Deep Reinforcement Learning for Pairs Trading: Evidence from Soybean Commodities

Abstract

This paper adopts a novel CA-DRL model in the pairs trading strategy of No.1 soybean futures, soymeal futures, and soyoil futures on the DCE (Dalian Commodity Exchange). Pairs trading strategies are comprised of two stages. First, the Co-integration method (CA) is applied to form pairs and the formation period for pairs is divided into one year, two years and three years. Second, the criteria for opening and closing positions are established using two methods to give the trading rules applied in the decision-making process, the Simple Threshold Method (ST) and Deep Reinforcement Learning (DRL). In this paper, we adopt four models, including the ST (Simple Threshold Method), the CA-ST (Co-integration method and Simple Threshold Method), the DRL and the CA-DRL (Co-integration method and Deep Reinforcement Learning). The experimental results show that the CA-DRL model has the best performance, whether the pair's formation period is one year, two years or three years, and China's non-genetically modified soybean futures can implement the pairs trading strategy between commodities.

Key words: Finance, pairs trading, DRL

1. Introduction

Pairs trading is a market-neutral strategy to obtain arbitrage profit when there is a co-integration relationship or high correlation between two stocks with statistical differences but similar characteristics (Huang, Huan, Xu, Zheng, & Zou, 2018). Due to its successful application in the 1980s, pairs trading has become a common arbitrage strategy and has been widely recognized. Gatev et al. (2006) first carried out a comprehensive study on the application of pairs trading strategy in the U.S. stock market. They pointed out that a simple pair trading strategy (PTS), namely the distance method (DM), can generate profits over an extended period. Much literature emerged following the work of Gatev et al. Follow-up research provides various insights for this field from different perspectives, always through the creation of other methods to identify potential pairs, including the Co-integration method (Vidyamurthy, 2004; Caldeira and Moura, 2013), time series method (Elliott et al., 2005; Cummins and Bucca, 2012), stochastic control (Jurek and Yang, 2007; Liu and Timmermann, 2013) and other methods (Xie and Wu, 2013; Wu, 2013; Stander et al., 2013; Xie et al., 2014).

The arbitrage strategy of pairs trading is to find a pair of asset portfolios with the same historical price change trend. Then, assuming that this equilibrium relationship will continue in the future, it is a question of continuously monitoring the spread between two assets; if it deviates from the historical average, investors short the overvalued asset and buy the undervalued asset. If their prices converge, investors can close their position to earn excess returns. The potential matching objects of pairs trading can be stocks, stock options or futures contracts. Ehrman (2012) proposed that pairs traders should fully consider and pay attention to the role of natural correlation before incorporating commodity futures into their portfolios. Natural correlations exist, not only between similar types of commodities, but also between the futures contracts for the same commodity in different expiration months. The spread of the Commodity Futures Portfolio with natural correlation often shows significant stability. Therefore, we choose to look at soybean futures, soyoil futures and soymeal futures as the asset selection objects for pairs trading. As they are the final products of soybean press, soyoil, soymeal and soybean have a very significant natural relationship.

China is the largest soybean-importing country globally, characterized by a low self-sufficiency rate and strong linkage between domestic and foreign soybean markets. About 87% of China's imported soybeans are used for soyoil pressing, and the remaining protein-rich soymeal is used in the feed industry. To reduce the price risk incurred by imported soybeans, on the one hand, China implemented the soybean revitalization plan to expand the planting area of domestic soybeans and on the other hand it insisted on promoting the diversification of soybean imports. American soybeans, and Brazilian soybeans came to China one after another. However, even with the record high of domestic and imported production this year, the price of soybeans keeps on rising. Simultaneously, in addition to the rising domestic price of soybeans, their international price is also high. At the beginning of November 2020, the price of American soybeans had been climbing for three months, from around \$8.70 / bushel to about \$10.50 / bushel; According to the relevant report issued by the U.S. Department of Agriculture in November 2020, the average price of American soybeans is expected to reach \$10.40 per bushel in 2020-2021. The significant fluctuation of the price of imported soybeans will cause big losses for Chinese soybean pressing enterprises.

Soybean futures are the most traded commodity futures in China. In terms of the trading volume of agriculture products futures and options, the Dalian Commodity Exchange (DCE) is the second largest soybean futures market after the Chicago Board of Trade (CBOT). In 2018 the trading volume of the No. 1 soybean futures contract of the DCE was 4.4 million tons, and in 2019 it was 3.7 million tons (Data resource: DCE). The DCE is the world's largest non-transgenic futures market for soybean. Up to now, there have been many kinds of research on the intra-commodity arbitrage strategy of soybean futures. Johnson et al. (1991), Barrett and Kolb (1995), Simon (1999) and Mitchell (2010) studied the intra-commodity arbitrage of soybean futures with squeezed margin or pricing relationship. Dunis et al. (2006), Wiles and Enke (2014) and Li et al. (2015) used a neural network to study the intra-commodity arbitrage of soybean futures. But without exception, research has focused on the soybean futures of CBOT. As the futures trading of the DCE matures, research literature on it is also increasing (Liu et al., 2016; Ruan et al., 2020). Brim (2020) applies the Deep Q Network (DQN) to the pairs trading strategy of the stock market to make a profit, applying the DRL method to pairs trading. Kim et al. (2019) use the deep reinforcement learning method with various trading and stop-loss boundaries to optimize the pairs trading strategy. The pairs are selected from the S & P 500 index stocks by using the Co-integration test.

At present, there is no such research to apply the DRL method to the pairs trading strategy of soybean and its derivatives futures. Therefore, this paper brings the soybean, soymeal and soyoil futures of the DCE into the asset selection pool of pairs trading, and applies the DRL algorithm to the pairs trading strategy. For the DRL, we make some improvements based on the traditional model. First, we use the Multi-Layer Perceptron (MLP) with a 7-layer neural network structure as the agent of the DRL algorithm. In addition, we use a traditional BP algorithm to train the MLP and the AdamOptimizer method to optimize the weight of the network. Moreover, to make the reward value of training positive or negative, and ensure there is no restriction on the size of the reward value, we use linear function as the activation function. In the early stage of the experiment, we train the experimental dataset repeatedly to avoid the problem that the connection weight of the MLP cannot converge when using a linear function. Furthermore, we use the pre-training method to prevent two potential problems in the actual training process: the massive amount of calculation, and the fact that the model can easily fall into local optimum.

By comparing with the traditional statistical arbitrage model (ST, CA-ST), we find that the DRL algorithm (DRL, CA-DRL) model performs better, whether the formation period for pairs is one year, two years or three years. Of all models, the CA-DRL performs best both in- and out-of-samples.

The contributions of this paper are as follows: On the one hand, for institutional and individual investors, this paper provides a quantitative investment strategy for reference, to spread the risk and obtain positive return when the market falls. Meanwhile, we use the DRL to optimize the trading signal of pairs trading. By taking a more reasonable way to select the optimal threshold of paired asset, we can decide whether to open or close the position at this time; On the other hand, for the whole futures market, the application of pairs trading strategy in China's soybean futures market is conducive to establish a more appropriate and effective price discovery mechanism. Investors' reasonable arbitrage will correct the deviation of the contract price and make the setting of the contract price more rational. At the same time, the research on the pairs trading of commodity futures can also help to improve the allocation efficiency of financial resources in the futures market.

2. Research data

This paper uses the daily closing price data of major contracts of No. 1 soybean, soymeal, and soyoil futures of the DCE. The data period is from January 9, 2006, to July 31, 2020. Meanwhile, we also use the main contract of No. 2 soybean futures to conduct the robustness test, for which the data cycle runs from January 4, 2016, to July 31, 2020. All contracts are trading on the same exchange and closing at identical times. A pairs trading strategy consists of two stages. The first stage is called the formation stage for pairs, and involves using appropriate methods to select the desired pairs from a pairs pool (total 3*2/2=3 pairs per period)), and using the selected pairs to construct the portfolio. Then, in the second stage, we make corresponding trading decisions for the selected pairs. The ST and DRL algorithm determine the trading rules adopted in the transaction. Finally, we use the trading rules for the out-of-sample period. Unlike in the study of Gatev, Goetzmann, and Rouwenhorst (2006), we formed a pair over 24 months (in-sample period) and trade over the next 12-month period (out-of-sample period). Simultaneously, to compare the arbitrage yield generated by using formation periods of different lengths of time, we also consider the pairs trading strategies involving matching within 12 and 36 months respectively.

Table 1: Data summary of commodities

Variable	Stationary	ADF	Std.Dev	Skewness	Kurtosis	Jarque-Bera
No.1 Soybean	0.000	0.090	0.012	0.021	7.905	0.000
Soymeal	0.000	0.025	0.016	-1.117	22.284	0.000
Soyoil	0.000	0.226	0.018	1.200	32.535	0.000
No.2 Soybean	0.000	0.079	0.023	-0.120	23.088	0.000

Note: If the p-value in the ADF test is lower than 1%, the commodity price series is considered stationary. The Std.Dev, Skewness, Kurtosis and Jarque-Bera tests are for the log return of commodities.

Table 1 summarizes the results of the statistical analysis of the data used in this paper. The results show that the daily closing price data of the central contracts of No. 1 soybean, soymeal, soyoil and No. 2 soybean futures are all non-stationary Series in the research period. The four groups of daily yield series do not conform to the normal distribution, and their kurtosis is much higher than the kurtosis of the normal distribution.

Table 2: Dickey-Fuller unit root test of first order difference series

Variable	Test statistics	at 0.01 level	at 0.05 level	at 0.10 level
Δsoybean_one	-41.975***	-3.43	-2.86	-2.57
Δsoymeal	-59.726***	-3.43	-2.86	-2.57
Δsoyoil	-18.928***	-3.43	-2.86	-2.57
Δsoybean_two	-11.180***	-3.43	-2.86	-2.57

Notes: *** significant at 0.01 level, * significant at 0.10 level.

As shown in Table 2, the four groups of daily closing price series are stable after the first difference. So far, our preliminary preparation for the data needed for this paper has been completed.

3. Trading models and methods

3.1 Pair formation methods

3.1.1 Co-integration method for Pairs formation

The CA takes the co-integration relationship as the theoretical basis for choosing the ideal pairs. Therefore, compared with the DM, the CA has more vital theoretical significance in selecting the pairs with trading opportunities. The spread between the two paired assets is generated by the actual error term δ_{ijt}

of their long-term relationship based on co-integration.The definition of the actual error term δ_{ijt} is given by formula (1):

$$\delta_{ijt} = -\alpha_{it} + \gamma\alpha_{jt} + \beta \tag{1}$$

Where, α_{it} and α_{jt} represent the price time series of asset i and asset j respectively, the co-integration coefficient γ is a non-zero real number. Therefore, as a linear combination of asset i price time series and asset j price time series, the spread δ_{ijt} is a stationary series. β is a constant. The Engle-Granger method was used to test the co-integration of all pairs.

(i) $E(\delta_{ijt})$ is independent of time t

(ii) $VAR(\delta_{ijt})$ is greater than 0 and independent of time t

(iii) $COV(\delta_{ijt},\delta_{ijv})$ is related to time t-s

Through the above co-integration analysis, we can find favourable trading opportunities from all possible pairs. However, compared with the DM, there is a defect in the application of the CA. That is, it can not generate a sort sequence containing all pairs as the DM can, so it is incapable of selecting the ideal trading opportunities according to the sort order.

3.2 Decision-making method

3.2.1 Simple Threshold Method

The ST refers to the fact that when the price difference deviates from more than two historical standard deviations σ, the transaction opens, and it closes when the mean regresses, the trading period ends, or the market withdraws (Gatev et al., 2006). P_{it} ,P_{jt} represents the standardized price series of assets i and j that constitute the paired portfolio. σ(.)2 is the sample variance. Therefore, the variance of the empirical spread can be calculated by formula (2):

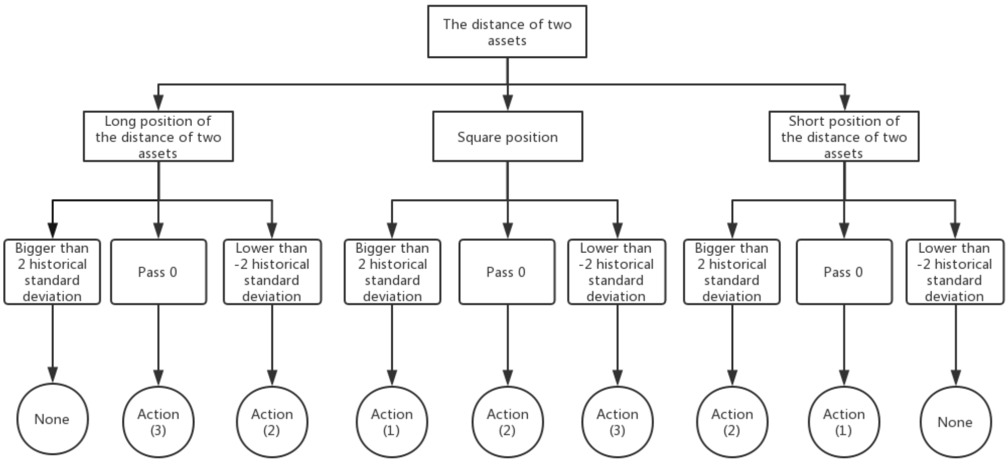
$$\sigma_{P_iP_j}{}^2 = \frac{1}{T}\sum_{t=1}^{T}\left(P_{it} - P_{jt}\right)^2 - \left(\frac{1}{T}\sum_{t=1}^{T}\left(P_{it} - P_{jt}\right)\right)^2 \tag{2}$$

The standardized spread between the original price series α_{it} and α_{jt} of asset i and asset j is as follows:

$$S_{\alpha_{it}\alpha_{jt}} = \frac{P_{it} - P_{jt}}{\sigma_{P_iP_j}{}^2} \tag{3}$$

Among these approaches, this paper uses the variance of the empirical spread $\sigma_{P_iP_j}{}^2$ as the historical standard deviation when judging the deviation degree of the spread between two assets. In making trading decisions, we use the value of $S_{\alpha_{it}\alpha_{jt}}$ as the basis of what kind of trading action to take.

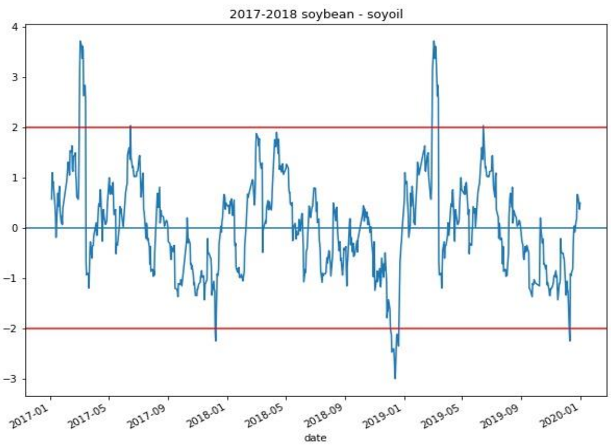
Figure 1 Decision-making process of transaction



Note: The initial position is the square position. 'None' means this situation cannot happen. Action (1) is the action to buy a_i with a value of V/2 , and sell a_j with a value of V/2 . Action (2) means that we do not make any transaction. Action (3) is to sell a_i with a value of V/2 and buy a_j with a value of V/2.

Figure 1 shows how to make trading decisions. The principle is shown in Figure 2.

Figure 2 A sample of individual pairs to explain the ST



Note: The figure above shows the trend of the spread relationship between soybean futures and soyoil futures in the period 2017-2018. The vertical axis is the standardized deviation.

where the horizontal axis represents time and the vertical axis represents the historical standard deviation of the two assets' spread series. When the spread sequence between the two assets is greater than the two historical standard deviations ($2S_{\alpha_{it}\alpha_{jt}}$) or less than ($-2S_{\alpha_{it}\alpha_{jt}}$), we open the trading position. When the spread returns to 0, we close the position.

However, there still exist some unavoidable problems in using the ST as the decision-making method to open or close the trading position. First, there is some subjectivity and randomness in the threshold value setting, such as opening the transaction when deviating from the two historical standard deviations adopted in this paper. The threshold value setting may not be reasonable, and the fact is that an appropriate threshold setting often has a significant impact on the profitability of the transaction. Second, we assume that the distance between the historical price series of two assets forming a pair follows a normal distribution, but sometimes this is not the case. Third, the Simple Threshold Method cannot carry out continuous operation. Therefore, compared with some methods that can conduct the ongoing process, the Simple Threshold Method tends to lose more profit opportunities under the same risk level. Finally, the Simple Threshold Method is only based on the distance between the two assets' price series as the basis for implementing the decision, without considering the impact of external information, so it will fall into the dilemma of incomplete information when making trading decisions.

3.3 DRL model in pairs transaction

3.3.1 Reinforcement Learning method

Reinforcement learning is a machine learning method, but it is different from other such methods.

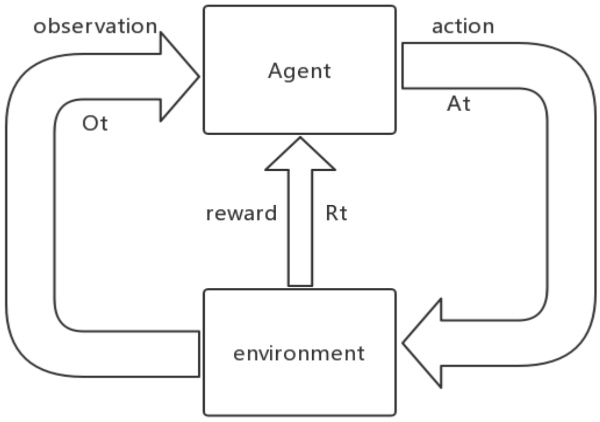
(i) In reinforcement learning there is no teacher and no label, and only reward.

(ii) Feedback on reinforcement learning is delayed and results cannot be returned immediately.

(iii) A series of actions and reward signals will continue a long time afterwards.

Reinforcement learning (RL) has four key elements: environment, reward, action and state. The problem to be solved is to find an optimal policy for a specific issue, which is composed of a series of actions to maximize the return of the model. This method has strong universality, so it has been applied in many research fields, such as game theory, cybernetics, swarm intelligence, statistics and genetic algorithms.

Figure 3 The basic structure of the RL algorithm



Note: The figure above shows the internal operation process of the RL algorithm.

As shown in Figure 3, the reinforcement learning algorithm's decision-making process needs to set an agent in advance. The agent can receive an observation in the current environment, and at the same time, it can receive a reward from the environment after executing an action. The environment object is an uncontrollable element. The agent does not know what kind of bonus the environment will give it when it acts. The environment only tells the agent the current environment state by providing an observation value, and at the same time, rewards the agent according to the possible results.

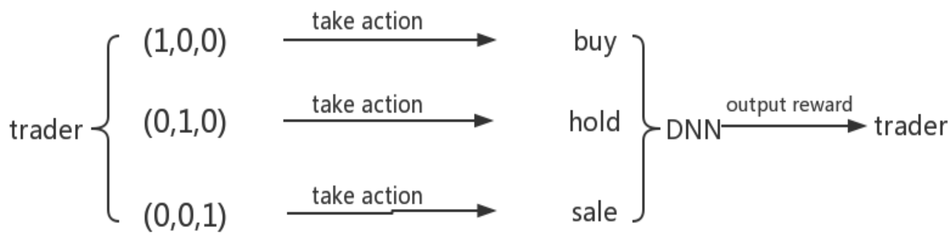
Deep reinforcement learning (DRL) is a new algorithm that combines deep learning with reinforcement learning to realize end-to-end learning from perception to action. Its operation mechanism is to input perceptual information, such as visual information, and then directly output actions through the deep neural network without hand-crafted work. Deep reinforcement learning has the potential to enable robots to learn one or more skills autonomously.

The following introduces the RL algorithm and shows how to apply the RL model to pairs trading strategy in this paper. Here we use the Co-integration method as the pairing method to select the ideal pairs in the asset pool composed of n different assets. The following formula can calculate the spread between paired assets a and b (a, b) at time t:

$$Spread^t{}_{(a,b)} = -\alpha_{at} + \varphi\alpha_{bt} + \beta \tag{4}$$

Where α_{at} and α_{bt} represent the price time series of assets a and b, respectively. According to the observed price difference, we can take corresponding trading actions. Here, we use $action^t{}_{(a,b)}$ to represent the trader's action to the paired combinations a and b based on the value of the spread at time t. The spread can be positive or negative, so there are three different choices for different spread observations:

Figure 4 The implementation process of action



Note: The figure above shows the execution process of the specific trading action. The final trading action will be passed into DNN and the reward value will be output.

As shown in Figure 4, if the output is (1, 0, 0), then the paired trader will perform the buy action, that is, at this time, $\text{action}^t{}_{(a,b)}$=buy. If the output is (0, 1, 0), $\text{action}^t{}_{(a,b)}$=hold, if the output is (0, 0, 1), $\text{action}^t{}_{(a,b)}$= sale. Different actions will be transmitted to the DNN neural network, and finally, a reward value will be fed back.

The environment at time t is $E^t{}_{(a,b)}$, $ST^t{}_{(a,b)}$ represents the pairing state composed of assets a and b at time t. The reward that the environment feeds back to the agent at time t is $R^t{}_{(a,b)}$, and we use $Dr^t{}_{(a,b)}$ to represent the daily income obtained. Then the internal processing mechanism of the RL algorithm can be described as follows. It is necessary to explain two variables. One is $ST^t{}_{(a,b)}$; when the agent performs different actions, the state of the environment will change accordingly. Therefore, three different actions will lead to three state values: long position, close position and short position. The environmental variable $E^t{}_{(a,b)}$ is composed of the X lags of the spread series of assets a and b at time t and the state value $ST^t{}_{(a,b)}$ at time t. The formula is as follows:

$$E^t{}_{(a,b)} = \left\{Spread^t{}_{(a,b)},\ Spread^{t-1}{}_{(a,b)} \cdots\cdots Spread^{t-x}{}_{(a,b)},\ ST^t{}_{(a,b)}\right\} \tag{5}$$

We assume that the agent processes a matching transaction composed of assets a and b. At time t, the agent performs an action ($Act^t{}_{(a,b)}$), and the action will be transferred to the environment variable $E^t{}_{(a,b)}$. At that moment, the state of the environment variable will change from $E^t{}_{(a,b)}$ to $E^{t+1}{}_{(a,b)}$. Meanwhile, the environment variable will feed back a reward value $R^{t+1}{}_{(a,b)}$ to the agent. In the research conducted in this paper, the total reward value consists of two parts: the daily income $Dr^{t+1}{}_{(a,b)}$ at time t+1, and the maximum weighted reward value estimated by the agent when the $Act^t{}_{(a,b)}$ was taken under the environment variable $E^t{}_{(a,b)}$ at time t.

$$R^{t+1}{}_{(a,b)}\left(E^t{}_{(a,b)}, Act^t{}_{(a,b)}\right) = Dr^{t+1}{}_{(a,b)} + \omega MAX_{Act^{t+1}{}_{(a,b)}} R^{t+2}{}_{(a,b)}\left(E^{t+1}{}_{(a,b)}, Act^t{}_{(a,b)}\right) \tag{6}$$

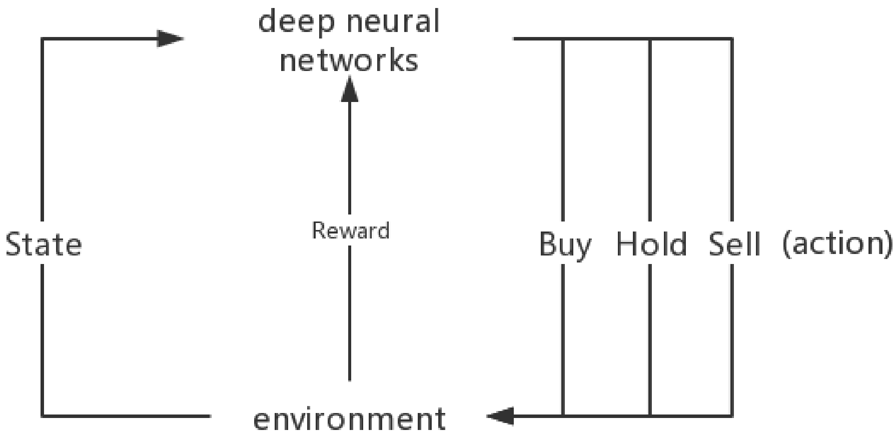
The coefficient ω represents the proportion of the maximum weighted return value in the total return value, that is, the degree of importance. If the value of ω is too close to 1, then DNN neural network will not converge. Here we choose the value of ω is 0.95.

3.3.2 DNN model in deep reinforcement learning (DRL)

3.3.2.1 Application of deep neural network in the RL model

The traditional RL algorithm uses Q-table as the agent to take the best action to get the maximum expectation of future reward in each given state. But Q-table has an obvious disadvantage. It must store the state-action value in the Q-table to calculate the update. For a more complex environment, when the state and action value need to be updated and stored frequently, the Q-table is no longer applicable. Based on this problem, we propose to use the deep neural network (DNN) as the agent part of the RL algorithm, and then deep reinforcement learning (DRL) is produced. The application of DNN in the RL model is shown in Figure 5.

Figure 5 Application of DNN in the RL model

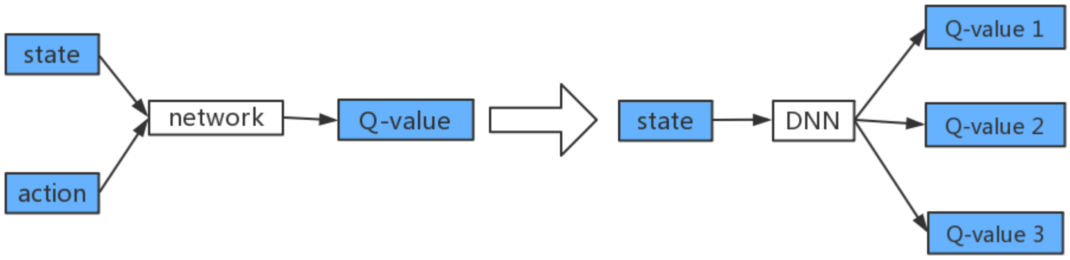


Note: The figure above shows the basic framework of the DRL algorithm. First, the DNN model is pre-trained by iteration. The role of the DNN model is to take different actions according to the input data and then trade in the environment. As a result, the state of the environment will change, and the reward value will be fed back to DNN.

Before DRL, the calculation process of the Q value is as shown in the figure below. We need to input the value of state and action, and then output the Q value. Therefore, there is the most significant disadvantage, which is that for each state, we need to carry out a forward calculation equal to the number of actions. That is, the calculation cost is proportional to the number of actions. In DRL, we only need to input the value of the state, including the current stock price and position information, and we can output the Q value, which is equivalent to only one forward calculation.

The differences in the calculation process of the Q value between the RL and the DRL are shown in Figure 6.

Figure 6 Calculation process of the DNN model in DRL

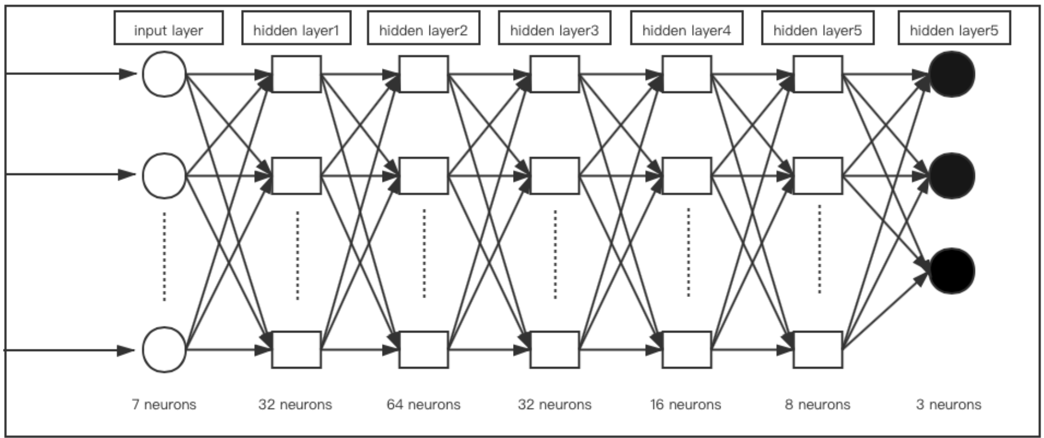


Note: The figure above shows the improvement of the DRL algorithm in the calculation of Q value compared with the traditional RL algorithm.

3.3.2.2 Introduction of the deep neural network model

In this paper, we use the Multi-Layer Perceptron with seven-layer structures as the DRL algorithm's agent. The MLP is a supervised learning algorithm. It learns non-linear functions by training the dataset and maps the input to the output. The MLP consists of three or more layers (input layer, output layer, and one or more hidden layers) (Oral et al., 2012). Each node in a layer is connected to each node in the next layer with a certain weight. Different numbers of cells are arranged in different layers. The number of cells contained in the input layer is equal to the number of parameters that affect the problem. The nodes of each layer only receive the input of the previous layer nodes and only pass their output to the nodes of the next layer (Tomassetti et al. 2009). The MLP has the ability to approximate any function (Principe et al., 2000). This indicates that the MLP can be broadly applied in non-linear change and function mapping problems.

Figure 7 The basic structure of the MLP model in DRL



Note: The figure above shows the basic architecture of the MLP algorithm with 7-layer neural network structure.

The basic framework of the MLP model in the DRL algorithm used in this paper is shown in Figure 7. The first level is the input level, which consists of seven neurons and represents the input with seven characteristic values, including five lags of the price difference sequence, one position status value, and one standard deviation value. There are five hidden layers between the input layer and the output layer, and each layer contains 32, 64, 32, 16, and 8 neurons respectively, which can provide enough space and depth for the agent to calculate the reward value according to the environment and state value. Finally, there is an output layer, which contains three neurons, representing the output of three characteristic values, which are three different reward values obtained after taking three different trading actions. The last layer is the output layer with three neurons, representing the output with three characteristic values, which are three different reward values obtained after taking three different trading actions.

In general, the MLP is trained by the BP algorithm. In this paper, based on network training with the BP algorithm, we also use AdamOptimizer, Gradient Descent (GD), and Stochastic Gradient Descent (SGD) to optimize the weight of the network. The experimental results show that the gradient descent and the random gradient descent methods are more likely to fall into local optimization than the AdamOptimizer methods. Therefore, the models used in this paper all adopt the AdamOptimizer method to optimize the weight in the network, which is a deformation of the gradient descent algorithm.

The activation function used in this paper is linear and there are two reasons for doing so. One is that the calculation result of the network output is the reward value corresponding to each action, and the final reward value can be positive or negative. Therefore, an activation function symmetrical to the origin is needed. In addition, this paper does not limit the cumulative reward value and therefore, the linear function satisfies these two requirements. However, there is a defect in using the linear activation function in that in the process of training, the linear function may cause the connection weight of the MLP in the DRL algorithm to be unable to converge. This leads to the problem that the estimated reward value of the MLP increases infinitely. However, it is worth noting that this situation rarely occurs, and we completely avoided it through repeated training of the experimental dataset in the early stage.

3.3.3 Implementation of the DRL algorithm in pairs trading

In the implementation of the DRL algorithm, we first extract training data from the sample in order. Because we compare the returns of formation periods of one year, two years and three years, we extract data for 243 days, 482 days and 726 days from the sample data. In the first step, we need to initialize the capital, denoted by K_0, and define the initial time t = 0. There are two parameters in the deep neural network at time t+1, one is the state ${ST^t}_{(a,b)}$ at time t, and the other is the X lags of the spread series of portfolios a and b at time t（${Spread^t}_{(a,b)}$, $\cdots\cdots {Spread^{t-x}}_{(a,b)}$）. The output of DNN determines the next action to be taken, whether to buy, hold or sell. Then, according to the action taken, the reward value is calculated, and the weight of the deep neural network (DNN) is adjusted. The reward value obtained by performing a single action at time t is calculated by formula (7):

$$Reward_t = K_t - K_{t-1} + \omega MAX\left(DNN\left({E^{t+1}}_{(a,b)} \middle| {action^{t+1}}_{(a,b)}\right)\right) \tag{7}$$

We repeat this process until time t = T. The total reward value obtained at time t can be calculated by formula (8):

$$Reward_T = (C_T - C_0 - Trading\ Fee)/Std \tag{8}$$

After calculating the total reward value of time t, we need to determine whether the total training times reach the target number N. If the target number N is not reached, we need to reinitialize the capital and time, and start the next iteration with the total time T. When the total training times reach the target number N, the training ends.

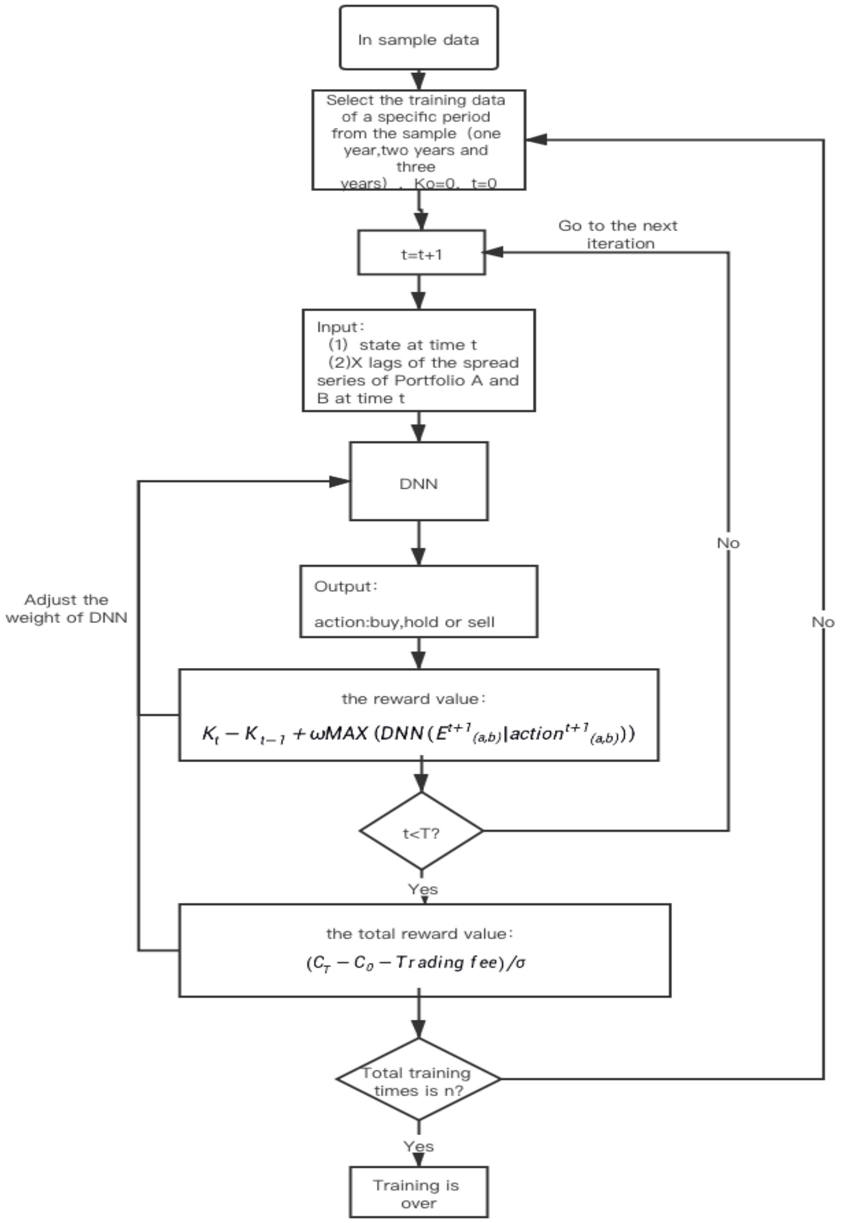
Next, we save the DNN model and use it as the agent of the DRL algorithm. At the end of the whole process, we continue to initialize the initial capital, time and training data. Then, based on the trained DNN model, we continue to train the model in the next training until the maximum number of training N is reached. The implementation process of the DRL algorithm in pairs trading is shown in Figure 8.

In the implementation of the DRL algorithm, we also use the pre-training method. We randomly select a group of pairs to train the model. Then, through the comparison of training effect, we select as the pre-training model of all other pairs the model that performs well both inside and outside the sample, and take the parameters of the pre-training model as the initial DNN connection weights of other pairs. The model of all other pairs runs 50 episodes based on the pre-training model.

Through the pre-training method, we can solve two problems in the actual testing process. One is the huge amount of computation. Because each pair has its trading model, it will create a huge amount of work if we want to train all the trading models. Through pre-training, the best model is selected and saved, then the models of all other pairs are trained based on the pre-training model. In this way, we can not only find a more suitable trading model for specific pairs but also reduce most of the calculations. Second, using the pre-training method can avoid the model falling into the local optimal problem. If the model falls into local optimum, the same agent in the DRL algorithm will make the same decision in different trading environments, such as carrying out a buy transaction in all environments. At the same time, if we use the random weight given by the DNN model for training, we will also fall into the dilemma of local optimization. Therefore, we need to adjust the weights of DNN constantly according to the feedback value in the process of pre-training, to effectively avoid the problem of invalid training.

The implementation process of the DRL algorithm in pairs trading is shown in Figure 8.

Figure 8 Implementation process of DRL algorithm



Note: The figure above shows the iterative process of the DRL algorithm used in this paper. The brain of DRL is DNN, and the environment variable is the historical price data of two commodity futures that form a pair.

4. Model performance

We use two methods to test the return of the pairs trading strategy for soybean, soymeal, and soyoil futures in the DCE. The first category is the statistical arbitrage method. We subdivide this category, mainly because there are differences in pairing selection methods. One of the models does not filter the original pairs, so all the pairs (3*2/2=3 in each trading year) will be selected to enter the pairs' transaction period of the next stage. The other model uses the Co-integration method to select appropriate pairs. By verifying the co-integration relationship between the two assets, the pairs satisfying the co-integration relationship are screened out. In the pairs' transaction period, the trading rules used in the two models are given by the Simple Threshold Method (ST), that is, the time point of opening and closing positions is controlled by the threshold, and the trading threshold selected in this paper is ± 2 standard deviations. Therefore, this paper studies two methods of Statistical Arbitrage, the ST (see Appendix A.1.1 for the raw data) and the CA-ST (see Appendix A.1.2 for the raw data).

The second category is the Deep Reinforcement Learning method. In this category, we study the model return of the DRL (see Appendix A.2.1 for the raw data) and the CA-DRL (see Appendix A.2.2 for the raw data). The DRL trades all pairs, while the CA-DRL selects the pairs that satisfy the co-integration relationship to enter the trading stage. The in-sample training of the DRL generates the trading rules for the two methods. We use the pre-training method here. First of all, we train the data of any pair, and then select the model with the best performance both inside and outside the sample to save as the pre-training model for all pairs in that year. After that, the in-sample training and out-of-sample fitting of all pairs in the year are carried out based on this pre-training model.

In addition, this study considers that the formation period for pairs (in-sample period) is one year, two years and three years respectively. The time of the pairs trading period (out-of-sample period) is one year. Below, this study compares and analyzes the arbitrage yield of different trading models under different formation periods.

4.1 Comparison of model return under a one-year formation period of pairs

In Table 3, we compare the four models' annual returns under the one-year formation period. It can be divided into two cases: in-sample and out-of-sample. We use six indicators to reflect the return of the models. The results are as follows:

Table 3 Model return under the one-year formation period

In sample	Maximum	Minimum	Mean	Std	Extremum	CV
ST	0.3415	0.0206	0.1854	0.0981	0.3209	0.5291
CA-ST	0.5583	0.4204	0.4006	0.1684	0.3351	0.4204
DRL	0.8513	-0.0962	0.2016	0.2150	0.9475	1.0668
CA-DRL	0.4886	0.3257	0.4078	0.0821	0.1642	0.2013
Out-of-sample	Maximum	Minimum	Mean	Std	Extremum	CV
ST	0.3021	-0.0469	0.0543	0.0857	0.3490	1.5783
CA-ST	0.2611	0.0164	0.1320	0.1003	0.2447	0.7598
DRL	0.3668	0.0584	0.1330	0.0802	1.3609	0.6030
CA-DRL	0.2139	0.1339	0.1616	0.0370	0.08	0.2290

Note: The first column is the highest annual return rate of each model, and the second is the lowest. The third column is the average annualized rate of return of each model. The next is the standard deviation of annualized rate of return, followed by extremum and coefficient of variation.

It can be seen from Table 3 that the CA-DRL has the highest average annualized rate of return, at 40.78%, followed by the CA-ST model. The highest rate of return in a single year comes from the DRL, which reaches 85.13%. However, the lowest annualized rate of return also comes from the DRL, which is

-9.62%, indicating that its rate of return is extremely unstable. This unstable feature can also be seen from the fact that the standard deviation, extreme value and coefficient of variation of DRL are the highest of all four models. Its yield fluctuates greatly. At the same time, the stability of the CA-DRL is also the best of the four models, which is reflected in the finding that the standard deviation values, extreme value and coefficient of variation of the CA-DRL are significantly lower than those of the other three models, and the minimum value of its yield is also greater than 0, indicating that the model yield can be stably positive. Considering the above indicators together, the CA-DRL performed best during the in-sample period when the formation period of pairs was one year.

In addition, in the interval outside the sample, the first is the maximum value and the minimum value. It can be noted that the highest annualized yield comes from the DRL model, which reaches 36.68%, and the lowest yield is -0.0469 from the ST model. However, from the average annualized yield, we can see that the average annualized yield of the CA-DRL model is the highest, at 16.16%, followed by the DRL, CA-ST and ST. Then, from the perspective of standard deviation, the CA-DRL also has the smallest standard deviation, indicating that its yield is the most stable of the four models, and the CA-ST model has the greatest volatility in yield. Finally, extreme value and coefficient of variation are compared. Similarly, of the four models, the value of these two indicators of the CA-DRL model is the smallest, and the dispersion of its yield is the lowest. Based on the above indicators, the CA-DRL model has the best model performance in the out-of-sample period when the formation period is one year and it can achieve a relatively stable positive rate of return.

4.2 Comparison of model return under a two-year formation period for pairs

In the case of the two-year formation period of pairs, we also use six indicators to reflect the return of the models in Table 4. The results are outlined below:

Table 4 Model return under the two-year formation period

In sample	Maximum	Minimum	Mean	Std	Extremum	CV
ST	0.2454	0.0601	0.1465	0.0551	0.1853	0.3761
CA-ST	0.3254	0.1906	0.2766	0.0526	0.1348	0.1902
DRL	0.4190	-0.0272	0.1568	0.1202	0.4462	0.7666
CA-DRL	0.6788	0.0683	0.3054	0.1924	0.6105	0.6300
Out-of-sample	Maximum	Minimum	Mean	Std	Extremum	CV
ST	0.1482	-0.0356	0.0444	0.0534	0.1838	1.2027
CA-ST	0.1721	-0.005	0.0592	0.0686	0.1771	1.1588
DRL	0.2703	-0.0058	0.1325	0.0733	0.2761	0.5532
CA-DRL	0.2751	0.02	0.1600	0.0754	0.2551	0.4713

Note: In this table, we compare the annual returns of the four models under a two-year formation period.

As shown in Table 4, we first consider the in-sample model return. From the maximum and minimum indicators, it can be seen that the highest annualized yield comes from the CA-DRL model, which is 67.88%, and the minimum annualized yield, from the DRL model, is -2.72%. The situation is the same as when the formation period of the pair is one year. The CA-DRL model still has the highest average annualized yield of 30.54%, followed by the CA-ST model at 27.66%. Finally, from the three indicators of standard deviation, extreme value and coefficient of variation, it can be seen that the model with the lowest degree of dispersion of yield is the CA-ST, followed by the ST model. On the contrary, the returns of two models (DRL and CA-DRL) with the Deep Reinforcement Learning method have great volatility, and their returns are not stable, but there are basically no negative returns. The volatility is large only because positive returns are sometimes large and sometimes small.

The second case concerns the out-of-sample. The largest annualized rate of return is from the CA-DRL, which is 27.51%, and the smallest is from the ST model, which is -3.56%. Then there is the average annualized rate of return, where the models are ranked in the order CA-DRL, DRL, CA-ST and ST from large to small. The order of the standard deviation is exactly the opposite to that of the average annualized rate of return, which indicates that although the average annualized rate of return of the CA-DRL is high, volatility is also high. However, from the overall standard deviation, the standard deviation of the four models is small, and the difference is not big. The biggest difference is only 0.02, which indicates that the stability of the four models is relatively flat. At the same time, the CA-DRL's coefficient of variation is the smallest, which indicates that the discrete degree of return data of the CA-DRL is the lowest, and the concentration trend is strong. Combined with the above indicators, the CA-DRL has the best model performance in the out-of-sample period when the formation period is two years.

4.3 Comparison of model return under a three-year formation period for pairs

When the formation period for pairs is three years, we study the model performance in-sample and out-of-sample. The results are shown in Table 5.

Table 5 Model return under the three-year formation period

In sample	Maximum	Minimum	Mean	Std	Extremum	CV
ST	0.1741	0.0614	0.1187	0.0405	0.1127	0.3412
CA-ST	0.2659	0.0993	0.1815	0.0597	0.1666	0.3289
DRL	0.2985	-0.0325	0.0909	0.1092	0.3309	1.2013
CA-DRL	0.4426	-0.0174	0.2131	0.1655	0.4600	0.7766
Out-of-sample	Maximum	Minimum	Mean	Std	Extremum	CV
ST	0.1245	-0.0579	0.0313	0.0572	0.1824	1.8275
CA-ST	0.1942	-0.1323	0.0390	0.1071	0.3265	2.7462
DRL	0.2342	-0.0400	0.1367	0.0767	0.2743	0.5611
CA-DRL	0.2701	0.0542	0.1715	0.0703	0.2159	0.4099

Note: In this table, we compare the annual returns of the four models under a three-year formation period.

We first consider the in-sample model return. The maximum annualized rate of return comes from the

CA-DRL model, which is 44.26%. The minimum value is -3.25%, which comes from the DRL model. The average annualized rate of return of the CA-DRL model is the highest, reaching 21.31%, but the volatility of its rate of return is also large because its standard deviation, extreme value and coefficient of variation are relatively high. On the contrary, although the average annualized rate of return of the two kinds of statistical arbitrage models is relatively low compared with the CA-DRL model, their rate of return is relatively stable, and the degree of dispersion is low. Among all the four models, the DRL model does not perform well. The average annualized rate of return is relatively low, and its stability is weak.

Considering the three-year formation period outside the sample, the largest annualized rate of return comes from the CA-DRL, reaching 27.01%, and the lowest annualized rate of return is -13.23%, coming from the CA-ST. The average annualized rate of return of the CA-DRL is the largest, reaching 17.15%, followed by the DRL. From the data of standard deviation, extreme value and coefficient of variation, the value of the CA-DRL is relatively low, which indicates that its stability is strong, and the discrete degree of return data is relatively weak. The CA-DRL model can realize positive returns more stably. Of the four models, the CA-DRL has the best performance.

Based on the above analysis, we can conclude that the CA-DRL is better than other models in the realization of return and regarding the stability of profit when the formation period for pairs is one year, two years and three years. The CA-DRL has the best performance of all four models.

4.4 Robustness check

We select the daily closing price data of the main contracts of No. 2 soybean, soymeal and soyoil futures for the robust test. The data cycle is from January 4, 2016, to July 31, 2020. Similarly, we also carry out a comparative study on the yield of pairs trading under ST (see Appendix B.1.1 for the raw data), CA-ST (see Appendix B.1.2 for the raw data), DRL (see appendix B.2.1 for the raw data), CA-DRL (see Appendix B.2.2 for the raw data). The results of the robustness test are shown in Table 6.

Table 6 Model return comparison of robustness test

ARR (in sample)	ST	CA-ST	DRL	CA-DRL
One-year	0.2420	0.4653	0.1976	0.5301
Two-year	0.1441	0.2260	0.0771	0.7424
Three-year	0.1389	0.2076	0.1581	0.2530
ARR (out-of-sample)	ST	CA-ST	DRL	CA-DRL
One-year	0.1841	0.1982	0.2359	0.2398
Two-year	0.2182	0.2087	0.1914	0.2676
Three-year	0.2252	0.3020	0.2708	0.5324

Note: ARR represents the average annualized rate of return of the four models. We divide the model return into two situations: in-sample and out-of-sample. In every situation, we study formation periods of one year, two years and three years. 'One-year', 'Two-year' and 'Three-year' above indicate that the formation period of pairs is one year, two years and three years.

From the results of the robustness test, the CA-DRL model has the highest average annualized rate of return. Therefore, the data of the robustness test shows that the performance of the CA-DRL model is still better than other models when the pairs trading strategy replaces No. 1 soybean futures by No. 2 soybean futures. This is consistent with the previous empirical conclusion.

5. Risk analysis

In this study, four different models are applied to the pairs trading strategy of soybean commodity futures. Those models are ST, CA-ST, DRL and CA-DRL. For the four models, we calculate the annualized return, volatility, Sharpe ratio and maximum withdrawal rate outside the sample, and consider the case when the formation period of pairs (in-sample period) is one year, two years and three years. The results of risk analysis are shown in Table 7.

Table 7 Model risk analysis under different formation periods of pairs

One-year（out-of-sample）	ST	CA-ST	DRL	CA-DRL
annualized return	5.10%	12.76%	13.03%	16.10%
volatility	0.79%	1.51%	0.69%	0.21%
Sharp Ratio	0.267	0.816	1.240	2.92
Maximum Drawdown	0.047	0	0.21	0
Two-year (out-of-sample)				
annualized return	4.31%	5.70%	13.02%	15.75%
volatility	0.31%	0.57%	0.58%	0.66%
Sharp Ratio	0.284	0.393	1.351	1.603
Maximum Drawdown	0.036	0.005	0.088	0
Three-year (out-of-sample)				
annualized return	2.97%	3.34%	13.40%	16.93%
volatility	0.36%	1.38%	0.64%	0.593%
Sharp Ratio	0.040	0.052	1.334	1.844
Maximum Drawdown	0.058	0.152	0.118	0

Note: This table shows the result of model risk analysis when the formation period is one year, two years and three years. We use four indicators to reflect the risk of the four models.

When the formation period is one year, the first result is the annualized return rate of the model. It can be seen that the return rate of the DRL model is higher than that of the statistical arbitrage model. The highest return rate of the model comes from the CA-DRL, at 16.10%. The lowest return rate of the model comes from the ST, at only 5.10%. The return of the CA-ST is 12.76%, which is close to that of the DRL. However, the volatility of the CA-ST is high, at 1.51%, which indicates that the return is unstable and has strong uncertainty. The volatility of the DRL is only 0.69% and 0.21%, which makes the return more stable. At the same time, the Sharpe ratio of the DRL and CA-DRL is also higher than that of the statistical arbitrage model, which indicates the higher excess return of the DRL and CA-DRL when taking certain risks. The

Sharpe ratio of the CA-DRL is the highest, at 2.92. Finally, there is the maximum withdrawal rate. The Maximum Drawdown of the CA-ST and CA-DRL models are both 0, which indicates that the annualized return rate is positive and there is no loss in any trading year. The Maximum Drawdown of the ST is 0.047, which means the maximum loss is 4.7%. The Maximum Drawdown of the DRL is 0.21, which means the maximum loss is 21%. This result shows that when the formation period of pairs is one year, the CA-DRL model not only has higher returns than other models, but also has more stable returns. It can ensure profits in all trading years, so it has better model performance.

When the formation period is two years, the annualized yield of the model based on the DRL algorithm is higher than that of the statistical arbitrage model. The CA-DRL still has the highest annualized yield of 15.75%. In terms of volatility, the volatility of the DRL and CA-DRL models is slightly higher than that of the ST and CA-ST, but the difference is not very large, and the highest volatility is only 0.66%, indicating that the overall income is relatively stable. The Sharpe ratio is the same as in the above situation, showing an increasing trend in the four models (ST, CA-ST, DRL and CA-DRL). The CA-DRL is the highest, at 1.603, indicating the higher excess return of the CA-DRL model under certain risks. The Maximum Drawdown of the DRL model is the highest, at 0.088, indicating that the maximum loss range is 8.8%. The Maximum Drawdown of the CA-DRL is the lowest, at 0, indicating that in each trading year, the annual return rate of the CA-DRL is positive, and it can make stable profits without losses. Generally speaking, when the formation period of pairs is two years, the performance of the CA-DRL is the best, and it is significantly better than other models.

In the case of the formation period of three years, the return of the model based on the DRL algorithm is still higher than that of the statistical arbitrage model. The highest return of the model is 16.93%, which comes from the CA-DRL model. In terms of volatility, the CA-ST model has the highest volatility rate of 1.38%, while the other models have little volatility. The Sharpe ratio of the CA-DRL is the highest, so it has the highest excess return when the risk is certain. Finally, in terms of the Maximum Drawdown, the maximum loss margin of the CA-ST is the highest, and the Maximum Drawdown of the CA-DRL model is 0. According to the performance of various indicators, when the formation period of pairs is three years, the CA-DRL model is the best model with the highest return and the lowest risk.

When we integrate the information above, we observe that, the CA-DRL has the highest average annualized rate of return in any case. Although its volatility is not always the lowest, except in the case of a one-year formation periods of pairs, the Maximum Drawdown of the CA-DRL is 0 whether the formation period is one year, two years or three years. This shows that the CA-DRL can achieve a positive yield in all trading years and has stable profitability. High volatility means that the positive yield is not stable enough, being sometimes high and sometimes low, but never negative. The CA-DRL has the highest Sharpe ratio in the out-of-sample case of all pair formation periods. Therefore, based on the above analysis, the ability of the CA-DRL when it comes to model return and risk control is significantly better than other models and compared with other models studied in this paper, CA-DRL is the optimal one.

6. Conclusion

This study mainly looks at the pairs trading strategy of No.1 soybean, soymeal and soyoil futures in the DCE. As the derivatives of soybean pressing, the three have a significant natural correlation, which can be considered as the asset selection object of paired trading. Through pairing the three commodity futures, a total of 3*2/2=3 pairs are formed in each cycle. We use the Co-integration method to select appropriate pairs from all possible pairing combinations. In the pairs trading period of the next stage, we make corresponding trading decisions for the selected pairs according to different trading rules. Those rules are generated by the Simple Threshold Method or the in-sample training of the DRL model. Therefore, this study takes a total of four models to conduct a comparative study on the model yield of the pairs trading strategy of the soybean and its derivatives futures. The four models are ST, CA-ST, DRL and CA-DRL. At the same time, taking into account the impact of the length of the formation period of the pair on the performance of the model, this paper also studies pairs formation periods of one year, two years and three years.

Of all the four models, the CA-DRL has obvious model advantages. Compared with the traditional statistical arbitrage method, using the DRL instead of the ST as a trading decision-making method has the following benefits. First, the setting of the threshold is objective. Second, it does not need to satisfy the assumption of normal distribution, and third the DRL is a method that can carry out continuous operation. Meanwhile, we use the MLP with a seven-layer structure as the agent of the DRL algorithm. Using the BP algorithm to train the MLP, we also adopt the AdamOptimizer method to optimize the weight in the network and ensure the stability of parameters by controlling the learning speed.

In addition, in the actual training process, the pre-training method is used to solve a large number of calculations, and the model quickly falls into the local optimal. On the other hand, compared with the DRL, the CA-DRL uses the Co-integration test to select the pairs meeting the co-integration relationship to enter the transaction stage. The two assets constituting these pairs often have a common random trend and have a long-term equilibrium relationship. Therefore, by studying and fitting the spread relationship between these pairs, they can have better prediction accuracy and model performance.

The empirical results show that the CA-DRL model can achieve better performance in the sub-cycle within the sample and outside the sample, regardless of whether the formation period is one year, two years or three years. The CA-DRL is superior to other models in the realization of average annual return and risk control. When the short-term soybean market fluctuates greatly, this model can provide a referable and usable investment tool for relevant institutional and individual investors, so as to avoid price risk and obtain investment income.

References

[1] Huang, B., Huan, Y., Xu, L., Zheng, L., & Zou, Z. (2018). Automated trading systems statistical and machine learning methods and hardware implementation: A survey. Enterprise Information Systems, 13(1), 132-144.

[2] Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. The Review of Financial Studies, 19(3), 797–827.

[3] Vidyamurthy, G. (2004). Pairs trading: Quantitative methods and analysis. John Wiley & Sons, Hoboken, N.J.

[4] Caldeira, J. F. and Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. Brazilian Review of Finance, 11(1):49–80.

[5] Elliott, R. J., Van Der Hoek*, John, and Malcolm, W. P. (2005). Pairs trading. Quantitative Finance, 5(3):271–276.

[6] Cummins, M. and Bucca, A. (2012). Quantitative spread trading on crude oil and refifined products markets. Quantitative Finance, 12(12):1857–1875.

[7] Jurek, J. W. and Yang, H. (2007). Dynamic portfolio selection in arbitrage. Working paper, Harvard University.

[8] Liu, J. and Timmermann, A. (2013). Optimal convergence trade strategies. Review of Financial Studies, 26(4):1048–1086.

[9] Xie, W. and Wu, Y. (2013). Copula-based pairs trading strategy. In Asian Finance Association (AsFA) 2013 Conference, Nanchang, Jiangxi, China.

[10] Wu, Y. (2013). Pairs trading:A copula approach. J. Derivatives Hedge Funds, 19(1), 12–30.

[11] Stander, Y., Marais, D., and Botha, I. (2013). Trading strategies with copulas. Journal of Economic and Financial Sciences, 6(1):83–107.

[12] Xie, W., Liew, Q.R., Wu, Y. and Zou, X. (2014). Pairs trading with copulas. Available at SSRN 2383185.

[13] Ehrman, Douglas S. (2012). The Handbook of Pairs Trading (Strategies Using Equities, Options, and Futures) || Futures and Currencies. 10.1002/9781119201526(), 213–224.

[14] Qingsong Ruan, Hao Cui, Liming Fan (2020). China's soybean crush spread: Nonlinear analysis based on MF-DCCA,Physica A: Statistical Mechanics and its Applications,Volume 554, 123899, ISSN 0378-4371.

[15] Johnson, R. L., Zulauf, C. R., Irwin, S. H. and Gerlow, M. E. (1991) The soybean complex spread: an examination of market efficiency from the viewpoint of a production process. J. Fut. Markts, 11, 25 – 37.

[16] Barrett, B. and Kolb, R. (1995). Analysis of spreads in agricultural futures. J. Fut. Markts, 15, 69 – 86.

[17] Simon, D. P. (1999). The soybean crush spread: empirical evidence and trading strategies. J. Fut. Markts, 19, 271 – 289.

[18] Mitchell J B. (2010). Soybean Futures Crush Spread Arbitrage: Trading Strategies and Market Efficiency. Journal of Risk & Financial Management, 3, 63-96.

[19] Dunis C L, Laws J, Evans B, John L, University M. (2006). Modelling and trading the soybean-oil crush spread with

recurrent and higher order networks: A comparative analysis. Artificial Higher Order Neural Networks for Economics & Business, 16, 193-213.

[20] Wiles, P. S. , & Enke, D. . (2014). Nonlinear modeling using neural networks for trading the soybean complex. Procedia Computer Science, 36.

[21] Li, H. T., Liu, X. J., Zhang, Y. B., Fu, Y. H., & Zheng, J. Y. (2015). The empirical analysis for the spread of soya oil and soybean meal based on wavelet neural network. International Journal of Economics & Finance, 7(6).

[22] QW Liu, & HH Sono. (2016). Empirical properties, information flow, and trading strategies of china's soybean crush spread. Journal of Futures Markets, 36(11), 1057-1075.

[23] Brim, A. (2020). Deep Reinforcement Learning Pairs Trading with a Double Deep Q-Network. 2020 10th Annual Computing and Communication Workshop and Conference (CCWC). IEEE.

[24] Taewook Kim & Ha Young Kim, (2019). "Optimizing the Pairs-Trading Strategy Using Deep Reinforcement Learning with Trading and Stop-Loss Boundaries," Complexity, Hindawi, vol. 2019, pages 1-20, November.

[25] Oral, M., Oral, E. L., & Ayd?N, A. (2012). Supervised vs. unsupervised learning for construction crew productivity prediction. Automation in Construction, 22(Mar.), p.271-276.

[26] Tomassetti, B., Verdecchia, M., & Giorgi, F. (2009). Nn5: a neural network based approach for the downscaling of precipitation fields – model description and preliminary results. Journal of Hydrology, 367(1-2), 14-26.

[27] Principle, J. C., Euliano, N. R., & Lefebvre, W. C. (2000). Neural and adaptive systems.

Appendices

Appendix A： Detailed yield data of trading model

A.1 Statistical arbitrage

A.1 .1 Simple Threshold Method

In this model, we select all the original pairs to trade in the next stage. The trading rules are given by the Simple Threshold method, and the detailed yield data are shown in the table below. Tables A.1, A.2 and A.3 show the model return when the pair formation period is one year, two years and three years respectively.

Table A.1 In-sample period and out-of-sample period is one year

Pair formation period	Average annual return (in-sample)	Average annual return (out-of-sample)
2006	0.0441	0.0063
2007	0.2906	0.3021
2008	0.3193	0.0712
2009	0.3415	-0.0394
2010	0.1923	0.0644
2011	0.1873	0.0250
2012	0.1598	-0.0469
2013	0.0591	0.1522
2014	0.1359	-0.0240
2015	0.0206	0.0207
2016	0.1855	0.0863
2017	0.1896	0.0505
2018	0.2497	0.0671
2019	0.2208	0.0245

Note: This table compares the average annual return of the ST under a one-year formation period for pairs. The first column is the formation period (in-sample period). The second is the average annual return within the sample. The third is the average annual return out-of-sample.

Table A.2 In-sample period is two years and out-of-sample period is one year

Pair formation period	Average annual return (in-sample)	Average annual return (out-of-sample)
2006-2007	0.2009	0.1482
2007-2008	0.2103	0.0127
2008-2009	0.2454	0.0352
2009-2010	0.1247	0.0347
2010-2011	0.1457	0.0833
2011-2012	0.0952	-0.0356
2012-2013	0.0736	0.1305
2013-2014	0.2140	-0.0315
2014-2015	0.1289	0.0360
2015-2016	0.0601	0.0519
2016-2017	0.1100	0.0876
2017-2018	0.1650	0.0140
2018-2019	0.1306	0.0103

Note: In this table, we compare the average annual return of the ST under a two-year formation period of pairs. The first column is the formation period (in-sample period), the second is the average annual return within the sample and the third is the average annual return out-of-sample.

Table A.3 In-sample period is three years and out-of-sample period is one year

Pair formation period	Average annual return (in-sample)	Average annual return (out-of-sample)
2006-2008	0.1624	0.0219
2007-2009	0.1741	0.0197
2008-2010	0.1543	-0.0434
2009-2011	0.0799	0.1051
2010-2012	0.1366	-0.0579
2011-2013	0.0665	0.1245
2012-2014	0.0849	0.0209
2013-2015	0.1601	0.1164
2014-2016	0.0614	0.0563
2015-2017	0.0791	-0.0094
2016-2018	0.1521	0.0310
2017-2019	0.1132	-0.0099

Note: In this table, we compare the average annual return of the ST under the three-year formation period for pairs. The first column is the formation period (in-sample period), the second is the average annual return within the sample and the third is the average annual return out-of-sample.

A.1.2 Co-integration Method

In the CA-ST model, we adopt the Co-integration method to screen the original pairs and select those that meet the co-integration relationship to enter the transaction. Its trading rules are also given by the ST. In the case of formation periods of one year, two years and three years, the detailed yield data are shown in the following table.

Table A.4 In-sample period and out-of-sample period is one year

Pair formation period	Average annual return (in-sample)	Average annual return (out-of-sample)
2010	0.4204	0.2611
2011	0.2232	0.1186
2014	0.5583	0.0164

Note: In this table, we compare the average annual return of the CA-ST under the one-year formation period of pairs. The first column is the formation period (in-sample period), the second is the average annual return within the sample and the third is the average annual return out-of-sample.

Table A.5 In-sample period is two years and out-of-sample period is one year

Pair formation period	Average annual return (in-sample)	Average annual return (out-of-sample)
2009-2010	0.3254	0.1721
2010-2011	0.3077	0.0317
2011-2012	0.1906	0.0254
2013-2014	0.3218	-0.0032
2014-2015	0.2958	0.1340
2017-2018	0.2180	-0.0050

Note: In this table, we compare the average annual return of the CA-ST under the two-year formation period for pairs. The first column is the formation period (in-sample period), the second is the average annual return within the sample and the third is the average annual return out-of-sample.

Table A.6 In-sample period is three years and out-of-sample period is one year

Pair formation period	Average annual return (in-sample)	Average annual return (out-of-sample)
2009-2011	0.2659	0.0213
2011-2013	0.1188	0.1444
2012-2014	0.0993	-0.1323
2015-2017	0.2343	-0.0232
2016-2018	0.2045	0.0293
2017-2019	0.1663	0.1942

Note: In this table, we compare the average annual return of the CA-ST under the three-year formation period of pairs. The first column is the formation period (in-sample period), the second is the average annual return within the sample and the third is the average annual return out-of-sample.

A.2 Model performance of the DRL

A.2.1 Take no account of co-integration relationship

In the DRL model, we select all the original pairs to enter the transaction. The optimal threshold of pairing assets is given by the DRL algorithm, and the formation period is also divided into three cases: one year, two years and three years. The detailed yield data are shown in the following table.

Table A.7 In-sample period and out-of-sample period is one year

In-sample period	Out-of-sample period	Asset 1	Asset 2	Average annual return in-sample	Average annual return out-of-sample
2006	2007	soybean	Soy oil	0.2955	0.2142
2006	2007	soybean	Soy meal	0.1181	0.0268
2006	2007	Soy oil	Soy meal	0.5428	0.0201
2007	2008	soybean	Soy oil	0.0351	0.0405
2007	2008	soybean	Soy meal	-0.0962	0.2713
2007	2008	Soy oil	Soy meal	0.4409	0.2241
2008	2009	soybean	Soy oil	0.5087	0.0897
2008	2009	soybean	Soy meal	0.1328	0.1463
2008	2009	Soy oil	Soy meal	0.3815	0.0896
2009	2010	soybean	Soy oil	-0.0869	0.3035
2009	2010	soybean	Soy meal	0.1156	0.0632
2009	2010	Soy oil	Soy meal	0.8513	0.0501
2010	2011	soybean	Soy oil	-0.0169	0.1214
2010	2011	soybean	Soy meal	0.2967	0.1718
2010	2011	Soy oil	Soy meal	0.2216	-0.0070
2011	2012	soybean	Soy oil	0.3765	0.0884
2011	2012	soybean	Soy meal	0.2308	0.0105
2011	2012	Soy oil	Soy meal	0.0411	0.1259
2012	2013	soybean	Soy oil	0.0568	0.0519
2012	2013	soybean	Soy meal	-0.0581	0.0843
2012	2013	Soy oil	Soy meal	0.3296	0.1811
2013	2014	soybean	Soy oil	0.1349	0.2175
2013	2014	soybean	Soy meal	-0.0033	-0.2157
2013	2014	Soy oil	Soy meal	-0.0239	0.5382
2014	2015	soybean	Soy oil	-0.0745	0.1601
2014	2015	soybean	Soy meal	0.0394	0.0567
2014	2015	Soy oil	Soy meal	0.0817	-0.0415
2015	2016	soybean	Soy oil	0.2384	0.0011
2015	2016	soybean	Soy meal	-0.0289	0.3608
2015	2016	Soy oil	Soy meal	0.1955	-0.1603
2016	2017	soybean	Soy oil	0.4115	0.0819
2016	2017	soybean	Soy meal	-0.0024	-0.0203
2016	2017	Soy oil	Soy meal	0.1508	0.1538
2017	2018	soybean	Soy oil	0.3903	0.2301
2017	2018	soybean	Soy meal	0.1683	0.1666
2017	2018	Soy oil	Soy meal	0.1794	0.2768
2018	2019	soybean	Soy oil	0.5478	0.2490
2018	2019	soybean	Soy meal	-0.0734	0.0383
2018	2019	Soy oil	Soy meal	0.4078	0.0238
2019	2020	soybean	Soy oil	0.5764	0.3677
2019	2020	soybean	Soy meal	0.2151	0.1757
2019	2020	Soy oil	Soy meal	0.2179	0.5571

Note: In this table, we compare the average annual return of the DRL under the one-year formation period of pairs. The first and second columns are the in-sample period and out-of-sample period, asset 1 and asset 2 represent the two assets that make up the pairs. The last two columns are the average annual return within the sample and out-of-sample.

Table A.8 In-sample period is two years and out-of-sample period is one year

In-sample period	Out-of-sample period	Asset 1	Asset 2	Average annual return in-sample	Average annual return out-of-sample
2006-2007	2008	soybean	Soy oil	0.1631	0.2402
2006-2007	2008	soybean	Soy meal	0.3480	0.3766
2006-2007	2008	Soy oil	Soy meal	0.2742	0.1001
2007-2008	2009	soybean	Soy oil	0.4357	0.1102
2007-2008	2009	soybean	Soy meal	0.5363	0.1148
2007-2008	2009	Soy oil	Soy meal	0.2850	-0.0342
2008-2009	2010	soybean	Soy oil	-0.1701	0.2144
2008-2009	2010	soybean	Soy meal	0.0732	0.1640
2008-2009	2010	Soy oil	Soy meal	0.1835	0.1043
2009-2010	2011	soybean	Soy oil	-0.0525	0.1696
2009-2010	2011	soybean	Soy meal	-0.0596	0.0577
2009-2010	2011	Soy oil	Soy meal	0.0651	0.0721
2010-2011	2012	soybean	Soy oil	0.2477	0.0200
2010-2011	2012	soybean	Soy meal	-0.1288	0.2202
2010-2011	2012	Soy oil	Soy meal	0.1881	0.1671
2011-2012	2013	soybean	Soy oil	0.1960	0.0945
2011-2012	2013	soybean	Soy meal	0.2677	0.1380
2011-2012	2013	Soy oil	Soy meal	0.2408	0.1517
2012-2013	2014	soybean	Soy oil	-0.0334	0.2222
2012-2013	2014	soybean	Soy meal	0.2215	-0.0688
2012-2013	2014	Soy oil	Soy meal	-0.2698	0.2189
2013-2014	2015	soybean	Soy oil	0.1693	0.0560
2013-2014	2015	soybean	Soy meal	0.1102	0.0612
2013-2014	2015	Soy oil	Soy meal	-0.0003	0.1192
2014-2015	2016	soybean	Soy oil	0.0431	0.1300
2014-2015	2016	soybean	Soy meal	0.1215	0.3101
2014-2015	2016	Soy oil	Soy meal	0.5604	0.2281
2015-2016	2017	soybean	Soy oil	0.1660	0.0542
2015-2016	2017	soybean	Soy meal	-0.1230	0.0161
2015-2016	2017	Soy oil	Soy meal	0.3521	-0.0876
2016-2017	2018	soybean	Soy oil	0.4074	0.1823
2016-2017	2018	soybean	Soy meal	0.2984	0.0179
2016-2017	2018	Soy oil	Soy meal	-0.1046	0.0442
2017-2018	2019	soybean	Soy oil	0.1136	0.2751
2017-2018	2019	soybean	Soy meal	0.2133	0.1167
2017-2018	2019	Soy oil	Soy meal	0.3552	-0.0198
2018-2019	2020	soybean	Soy oil	0.1067	0.3508
2018-2019	2020	soybean	Soy meal	0.4382	0.1683
2018-2019	2020	Soy oil	Soy meal	-0.1240	0.2919

Note: In this table, we compare the average annual return of the DRL under the two-year formation period of pairs.

Table A.9 In-sample period is three years and out-of-sample period is one year

In-sample period	Out-of-sample period	Asset 1	Asset 2	Average annual return in-sample	Average annual return out-of-sample
2006-2008	2009	soybean	Soy oil	0.3616	0.2013
2006-2008	2009	soybean	Soy meal	0.1504	0.1084
2006-2008	2009	Soy oil	Soy meal	0.1191	0.2326
2007-2009	2010	soybean	Soy oil	-0.1017	0.1975
2007-2009	2010	soybean	Soy meal	0.4903	0.1522
2007-2009	2010	Soy oil	Soy meal	0.3491	0.1652
2008-2010	2011	soybean	Soy oil	0.5860	0.3445
2008-2010	2011	soybean	Soy meal	0.3447	0.2703
2008-2010	2011	Soy oil	Soy meal	-0.0353	0.0878
2009-2011	2012	soybean	Soy oil	0.0101	0.1500
2009-2011	2012	soybean	Soy meal	0.0813	0.0862
2009-2011	2012	Soy oil	Soy meal	-0.1250	0.1990
2010-2012	2013	soybean	Soy oil	-0.0206	0.1024
2010-2012	2013	soybean	Soy meal	0.0121	0.0659
2010-2012	2013	Soy oil	Soy meal	0.0032	0.3865
2011-2013	2014	soybean	Soy oil	0.1255	0.1650
2011-2013	2014	soybean	Soy meal	-0.0307	0.0710
2011-2013	2014	Soy oil	Soy meal	-0.1922	0.3922
2012-2014	2015	soybean	Soy oil	-0.1235	-0.0984
2012-2014	2015	soybean	Soy meal	0.2154	-0.0050
2012-2014	2015	Soy oil	Soy meal	-0.1316	-0.0168
2013-2015	2016	soybean	Soy oil	0.0256	0.1197
2013-2015	2016	soybean	Soy meal	-0.0364	0.2026
2013-2015	2016	Soy oil	Soy meal	0.1445	0.1788
2014-2016	2017	soybean	Soy oil	0.0256	0.0434
2014-2016	2017	soybean	Soy meal	-0.0653	0.0841
2014-2016	2017	Soy oil	Soy meal	0.5028	0.3757
2015-2017	2018	soybean	Soy oil	-0.0617	0.0921
2015-2017	2018	soybean	Soy meal	0.0571	0.1234
2015-2017	2018	Soy oil	Soy meal	0.4214	0.1675
2016-2018	2019	soybean	Soy oil	-0.0991	-0.1803
2016-2018	2019	soybean	Soy meal	0.2237	0.0805
2016-2018	2019	Soy oil	Soy meal	-0.0819	0.1875
2017-2019	2020	soybean	Soy oil	0.1821	0.3127
2017-2019	2020	soybean	Soy meal	-0.0489	-0.0570
2017-2019	2020	Soy oil	Soy meal	-0.0038	-0.0664

Note: In this table, we compare the average annual return of the DRL under the three-year formation period of pairs. The first and second columns are the in-sample period and out-of-sample period, asset 1 and asset 2 represent the two assets that make up the pairs. The last two columns are the average annual return within the sample and out-of-sample.

A.2.2 Take account of co-integration relationship

The difference between the CA-DRL and the DRL is whether or not to adopt the Co-integration method to screen the original pairs. The CA-DRL selects the pairs that meet the co-integration relationship to enter the transaction, and then the DRL algorithm is used to give the optimal threshold of pairing assets. Table A.10 shows all pairs that satisfy the co-integration relationship. Table A.11 shows the detailed return data of the four models when the formation period is one year, two years and three years.

Table A.10 Pairs satisfying the co-integration relationship

In-sample period	Asset 1	Asset 2	P-value	parameter1	parameter2
2010	soybean	Soy oil	0.00065	0.27115	1929.19199
2011	soybean	Soy oil	0.02005	0.19434	2605.11888
2011	soybean	Soy meal	0.03325	0.43066	3140.35127
2014	Soy oil	Soy meal	0.01762	1.45542	1087.89195
2006-2007	soybean	Soy oil	2.31898×10^{-8}	0.32110	953.96975
2006-2007	Soy oil	Soy meal	1.22574×10^{-6}	2.85504	-325.99385
2009-2010	soybean	Soy oil	0.00083	0.32554	1438.10367
2010-2011	soybean	Soy oil	0.00001	0.24080	2162.04992
2011-2012	soybean	Soy meal	0.00263	0.33451	3431.04798
2013-2014	Soy oil	Soy meal	0.01544	1.98534	-466.17957
2014-2015	Soy oil	Soy meal	0.00728	0.67215	3834.20158
2017-2018	soybean	Soy oil	0.00137	0.36653	1623.00543
2009-2011	soybean	Soy oil	0.04501	0.28892	1703.60334
2011-2013	soybean	Soy meal	0.03951	0.27226	3628.48567
2012-2014	soybean	Soy oil	0.04932	0.06105	4108.49344
2015-2017	Soy oil	Soy meal	0.00925	0.80349	3687.83425
2016-2018	soybean	Soy oil	0.00299	0.31151	1893.70939
2017-2019	soybean	Soy oil	0.01708	0.32388	1788.90368

Note: In this table, we list the pairs that satisfy the co-integration relationship.

Table A.11 Model return of pairs satisfying co-integration relationship

In-sample period	Out-of-sample period	Asset 1	Asset 2	Average annual return in-sample	Average annual return out-of-sample
2010	2011	soybean	Soy oil	0.3257	0.2139
2011	2012	soybean	Soy oil	0.4899	0.0347
2011	2012	soybean	Soymeal	0.3270	0.2394
2014	2015	Soy oil	Soymeal	0.4886	0.1339
2006-2007	2008	soybean	Soy oil	0.1692	0.2402
2006-2007	2008	Soy oil	Soymeal	0.2554	0.1001
2009-2010	2011	soybean	Soy oil	0.2730	0.1696
2010-2011	2012	soybean	Soy oil	0.3248	0.0200
2011-2012	2013	soybean	Soymeal	0.1278	0.1380
2013-2014	2015	Soy oil	Soymeal	0.6788	0.1192
2014-2015	2016	Soy oil	Soymeal	0.4531	0.2281
2017-2018	2019	soybean	Soy oil	0.0683	0.2751
2009-2011	2012	soybean	Soy oil	0.3776	0.1500
2011-2013	2014	soybean	Soymeal	0.0359	0.1264
2012-2014	2015	soybean	Soy oil	0.2240	0.0542
2015-2017	2018	Soy oil	Soymeal	0.4426	0.2104
2016-2018	2019	soybean	Soy oil	0.2160	0.2178
2017-2019	2020	soybean	Soy oil	-0.0174	0.2701

Note: In this table, we compare the average annual return of the CA-DRL under the one-year, two-year and three-year formation periods. The first and second columns are the in-sample period and out-of-sample period, asset 1 and asset 2 represent the two assets that make up the pairs. The last two columns are the average annual return within the sample and out-of-sample.

Appendix B：Detailed yield data of robust test

In the robustness test, we use the No.2 soybean futures of the DCE instead of the No.1 soybean futures for arbitrage, and all the models remain unchanged.

B.1 Statistical arbitrage

B.1 .1 Simple Threshold Method

In the ST model, we trade all the original pairs, and the trading rules are given by the ST. Table B.1 lists the detailed return data of the models in all formation periods for pairs.

Table B.1 Model return of the ST

Pair formation period	Pairs trading period	Average annual return in-sample	Average annual return out-of-sample
2016	2017	0.2449	0.0942
2017	2018	0.1109	0.0615
2018	2019	0.2218	0.1377
2019	2020	0.3903	0.4429
2016-2017	2018	0.1465	-0.0314
2017-2018	2019	0.1060	0.3063
2018-2019	2020	0.1798	0.3798
2016-2018	2019	0.1038	0.3067
2017-2019	2020	0.1740	0.1437

Note: In this table, we compare the average annual return of the ST under the one-year, two-year and three-year formation periods. The first and second columns are the in-sample period and out-of-sample period. The last two columns are the average annual return within the sample and out-of-sample.

B.1 .2 Co-integration Method

The CA-ST uses the Co-integration method to screen the pairs, and the trading rules are given by the ST. When the formation period for pairs is one year, two years and three years, the detailed yield data are shown in Table B.2.

Table B.2 Model return of the CA-ST

Pair formation period	Pairs trading period	Average annual return in-sample	Average annual return out-of-sample
2019	2020	0.4653	0.1982
2018-2019	2020	0.2260	0.2087
2016-2018	2019	0.1423	0.4879
2017-2019	2020	0.2728	0.1160

Note: In this table, we compare the average annual return of the CA-ST under the one-year, two-year and three-year formation periods. The first and second columns are the in-sample period and out-of-sample period. The last two columns are the average annual return within the sample and out-of-sample.

B.2 Model performance of the DRL

B.2.1 Take no account of co-integration relationship

The DRL selects all original pairs to enter the pairs' transaction period and gives the optimal threshold of pairing assets by the DRL algorithm. The detailed yield data are shown in Table B.3.

Table B.3 Model return of the DRL

In-sample period	Out-of-sample period	Asset 1	Asset 2	Average annual return in-sample	Average annual return out-of-sample
2016	2017	Soybean2	Soy oil	0.4130	0.2156
2016	2017	Soybean2	Soy meal	0.1776	0.0414
2016	2017	Soy oil	Soy meal	0.3858	0.3949
2017	2018	Soybean2	Soy oil	0.1072	0.1502
2017	2018	Soybean2	Soy meal	0.1829	0.0576
2017	2018	Soy oil	Soy meal	0.6785	0.1754
2018	2019	Soybean2	Soy oil	-0.2602	0.2947
2018	2019	Soybean2	Soy meal	0.1151	0.1706
2018	2019	Soy oil	Soy meal	-0.0209	-0.0128
2019	2020	Soybean2	Soy oil	0.2431	0.2398
2019	2020	Soybean2	Soy meal	0.0505	0.5557
2019	2020	Soy oil	Soy meal	0.2990	0.5475
2016-2017	2018	Soybean2	Soy oil	0.0747	0.0495
2016-2017	2018	Soybean2	Soy meal	0.3045	0.0614
2016-2017	2018	Soy oil	Soy meal	0.0514	0.0678
2017-2018	2019	Soybean2	Soy oil	0.1476	0.5620
2017-2018	2019	Soybean2	Soy meal	-0.1345	0.0423
2017-2018	2019	Soy oil	Soy meal	-0.0192	-0.0198
2018-2019	2020	Soybean2	Soy oil	-0.0049	0.1943
2018-2019	2020	Soybean2	Soy meal	0.1425	0.3408
2018-2019	2020	Soy oil	Soy meal	0.1314	0.4240
2016-2018	2019	Soybean2	Soy oil	0.0822	0.6222
2016-2018	2019	Soybean2	Soy meal	0.0995	0.1251
2016-2018	2019	Soy oil	Soy meal	-0.0826	-0.0207
2017-2019	2020	Soybean2	Soy oil	0.4648	0.4426
2017-2019	2020	Soybean2	Soy meal	0.1339	0.3215
2017-2019	2020	Soy oil	Soy meal	0.2508	0.1343

Note: In this table, we compare the average annual return of the DRL under the one-year, two-year and three-year formation period. The first and second columns are the in-sample period and out-of-sample period, asset 1 and asset 2 represent the two assets that make up the pairs. The last two column are the average annual return within the sample and out-of-sample.

B.2.2 Take account of co-integration relationship

The CA-DRL adopts the Co-integration method to select pairs, and then the DRL algorithm gives the optimal threshold of pairing assets. All pairs that meet the co-integration relationship are shown in Table B.4. The detailed return data of the models under formation periods of one year, two years and three years are shown in Table B.5.

Table B.4 Pairs satisfying co-integration relationship

In-sample period	Asset 1	Asset 2	P-value	parameter1	parameter2
2019	Soybean2	Soy oil	0.00913	0.44300	613.03772
2018-2019	Soybean2	Soy oil	0.00346	0.28299	1598.36707
2018-2019	Soybean2	Soy meal	0.00522	0.38601	2049.67776
2016-2018	Soybean2	Soy oil	0.03862	0.51219	543.18011
2017-2019	Soybean2	Soy oil	0.01578	0.45938	732.93258

Note:In this table, we list the pairs that satisfy the co-integration relationship.

Table B.5 Model return of the CA-DRL

In-sample period	Out-of-sample period	Asset 1	Asset 2	Average annual return in-sample	Average annual return out-of-sample
2019	2020	Soybean2	Soy oil	0.5301	0.2398
2018-2019	2020	Soybean2	Soy oil	0.8691	0.1943
2018-2019	2020	Soybean2	Soy meal	0.6157	0.3408
2016-2018	2019	Soybean2	Soy oil	0.2608	0.6222
2017-2019	2020	Soybean2	Soy oil	0.2452	0.4426

Note: In this table, we compare the average annual return of the CA-DRL under the one-year, two-year and three-year formation periods for pairs. The first and second columns are the in-sample period and out-of-sample period, asset 1 and asset 2 represent the two assets that make up the pairs. The last two columns are the average annual return within the sample and out-of-sample.