship

Research on commodity futures first began abroad. Working (1949) has the earliest mention of futures arbitrage on the futures storage. It is because of the existence of arbitrage that futures prices are more complex than other financial products. Working believes that the core of hedging is whether or not to find the changes of basis to seek profit, that is, the changes in the price difference between the futures market and the spot market to find opportunities for hedging. Black (1976) explicitly points out that the futures market is forward-looking, and futures prices include future expectations of spot prices.

In terms of capital flow, James and Richard (2001) found strong evidence of “money flow momentum,” in that lagged money flows can be used to predict future money flows. Furthermore, they found that money flows appeared to predict cross-sectional variation in future returns. Andrea and Owen (2007) used mutual fund flows as a measure of individual investor sentiment for different stocks, and found that high sentiment predicts low future returns.

Domestic research on commodity futures mainly focuses on quantitative trading strategies. Anhua (2005) makes an empirical study of the above three hedging trading strategies by using the historical transaction data of China’s soybean futures, and finally found that hedging strategy based on traditional hedging strategies was the least effective while the one based on HKM strategy was the most effective. Minimum variance strategy and HKM strategy were both more effective than the traditional strategy to reduce the risk of hedging. Yiqian (2013) found that the benefits of the price momentum strategy came from the large trend of large commodity bull and bear market and the plate trend caused by seasonal factors. Industrial products are closely linked to the macro-economy, and the cycle of price fluctuations is relatively long. The large trend of more than 1 year often appear, but such phenomena in the plate trend are fewer. Zhihong (2017) constructed a “short-term” quantitative investment strategy based on the Hurst Index, Turtle Rules, and Bollinger Bands, empirically showing that the strategy can obtain excess returns. Xiaojian and Qianqian (2018) constructed a paired trading strategy based on the OU process. By seamlessly splicing the data of different main contract data of commodity futures, a combination of selections can achieve a higher rate of return; whether in success rate, profitability or the possible loss of investment, the correlation pairing method is better than the SSD pairing method.

In general, domestic and foreign scholars’ research mainly focuses on the arbitrage of commodity futures. The analysis of capital flow is also limited to the stock market, and the method of machine learning is rarely used to analyze capital flow. This paper mainly uses SVM to study capital flow of commodity futures market. Firstly, the principal component analysis method is used to construct the principal component factor based on the K-line basic market data and the K-line index data. Then the above factors are cross-validated using the Holdout verification form to generate the training set and test set of the support vector machine. Then, we use the genetic algorithm to optimize the penalty parameters and kernel functions of SVM and obtain the highest accuracy of classified prediction of capital flow. Finally, this paper uses the traversal algorithm to find the time window with the highest accuracy of SVM classified prediction of capital flow.

2. Theoretical Models

2.1 Technical Analysis

The technical analysis method is a method used to analyze the trend according to the change rule of the price chart and the K-line technical indicators. In this paper, the support vector machines method are used to analyze the capital flow of commodity futures. What’s more, the K-line technical indicators of the technical analysis method are introduced to construct a set of new input vectors. The selected technical indicators and the relevant calculation formulas are shown in Table 1:

Table 1. Indicators and description

Indicator	Formula	Terminology
MA	$MA_n = \frac{\sum_{i=1}^n C}{n}$	MA_n is the simple average of the closing price of the n day; C is the closing price of the n day; n is the number of trading days.
MACD	$\begin{cases} MACD = EMA_{12} - EMA_{26} \\ EMA_n = \frac{EMA_n \times (n-1)}{(n+1)} + \frac{C \times 2}{(n+1)} \end{cases}$	EMA_n is The moving average of the index of the n days; EMA' is the moving average of the index of the $(n-1)$ day; C is the closing price of the n day.

<i>BIAS</i>	$BIAS = \frac{(C - MA_n)}{MA_n}$	<i>C</i> is the closing price of the <i>n</i> day. <i>MA_n</i> is the simple average of the closing price of the <i>n</i> day.
<i>OBV</i>	$OBV = OBV' + \text{sgn} \times V$	<i>OBV'</i> is the value of <i>OBV</i> of the previous day; <i>sgn</i> determines a value is positive or negative. <i>V</i> is the volume of futures traded today (the number of futures contracts) When <i>sgn</i> =+1, today's closing price is larger than yesterday's closing price. When <i>sgn</i> =-1, today's closing price is less than yesterday's closing price.
<i>CCI</i>	$CCI = \frac{\left(\frac{H+L+C}{3} - MA_n\right)}{\left(\frac{\sum_{i=1}^n (MA_n - C)}{n}\right) / \varphi}$	<i>H</i> is the highest price of the <i>n</i> day; <i>L</i> is the lowest price of the <i>n</i> day; <i>C</i> is the closing price of the <i>n</i> day; <i>φ</i> is a coefficient equal to 0.015.
<i>RSI</i>	$RSI = \left[\frac{\sum_{i=1}^n \Delta C_n^\alpha}{\sum_{i=1}^n \Delta C_n^\alpha + \sum_{i=1}^n \Delta C_n^\beta} \right] \times 100$	$\sum_{i=1}^n \Delta C_n^\alpha$ is the sum of the rising price changes in the <i>n</i> day; $\sum_{i=1}^n \Delta C_n^\beta$ is the sum of the falling price changes in the <i>n</i> day.
<i>KDJ</i>	$\begin{cases} RSV(n) = \frac{C-L}{H-L} \times 100 \\ K = 2/3K' + 1/3RSV \\ D = 2/3D' + 1/3K \\ J = 3D - 2K \end{cases}$	<i>C</i> is the closing price of the <i>n</i> day. <i>L</i> is the lowest price of the <i>n</i> day. <i>H</i> is the highest price of the <i>n</i> day. <i>D'</i> is the value of <i>D</i> of the previous day.

2.2 Support Vector Machine

In support vector machine, the four most commonly used functions at present are Sigmoid kernel function, linear kernel function, polynomial kernel function, and Gaussian radial basis kernel function. Because different kernel functions have different characteristics, the performance of classification tests using different kernel functions may be different.

(1) Sigmoid kernel function: $K(x, x') = \tanh(g \varphi(x, x') + c) \quad g \geq 0, \varphi(x) > 0, c < 0$

(2) linear kernel function: $K(x, x') = x^T x'$

(3) polynomial kernel function: $K(x, x') = (g x^T x' + c)^d, \quad g \geq 0, c \geq 0$

(4) Gaussian radial basis kernel function: $K(x, x') = \exp\{-g|x-x'|^2\}$

This paper compares these four kernel functions and chooses a support vector machine model based on Gaussian radial basis kernel function.

2.3 Genetic Algorithm

Genetic algorithm is a simulated evolutionary algorithm and it is an effective search algorithm to solve the optimization problem. In the optimization process, because the genetic algorithm basically does not need other auxiliary condition information and the knowledge of the search space, it has the advantages of self-adaptation and self-learning. So compared with many optimization algorithms, the genetic algorithm can solve more optimization problems, such as Non-differentiable, discontinuous, stochastic optimization and other issues. The basic flow of genetic algorithms is shown in Figure 1.

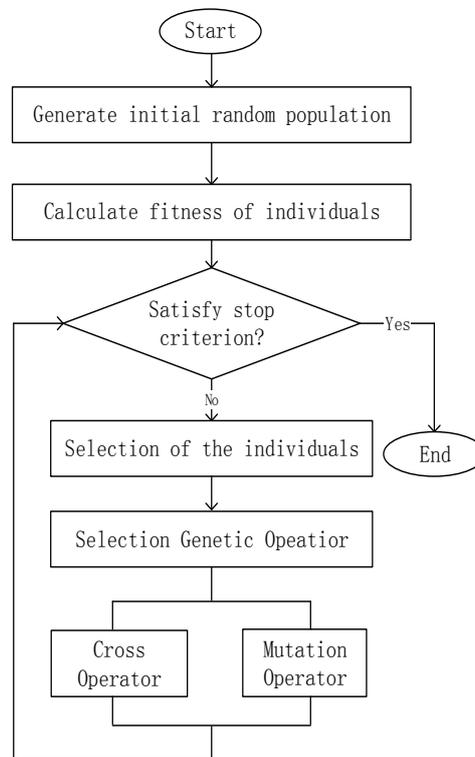


Figure 1. Flowchart of genetic algorithm

Step 1: Initialization: The random method generates the initial population and genetically encodes the population.

Step 2: Fitness evaluation: The fitness of each individual is calculated. When the fitness of the individual is stable, the chromosomes (solutions) carried by individuals in the current species population meet our requirements for optimal solutions.

Step 3: Individual selection: The individual choices are adaptive value proportional selection method, ranking selection method, and league selection method. The adaptive value proportional selection method is the nearest selection method in individual selection. Therefore, this method is selected by this method and its formula is as follows:

$$\rho_i = \frac{f(x_i)}{\sum_{i=1}^n f(x_i)} \quad (1)$$

Step 4: Crossover operation: By randomly selecting two individuals in the population, the chromosomes are assigned positions in accordance with a certain probability, so that chromosomes carried by the next generation of individuals become more excellent.

Step 5: Mutation operation: by using a certain probability to replace the coding of certain positions on the chromosome coding of a single individual with other codes to simulate gene mutations in the natural world, and to improve the ability of the genetic algorithm in local optimization.

3. Empirical Analysis

3.1 Data Resources

This paper uses the main contract of silver in Shanghai Futures Exchange as the analysis object, using the commodity futures minute data from May 10, 2012 to December 31, 2013 to construct the capital flow model of commodity futures based on SVM.

In this paper, the basic market data of K-line are selected, including the opening price (O), closing price (C), highest price (H), lowest price (L), exchange volume (V), turnover (T), open interest (P). the technical indicators' data are selected, including MA, MACD, BIAS, OBV, etc. The data are from the fifth Teddy Cup data Mining Contest.

3.2 Principal Component Analysis

According to the correlation analysis of K-line basic market data and K-line technical indicators' data, this paper finds that there exists a certain correlation between data. Therefore, this paper uses the principal component analysis method to reduce the input vectors of the K-line basic market data and K-line technical indicators' data respectively, and obtains the principal component factors. The sum of the variance contribution rate of these factors is in conformity with the standard requirements. The variance contribution rate is shown in Figure 2 and Figure 3.

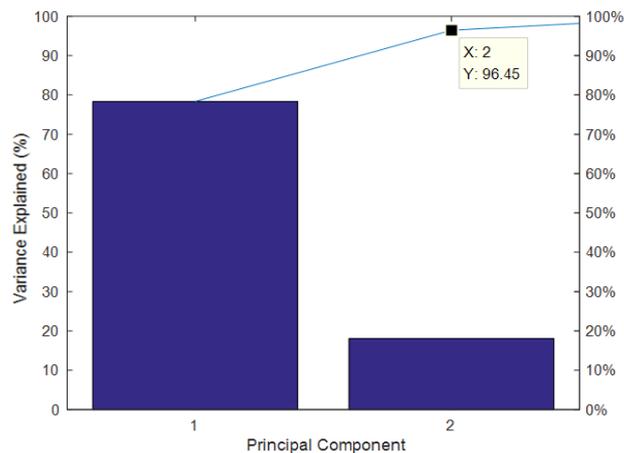


Figure 2. The principal component analysis of the K-line basic market data

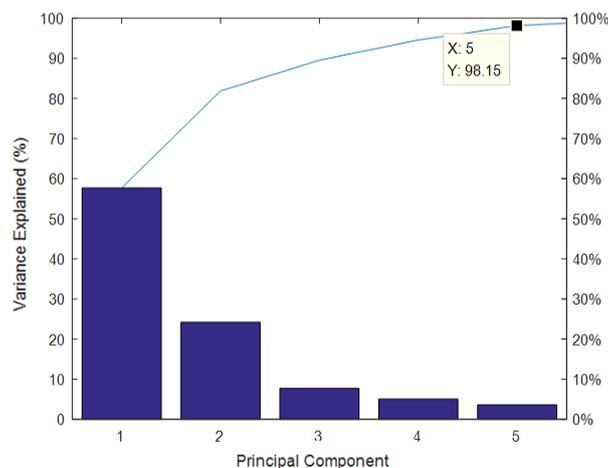


Figure 3. The principal component analysis of K-line technical indicator's data

In this paper, based on the condition that the variance contribution rate is greater than 95%, the original data for the SVM input vectors, we can obtain the K-line basic market principal component input vectors and the K-line technical indicators' principal component input vectors.

3.3 Cross Validation and Maximum-Minimum Normalization

The basic idea of cross validation is to group data sets by some standard, with one part as a training set and the other as a test set. The first step is to train the classifier using the training set, and then utilize the test set to test the training model and evaluate the classifier. This paper uses the Holdout verification to cross-validate the principal component input vectors, generating the training set and the test set.

In addition, in order to eliminate the dimensional impact between the indicators, the training set and the test set are subjected to the maximum-minimization normalization process. Take the input components of the principal components of the K-line technical indicators as an example. The normalization process of the training set and the test set is shown in Table 2 and Table 3.

Table 2. Normalization of the training set of technical indicators' principal component input vectors

Factor ₁	Factor ₂	Factor ₃	Factor ₄	Factor ₅
0.306015	1.646441	-0.875570	-0.044723	-0.221829
0.456310	-0.256457	0.734438	0.312558	-0.168353
-0.527510	0.413289	-0.576738	0.529379	0.690191
...
-0.655406	-2.096482	1.811881	-0.760243	0.543435
-0.908992	-1.323485	0.915296	-0.453550	0.932429
-0.834858	-1.880137	1.404727	-1.224413	0.561279

Table 3. Normalization of the test set of technical indicators' principal component input vectors

Factor ₁	Factor ₂	Factor ₃	Factor ₄	Factor ₅
0.483746	1.258975	-0.604903	-0.052598	-0.395255
0.342224	1.645443	-0.849691	-0.078249	-0.236997
0.405560	1.559991	-0.691992	-0.100916	-0.205009
...
-0.136623	-1.922203	1.877768	-0.771144	0.183785
-0.317444	-2.190380	2.083419	-0.903290	0.267445
-1.012070	-1.440633	0.837212	-0.881999	0.719244

3.4 Construction of the Capital Flow Model

After the SVM input vectors and output vectors are obtained, the SVM can be trained and tested.

In order to improve the classification accuracy, the genetic algorithm is used to optimize the penalty parameter *c* and kernel function *g* in SVM. Thanks to MATLAB, the relevant parameters are worked out and shown in Table 4, the model fitness is shown in Figure 4 and Figure 5.

Table 4. Comparison of capital flow models

Model	Capital Flow Model Based on K-Line	Capital Flow Model Based on K-Line Technical
	Basic data	Indicator's data
<i>c</i>	1.6896	22.7228
<i>g</i>	426.0239	0.1869
Accuracy of the training set	69.6203%	67.1958%
Accuracy of the test set	68.8037%	66.0142%

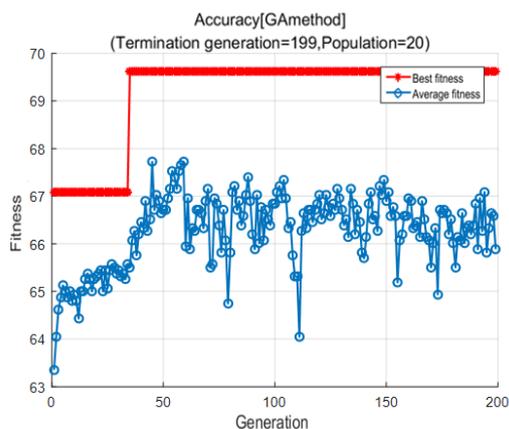


Figure 4. The capital flow model based on K-line basic data

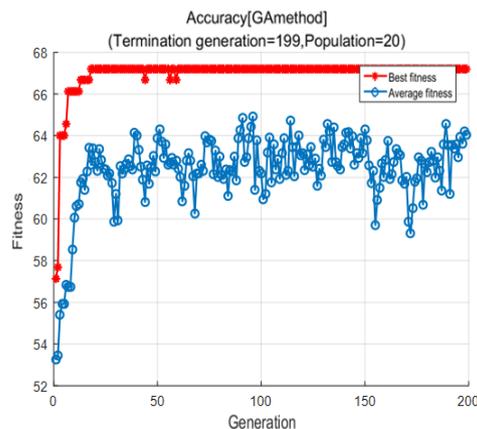


Figure 5. The capital flow model based on K-line technical indicator's data

It is obvious that the capital flow model based on K-line basic data and the capital flow model based on K-line technical indicators' data both have higher prediction accuracy rates, that is to say, these two models can well portray the capital flow of the futures market.

3.5 Optimization of Capital Flow Model

The data of the capital flow models established in 4.2 has a large amount scale and a long time period is also longer both in the training set and the test set. Therefore, it takes more time to calculate and construct the strategy. What's more, in reality, the capital flow in the futures market changes rapidly, and the applicability of the static capital flow models may be affected. Therefore, from the perspective of improving the speed of model calculation and shortening the cycle of model training data, traversal algorithms is used in this paper. The data cycle of the SVM-based capital flow model is optimized to find the cycle with the highest accuracy.

Taking n as the input scroll time window length of the input vectors, the optimal scroll time window length based on the principal component input vectors of K-line basic data and the K-line technical indicators' data can be obtained. The results are shown in Figure 6 and Figure 7.

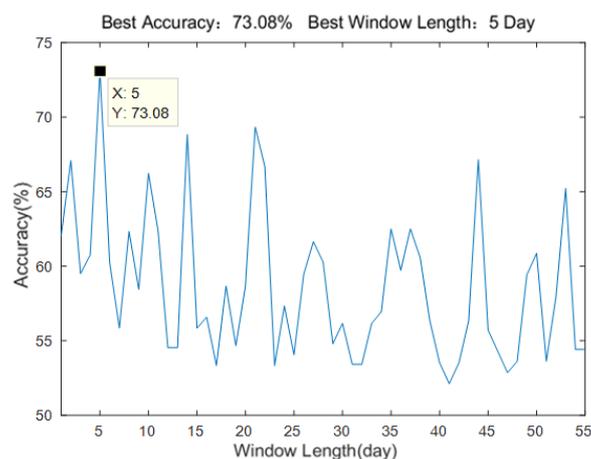


Figure 6. The optimal rolling time window length based on the principle component input vectors of K-Line basic data

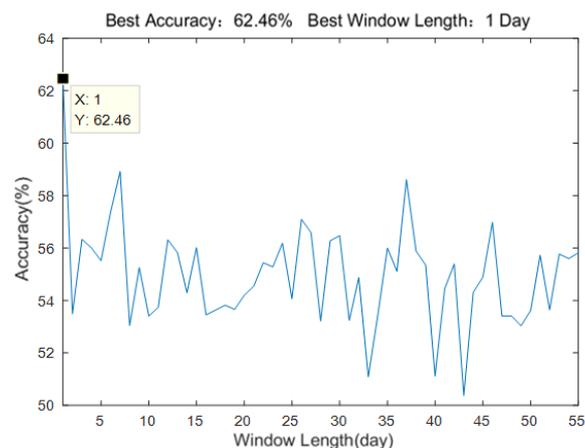


Figure 7. The optimal rolling time window length based on the principle component input vectors of K-Line technical indicators' data

It can be seen from Figure 6 that the principal component input vectors of K-line basic data has a maximum accuracy of 73.08% for classification prediction and an optimal rolling window length of 5 days. The principal component input vectors of K-line technical indicators' data has a maximum classification accuracy of 62.46% and an optimal rolling window length of 1 day.

4. Conclusion

The research of this paper has certain academic significance and practical significance.

Academic significance: 1) Existing studies rarely use SVM to analyze the flow of commodity futures funds. The study in this paper has increased the application of this aspect. In this paper, the kernel function is used to construct the model of the capital flow of commodity futures, instead of simply using a certain kind of kernel function to build the model. Through the optimization of the parameters of the genetic algorithm and the traversal algorithm window optimization, to build a capital flow model with stronger applicability and practicality; 2) Compared to the traditional method of forecasting the value of the commodity futures closing price, this paper uses the advantage of statistical classification of the support vector machine to predict the ups and downs of the closing price of commodity futures, which is not limited to the dependence of traditional methods on the accuracy of absolute numerical prediction.

Practical significance: This paper uses a support vector machine, principal component analysis, traversal algorithms, etc. to build a classification prediction model, based on which a model suitable for analysis of the capital flow of commodity futures is constructed. The application of this classification prediction model is not limited to commodity futures, it can also provide certain guidance for the production and business plan of the company, government economic policies, etc. It is beneficial for enterprises, self-employed individuals, etc. to use the futures market to hedge, preserve value and guide production. Taking agricultural product futures market as an example, this model helps to comprehensively and objectively grasp the supply changes, price changes and future trends of agricultural products, scientifically and rationally study and formulate agricultural development plans, guide agricultural standardized production and industrialized operations, and promote the modern development of agriculture.

References

- Anhua, Z. (2005). *Empirical research on hedging strategy of commodity futures*. Wuhan University, 2005.
- Bennett, J. A., & Sias, R. W. (2001). Can Money Flows Predict Stock Returns? *Financial Analysts Journal*, 57(6), 64-77. <https://doi.org/10.2469/faj.v57.n6.2494>
- Carmona, R., & Durrleman, V. (2003). Pricing and Hedging Spread Options. *SIAM Review*, 45(4), 627-685. <https://doi.org/10.1137/S0036144503424798>
- Frazzini, A., & Lamont, O. A. (2008). Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics*, 88(2), 299-322. <https://doi.org/10.1016/j.jfineco.2007.07.001>
- Working, H. (1949). The Theory of Price of Storage. *The American Economic Review*, 39(6), 1254-1262.
- Xiaojian, Y., & Qianqian, Z. (2018). Matching Trading Strategy of Commodity Futures Market Based on OU Process]. *Southern Finance*, 1-9. <http://kns.cnki.net/kcms/detail/44.1479.F.20180323.1727.002.html>
- Yiqian, W. (2013). *Research on Commodity Futures Trading Strategy Based on Momentum Effect*. Fudan University, 2013.
- Zhihong, L. (2017). *Research on the "short period" quantitative investment strategy of domestic commodity futures*. Zhejiang University, 2017.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).