

Design of Cross-Product Arbitrage Strategy in Forward Market

Lieyuan Huang

Chengdu Tianfu School, Chengdu, China

Email: lyhuang0610@163.com

How to cite this paper: Huang, L. Y. (2022). Design of Cross-Product Arbitrage Strategy in Forward Market. *Technology and Investment*, 13, 144-162.

<https://doi.org/10.4236/ti.2022.134010>

Received: September 20, 2022

Accepted: November 5, 2022

Published: November 8, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

All investors are speculators. They profit from longing an asset and selling it at a higher price or shorting an asset and buying it at a lower price. This is the fundamental concept of arbitrage. Although it sounds simple, arbitrage does not always work. Therefore, researchers have developed systematic and scientific statistical arbitrage approaches for investigation. In this article, we dived into forming pair trading portfolios by using the cointegration analysis method. The objects we investigated are egg, corn, and soybean meal in the future market of China. In the forming stage of the strategy, we proved the existence of a cointegration relationship among the three pairs, namely the egg-corn pair, the egg-soybean meal pair (and the corn-soybean pair. In the back-test study, both the egg-corn pair and egg-soybean meal pair are profitable.

Keywords

Future Market, Arbitrage, Cointegration, Cross-Product Arbitrage Strategy

1. Introduction

The future contract allows traders to lock the prices of underlying commodities or assets. Future contracts have lower transaction cost, more flexible operating system, and future contracts are typically highly leveraged which would create more potential profits or risks, which attracts many speculators and institutional investors. Three major exchanges in China were established in the 1990s, respectively Zhengzhou Commodity Exchange (ZCE) in 1990, Dalian Commodity Exchange (DCE) in 1993, and Shanghai Futures Exchange (SHFE) in 1999. In specific, SHFE trades metals and energy mostly, while ZCE and DCE mainly trade agricultural commodities. One of the features of China's commodity future market is that a large proportion of trade occurs domestically even though China

is one of the largest commodity importers in the world. In addition to that, as a means of curbing excessive speculation and preventing distortions in the spot market, the Chinese government implements stringent price limits and position limits according to [Ao, J. & Chen, J. \(2020\)](#). As a result of frequent government intervention, the development of future market is sluggish. In our research, we found that the egg has great market potential as the demand is numerous in China and the egg market is not overly saturated. There are roughly around 3.95 billion laying hens in the world and China accounts for 35% of that volume according to [Wu, Y.F. & Fu, Q. \(2018\)](#). Nowadays, increasing numbers of specialized and scaled hen breeders have shown a promising future for the egg market. There are approximately 67,500 firms related to egg production in China in 2022. Meanwhile, the price fluctuation of eggs is evident since egg is easily affected by climate, pandemic, etc. According to [Wu, Y.F. & Fu, Q. \(2018\)](#), they used descriptive statistics and BP models to analyze the price fluctuation of eggs in both the long-run and short-run. This high volatility led to uncertainty of breeders' profits and given the increasing numbers of people participating in the egg industry, we can foreshadow that the numbers of traders in egg future contract will increase to avoid risks. Based on the context of egg spot and future market, we managed to use corn and soybean meal, two highly related products to form arbitrage pairs. In the production of eggs, solely chicken fodder takes accounts for around 60% - 70%. Most importantly, corn and soybean meal are the essential ingredient where corn takes about 64% of fodder while soybean meal takes around 26% based on research of [Kaiqi, Z., Ronghua, J. & Zhinan, L. \(2019\)](#).

The figure below demonstrates the total future turnover of three agricultural futures, namely the egg, soybean meal, and corn. The turnover data started from 2013 to 2022, and is collected from CSMAR ([Figure 1](#)).

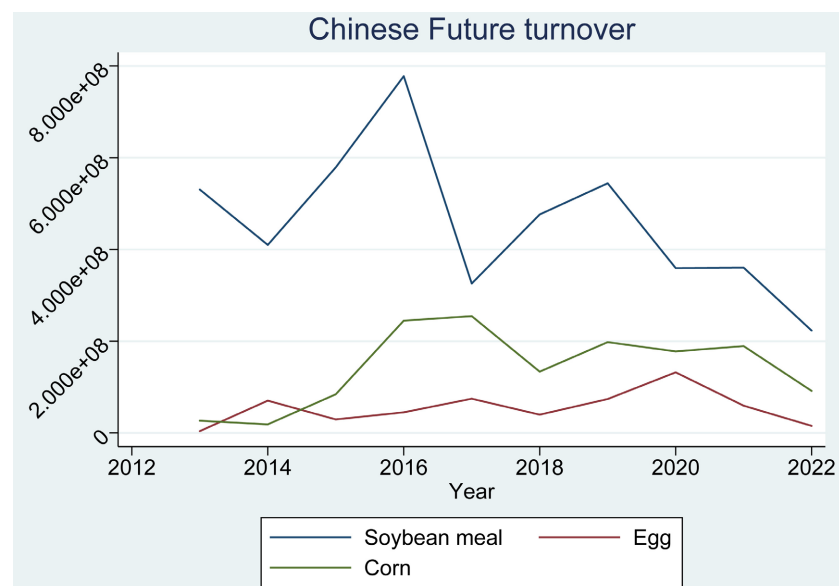


Figure 1. Chinese future turnover.

Soybean meal had the greatest volume in turnovers and presented the most volatile trends. It peaked around 2016 with turnovers of around 8.0e. In contrast, egg, and corn futures turnover showed a relatively unpopular market where the turnover for corn fluctuated between 2.0e. Moreover, since the egg market was listed on the future market in 2013, the turnover had been moving between 1.0e. Nevertheless, all three futures turnovers hit the bottom in the first two quarters of 2022.

According to the publication of the pair trading articles by [Gatev et al. \(2006\)](#), which are frequently implemented in financial markets, we intend to find the pairs that possess the characteristics of mean reversion. In other words, the pair which used to share synchronizing prices will eventually return to the average price of the entire data set. On the premise of this property, we adopted cointegration and time series analysis approach to form arbitrage strategies and empirical models. However, time series analysis is often considered to be non-stationary which is a stochastic process and seems to be the nature of economics. Nevertheless, by using the cointegration method we can testify the stationarity of a time series. In a short version, there are two steps. First, construct linear functions for pairs and run regression on the time series. Second, use the ADF test for testing the stationarity of estimated residuals of the time series. The fundamental tick is to construct a linear stationary time series from two non-stationary time series. The cointegration underlies the linear relationship of two non-stationary time series to be traded as one asset and was first discovered by [Engle and Granger \(1987\)](#). Cointegration can be applied in many areas of study, for instance, [Yang, J., Li, Z. & Wang, T. \(2021\)](#) conduct their result of price discovery function of future market using cointegration. In this article, we abstracted our data with two trading frequencies from the Wind database, respectively five minutes of high-frequency trading data and daily trading data. The results indicated that three pairs, namely the egg-corn pair, the egg-soybean meal pair, and the corn-soybean pair, have a cointegration relationship. The importance after we find the cointegrated pairs is to confirm an estimated hedge ratio, and the parameter of selling out and buying in, etc. Implement different indexes and adopt JoinQuant a quantitative back-test platform, performing back-test data we have concluded different profitable results in two frequencies.

The difference in this paper is that regardless of the prosperity of the commodity future market in China, it is still unsophisticated comparing to western countries. There are only a handful of people who investigated these three products through the cointegration and time series analysis approach and designed a strategy based on it. [Kaiqi, Z., Ronghua, J. & Zhinan, L. \(2019\)](#) studied these three commodities for hedging purpose in a similar approach. In that paper, they estimated the hedging ratio through OLS, B-VAR, and ECM models to analyze the effect of hedging. Or like [Xu, H. \(2017\)](#) used the Kalman Filter and Markov Chain Monte Carlo to design arbitrage trading strategy. On the other hand, we devoted to serve the arbitrage purpose. Furthermore, the highlight of this article is that we used both daily frequency data and five mi-

notes high-frequency data, which will deliver us varied results from data in the same period.

To be more comprehensive, the following article is arranged in this sequence. Section 2 will articulate the designing of arbitrage strategy step-by-step from the examination of the cointegration relationship to the settings of the back-test. Section 3 will thoroughly demonstrate the statistical results and the most important indexes in the back-test results. Section 4 summarize the results of our back-test and evaluate the main contribution and innovation of this paper.

2. Methodology

In this paper, we investigated the possibility of two-by-two arbitrage combinations of egg, soybean meal, and corn, as soybean meal and corn are the two most important ingredients in chickens' fodder and account for the major costs in the production of egg. Stübinger, J. & Bredthauer, J. (2017) they construct statistical arbitrage strategy based on different approaches, however, our core strategy is applying statistical arbitrage pair trading with different frequency data. In statistical arbitrage, the arbitrage opportunity occurs as a consequence of market inefficiency. According to Gatev et al. (2006), pair trading is a quantitative arbitrage strategy which is based on two steps. First, find historically price-synchronous futures that moved together. Then, upon divergence occur, long undervalued future while short overvalued future to form a hedge or arbitrage portfolio. In Shen, L., Shen, K., Yi, C., & Chen, Y. (2020), they have explained and demonstrate it thoroughly. We collected our data from Dalian Commodity Exchange on the Wind database and to be accurate, we selected five minutes closing price of the three commodities from 12/31/2021 9:05 to 5/31/2022 14:59 as the first testing sample and daily closing price of the three commodities from 12/31/2021 to 5/31/2022 as the second testing sample. The data we selected would be used to testify the cointegration relationship between two commodities and if the relationship exists, we will adopt JoinQuant, the quantitative trading platform, to back-test the pairs in the same period.

2.1. Unit Root Test

The assumptions we made before examining the cointegration relationship were both pairs are non-stationary time series and of order 1. Presume that two time series X_t and Y_t are integrated of order 1, denoted as $I(1)$. The Cointegration relationship only exists if the two-time series can combine to a linear function of $z_t = Y_t - \alpha X_t$. Henceforth, we need to examine whether time series in pairs are $I(1)$.

In order to find the cointegration relationship, we first need to form an arbitrage combination.

The variables used in the combination are listed as follows (Table 1).

In this paper, we first analyze the correlation of the trading assets by using STATA. Then for each variable, we apply the Augmented dickey-fuller test for

Table 1. Variable definition.

Variable	Definition
EC	Egg closing price
BC	Soybean meal closing price
CC	Corn closing price

unit root as a method to avoid spurious regression. This approach is designed for an initial judgment upon the stability of time series. If there is a unit root presented in the time series, it implies that the time series is non-stationary and the combinations are not cointegrated. Hence, the combinations cannot be paired. Alternatively, if the null hypothesis is rejected and the combinations show cointegration, it implies that the time series is stationary and the combinations can be paired.

The sample model is as follows:

$$Y = \alpha + \beta X + e \quad (1)$$

$$H_0 : \beta = 1 \longleftrightarrow H_a : \beta < 1 \quad (2)$$

If reject the null hypothesis, there is no unit root:

$$ADF = \frac{\hat{\beta} - 1}{SE(\hat{\beta})} \quad (3)$$

Thus, the time series is stationary.

In this paper, our Cointegration models are as followed:

$$EC_t = \alpha + \beta BC_t + e_1, \quad t = 1, 2, \dots, T \quad (4)$$

$$EC_t = \alpha + \beta CC_t + e_2, \quad t = 1, 2, \dots, T \quad (5)$$

$$BC_t = \alpha + \beta CC_t + e_3, \quad t = 1, 2, \dots, T \quad (6)$$

2.2. Cointegration Test

After testing for stationary time series, based on the study of [Engle, R. F. & Granger, C. W. \(1987\)](#) and [Hendry, D. F. & Juselius, K. \(2000\)](#), we use OLS regression to obtain the residuals, denoted as e_x . In our final step to prove whether the two commodities have cointegration relationship, we run ADF testing on the residuals. If they are stationary time series, then the relationship exists. Functions (4), (5), (6) and OLS regression were used to calculate the residual term: e_1, e_2, e_3

$$e_x = Y - \alpha - \beta X \quad (7)$$

The cointegration relationship is the prerequisite for us to find the ratio of pairs trading and the spread series of two commodities.

The mean spread series is defined as: $MSpread = e_x = Y - \alpha - \beta X$.

JoinQuant

To back-test our cointegrated future pairs, we need to find the ratio of trading.

We did this on the foundation of OLS regression. OLS regression will provide the constant value of β term that indicates the ratio. Applying the actual ratio of fodder and egg in egg-laying chicken breeding enterprises, we found that 2.1 kg of fodder can produce approximately 1 kg of egg. Thus, as mentioned previously about the ingredient of fodder, we concluded that for each 500 kg of egg, a breeder enterprise needs to use 315 kg of soybean meal and 735 kg of corn. Given that information, we established a parameter to adjust the ratio of trading to correspond to the realistic laying hen industry. Eventually, our trading ratios are 50:24 for egg and soybean meal, and 16:40 for egg and corn. We will not investigate the pair of corn and soybean meal, as our priority is to form a strategy based on egg.

On JoinQuant, we need several basic indexes to set up a back-test including commission charge for buy, sell and close, initial margin, initial capital (**Table 2**).

In addition to the basic indexes, we need to set up a series of signals for taking the position, closing the position, and stopping loss. We used the Z-value in the normal distribution (denoted as Q) to times the standard deviation σ from Ms_{spread} as the threshold for trading signals. According to previous study, [Vidyamurthy \(2004\)](#), who implied the optimal threshold where open a position is 0.75σ , close a position as 2σ and stop-loss as 0. Moreover, based on the studies of [Gatev et al. \(2006\)](#), who set a fixed threshold defined two unconditional standard deviations from sample spread. However, we noticed that in our work, most historical spread will either not reach 2 standard deviations or performs low profitability at $Q = 0.75$. Therefore, we have implemented hundreds of times of adjustments to settle the range of Q value and the signal for close position and stop-loss for each pair. The range of Q value will be between 0.45 - 0.75 and the signal for close position and stop-loss for each pair will be listed in next table (**Table 3** and **Table 4**).

Table 2. Trading index.

Index	Value
Buy and sell commission charge	0.000023
Close commission charge	0.000023
Initial margin	15%
Initial capital	1000000

Table 3. Threshold definition table.

Signals	Threshold
Take position up	$Q\sigma$
Take position down	$Q\sigma$
Close position	Absolute value [$<x$]
Stop-loss	$Y\sigma$

Table 4. Signal value.

Frequency	Pair	X	Y
5 minutes	Egg-Soybean meal	0.2	1.1
	Egg-Corn	0.113	1.15
Daily	Egg-Soybean meal	0.2	1.8
	Egg-Corn	0.113	1.3

As our back-test data are in a frequency of 5-min bar and daily spread series, to avoid losses induced by divergence, we required comparison to previous period and the current period. $MSpread_t$ as current spread series and $MSpread_{t-1}$ as the previous period. The logic of the entire strategy is as followed (**Figure 2**).

For the result shown in JoinQuant, there are various means to judge the feasibility of the strategy. We will mainly assess the feasibility of strategy through indexes, Total return, Total annualized return, Benchmark volatility, Division Abnormal profit, Sharp ratio, Maximum drawdown (**Table 5**).

3. Result

3.1. Statistic Result

- **Daily frequency**

1) Correlation analyzing

We concluded the possibility of mean reversion and arbitrage through our graph (**Figure 3**) and correlation matrix (**Table 6**). The long-run trend for the three markets is similar regardless of the different levels of fluctuation. The graph shows the BC has the most evident fluctuation, then EC, and lastly the least fluctuated CC. Corn takes account for the greatest proportion of cost in egg production and therefore they are related the strongest with 0.908. CC and BC as two highly related agricultural products who share a relatively strong correlation of 0.768. Finally, EC and BC are correlated by 0.606.

2) Unit root test

After establishing the existence of the possibility of mean reversion and arbitrage, we use the ADF test to identify the stability of our time series and avoid spurious regression.

ADF test (**Table 7**) indicates that three variables are insignificant and fail to reject the null hypothesis. Therefore, the unit root exists in the original form and the time series are non-reposeful. We then constructed difference equations to testify whether the variables are first-order integrated (**Figures 4-6**).

ADF test (**Table 8**) for unit root subjecting first order difference equations shows that they are first-order integrated and stationary. Their P -values are all approximate 0 which is very significant P -value and reject the null hypothesis, where unit root does not exist.

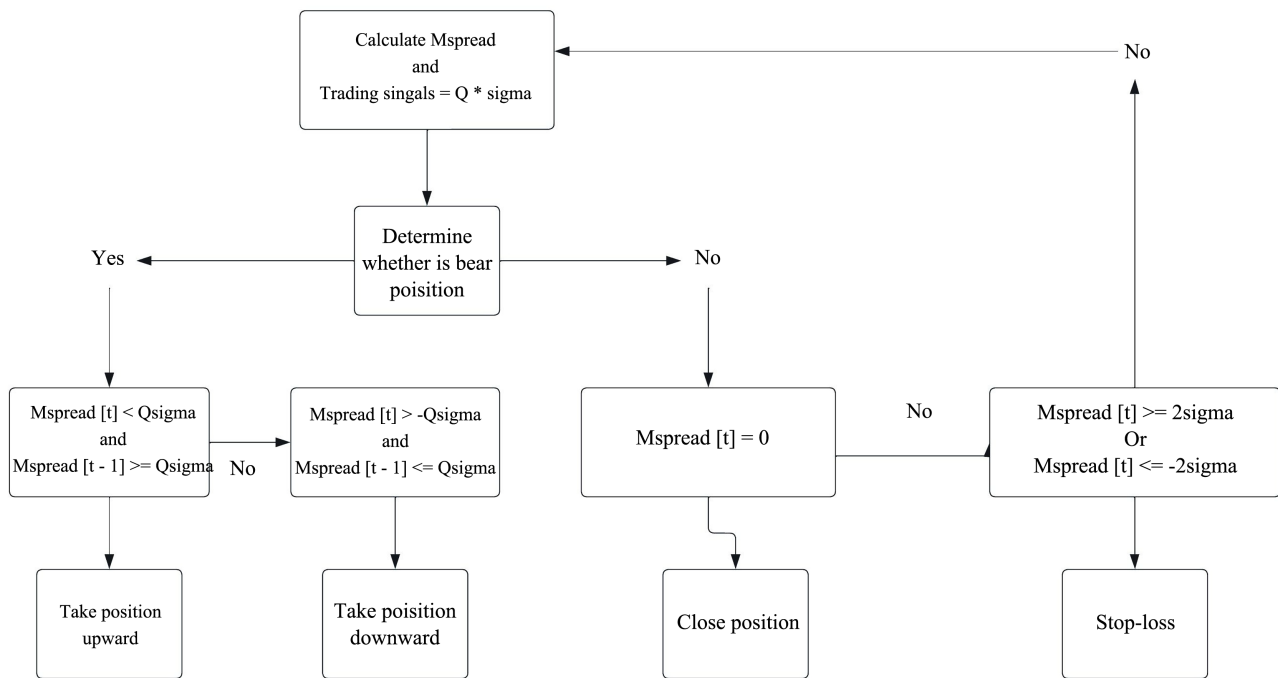


Figure 2. Strategy logic.

Table 5. JoinQuant index.

Evaluation indicator	Formula	Description
Total return	$\frac{v_{end} - v_{start}}{v_{start}} \times 100\%$	v_{start} : The total position and cash in the beginning of back-test. v_{end} : The remaining position and cash after back-test.
Total annualized return	$\left((1 + R)^{\frac{250}{n}} - 1 \right) \times 100\%$	R : Revenue. n : Days of strategy execution.
Benchmark volatility	/	Corn dominant contract.
Division Abnormal profit	$\frac{R_{total} + 1}{R_{bench} + 1} - 1$	R_{total} : Total revenue. R_{bench} : Benchmark revenue.
Sharp ratio	$\frac{R_{annual} - R_{free}}{v_{return}}$	R_{annual} : Annual Yield. R_{free} : Risk-free Yield. v_{return} : Volatility of return.
Maximum drawdown	$\frac{Max(v_x - v_y)}{v_x}, y > x$	v_x, v_y : Total position and cash on sometime.

Table 6. Correlation-daily.

Variables	(1)	(2)	(3)
(1) EC	1.000		
(2) BC	0.606	1.000	
(3) CC	0.908	0.768	1.000

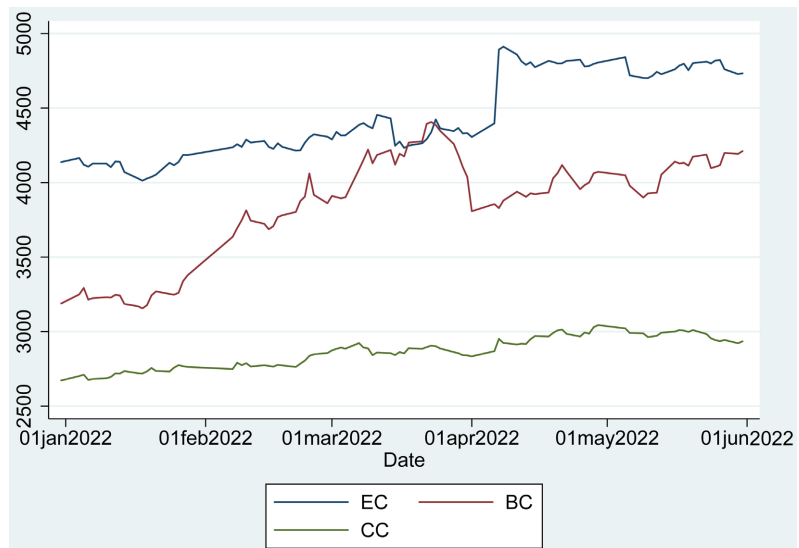


Figure 3. Correlations.

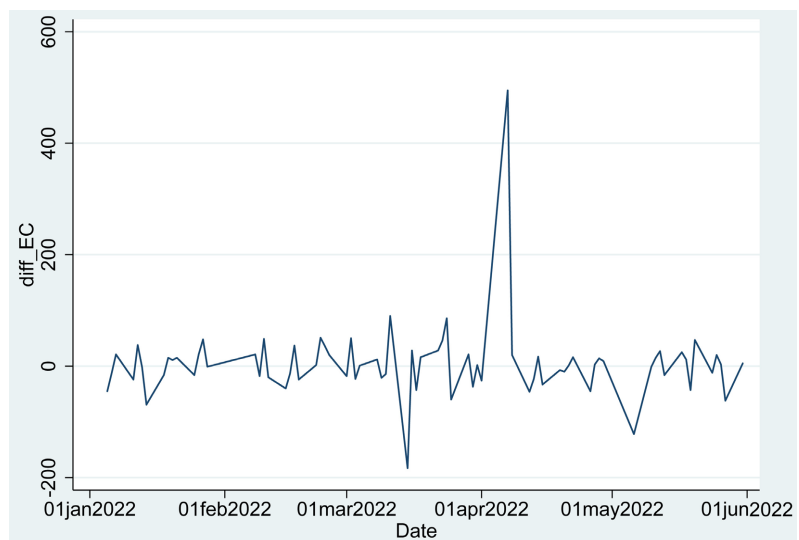


Figure 4. EC I (1) line graph-daily.

Table 7. Dickey-fuller test for unit root-daily.

Variable	<i>P</i> -value	<i>P</i> < 0.1
EC	0.370	Insignificant
BC	0.311	Insignificant
CC	0.264	Insignificant

Table 8. Dickey-fuller test for unit root in I (1).

Variable	<i>P</i> -value	<i>P</i> < 0.1
Diff_EC	0.000	Significant
Diff_BC	0.000	Significant
Diff_CC	0.000	Significant

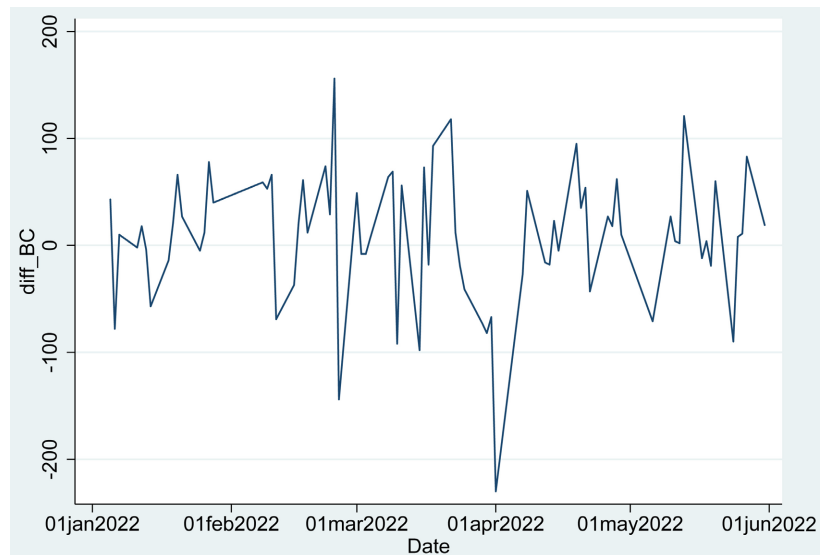


Figure 5. BC I (1) line graph-daily.

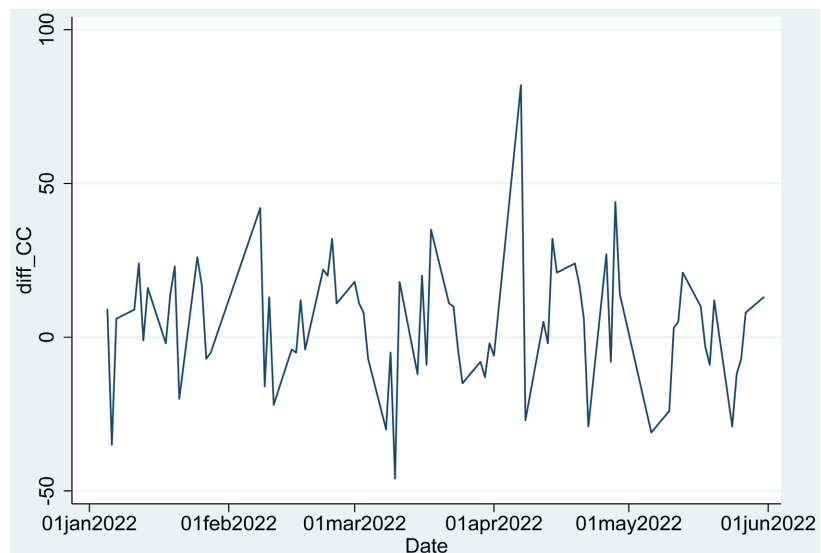


Figure 6. CC I (1) line graph-daily.

3) Cointegration test

Residual constant term is as followed (**Table 9**).

As if the ADF test (**Table 10**) indicates the corresponding residuals are first-order integrated time series, then the corresponding variables will have a cointegration relationship that allows us to build arbitrage strategies.

Thereafter, we showed all three combinations have cointegration relationships and proved the existence of mean reversion and arbitrage opportunity.

Then, the following is to find the Mspread in to determine the trading signals (**Figures 7-9**).

Based on **Table 11**, Mspreads have 97 number of observation and the approximate means are zero, Mspread 2 has the smallest standard deviation of 118.116 followed by 223.844 (Mspread 1) and 227.466 (Mspread 3). The price

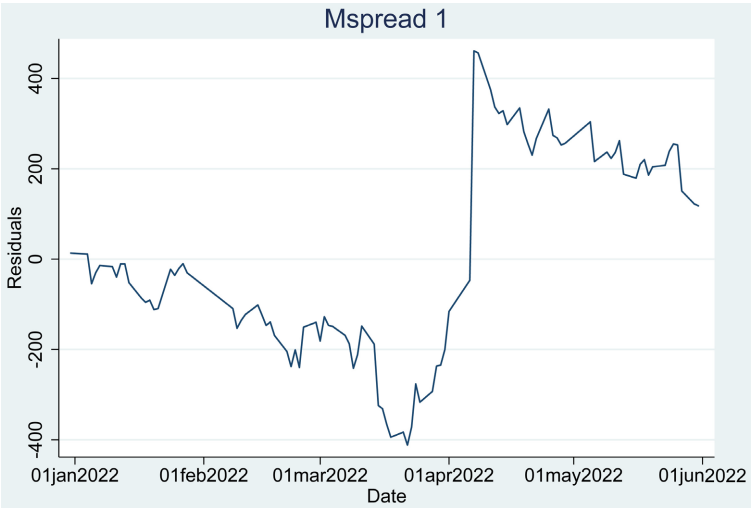


Figure 7. Mspread 1.1.

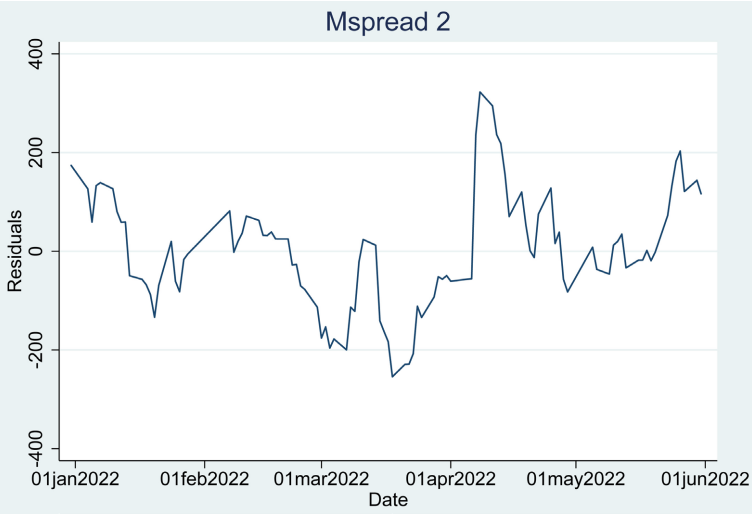


Figure 8. Mspread 1.2.

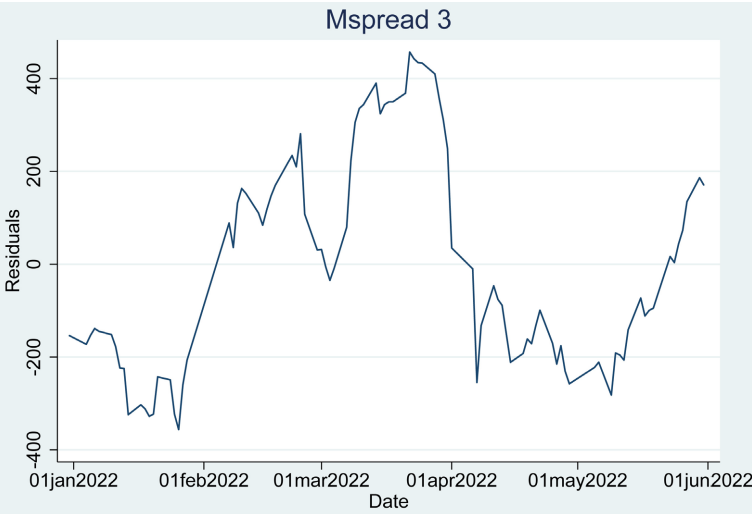


Figure 9. Mspread 1.3.

Table 9. Residual constant term-daily.

Function	α	β
(4)	2593.059	0.48
(5)	-2674.908	2.485
(6)	-3753.012	2.656

Table 10. Dickey-Fuller test for unit root of residual in I (1).

Variable	<i>P</i> -value	<i>P</i> < 0.1
Diff_ e_1	0.000	Significant
Diff_ e_2	0.000	Significant
Diff_ e_3	0.000	Significant

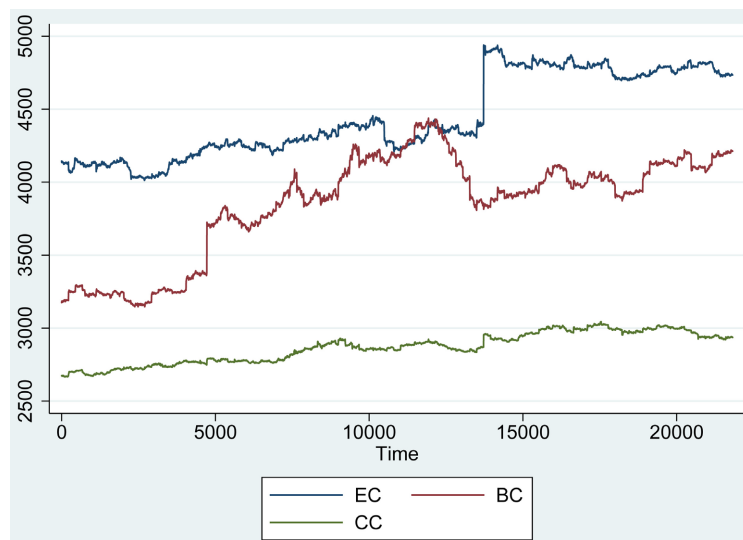
Table 11. Descriptive statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
e_1	97	0	223.844	-411.725	461.092
e_2	97	0	118.116	-254.273	322.764
e_3	97	0	227.466	-356.396	457.271

fluctuation for three combinations is around the mean value 0 as the graphs indicated. This further satisfies the condition for mean reversion and arbitrage.

- **Five minutes frequency**

- 1) **Correlation analyzing**

**Figure 10.** Correlation-5-min.

Based on the correlation graph (Figure 10) from 12/31/2021 9:05 to 5/31/2022 14:59, the trend of the three markets is similar in the long-run which indicates

EC, BC, and CC have relatively strong correlations and the two-by-two combination attributed with the relatively stable price difference. Considering the cost distribution of egg production spent on corn (64%), it is reasonable to suggest that the correlation between EC and BC has a relatively significant correlation. To analyze the data from the graph, our results of correlation are as shown (Table 12).

According to the matrix of correlation (Table 12), the among three two-by-two combination EC and CC have the strongest correlation coefficient (0.909), while BC and CC followed with 0.770, and EC and BC with 0.604. Using both graphic and correlation analyses, we have established the possibility of mean reversion and arbitrage opportunity.

2) Unit root test

Applying Stata to run the ADF test (Table 13) indicates that BC and CC are significant and reject the null hypothesis whereas *P*-value for EC is insignificant, which fails to reject the null hypothesis. Therefore, BC and CC are stationary time series while EC is non-reposeful time series and has presence of unit root.

Next step is to exam whether EC is integrated of order 1.

The graph (Figure 11) and the ADF test (Table 14) show that EC is first-order integrated and has no unit root; it rejects the null hypothesis. EC is a stationary time series after diff. This allows us to examine the cointegration relationship.

3) Cointegration test

The constant is residual term is as followed (Table 15).

To testify for the cointegration relationship of three combinations, we proceeded an ADF test on the residual term (Table 16).

Table 12. Correlation-5-min.

Variables	(1)	(2)	(3)
(1) EC	1.000		
(2) BC	0.604	1.000	
(3) CC	0.909	0.770	1.000

Table 13. Dickey-fuller test for unit root-5-min.

Variable	<i>P</i> -value	<i>P</i> < 0.1
EC	0.218	Insignificant
BC	0.100	Significant
CC	0.074	Significant

Table 14. EC dickey-Fuller test for unit root in I (1).

Variable	<i>P</i> -value	<i>P</i> < 0.1
Diff_EC	0.000	Significant

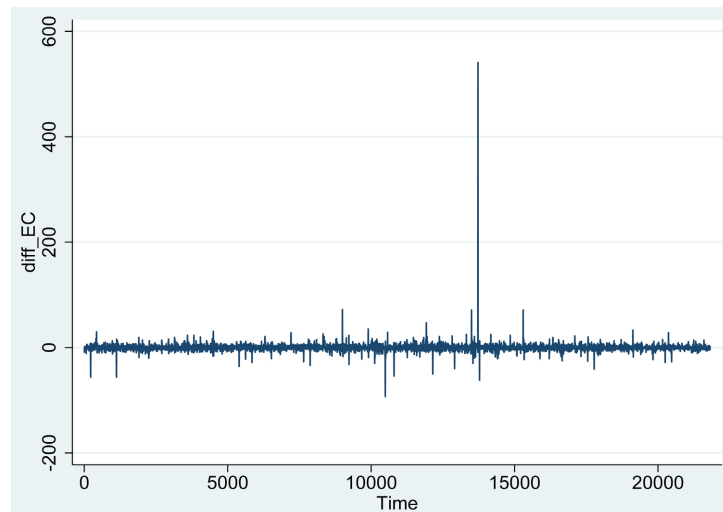


Figure 11. EC I (1) line graph.

Table 15. Residual constant term-5-min.

Function	α	β
(4)	2605.575	0.478
(5)	-2688.968	2.49
(6)	-3789.436	2.667

Table 16. Dickey-Fuller test for unit root-5-min.

Variable	<i>P</i> -value	<i>P</i> < 0.1
e_1	0.109	Insignificant
e_2	0.002	Significant
e_3	0.091	Significant

It is inferable that e_1 is a non-reposeful time series in its original function since Egg is a first-order integrated time series. Henceforth, we implemented the first order difference to e_1 (Table 17).

Thereafter, we showed all three combinations have cointegration relationships and proved the existence of mean reversion and arbitrage opportunity.

To find the trading signals, we need to calculate the Mspread (Figures 12-14).

Descriptive statistics

Based on Table 18, Mspreads have 4395 number of observation and the approximate means are zero. Mspread 2 has the smallest standard deviation of 117.148 followed by 224.121 (Mspread 1) and 226.745 (Mspread 3). The price fluctuation for the three combinations is around the mean value 0. This further satisfies the condition for mean reversion and arbitrage.

3.2. Stimulation Result

- Five minutes frequency

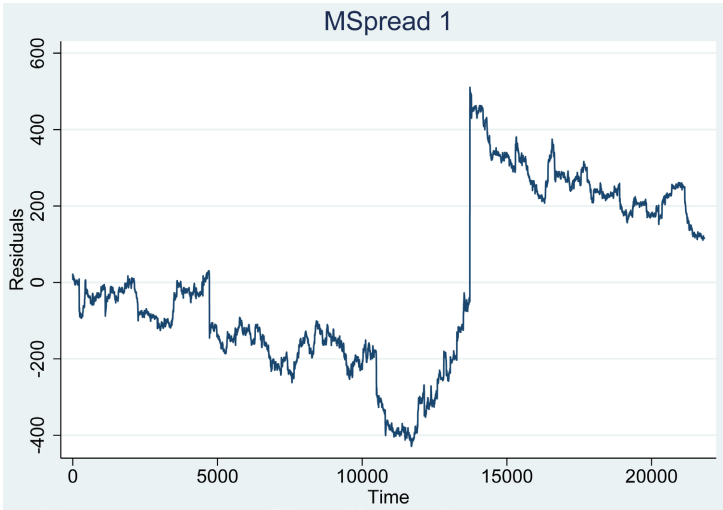


Figure 12. Mspread 2.1.

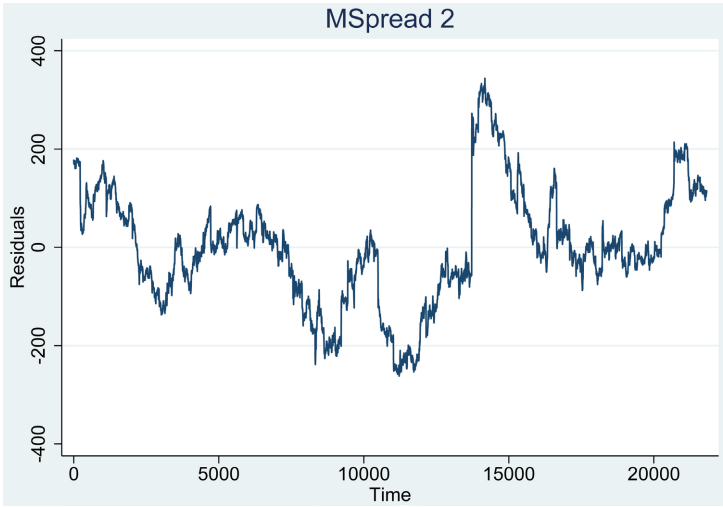


Figure 13. Mspread 2.2.

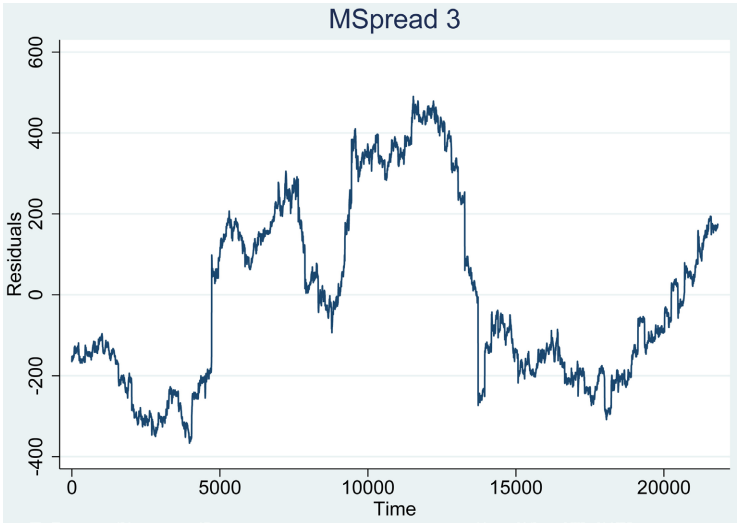


Figure 14. Mspread 2.3.

By abstracting data in sample to back-test our strategy while applying threshold from the adjustment phase, as well as setting corn dominant contract as a benchmark and frequency as 5-min every bar, we have back-tested our three combinations as followed (**Table 19**).

To interpret our results, we found from our table that 0.5 for EB pair and 0.45 for EC pair maximized the profit. Especially the result for EB, gaining 14.69% rate of return and 42.38% annual rate of return and this outperform all other pairs.

- **Daily frequency**

Daily frequency in comparison to 5-mins high frequency showed less sensitivity. The changes in Q value affected the result insignificantly as a result of less population of data. Nevertheless, the results of two pairs in the same range shared some similarities as below (**Table 20**).

Table 17. Dickey-fuller test for unit root in I (1).

Variable	P-value	P < 0.1
Diff_e ₁	0.000	Significant

Table 18. Descriptive statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Mspread 1	4365	0	224.121	−428.929	510.366
Mspread 2	4365	0	117.148	−262.074	343.776
Mspread 3	4365	0	226.745	−366.53	490.457

Table 19. Back-test 5-min.

Threshold	Pair	Yield	Annual Yield	Benchmark Yield	Division abnormal profit	Sharp ratio	Max drawdown
0.45	EB	14.11%	40.53%	9.68%	4.04%	2.282	4.57%
	EC	4.01%	10.67%	9.68%	−5.17%	0.685	5.39%
0.5	EB	14.69%	42.38%	9.68%	4.57%	2.284	4.57%
	EC	3.84%	10.21%	9.68%	−5.32%	0.637	5.40%
0.55	EB	13.41%	38.32%	9.68%	3.41%	2.047	4.57%
	EC	3.59%	9.51%	9.68%	−5.55%	0.565	5.41%
0.6	EB	11.18%	31.40%	9.68%	1.37%	1.619	4.96%
	EC	3.28%	8.68%	9.68%	−5.83%	0.471	5.93%
0.65	EB	10.16%	28.33%	9.68%	0.44%	1.391	5.30%
	EC	3.23%	8.53%	9.68%	−5.88%	0.448	6.05%
0.7	EB	12.09%	34.18%	9.68%	2.19%	1.620	4.57%
	EC	2.92%	7.69%	9.68%	−6.17%	0.364	6.26%
0.75	EB	11.10%	31.17%	9.68%	1.30%	1.461	4.57%
	EC	1.83%	4.78%	9.68%	−7.16%	0.072	6.27%

Table 20. Back-test day.

Threshold	Pair	Yield	Annual Yield	Benchmark Yield	Division abnormal profit	Sharp ratio	Max drawdown
0.45	EB	16.61%	48.58%	9.68%	6.32%	2.419	4.63%
	EC	4.33%	11.55%	9.68%	-4.87%	1.015	4.31%
0.5	EB	16.61%	48.58%	9.68%	6.32%	2.419	4.63%
	EC	0.07%	0.18%	9.68%	-8.76%	-0.450	5.91%
0.55	EB	16.21%	47.30%	9.68%	5.96%	2.347	4.63%
	EC	0.07%	0.18%	9.68%	-8.76%	-0.450	5.91%
0.6	EB	16.21%	47.30%	9.68%	5.96%	2.347	4.63%
	EC	-0.39%	-1.00%	9.68%	-9.18%	0.471	5.93%
0.65	EB	16.21%	47.30%	9.68%	5.96%	2.347	4.63%
	EC	-0.39%	-1.00%	9.68%	-9.18%	-0.575	5.91%
0.7	EB	15.64%	45.44%	9.68%	5.44%	2.243	4.63%
	EC	-0.39%	-1.00%	9.68%	9.18%	-0.575	5.91%
0.75	EB	15.64%	45.44%	9.68%	5.44%	2.243	4.63%
	EC	-0.39%	-1.00%	9.68%	9.18%	-0.575	5.91%

Once Again in our table, when the threshold of taking positions is set at 0.5, the profitability of the pair of egg-soybean meal outperforms every other pair with a 16.61% rate of return and 48.58% of rate of annual return. Condition is slightly different for the pair of egg-corn as the only Q value that would generate profit above 1% is 0.45 with a 4.33% rate of return and 11.55% rate of annual return.

4. Conclusion

A statistical arbitrage strategy designed for the three commodity future contracts is discussed in this paper. We used the cointegration method to confirm that these three future contracts have the characteristics of mean-reverting. These three future contracts were used to form pairs to arbitrage. We first set the trading ratio as 50:24 for egg-soybean meal, and 16:40 for egg-corn. Then we specifically focus on the threshold trading on the back-test platform. We found when the Q (t-value) is at 0.5, the performance of the egg-soybean meal pair in 5-min every bar excels yielding 14.69%. Egg-soybean meal pair in daily frequency is optimal at the Q value = 0.45 with a yield of 16.61%. Meanwhile, the egg-corn pair in both 5-mins and daily frequency yield 4.01% and 4.33% respectively. It performs best when Q (t-value) is 0.45. Although daily frequency results outperformed 5-mins results, this is because our trading signals may not be optimal. Since the 5-mins back-test presents better flexibility.

This paper shows that despite China's prosperity, the commodity futures

market remains unsophisticated in comparison to western countries. Cointegration and time series analysis have been used only by a few people to investigate these three products and design a strategy. These three commodities were studied by Kaiqi, Z., Ronghua, J. & Zhinan, L. (2019) in a similar manner for the purpose of hedging purposes. To analyze the effect of hedging, they used OLS, B-VAR, and ECM models. Alternatively, we are dedicated to arbitrage. As well, this article has the advantage of combining daily frequency data with five minutes high-frequency data, resulting in a variety of results from data acquired in the same period.

Nonetheless, this paper still has many flaws. The threshold is set in an invariant trading trigger which is rather inefficient as in a normal distribution mean, since a certain percentage of the population is neglected during the back-test. Hence in the future, we will implement machine learning to compare different approaches for selecting thresholds. For instance, the artificial neural network thresholds used in Roa, A. A. (2018) may make our strategy more precise and accurate. Furthermore, Zhao, Z., Zhou, R., & Palomar, D. P. (2019) research on unified optimization framework is also intriguing.

In conclusion, we illustrated the feasibility of pair trading for egg, corn, and soybean meal future contracts. Investors and arbitrageurs can use this article as a reference to apply to their own situations.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Engle, R. F., & Granger, C. W. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica: Journal of the Econometric Society*, 55, 251-276. <https://doi.org/10.2307/1913236>
- Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs Trading: Performance of a Relative-Value Arbitrage Rule. *The Review of Financial Studies*, 19, 797-827. <https://doi.org/10.1093/rfs/hhj020>
- Hendry, D. F., & Juselius, K. (2000). Explaining Cointegration Analysis: Part 1. *The Energy Journal*, 21, No. 1. <https://doi.org/10.5547/ISSN0195-6574-EJ-Vol21-No1-1>
- Shen, L., Shen, K., Yi, C., & Chen, Y. (2020). An Evaluation of Pairs Trading in Commodity Futures Markets. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 5457-5462). IEEE. <https://doi.org/10.1109/BigData50022.2020.9377766>
- Stübinger, J., & Bredthauer, J. (2017). Statistical Arbitrage Pairs Trading with High-Frequency Data. *International Journal of Economics and Financial Issues*, 7, 650-662.
- Xu, H. (2017). *High Frequency Statistical Arbitrage with Kalman Filter and Markov Chain Monte Carlo*. Master's Thesis, University of Waterloo.
- Kaiqi, Z., Ronghua, J. & Zhinan, L. (2019). Industry Chain Hedging Plan of Laying Hens Breeding Company. *Journal of China Agricultural University*, 24, 219-229.
- Wu, Y.H., & Fu, Q. (2018). Analysis of Egg Price Fluctuation and Cause. *The Journal of Agricultural Science*, 10, 581-587. <https://doi.org/10.5539/jas.v10n11p581>

- Zhao, Z., Zhou, R., & Palomar, D. P. (2019). Optimal Mean-Reverting Portfolio with Leverage Constraint for Statistical Arbitrage in Finance. *IEEE Transactions on Signal Processing*, 67, 1681-1695. <https://doi.org/10.1109/TSP.2019.2893862>
- Roa, A. A. (2018). *Pairs Trading: Optimal Thershold Strategies*. Master's Thesis, Universidad Complutense de Madrid. <https://deanstreetlab.github.io/papers/papers/Statistical%20Trading/Pairs%20Trading%20-%20Optimal%20Threshold%20Strategies.pdf>
- Yang, J., Li, Z., & Wang, T. (2021). Price Discovery in Chinese Agricultural Futures Markets: A Comprehensive Look. *Journal of Futures Markets*, 41, 536-555. <https://doi.org/10.1002/fut.22179>
- Ao, J., & Chen, J. (2020). Price Volatility, the Maturity Effect, and Global Oil Prices: Evidence from Chinese Commodity Futures Markets. *Journal of Economics and Finance*, 44, 627-654. <https://doi.org/10.1007/s12197-019-09497-1>
- Vidyamurthy, G. (2004). *Pairs Trading: Quantitative Methods and Analysis*. John Wiley & Sons.