

# Research on the Intertemporal Arbitrage Strategy of CSI 300 Stock Index Futures

Chen Liwen<sup>1,\*</sup>, Liu Huangjin<sup>2</sup>

<sup>1,2</sup> *School of Economic & Management, Nanjing University of Science and Technology, Nanjing, China*  
*\*Chen Liwen. E-mail: chendi95@126.com*

## ABSTRACT

With the great uncertainty of the stock market, stock index futures undertake the function of risk aversion, speculation and arbitrage. Intertemporal Arbitrage method is universally adopted in CSI 300 stock index futures. This paper analyses the theories of statistical arbitrage and high-frequency arbitrage firstly. Then the high-frequency data of 1 minute CSI 300 index future prices are used to conduct an intertemporal arbitrage strategy. Co-integration analysis uses historical data to analyze the long-term spread equilibrium relationship of different term contracts, so as to use the cointegration relationship to predict future spreads. When the market spread deviates from the equilibrium interval, an inter-temporal arbitrage strategy can be established. The backtesting results with the sample data and 60-minute pricing data prove that the strategy is feasible and obtains a certain arbitrage profits. The empirical result offers a practical plan for the research on futures arbitrage.

**Keywords:** *intertemporal arbitrage strategy, statistical arbitrage, cointegration test, high-frequency data, pairing transaction*

## 1. INTRODUCTION

The stock index futures are standardized contracts with the CSI 300 index as the target, which covers the integrated price of 300 the best performed stocks in A stock market of China, thus having strong representation and response to the market cyclical fluctuations timely. Intertemporal spread arbitrage benefits from building and hedging the same amount, opposite positions, different months, and the same underlying futures contracts. It is divided into bull market arbitrage, bear market and butterfly arbitrage. Therefore, CSI 300 index futures can effectively avoid risks, protect investors' stable returns, restrain the excessive volatility of stock market prices, and play the real role of 'barometer' of economy.

Cornell B. and French K.R.(1983)<sup>[1]</sup>, under the assumption of a perfect market, derived a simplified futures pricing model in which the futures price equals the holding cost of the spot price plus the spot-to-future delivery date.

However, in practice, it is found that the market price is often different from the theoretical price. Then the perfect market assumption is continuously expanded, and the real market factors are gradually introduced, and the model is revised to make it more in line with the market conditions. Modest and Sundaresan(1983)<sup>[2]</sup> found the existence of no-arbitrage interval by transaction cost and short-selling mechanism, plus the assumption of dividend factors and non-random changes in dividends. Based on the cost-holding pricing model, no arbitrage interval was obtained. C. J. Granger(1987)<sup>[3]</sup> proposed a cointegration method to solve the problem that non-stationary time series cannot use

classical regression and avoid pseudo-regression. Then the statistical method plus cointegration theory became the main method to study the series of stock index futures spreads.

Klemkosky and Lee (1991)<sup>[4]</sup> consider the different transaction costs and lending rates, taxation, dividend yield and other factors to construct a spot and futures portfolio strategy to obtain no arbitrage interval prices, thus deriving the no-arbitrage pricing range. Neal(1995)<sup>[5]</sup> used high-frequency data to empirically analyze the cost-holding model and found that there is a deviation between the futures theoretical price and the actual price. The arbitrage opportunity exists, but it is difficult to capture. Henker and Martens (2005)<sup>[6]</sup> illustrate that arbitrage-free sensitivity reflects changes in the cost of a transaction.

Zhong Qunqiu and Xi Juncheng (2008)<sup>[7]</sup> used the CSI 300 simulation transaction data to study the inter-temporal arbitrage model based on cointegration and the spread arbitrage strategy based on holding cost. The former can get better results because the latter is in the parameters. There is a lack of time, control and controllability. Lirong Zhu, Xin Su and Yong Zhou (2015)<sup>[8]</sup> used the cointegration model, error correction model and GRACH model to study the arbitrage strategy of commodity futures, and considered the situation of opening, closing and stop loss, and considered the income under the three models. The rate gap is minimal. Zhu Huasheng, Zhou Zhizhong(2018)<sup>[9]</sup> found that the restricted trading rules amplified the volatility of stock index futures and weakened the risk management function. Reasons of this elimination were analysed through EGARCH model and practical advices were supported. Miao Yu, Zhu Jiaming(2019)<sup>[10]</sup> optimized the traditional assumption that the variance is constant through GARCH model.

The mainstream theory is divided into two aspects. The first is mainly based on the theory of holding costs. The value of a futures contract is the current value of the underlying asset plus the final value of a series of holding costs. When the spot price and the futures price difference fall outside the long-term equilibrium spread range, the position-opening operation can be carried out. When the price difference returns to the reasonable price range, the contract is closed, thereby completing the complete risk-free arbitrage. The cost of ownership theory is mainly applied to the current market. The other is a statistical arbitrage based on cointegration. Statistical arbitrage detects the long-term equilibrium relationship of contracts of different maturities. When the contract spread deviates from the stable equilibrium relationship, it indicates that the contract value is overvalued or undervalued, and the arbitrage operation is carried out by using abnormal market fluctuations to achieve risk-free arbitrage.

This paper studies the arbitrage method of using spreads between stock index futures contracts. Comparing the theory of pricing of stock index futures and Statistical arbitrage models, 1 minute high frequency data of CSI 300 index futures IFL0 and IFL1 contracts were selected for research and co-integration test were used to build up arbitrage strategy. Based on the principle of convergence of the research object, the trading rules were set up by gap distribution, Finally sample data is used back to measure the feasibility of model in effect.

## 2. RESEARCH BACKGROUND

### 2.1 Theory Explanation

#### 2.1.1 The principle of statistical arbitrage

Statistical arbitrage is based on the linkage effect between related contracts. Specific steps: By tracking historical data, analyzing whether the price time series is stable, whether the existence of equilibrium regression is present, and then analyzing the long-term equilibrium relationship of the price, predicting future price changes based on the characteristics of the price difference series; using the model to track market data When the market price deviates from the equilibrium no-arbitrage interval, it can only carry out quantitative operations and reverse operations, and strive to obtain risk-free profits.

#### 2.1.2 Definition of statistical arbitrage

Statistical arbitrage is defined as:

- (1)  $V(0) = 0$
- (2)  $\lim_{t \rightarrow \infty} E[v(t)] > 0$
- (3)  $\lim_{t \rightarrow \infty} \frac{Var[v(t)]}{t} = 0$
- (4)  $\lim_{t \rightarrow \infty} P[v(t)] = 0$

Which  $V(t)$  represents the cumulative discounted value. By definition, statistical arbitrage satisfies four conditions: zero initial investment cost, positive expected discounted profit, and the probability of loss converges to zero. If the probability of loss is zero, the time-averaged variance converges to zero.

The arbitrage operation of statistical arbitrage does not stick to the macroscopic changes in the whole market, and pays more attention to the linkage effect between related assets. Therefore, the arbitrage accuracy based on statistical tools is higher, and the arbitrage opportunities can be found in time to carry out continuous and stable arbitrage operations.

### 2.2 Data Selection

The CSI 300 stock index futures contract currently has a total of four contract types, namely the current month contract, the next month contract and the subsequent two quarterly contracts. This paper selects 1 minute high frequency data, which is the 1 minute closing price of the trading period of January 10, 2019, 9:31-2019, April 10, 15:00, and the sample capacity is 13,920.

## 3. DATA ANALYSIS

### 3.1 Analysis of contract correlation

In the first step, the logarithm of the four price series is taken, the absolute value of the data is reduced, and the collinearity, heteroscedasticity of the model are also weakened, and correlation analysis is performed. In the second step, comparing the correlation coefficients between the two pairs, it can be seen that the IFL0 current month contract and the IFL1 next month contract have the highest correlation, so the two contracts are selected for cointegration test. The results of the contract correlation analysis are shown in Table 1.

**Table 1** Correlation Coefficient

Data	LNIFL0	LNIFL1	LNIFL2	LNIFL3
LNIFL0	1.0000			
LNIFL1	0.9994	1.0000		
LNIFL2	0.9968	0.9983	1.0000	
LNIFL3	0.9925	0.9947	0.9983	1.0000

### 3.2 Data stability test

Next, ADF method is used to check the unit root and determine whether the IFL0 current month contract and the IFL1 current month contract are stable.

According to the ADF test result of LNIFL0 and LNIFL1, the t values are significantly greater than the 10% threshold, and the corresponding adjoint probability P values is

significantly greater than 0.05, and the original hypothesis has a unit root, so the original price sequence is non-stationary. Perform a first-order difference on the original sequence and then perform an ADF unit root test. The results show that the first-order differential price series passes the t-test of the significance level of 1%, which is obviously smaller than the critical value, and meets the requirement that the P value is less than 0.05. The original hypothesis is rejected, indicating that the differential price series meets the requirements of wide stability.

**Table 2** Regression Equation Results

Explained variable	Explanatory variables	Regression coefficients	Standard deviation	T-value
LNIFL1	C	-0.367	0.002625	-139.8
	LNIFL0	1.044	0.000316	3307.3
R-square	0.998729			

Next, the resulting residual sequence is tested for stationarity, and the residual sequence after the equation regression is a stationary sequence at the 1% significance level. The test results are shown in table 3.

**Table 3** ADF Test of Residual

	T-value	1%	5%	Smooth or not
Residual	-8.5025	-3.9587	-3.4101	Yes

The co-integration analysis of the 1-minute high-frequency logarithmic price series of the futures IFLO and IFL1 is obtained. Through the first-order difference of the price series, the unit root test passes, and the sequence is wide and stable, avoiding the pseudo-regression. In the ADF test of the residual sequence after regression, the stationary sequence was a 1% significance level. There is a cointegration relationship in the proof contract. The regression equation obtained by the OLS method is:

$$LNIFL1 = 1.0440 \times LNIFL0 - 0.3670 \quad (3307.3) \quad (-139.8)$$

Due to this kind of equilibrium relationship, the coefficient of LNIFL0 in the regression model can be used as the hedging ratio of the transaction, and 1 unit of IFLO requires 1 unit of IFL1 for reverse hedging.

### 3.4 Characteristic analysis of spread sequence

The basic statistics of the residual sequence were analyzed, and the mean and standard deviation were found to be 1.25e-15 and 0.001548 respectively. The results are shown in Table 4.

### 3.3 Co-integration relationship test

After the relevant contract price series satisfies the first-order single-step process, the EG two-step method is used to judge the long-term equilibrium relationship between the two. The results show that the linear regression equation has a good fitting effect. The specific test results are shown in Table 2

**Table 4** Statistics of the Residual Sequence

Residual	Mean	Standard deviation	JB-value	P-value
Spread	1.25e-15	0.001548	540236.4	0.0000

The decentralized spread is represented by Msspread, which can be used to derive the decentralized spread sequence:

$$\begin{aligned} Msspread &= \text{Spread} - \text{mean}(\text{Spread}) \\ &= LNifl1 - 1.0440LNifl0 + 15.3670 \\ &\quad - 1.25e \end{aligned}$$

## 4. MODELLING

### 4.1 Formulation of trading strategy

The residual sequence modeling includes various modeling methods such as Pandit-Wu model, ARMA-GARCH model, nonparametric method and mixed normal distribution. This paper adopts the nonparametric method commonly used in academia.

The standard deviation of the residual sequence is  $\sigma$ , the arbitrage opening threshold, the closing threshold and the stop threshold are respectively  $K1$ ,  $K2$ ,  $K3$  and the upper and lower boundaries of the transaction are the standard deviation of the spread series multiplied by the threshold,  $K_n \times \sigma$ . Vidyamurthy G. (2004)<sup>[11]</sup> argues that if the decentralized spread sequence is a white noise sequence,  $\pm 0.75 \times \sigma$  is the optimal open trigger signal. In this paper, the opening threshold is  $\pm 0.75 \times \sigma$ , and the closing threshold is the price difference returned to the equilibrium spread ( $-0.75 \times 0.001548$ ,  $0.75 \times 0.001548$ ), that is ( $-0.001611, 0.001611$ ) to close the position. The stop loss threshold is set to 2 standard deviations.

(1) The opening signal of the bull market arbitrage strategy. When  $MS_{spread} > +3/4 \times \sigma$ , the price difference deviates upward to a certain extent, and 1 IFL0 contract is sold while selling 1 IFL1 contract. The bull market intertemporal arbitrage forecast spreads, buying into the near month to sell the distant month, when the spread between the two contracts returns to the equilibrium level, investors close the position and take risk-free returns.

(2) The opening signal of the bear market intertemporal arbitrage.

When  $MS_{spread} < -3/4 \times \sigma$ , the price difference deviates downward to a certain extent, and 1 IFL0 contract is sold while buying 1 IFL1 contract. The bear market arbitrage strategy predicts a decrease in the spread, and sells the near-month to buy the distant month. When the price difference returns to the equilibrium interval, the reverse closing operation is performed. At this point, the actual price of the forward contract is undervalued, or the price of the near-month contract is overvalued.

(3) Close the signal.

When  $|MS_{spread}| < 3/4 \times \sigma$ , the price difference sequence returns to the  $\pm 3/4 \times \sigma$  no-arbitrage interval and the clearance is cleared.

(4) Stop loss signal.

When the spread is further expanded to an area other than  $\pm 2\sigma$ , the stop loss is immediately closed to avoid the loss caused by the  $MS_{spread}$  not converge as scheduled.

#### 4.2 Strategy performance test

After the arbitrage model is implemented programmatically, the 60-minute spread data from April 1 to May 3, 2019 is selected for backtesting, and the arbitrage strategy is obtained. The price difference sequence of the CSI 300 stock index futures in the sample has 62 triggers to open the position, 7 stop trigger signals, and more arbitrage opportunities. As shown in Figure 1.

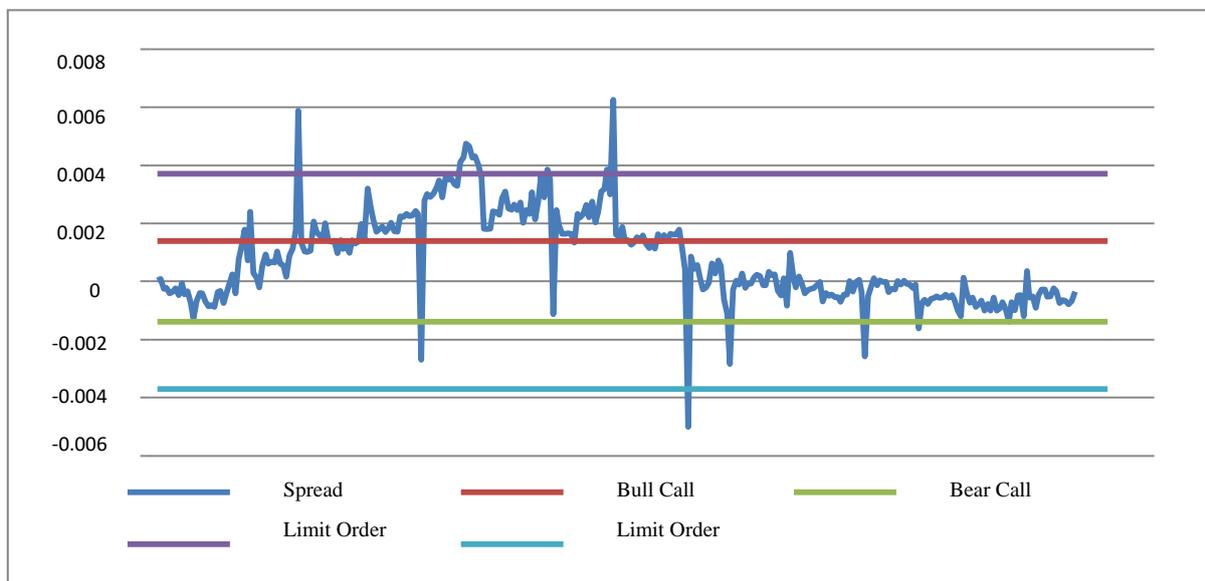


Figure 1 Results of arbitrage strategy for CSI 300 stock index futures

Since the average arbitrage time of the model established by using 60-minute high-frequency data is too short, the data backtest from April 1 to May 1, 2019 is selected, and the following cost calculation method is used to specify the arbitrage performance.

Assume that the IFL0 and IFL1 contract points are A and B, respectively. The points when closing the position are A1 and B1.

$$\begin{aligned} \text{Position cost} &= \text{margin} + \text{handling fee} \\ &= \text{contract value} \times (8\% + 0.01\%) \\ &= (A + B) \times 300 \times (8\% + 0.01\%) \end{aligned}$$

Earn the difference when you open the position and close the position:

$$[(A - A_1) + (B_1 - B)] \times 300 - (A_1 + B_1) \times 300 \times 0.01\%$$

Downward position to close the position to earn the spread is:

$$[(A_1 - A) + (B - B_1)] \times 300 - (A_1 + B_1) \times 300 \times 0.01\%$$

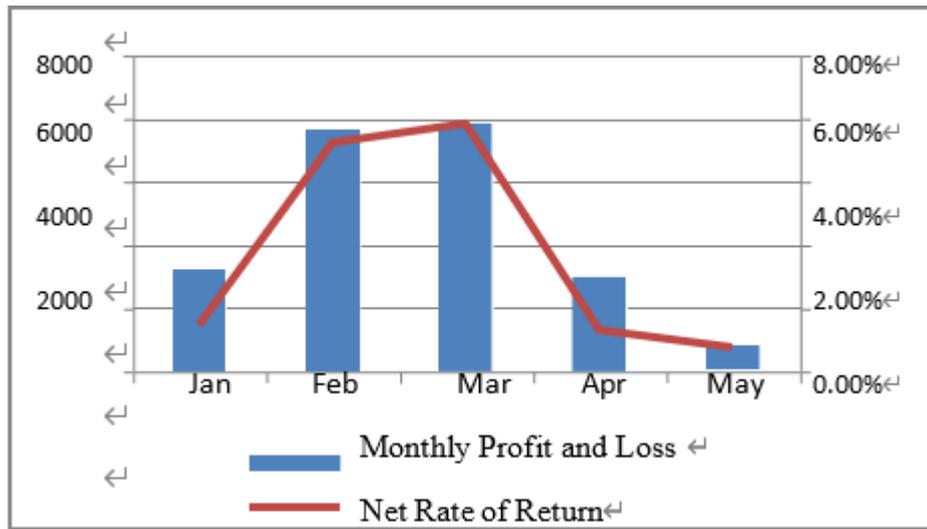
According to the daily data backtest, the cumulative net rate of return is 22.46%, resulting in 6 down arbitrage opportunities and 3 upward arbitrage opportunities. The single rate of return is up to 17.4%. The return space is smaller than other cross-species transactions. The number of high-frequency data samples in futures is large, and the threshold for opening and closing is small due to the small standard deviation of residuals, so there are not many arbitrage opportunities.

The 60-minute high-frequency data was brought into the program for measurement. From the back-test results, it was found that 69 inter-period arbitrage was carried out, and the cumulative yield reached 23.66%, as shown in Table 5.

**Table 5** Trading Results on 60minutes data

Statistical indicators	Number
Number of sample data	424
Number of operations	69
Cumulative profit points	165.7
Accumulated loss points	65
Net profit points	107.7
Cumulative net rate of return	23.66%

From the monthly profit and loss analysis, it can be seen that there is no monthly loss in the month. Compared with the daily data, the high-frequency data back-testing arbitrage in the sample is more frequent, and a higher cumulative yield is obtained, as shown in Fig. 2.



**Figure 2** Monthly profit and net rate of return

From the actual backtest results of the intertemporal arbitrage model, from January 1 to May 3, 2019, the Shanghai and Shenzhen 300 stock index futures 60 minutes of high frequency data samples backtest, a total of 62 trigger opening signals; sample data Backtesting, 9 times to trigger the opening signal, the cumulative net rate of return reached 23.66%, the loss point is significantly smaller than the profit point. The data backtest from April 1 to May 1, 2019 can specify the operation process of the arbitrage strategy, indicating that the model has certain arbitrage space and feasibility in China's stock index futures market. In general, the intertemporal arbitrage model based on statistical arbitrage and cointegration theory can provide some suggestions and references for investors

**5. CONCLUSION**

This paper uses the high-frequency data of the 1 minute closing price of the CSI 300 stock index futures to explore the arbitrage opportunities between the two highly relevant different expiring contracts with co-integration relationship and construct a corresponding statistical arbitrage mechanism. Through empirical backtesting, when the opening position is set to 0.75 times standard deviation, the stop loss point is 2 times standard deviation, and the closing position is the equilibrium price range, there are a large

number of statistical arbitrage opportunities. This model provides a feasible idea for investors.

**REFERENCES**

[1] Cornell B., French K. The pricing of stock index futures[R]. *Journal of Future Markets*, 1983,3(01):1~14

[2] Klemkosky. R.C. and J.H. Lee. The Intraday Ex Post and Ex Ante Profitability Index Arbitrage[J]. *The Journal of Futures Markets*, 1991:291~311

[3] D.M. Modest and M.Sundaresan. The relationship between spot and futures prices in stock index futures markets: Some preliminary evidence[J]. *The Journal of Futures Markets*, 1983,3(1):15~41

[4] Engle R.F., and Grange C. W. J. Co-integration and Correction: Representation, Estimation and Testing [J]. *Econometrica*, 1987,55(2):251~276.

[5] Burgess N. Statistical arbitrage models of the FTSE100[R]. *Computational Finance*,1999,The MIT Press,2000:297~312

[6] Qiu Zhongqun, Cheng Xijun. Inter-temporal arbitrage strategy model of stock index futures based on cointegration[J]. *Systems Engineering*, 2008, (12): 26-29.

[7] Yang Xiaoqiang. Inter-temporal arbitrage analysis of stock index futures based on cointegration theory [N]. Futures News, 2008-06-03 (003).

[8] Li Shiwei. Research on Intertemporal Arbitrage of Shanghai and Shenzhen 300 Stock Index Futures Based on Cointegration Theory[J]. Journal of China Jiliang University,2011,(2):198-202.DOI:10.3969/j.issn.1004-1540.2011.02.021.

[9] Zhu Huasheng, Zhou Zhizhong. The Influence of High Frequency Trading on Stock Index Futures Market in China: Based on an Empirical Analysis of Regulation Adjustment of Stock Index Futures Market in Recent Years[J]. Shanghai Management Science, 2018,40(04):37-41.

[10] Miao Yu, Zhu Jiaming. Analysis on CSI 300 Stock Index Futures Arbitrage Strategy: Based on GARCH empirical model[J]. Journal of Beijing Institute of Graphic Communication, 2019,27(05):77-80.

[11] Vidyamurthy G. (2004) .Pairs Trading: Quantitative Methods and Analysis[R]. Hoboken, New Jersey: John Wiley&Sons.