An Algorithm for Spread Arbitrage Process in the CSI-300 Futures Market

Jiayue Xu
Glorious Sun School of Business and Management,
Donghua University
Shanghai, China, +86 13764267866
yuejiaxu@gmail.com

Hong Fan
Glorious Sun School of Business and Management,
Donghua University
Shanghai, China, +86 15201987180
hongfan@dhu.edu.cn

***Abstract*—With the development of modern capital markets, the stock index futures have become one of the most active investment instruments in the world. The arbitrage of stock index futures actually plays a very important role in making the price of stock index futures be rational and activating the market. In the present paper, we propose an algorithm of spread arbitrage process, which aims to improve the statistical arbitrage strategy of pairs trading based on co-integration. Firstly, the long-term relationships among index futures are detected by co-integration tests. Then the algorithm is applied to the IF1207 and IF1208 contracts of CSI300 futures in China to test its performance. The empirical results suggest that the average profit from spread arbitrage is statistically significant and the rates of return of spread arbitrage are very attractive.**

Keywords-CSI-300 stock futures; Algorithm of spread arbitrage; Statistic arbitrage; Co-integration model

I. INTRODUCTION

With the development of modern capital markets, the stock index futures have become one of the most active investment instruments in the world. The CSI-300 stock index futures contract started to be traded on April 16, 2010. As the first index futures contract in China, it's not only a new transaction type in China's futures market, but also a representative of an effective risk avoiding instrument. Making full and efficient use of the CSI-300 stock index futures has become an issue of common concern in this field.

Many researches have been done in this field. Burgess(1999) proposes three steps of using statistical arbitrage method; in the first step, a portfolio which is composed of long and short positions is build up; then co-integration regression is made to create error correction mechanism; finally, trading system is implemented and the predicable component of returns is explored[1]. Ganapathy Vidyamurthy (2004) divides pairs trading into two parts: statistical arbitrage and risk arbitrage. He holds the view that statistical arbitrage is based on relative pricing theory[2]. Qiu and Cheng (2008) test the validity and efficiency of statistical arbitrage strategy model on co-integration by the simulating trading data, and the result indicates that there also has arbitrage space in index futures at home[3].

In the present paper, according to the theory of statistical arbitrage, we propose an algorithm to make improvements on conventional strategy of pairs trading based on co-integration. In the algorithm, we create a signal index that indicates the beginnings of transaction and stop-loss points. Subsequently, we use intraday high frequency data (5 minutes) to test the algorithm. The empirical results suggest that more arbitrage opportunities and more profits can be achieved by the algorithm.

The rest of paper is organized as follows. Section 2 presents the definition and theory of statistical arbitrage. Section 3 proposes the algorithm of statistical arbitrage strategy. Section 4 shows the empirical results. Section 5 concludes with some remarks on potential further research.

II. STATISTICAL ARBITRAGE

A. Mathematical Definition

S. Hogan, R. Jarrow and M.Warachka (2004) give the mathematical definition of statistical arbitrage: A statistical arbitrage is a zero initial cost, self-financing trading strategy $\left(\mathrm{x}\left(\mathrm{t}\right):\mathrm{t}>0\right)$ with cumulative discounted value $v\left(\mathrm{t}\right)$ so that:

(i) $v\left(0\right)=0$

(ii) $\lim_{t\to\infty} E\left[v\left(t\right)\right]>0$

(iii) $\lim_{t\to\infty} P\left[v\left(t\right)<0\right]=0$

(iv) $\lim_{t\to\infty} \frac{Var\left[v\left(t\right)\right]}{t}=0$, if $\forall t<\infty, P\left[v\left(t\right)<0\right]>0$

A statistical arbitrage satisfies four conditions (i) it is a zero initial cost and self-financing trading strategy, that in the limit has (ii) positive expected discounted profits, (iii) a probability of a loss converging to zero, and (iv) a time-averaged variance converging to zero if the probability of a loss not become zero in finite time.

B. Principle

Statistical arbitrage holds that if two highly correlated assets maintain this favourable correlation in the future, then the deviation of prices between assets can be corrected when this deviation happens which aims to create arbitrage opportunities. From the practice of statistical arbitrage, investors should buy the relatively poor assets and sell the good assets when two assets deviate from each other, and then close positions when the deviation gets corrected.

The underlying basic concept behind the principle of statistical arbitrage is mean-reversion which represents the price of asset will return to its long-term equilibrium value and the price series of asset are stationary processes in practice. If the series are stationary, we can create a signal discovery mechanism indicating whether the asset price deviates from its long-term equilibrium value and whether arbitrage opportunities exist or not. In the next section, we will propose an algorithm of the signal discovering.

C. *Algorithm of Arbitrage Process*

As a consequence, firstly we should take a test of co-integration on different contracts in order to confirm whether this relationship exists. If there is a co-integration relationship, the prices of contracts will maintain equilibrium in the long run, and will not deviate from each other without limit, resulting in the arbitrage to be possible. On the contrary, if there is not a co-integration, the prices of contracts will be likely to deviate from each other without limit, leading to huge risks for arbitrage. Hence we can achieve arbitrage by applying a trading rule only if there is a long-term equilibrium between assets.

The algorithm of arbitrage process can be summarized as following steps:

Suppose that the 5 minute's price series of CSI-300 near month contracts and far month contracts in every trading day are $\{IFL0_t, 1 \le t \le T\}$ and $\{IFL1_t, 1 \le t \le T\}$. Their logarithmic price series are $\{Ln(IFL0_t), 1 \le t \le T\}$ and $\{Ln(IFL1_t), 1 \le t \le T\}$. The algorithm steps are as follows:

(1) Select the first T_1 pairs of data for training to get some t trading parameters. At the beginning, we employ Engle-Granger two steps method for co-integration test of these data: if they are not co-integrated, algorithm ending; if they are co-integrated, takes Step (2).

(2) Thinking about the simple regression model $Ln(IFL1_t) = \beta \times Ln(IFL0_t) + \varepsilon_t, 1 \le t \le T_1$. Using OLS to estimate β and calculate the spread $sp_t = Ln(IFL1_t) - \beta \times Ln(IFL0_t), 1 \le t \le T_1$. Calculate the mean μ and standard deviation σ.

(3) Setting parameters such as enter market threshold λ, closing position threshold δ and stop-loss threshold χ. Begin loop: $t \leftarrow T_1 + 1$. ($open = 0$ means short position, $open = 1$ means long position).

(4) When $t = T$, if $open = 1$, force liquidation, $open \leftarrow 0$;

(5) When $t < T$, take step (5).

(6) Calculate $sp_t = Ln(IFL1_t) - \beta \times Ln(IFL0_t)$.

(7) When $open = 0$:

If $\lambda\sigma < sp_t - \mu \le \chi\sigma$ (case A), then buy near month contracts and sell far contracts. $open \leftarrow 1$;

If $-\chi\sigma < sp_t - \mu \le -\lambda\sigma$ (case B), then sell near month contracts and buy far contracts. $open \leftarrow 1$;

When $open = 1$:

If $\lambda\sigma < sp_t - \mu \le \chi\sigma$ (case A), until if $sp_t - \mu \le \delta\sigma$, $open \leftarrow 0$;

If $\lambda\sigma < sp_t - \mu \le \chi\sigma$ (case A), until if $sp_t - \mu > \chi\sigma$, $open \leftarrow 0$, stop loss, break;

If $-\chi\sigma < sp_t - \mu \le -\lambda\sigma$ (case B), until if $sp_t - \mu \ge -\delta\sigma$, $open \leftarrow 0$;

If $-\chi\sigma < sp_t - \mu \le -\lambda\sigma$ (case B), until if $sp_t - \mu < -\chi\sigma$, $open \leftarrow 0$, stop loss, break;

Else $t \leftarrow t+1$, back to (4).

III. DATA AND METHODOLOGY

A. *Data Selection*

Five minutes synchronous prices of contracts IF1207 and IF1208 from CSI300 futures are used in the empirical analysis; meanwhile these contracts are dealt with by a logarithmic process. The 432 pairs of prices observations from 18 June 2012 to 28 June 2012 are obtained from Dazhihui Stock Exchange System. Data series are firstly analyzed by using Eviews6.0.

Figure 1 presents the plot of synchronous logarithmic prices of contracts IF1207 and IF1208. Obviously, two series are unstationary, and all prices move in the same pattern, suggesting they are contemporaneously highly correlated and probably have a long-term equilibrium relationship.

Figure 1. Price of Contract IF1207 and IF1208

B. *Correlation Analysis:*

Correlation is a statistical method describing closeness of relationships among variables. The quantitative indicator of this closeness is correlation coefficient. In this dissertation, simple linear correlation method is used to calculate correlation coefficient between two contracts. The formula is as follows:

$$r=\frac{\sum\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}$$

x represents contract IF1207, y represents contract IF1208.

According to the data, $r=0.99986$, it is obvious that IF1207 is strongly correlated with IF1207 and they have positive correlation. Their prices move the same way with each other. Therefore, arbitrage opportunities exist between IF1207 and IF1208.

C. *ADF Test*

Additionally, this possibility can be tested by ADF Unit Root Test, followed by the results in Table I. T-statistics of ADF test of LN(IF1207) and LN(IF1208) are larger than those of test critical values at 1% level, 5% level and 10% level. T-statistics of ADF test of D(Ln_IF1207) and D(Ln_IF1208) are larger than those of test critical values at 1% level, 5% level and 10% level while Prob. value is approximate to 0. So the first two different sequences in the 1% significance level mean we should reject the original assumption of the existence of a unit root. The log-prices series of two contracts both have a unit root while their first order difference sequences both have not, suggesting that the log prices series are first-order integration $I(1)$ series.

TABLE I. AUGMENTED DICKEY-FULLER UNIT ROOT TEST

Null Hypothesis: Ln(IF1207) has a unit root			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.735095	0.3978
Test critical values:	1% level	-2.570326	
	5% level	-1.941558	
	10% level	-1.616207	
Null Hypothesis: Ln(IF1208) has a unit root			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.770643	0.3820
Test critical values:	1% level	-2.570326	
	5% level	-1.941558	
	10% level	-1.616207	
Null Hypothesis: D(LN_IF1207) has a unit root			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-14.19869	0.0000
Test critical values:	1% level	-2.570338	
	5% level	-1.941560	
	10% level	-1.616206	
Null Hypothesis: D(LN_IF1208) has a unit root			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-14.06122	0.0000
Test critical values:	1% level	-2.570338	
	5% level	-1.941560	
	10% level	-1.616206	

D. *Co-integration Test*

Then we can determine whether they are co-integrated by E-G co-integration test[4].

First of all, we estimate the regression equation using ordinary least square method, taking log-price of IF1207 as dependent variable and log-price of IF1207 as independent variable. Estimated regression model is

$$Ln\left(IF1208\right)=-0.001346+1.000355Ln\left(IF1207\right)+\varepsilon_t$$

Secondly, we take unit root test on residual sequences $\{\varepsilon_t\}$. The result is given in Table II. We can easily find that the residual series is stationary without unit root, which means the two series of log-price for IF1207 and IF1208 are co-integrated and have a long-term equilibrium relationship. As a result, the short time price deviation can be considered as arbitrage opportunity. According to the regression model, the proportion of IF1207 and IF1208 in portfolio is 1: 1.000355 (since the number of contract is integer in the real transaction, this proportion can be assumed as 1:1 approximately). The log-spread is given as $sp_t\text{=}Ln\left(IFL1_t\right)\text{-}Ln\left(IFL0_t\right)$, and the residual $\varepsilon_t\text{=}sp_t\text{-}\mu$ with the plot being presented in Figure 2.

TABLE II. AUGMENTED DICKEY-FULLER UNIT ROOT TEST ON RESIDUAL

Null Hypothesis: RESID01 has a unit root			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.862538	0.0000
t critical values:	1% level	-2.570364	
	5% level	-1.941563	
	10% level	-1.616204	

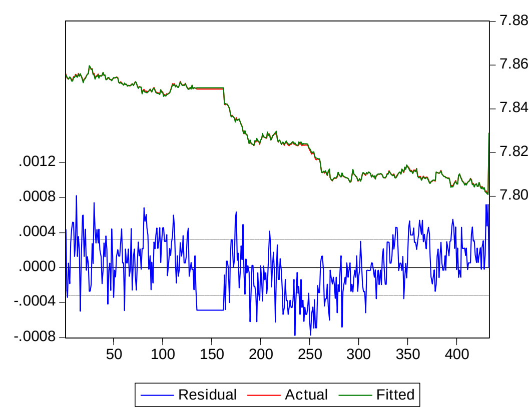

Figure 2. Logarithmic spread of IF1207 and 1208

E. *Trading Signals and Stop-loss Point*

Suppose that enter market threshold λ=1.286 (90% quantile of standard normal distribution, ensure the arbitrage trading will not loss at a possibility of 90%), close position threshold δ=0.991 (84% quantile of standard normal distribution) and stop-loss threshold $\chi=2.1$ (99% quantile of standard normal distribution).

IV. Arbitrage Analysis and Empirical Results

The 432 pairs of five minutes synchronous prices of contracts IF1207 and IF1208 from CSI300 futures observations from 18 June 2012 to 28 June 2012 are used as sample interval to test the performance of the arbitrage algorithm

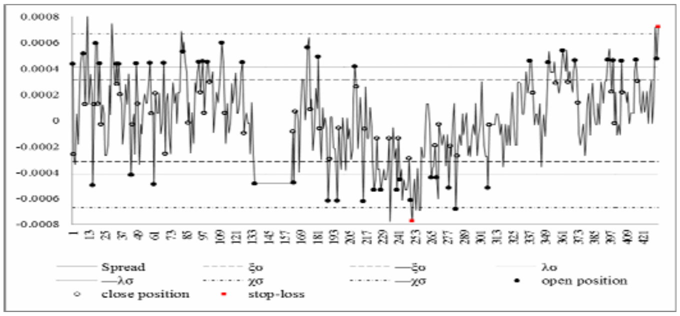

Figure 3. Transaction Timing

Table III reports the return performance under the given strategy, and Figure 4 presents the plot of period returns and accumulated returns curve in sample (transaction costs are not taken into account temporarily).

We can see that the statistical arbitrage based on co-integration theory gains a 0.8145% of success ratio, and only loses twice in 54 trading times. According to the Fig. 4, these two lost are attributed to instantaneously drastic fluctuation of the spread, which means spread volatility exceeds the stop-loss boundary results in loss. That proves the setup of stop-loss threshold value is very significant for the success rate of arbitrage. Though the arbitrage opportunities are limited, we can gain an abnormal return only seizing one tenth of opportunities.

TABLE III. PERFORMANCE OF THE ARBITRAGE STRATEGY

Times of arbitrage	45	Average return rate	0.0181%
Times of bull arbitrage	27	Accumulated return rate	0.8145%
Times of bear arbitrage	18	Return rate per day	0.1018%
Times of stop-loss trade	2	Return rate per year	About 30.54%

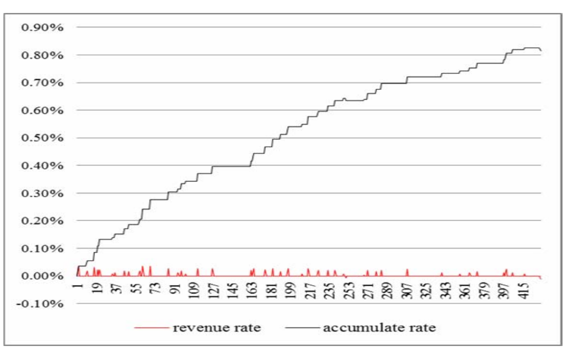

Figure 4. Returns Curve

V. Conclusion

In this paper, we proposed an algorithm of spread arbitrage process in order to improve the statistical arbitrage strategy of pairs trading based on co-integration. The result of empirical study suggested that the average profit from spread arbitrage was statistically significant and the rates of return of spread arbitrage were very attractive. Co-integration tests were applied to identify whether two or more index futures were good substitutes for spread arbitrage. The co-integration relation was found between two CSI-300 futures contracts, IF1207 and IF1208. This indicated that there was a long-term equilibrium level between correlated index futures, and the spread derive from co-integration relationship were stationary. Trading rules and filters were set to take advantage of the disequilibrium and to test the profitability of spread arbitrage of the index futures. The rates of return of spread arbitrage were found to be highly attractive under certain assumptions. However, this research only has performed in-sample simulation. The research on index futures spread arbitrage will be more complete and meaningful if in-sample coefficients could have been used to test the profitability of out-of sample data.

Acknowledgment

We acknowledge the support from the National Natural Science Foundation of China under Grant Nos.70971021 and 61075105, and the support from Innovation Program of Shanghai Municipal Education Commission under Grant No.12ZS055.

References

[1] N. Burgess, “Statistical Arbitrage Model of the FTSE 100. Computational Finance”, Computational Finance, 1999

[2] G. Vidyamurthy, Pairs Trading: Quantitative Methods and Analysis. New York: Wiley & Sons, 2004

[3] Z. Qiu and X. Cheng, “Stock Index Futures Arbitrage Model Based on Integration”, System Engineering, 2006, 26(12), pp 26-29

[4] R. Engle and C. Grange, “Cointegration and Error Correction: Representation Estimation and Testing”, Econometrical, 1987, 55(2). pp 251-276

[5] Honan S and Jarrow R, “Teo M and Warachka M, Testing Market Efficiency using Statistical Arbitrage with Application to Momentum and Value Strategies”, Journal of Financial Economics, 73(3), pp 525-565