The Statistical Arbitrage Study of CSI 500 Stock Index Futures Based on Intraday Effect

Jianwen Zhang, Guoqiang Tang*, Qiaofen Miao, Jingling Yang

College of Science, Guilin University of Technology, Guilin, China
Email: 1020180934@glut.edu.cn, *tanggq@glut.edu.cn

Abstract

Taking the CSI 500 stock index futures as the research object, the regression model of dummy variables of five indicators, including high-frequency return rate, volume change rate and near and far month contract price, was established. Then test whether these five indicators are affected by intraday effect and carry out statistical arbitrage based on intraday effect of spread. The empirical results show that the CSI 500 stock index futures have obvious intraday price fluctuations within 15 minutes of opening, and the intraday effect of the near-month contract is more significant than the far-month contract. The arbitrage strategy based on the intraday effect of spreading all the sample of both inside and outside can achieve higher success rate and yield, which is suitable for the short-term arbitrage. In actual trading, given the known probability of intraday profit, the intraday arbitrage method can provide reference for trading operation and risk aversion, so as to avoid losses caused by missed arbitrage opportunities.

Keywords

Intraday Effect, Virtual Variable Regression Model, Stock Index Futures, Statistical Arbitrage

1. Introduction

The intraday effect is a kind of calendar effect. The calendar effect refers to the market vision related to date and time. Its universal existence has been confirmed by numerous scholars, which has a certain impact on financial market investment and academic research. According to the research ideas and methods of experts and scholars, there are roughly three types of tests for the intraday effect. The first category, descriptive statistics, some scholars based on the time series of financial variables, as well as skewness, kurtosis, normality, fluctuation

interval and mean and other factors to analyze the existence of intraday effect or not. For example, Wu Gang-hui (2014) divided the trading time into 18 time periods and calculated the average value of six indicators such as volume, depth, absolute spread and relative spread in each time period to discuss the intraday effect characteristics [1]. For example, Wang Weiguo *et al.* (2015) believe that time series data with intraday effects must be adjusted to more accurately estimate and analyze the regression model [2]. The second category, Non-Parametric Statistical Methods, Wang Shun (2014) used the modified LM non-parametric test to identify the calendar effect of stock index futures and spot volatility, and found that their volatility showed double "W" and oblique "L" type [3]. Lin Xiang-you *et al.* (2015) tested the possible intra-week effect, maturity effect and double calendar effect of the securities market by Wilcoxon-Mann-Whitney nonparametric test method, AR model with dummy variables, and double difference mode [4]. Zhao Xiujuan *et al.* (2015) analyzed 1 minute data and found that the absolute yield of the Shanghai and Shenzhen 300 stock index futures was "LM" type, and the transaction volume was "WV" type [5]. The third category, the virtual variable regression method, is also the most used method adopted by most scholars. In the empirical process, the least squares regression model, ARMA model or GARCH model is selected for the distribution characteristics of the variable sequence, and time-dependent dummy variables are added to the model. According to the parameter significance of the dummy variable, the difference of the research target in each time period is determined. For example, Guo Yan-feng (2013) used both OLS and GARCH models to test the intraday, overnight and intraday effects of the Shanghai-Shenzhen 300 stock index futures yield sequence [6]. Yang Hua-qing (2015) used the AR-GARCH model with dummy variables to test the calendar effect in the closing price, trading volume and open interest of the Shanghai and Shenzhen 300 stock index futures, and found that the intraday effect mainly appeared in the market opening and closing stage [7]. For example, Zhou Shijun (2016) used high frequency to achieve volatility, considering the impact of overnight earnings, constructing Copula-Realized GARCH model to estimate the optimal hedging ratio of stock index futures [8]; Yang Yang (2016) proposed theoretical analysis of stock index futures affecting the spot Pathway, and establish the GARCH model empirical study on the impact of SSE 50 and CSI 500 stock index futures on the volatility of the spot market [9].

Statistical arbitrage refers to the use of quantitative analysis methods to build a portfolio based on statistical correlation theory, independent of economic theory, to achieve a certain degree of risk aversion, and obtain stable excess returns. The key to statistical arbitrage is the choice of arbitrage strategy. The strategy of statistical arbitrage is very rich, mainly including arbitrage strategy based on co-integration, strategy based on neural network model and strategy based on principal component analysis. The strategy based on co-integration is widely used and widely studied. For example, Hao Ning (2016) uses statistical

arbitrage to establish a long-term equilibrium co-integration equation to analyze the arbitrage space of CSI 500 stock index futures [10]. Yan Liang-wen *et al.* (2016) based on the co-integration and conditional heteroscedasticity model theory, searching for the best arbitrage and risk measure by standard residual sequence, thus establishing an optimal arbitrage scheme, and achieving better results than the threshold method based on confidence level determination [11]. Zhang Zhen (2017) proposed a dynamic prediction interval, and statistical arbitrage through a co-integration model of dynamic thresholds [12].

The CSI 500 stock index futures is a futures contract with the CSI 500 Index as the subject matter. Since its listing on April 16, 2015, it has been less than three years. At present, the domestic research on the CSI 500 stock index futures mainly focuses on price discovery and risk. Research was on transfer and other functions, impact on stock market stocks, spot market and statistical arbitrage. For example, Xu Jinjian (2016) introduced a virtual variable GARCH model to empirically analyze the impact of stock index futures on the index volatility [13]. Bai Hao (2017) studied the impact of the CSI 500 stock index futures trading on its underlying index price fluctuations [14]. There are also a few scholars who have studied the calendar effect of CSI 500 stock index futures, but relatively few. For example, Wu Xiaohua (2016) found that the Shanghai and Shenzhen 300 stock index futures have a weekly effect by establishing a conditional heteroscedasticity model, while the CSI 500 stock index futures and the SSE 50 stock index futures are not affected by the week effect [15].

According to the sample selection method of the CSI 500 Index, the CSI 500 stock index futures with the CSI 500 Index as the target can reflect the development status of the small and medium-sized companies. Secondly, since the listing time has only been three years, there is relatively little research on the CSI 500 stock index futures. Therefore, it is of great significance to conduct the intraday effect and statistical arbitrage research on the intraday trading data for the CSI 500 stock index futures. I hope that through the empirical research of this article, the financial investors will be referred to the investment information and expand the research scope of the CSI 500 stock index futures.

2. Theory and Modeling Steps

2.1. Principle of Calendar Effect

Calendar effects are a kind of market vision. In financial markets, the cyclical trend of financial indicators such as volatility, trading volume, and trading frequency is related to the date, including seasonal effects, monthly effects, and intra-week effects, due date effect, and so on [16]. The intraday effect is one of them, which refers to the fact that financial variables have significant fluctuations during certain periods of the trading day.

Specifically, the seasonal effect refers to a phenomenon in which abnormal changes in sequence objects are related to seasonal factors. The week effect, also known as the intra-week effect, means that the change in the economic indicator

on a certain day of the week is significantly different from the other days. Similar to the week effect, there is also a workday effect. Obviously, the workday effect, that is, the sequence object, differs between weekdays and non-workdays. The maturity effect, also known as the delivery date effect, refers to the abnormal fluctuation of the yield, volatility or volume of the underlying index on the settlement date of the financial product [17]. The intraday effect refers to the significant fluctuations in financial variables during certain periods of the trading day.

2.2. Test Model of Intraday Effect

This paper uses the virtual variable regression method to test the intraday effect of the CSI 500 stock index futures trading data indicators. To prevent multicollinearity, a session dummy variable is set every 5 minutes. Intraday trading 4 hours for a total of 48 5 minute periods, reducing the total number of dummy variables by one, excluding the last 5 minutes. That is to say, 47 dummy variables and constant terms are introduced in the regression model for research.

In order to prevent pseudo-regression and ensure the reliability of parameter test and estimation results, the stationary test, autocorrelation test and heteroscedasticity test should be performed on the test variables respectively. Then select the appropriate model based on the test results. When the sequence satisfies smoothness, no autocorrelation, and heteroscedasticity, select the OLS model with dummy variables, *i.e.*:

$$X_t = \alpha_0 + \alpha_1 h_1 + \alpha_2 h_2 + \cdots + \alpha_{46} h_{46} + \alpha_{47} h_{47} + \varepsilon_t \tag{1}$$

where, X_t represents the variable of intraday effect test, $\alpha_i, i = 0,1,2,\cdots 47$ represents constant terms and coefficients, h_i is a dummy variable, ε_t is a random error term. In order to prevent the occurrence of multicollinearity, the dummy variable for testing the intraday effect is one less than the actual one, that is, the first 5 minute period in each trading day is not included. When at 9:30-9:35, $h_1 = 1$ $h_2 = h_3 = \cdots = h_{47} = 0$. When at 9:35 - 9:40, $h_2 = 1$ $h_1 = h_3 = \cdots = h_{47} = 0$. Other definitions and so on. When $h_1 = h_2 = h_3 = \cdots = h_{47} = 0$, the significance of the constant term is the significance of the last 5 minutes period.

When the sequence satisfies smoothness, no autocorrelation, but heteroscedasticity, choose the GARCH model with dummy variables.

$$\begin{cases} X_t = \alpha_0 + \alpha_1 h_1 + \alpha_2 h_2 + \cdots + \alpha_{47} h_{47} + \varepsilon_t, \\ \sigma_t^2 = \beta_0 + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2, \end{cases} \tag{2}$$

where, the above equations are mean equation and conditional variance equation with dummy variables respectively. X_t is the variable of intraday effect test, α_i, β_i represents constant term and variable coefficient, ε_t is a random error term. $h_i, i = 1, 2, \cdots, 47$ is the dummy variable corresponding to the 5-minute time period of the day.

When the sequence is stable and there is autocorrelation, but there is no heteroscedasticity, choose the ARMA model with dummy variables, namely:

$$
\begin{cases}
f(t) = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\
X_t = f(t) + \alpha_0 + \alpha_1 h_1 + \alpha_2 h_2 + \cdots + \alpha_{47} h_{47} + e_t
\end{cases}
\tag{3}
$$

where, $f(t)$ represents the ARMA(p,q) model, p is the autoregressive order, q is the moving average order, $\phi_i (i = 1, 2, \cdots, p)$ and $\theta_i (i = 1, 2, \cdots, q)$ represent the coefficients of the autoregressive term and the moving average term, respectively. Other symbols have the same meaning as above.

When the sequence is stable and has autocorrelation and heteroscedasticity, ARMA-GARCH model with dummy variables is selected, that is:

$$
\begin{cases}
X_t = f(t) + \alpha_0 + \alpha_1 h_1 + \alpha_2 h_2 + \cdots + \alpha_{47} h_{47} + \varepsilon_t \\
\sigma_t^2 = \beta_0 + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2
\end{cases}
\tag{4}
$$

where, $f(t)$ represents the ARMA model, the above equations are the mean value equation and conditional variance equation with dummy variables and ARMA model added. Other symbols have the same meaning as above. When the sequence is not stable, the differential processing is performed first, and then the autocorrelation and heteroscedasticity are performed under the condition that the smoothness is satisfied, and then the appropriate model category is selected.

2.3. Arbitrage Mechanism Based on Intraday Effect of Spread

In the existing literature, statistical arbitrage mainly uses setting thresholds for closing positions and closing thresholds. Arbitrage based on the predicted standard residual sequence, the accuracy of the model prediction and the operating speed of the actual transaction affect the possibility and size of the arbitrage. According to the intraday effect test of the spread, there is a significant fluctuation in the spread during the intraday effect period. If the position can be closed by a known time difference, the arbitrage success rate and profitability can be increased. Combine this idea with the actual establishment of arbitrage mechanism; the main trading strategies are as follows.

Arbitrage is carried out according to the intraday period of the intra-sample spread and the coefficient of the corresponding dummy variable, when the significant dummy variable coefficient in the model is positive, carry out forward arbitrage for the intraday period corresponding to this variable. Start a position at the beginning, that is, buy a near-month contract, sell a distant month contract, and close the position at the end, that is, sell the near-month contract and buy the far-month contract. Assume that the spread between the near and far-month contracts IC1709 and IC1712 detects the intraday effect at 9:55 - 10:00 am, and the dummy variable coefficient is positive. Then, according to the intraday effect arbitrage strategy, the forward transaction is carried out, and the position is opened at 9:55, and the position is closed at 10:00. In actual opera-

tion, the arbitrage effect may be poor, and the arbitrage plan needs to be corrected. The specific measures are as follows:

1) Select the best intraday effect test model in the sample. According to the period and the variable coefficient of the significant dummy variables in the model, the arbitrage of the near and far-month contract closing price data in the sample is arbitrarily, and the success rate and total success rate and profitability of each arbitrage interval are counted;

2) When there are more significant variables, that is, there are many arbitrage opportunities in the day, and there are adjacent arbitrage intervals in the same arbitrage direction, try to merge the adjacent arbitrage periods, reduce the cost of handling fees, and observe whether the arbitrage effect is improved;

3) When the success rate of some arbitrage intervals is not high and the profit of the arbitrage interval is low, delete these arbitrage opportunities.

2.4. Modeling Step

This paper establishes a virtual variable model to explore whether the intraday trading of CSI 500 stock index futures has an intraday effect. If it exists, build a statistical arbitrage strategy on this basis. The modeling steps are as follows:

1) Data preprocessing. Using the 5-minute data of the day to generate logarithmic rate of return (high-frequency yield), volume change rate, volume change rate, price fluctuation range, and near-distance monthly contract price difference, and 47 time-varying dummy variables;

2) Stationarity, autocorrelation and heteroscedasticity test. Performing stationarity tests on five test variables of near and far-month contracts in different samples, and smoothing non-stationary sequences; Secondly, autocorrelation test and ARCH effect test are respectively carried out;

3) Selection test model. Select the appropriate test model based on the results of stationarity, autocorrelation and ARCH effect test;

4) Model fitting and test result analysis. Construct a model for the test variables separately, and judge the intraday effect form according to the parameter significance analysis of the dummy variables in the fitted optimal model;

5) Build statistical arbitrage strategies and implementation fixes. According to the intraday effect of the price difference test, the corresponding statistical arbitrage strategy is constructed. The data in the sample was used for verification, and the arbitrage interval with unsatisfactory results was modified.

3. Empirical Research

3.1. Data Sources

In order to understand the intraday effect of CSI 500 stock index futures more comprehensively and enhance the reliability of the results, from the perspective of sampling points and near and far-month contracts. Select IC1705 (near-month contract) and IC1706 (far-month contract) from April 7 to April 20, 2017 (10

trading days), and IC1709 (near-month contract) and IC1712 (far-month contract) in 2017 From July 7th to July 20th (10 trading days) as two comparison samples. A total of 3840 high-frequency 5-minute trading data were used for the intraday effect study. In the statistical arbitrage strategy study, the closing price of IC1709 (near-month contract) and IC1712 (far-month contract) from July 7 to July 20, 2017 (10 trading days) is used as sample data in 2017. The closing price of solstice on July 21, 2017 (5 trading days) is used as the out of sample forecast data. The data in this article comes from the access to the letter futures trading software. Example data is shown in Table 1.

3.2. Test of Intraday Effect

3.2.1. Data Preprocessing

This paper chooses the high-frequency yield, trading volume, position, spread and price fluctuation of stock index futures as the research object. The price volatility is defined by the highest price and the lowest price. Taking the logarithm can make the data more stable, does not change the correlation between the data, and the logarithmic rate of return satisfies the additivity, which has better statistical characteristics than the simple rate of return, so the rate of return is in logarithmic form [18]. Taking into account the small change in 1 minute during the day, the 5-minute trading data is used to create a series of profitability, volume change rate, position change rate, spread sequence and price fluctuation range, in which the rate of change is expanded by 100 times and converted into a percentage sequence. The specific formula is as follows.

$$
\begin{cases}
lr_t = \left(\ln p_t - \ln p_{t-1}\right) \times 100 \\
dvol_t = \dfrac{vol_t - vol_{t-1}}{vol_{t-1}} \times 100 \\
dope_t = \dfrac{ope_t - ope_{t-1}}{ope_{t-1}} \times 100 \\
jc_t = p_{1t} - p_{2t} \\
vp_t = \ln\left(hp_t\right) - \ln\left(lp_t\right)
\end{cases}
\tag{5}
$$

where, lr_t is logarithmic yield, $dvol_t$ is the volume change rate, $dope_t$ is the rate of change in position, jc_t represents the sequence of spreads, vp_t represents the price fluctuation range; p_t and p_{t-1} are the closing price of stock index futures at t, $t-1$, vol_t and vol_{t-1} are the volume at t, $t-1$, ope_t and ope_{t-1} are the amount of positions at t, $t-1$, p_{1t} represents the time t of the current month contract, p_{2t} is the closing price at time t of the far month contract, hp_t and lp_t represent the highest and lowest prices at time t.

3.2.2. Variable Verification and Model Construction

In this paper, we first test the stability, autocorrelation and heteroscedasticity of the five intraday effect test variables, and select the appropriate model based on the test results. The test results are shown in the table.

As shown in Table 2, ADF unit root is used to test the logarithmic return rate,

volume change rate, position change rate, price fluctuation range and the stability of the spread sequence of the near and far month contract of China securities 500 stock index futures. The results show that all variables reject the null hypothesis except that the spread sequence accepts the existence of the unit root null hypothesis, and the sequence is stable. Therefore, differential treatment of the spread is required before regression analysis. The log return autocorrelation test p value of the IC1705, IC1706 and IC1712 contracts in the sample is greater than 0.05, accept the null hypothesis that there is no autocorrelation, and the other sequences have obvious autocorrelation. As can be seen from the ARCH effect test results in the above table, only the p value of the change rate sequence of IC1705, IC1706 and IC1709 contracts is greater than 0.05, the null hypothesis is accepted and there is no heteroscedasticity, and the rest samples have ARCH effect to varying degrees.

Among them, there were 27 zero trading volumes in the far month contract IC1712 of July for CSI 500 stock index futures in 5 minutes, making it impossible for the sequence to test autocorrelation and ARCH effect. The zero trading volume distribution was analyzed independently. The specific distribution periods are shown in Table 3.

According to the statistical analysis of the 5 minutes of zero trading volume distribution of IC1712 stock index futures within a day, it can be seen that two of the 27 times without trading are 10:35 - 10:40, 11:20 - 11:25 and 11:25 - 11:30, as well as 13:00 - 13:05, 13:15 - 13:20 and 14:20 - 14:25 in the afternoon. The periods with zero trading volume once are mainly distributed in the afternoon trading time, especially in the half hour from 13:25 to 13:55. To sum up, the probability of zero trading volume in IC1712 stock index futures within 10 minutes before the closing bell in the morning and 20 minutes before the opening bell in the afternoon is relatively high.

According to the test results in Table 2, an appropriate model is selected. The preliminary model categories are shown in Table 4.

Table 1. Data example.

	date	time	closing price
1	2017/4/7	935	6430
2	2017/4/7	940	6429
3	2017/4/7	945	6439.2
4	2017/4/7	950	6435
...			...
477	2017/4/20	1445	6288.8
478	2017/4/20	1450	6280.8
479	2017/4/20	1455	6281.2
480	2017/4/20	1500	6278.8

Table 2. Stability, autocorrelation and p value of ARCH effect test results of variables.

index	contract	stationarity test	autocorrelation test	ARCH effect test	index	contract	stationarity test	autocorrelation test	ARCH effect test
logarithmic yield	IC1705	0.01	0.5847	3.55E−05	position change rate	IC1705	0.01	<2.2E−16	<0.2E−16
	IC1706	0.01	0.2688	0.002402		IC1706	0.01	<2.2E−16	1.05E−07
	IC1709	0.01	0.000883	7.06E−05		IC1709	0.01	<2.2E−16	0.01792
	IC1712	0.01	0.2436	3.11E−12		IC1712	0.01	4.78E−13	2.53E−11
volume change rate	IC1705	0.01	7.64E−06	0.5234	price fluctuation	IC1705	0.01	2.00E−13	1.57E−10
	IC1706	0.01	1.19E−09	0.3437		IC1706	0.01	1.00E−11	8.61E−12
	IC1709	0.01	6.23E−11	0.6447		IC1709	0.01	<2.2E−16	<0.2E−16
	IC1712	0.01	NA	NA		IC1712	0.01	<2.2E−16	<0.2E−16
spread	April	0.06839	<2.2E−16	<2.2E−16	spread	July	0.4484	<2.2E−16	<0.2E−16

Table 3. IC1712 stock index futures within 5 minutes of the volume distribution period.

times	periods of time					
	10:00 - 10:05	10:40 - 10:45	10:45 - 10:50	11:00 - 11:05	11:15 - 11:20	
1	13:10 - 13:15	13:25 - 13:30	13:35 - 13:40	13:40 - 13:45	13:45 - 13:50	
	13:05 - 13:10	14:00 - 14:05	13:50 - 13:55	14:10 - 14:15	14:50 - 14:55	
2	10:35 - 10:40	11:20 - 11:25	11:25 - 11:30	13:00 - 13:05	13:15 - 13:20	14:20 - 14:25

Table 4. Categories of the initial model for each indicator variable.

index	contract	model category	index	contract	model category
logarithmic yield	IC1705	GARCH	Position change rate	IC1705	ARMA-GARCH
	IC1706	GARCH		IC1706	ARMA-GARCH
	IC1709	ARMA-GARCH		IC1709	ARMA-GARCH
	IC1712	GARCH		IC1712	ARMA-GARCH
volume change rate	IC1705	ARMA	Price fluctuation	IC1705	ARMA-GARCH
	IC1706	ARMA		IC1706	ARMA-GARCH
	IC1709	ARMA		IC1709	ARMA-GARCH
	IC1712	——		IC1712	ARMA-GARCH
spread	IC1705-IC1706	ARIMA-GARCH	spread	IC1709-IC1712	ARIMA-GARCH

3.2.3. Analysis of the Intraday Effect Situation

The preliminary model category in Table 4 was compared, dummy variables were added to establish the model, and the model order identification, parameter estimation and model diagnosis of each indicator variable were conducted respectively. The comprehensive model is selected by comprehensive AIC criterion, BIC criterion, parameter saliency and model residual analysis. The time period in which the intraday effect exists and the corresponding dummy variable coefficient are arranged as shown in Tables 5-9.

Table 5. Logarithmic rate of return intraday effect test parameter estimation results.

contract	intraday effect period	coefficient	p	contract	intraday effect period	coefficient	p
IC1705	9:35 - 9:40	−0.09922	0.00672	IC1706	9:30 - 9:35	0.07836	0.046875
	10:05 - 10:10	−0.09322	0.008344		9:35 - 9:40	−0.11553	0.003324
	13:20 - 13:25	−0.09174	0.012883		13:20 - 13:25	−0.07697	0.038694
	13:55 - 14:00	0.072919	0.070689		10:05 - 10:10	−0.06285	0.077906
IC1709	9:30 - 9:35	−0.13929	0.018972	IC1712	9:55 - 10:00	−0.10723	0.044437
	10:20 - 10:25	−0.06442	0.000138		10:30 - 10:35	0.103596	0.034577
	10:40 - 10:45	−0.22498	0.004304		14:40 - 14:45	−0.09625	0.061231
	10:45 - 10:50	−0.12455	0.034558		14:45 - 14:50	0.115593	0.043447
	11:25 - 11:30	−0.12324	0.000179				
	13:30 - 13:35	−0.09796	0				

Table 6. Estimation results of intraday effect test parameters of volume change rate.

contract	intraday effect period	coefficient	p	contract	intraday effect period	coefficient	p
IC1705	10:15 - 10:20	−81.1126	0.0451	IC1706	10:15 - 10:20	−75.6245	0.0733
	13:00 - 13:05	−102.163	0.0116		11:15 - 11:20	−86.9353	0.0394
	13:05 - 13:10	−69.1094	0.0879		13:00 - 13:05	−115.709	0.0072
	14:25 - 14:30	−76.6795	0.0578		13:30 - 13:35	−86.2141	0.0412
	11:00 - 11:05	−109.441	0.0002		14:25 - 14:30	−61.313	0.0373
	11:15 - 11:20	−81.0439	0.0059		13:30 - 13:35	−60.6029	0.0396
IC1709	9:50 - 9:55	−73.7594	0.0122		14:05 - 14:10	−59.6402	0.0428
	13:00 - 13:05	−71.3882	0.0153		10:25 - 10:30	−55.3037	0.0603
	13:45 - 13:50	−66.546	0.0238		11:25 - 11:30	−53.1169	0.0712

Table 7. Change in the volume of positions.

object	period of significant intraday effect					
population	9:30 - 9:35	9:35 - 9:40	9:40 - 9:45	9:45 - 9:50	9:50 - 9:55	
recent months contract	9:30 - 9:35	9:35 - 9:40	9:40 - 9:45	9:45 - 9:50	9:50 - 9:55	9:55 - 10:00
	10:00 - 10:05	10:05 - 10:10	10:30 - 10:35	11:20 - 11:25	13:00 - 13:05	
far-month contract	9:30 - 9:35	9:35 - 9:40	9:40 - 9:45	9:45 - 9:50	9:50 - 9:55	14:20 - 14:25

Table 8. The intraday effect significant period of the same price difference sequence under two samples.

	virtual variable	intraday effect period	coefficient	p	July sample	intraday effect period	coefficient	p
April sample	h_{35}	13:50 - 13:55	−5.248	0.001		13:50 - 13:55	5.6262	0.0288
	h_{32}	13:35 - 13:40	−4.849	0.004		13:35 - 13:40	7.1880	0.0067
	h_{40}	14:15 - 14:20	−4.606	0.004		14:15 - 14:20	4.8469	0.0405
	h_{33}	13:40 - 13:45	−3.693	0.028		13:40 - 13:45	10.485	0.0000
	h_{29}	13:20 - 13:25	−3.595	0.034		13:20 - 13:25	6.5375	0.0158
	h_{41}	14:20 - 14:25	−3.204	0.041		14:20 - 14:25	5.3476	0.0207
	h_{37}	14:00 - 14:05	−3.173	0.055		14:00 - 14:05	7.229	0.0038

Table 9. Summary of periods with significant intraday effect of price fluctuation amplitude series.

object	period of significant intraday effect						
population	9:30 - 9:35	9:35 - 9:40	9:40 - 9:45				
recent months contract	9:30 - 9:35	9:35 - 9:40	9:40 - 9:45	9:45 - 9:50	9:50 - 9:55	10:10 - 10:15	13:55 - 14:00
far-month contract	9:30 - 9:35	9:35 - 9:40	9:40 - 9:45	9:45 - 9:50	10:05 - 10:10		
April sample	9:30 - 9:35	9:35 - 9:40	9:40 - 9:45	13:55 - 14:00			
July sample	9:30 - 9:35	9:35 - 9:40	9:40 - 9:45	9:45 - 9:50	9:50 - 9:55	9:55 - 10:00	10:00 - 10:05

The imaginary coefficient of the dummy variable and the corresponding parameter estimated in the above table indicate that the CSI 500 stock index futures yield does have an intraday effect. Different days, the degree of intraday effect is different. Among them, the near month IC1709 of the sample in July has a significant negative yield at 9:30 to 9:35, 10:40 to 10:45, 10:45 to 10:50 and 11:25 to 11:30, and the far month IC1712 has a significant positive yield at 10:30 to 10:35 and 14:45 to 14:50. The intraday effect of the front-month contract is significant during 13:30 - 13:35, and the yield is negative.

According to the test results, the volume change rate of the sample contract in the near month of July changed significantly in different periods of the day. In the April sample, when 10:15 - 10:20 and 13:00 - 13:05, there are intraday effect in the near and far month contracts; In general, the intraday effect of the front-month contract is significant between 13:00 - 13:05 and 14:25 - 14:30, and the coefficient of the corresponding dummy variable is negative, indicating a decrease in trading volume.

Table 7 shows the periods in which the change rate of position holding has obvious changes on the whole and in the near- and far-month contracts. As shown in the above table, the change rate of position holding is 9.30-9:55, that is, there is a significant fluctuation within 25 minutes of opening; The change rate of positions under the recent month contract is mainly at 9:30 - 10:10 and 13:00 - 13:05, that is, the intraday effect is 40 minutes after the opening of the morning and 5 minutes after the opening of the afternoon. In addition to the 25-minute opening of the far-month contract in the morning, the significant time period in the afternoon is 14:20 - 14:25.

According to the test results, the price difference sequence has more time periods during which the fluctuation range is larger during the day. From Table 7, the virtual variables that are significant under both samples are h_{35}, h_{32}, h_{40}, h_{33}, h_{29}, h_{41} and h_{37}. That is, the interval in which the intra-day effect exists is mainly distributed at 13:20 - 13:25, 13:35 - 13:45, 13:50 - 13:55, 14:00 - 14:05 and 14:15 - 14:25. Comparing the dummy variable coefficients, there is a negative spread in the April sample during the significant period of the intraday effect, while the July sample has a positive spread and the spread is more volatile.

As can be seen from Table 9, the price fluctuation range of the CSI 500 stock index futures is between 9:30 and 9:45. There are obvious intraday fluctuations in the opening 15 minutes, of which the near-month contract is more significant

than the far-month contract, and the dummy variable coefficient is significant every 5 minutes during the 25 minutes of opening. The time period in which the intraday effect exists in the July sample is mainly concentrated at 9:30 - 10:05, which is within 45 minutes of opening. The April sample was relatively less affected by the intraday effect.

3.3. Statistical Arbitrage Based on the Intraday Effect of Spreads

It is known from the above that the spread sequence has significant intraday fluctuations, so the direction and time of arbitrage are made according to the intraday effect test model of the coefficient of significant variables in the spread sequence (see Table 10 for details). There are 20 intraday effect periods, that is, there are 20 arbitrage opportunities every day. In the five time periods of 10:25 - 10:30, 10:45 - 10:50, 10:55 - 11:00, 14:00 - 14:05 and 14:55 - 15:00, the variable coefficient is negative, and reverse arbitrage can be adopted. Similarly, positive arbitrage is adopted for 15 time periods such as 9:30 - 9:35, 9:35 - 9:40 and 13:20 - 13:25.

According to this arbitrage design, carry out simulated arbitrage on the data in the sample. The success rate of arbitrage in each period is shown in Table 11.

As shown in Table 11, simulated arbitrage was carried out on the 10-day data in the sample, a total of 200 positions were built and 123 positions were closed for profit. In general, there is a 61.5% arbitrage success rate, in which 80% success rate can be obtained within the three ranges of 14:00 - 14:05, 14:30 - 14:35 and 14:55 - 15:00. The arbitrage effect is slightly worse in the 13:30 - 13:35 and 14:35 - 14:40 intervals, with a success rate of 40%. In the case of excluding fees, a total of 361.8 points of profit, a single maximum profit of 22 points, a single maximum loss −14.2 points. In order to reduce the expense of handling fee and increase the possibility of profit, the adjacent periods with the same arbitrage direction in the 20 periods with intraday effect were merged and corrected, and the in-sample and out-sample arbitrage was conducted again to observe whether the arbitrage effect was improved. The results are shown in Table 12.

After 20 significant periods of consolidation correction, reduced to 12 arbitrage interval. After the modified arbitrage design in the 10-day sample, the arbitrage effect will be improved, and the total success rate will increase to 69.2%. After deducting the handling fee, the profit will be 41951.54 yuan, of which the maximum single profit is 4284.81 yuan, the single maximum. The loss is −2877.28 yuan. If using the same arbitrage strategy to simulate the data outside the sample, the arbitrage total output power of the sample outside the sample can reach 83.3%, the profit is 7629.28 yuan, the maximum profit for a single arbitrage is 1521.907 yuan, and the maximum loss is 1799.34 yuan. At 9:30 - 9:40, 9:55 - 10:00, 10:25 - 10:30, 13:20 - 13:45, 13:50 - 13:55, 14:00 - 14:05 and 14:55-15:00 These 7 time slots are arbitrage, the success rate is 100%, but the arbitrage fails in the 14:30-14:45 period. After expanding the sample out-of-sample forecast, the success rate is reduced, and the profit increase is not large. The 5-day success rate outside the sample is 66.7%, the total profit is 15424.53 yuan,

and the single maximum profit is 3121.601 yuan. The maximum loss is the same as the 2-day result outside the sample. The arbitrage success rate is 100% profitable in the two intervals of 13:50 - 13:55 and 14:00 - 14:05, and the success rate of other arbitrage intervals is generally reduced.

Table 10. Arbitrage success rate in the sample.

intraday effect period	arbitrage direction	buying time	closing time	arbitrage success rate in the sample
9:30 - 9:35	positive	9:31	9:35	60%
9:35 - 9:40	positive	9:35	9:40	70%
9:55 - 10:00	positive	9:55	10:00	70%
10:25 - 10:30	negative	10:25	10:30	70%
10:45 - 10:50	negative	10:45	10:50	60%
10:55 - 11:00	negative	10:55	11:00	60%
13:20 - 13:25	positive	13:20	13:25	70%
13:25 - 13:30	positive	13:25	13:30	60%
13:30 - 13:35	positive	13:30	13:35	40%
13:35 - 13:40	positive	13:35	13:40	50%
13:40 - 13:45	positive	13:40	13:45	70%
13:50 - 13:55	positive	13:50	13:55	50%
14:00 - 14:05	negative	14:00	14:05	80%
14:15 - 14:20	positive	14:15	14:20	60%
14:20 - 14:25	positive	14:20	14:25	50%
14:30 - 14:35	positive	14:30	14:35	80%
14:35 - 14:40	positive	14:35	14:40	40%
14:40 - 14:45	positive	14:40	14:45	60%
14:50 - 14:55	positive	14:50	14:55	70%
14:55 - 15:00	negative	14:55	15:00	80%

Note: Due to the inability to obtain the opening price at 9:30 am, the arbitrage interval is 9:30 - 9:35, and the position is opened at 9:31.

Table 11. Day effect arbitrage design.

significant variables	corresponding period of time	variation coefficient	p	arbitrage direction	buying time	closing time
h_1	9:30 - 9:35	3.989318	0.035948	positive	9:31	9:35
h_2	9:35 - 9:40	5.284662	0.007923	positive	9:35	9:40
h_6	9:55 - 10:00	5.502699	0.015617	positive	9:55	10:00
h_{12}	10:25 - 10:30	−6.10354	0.016757	negative	10:25	10:30
h_{16}	10:45 - 10:50	−5.48434	0.039483	negative	10:45	10:50

Continued

h_{18}	10:55 - 11:00	−5.664	0.035961	negative	10:55	11:00
h_{29}	13:20 - 13:25	6.53757	0.015836	positive	13:20	13:25
h_{30}	13:25 - 13:30	9.536892	0.000401	positive	13:25	13:30
h_{31}	13:30 - 13:35	6.508251	0.015065	positive	13:30	13:35
h_{32}	13:35 - 13:40	7.188046	0.006782	positive	13:35	13:40
h_{33}	13:40 - 13:45	10.48504	0.000067	positive	13:40	13:45
h_{35}	13:50 - 13:55	5.626211	0.028821	positive	13:50	13:55
h_{37}	14:00 - 14:05	−7.229	0.003851	negative	14:00	14:05
h_{40}	14:15 - 14:20	4.846947	0.040511	positive	14:15	14:20
h_{41}	14:20 - 14:25	5.34761	0.020729	positive	14:20	14:25
h_{43}	14:30 - 14:35	5.084704	0.020153	positive	14:30	14:35
h_{44}	14:35 - 14:40	4.211954	0.0468	positive	14:35	14:40
h_{45}	14:40 - 14:45	4.974627	0.014848	positive	14:40	14:45
h_{47}	14:50 - 14:55	3.762071	0.043918	positive	14:50	14:55
h_{48}	14:55 - 15:00	−14.401	0	negative	14:55	15:00

Table 12. Adjusted in sample arbitrage results.

arbitrage direction	buying time	closing time	in-sample	2 days out of sample success rate	5 days out of sample success rate
positive	9:31	9:40	70.0%	100%	60.0%
positive	9:55	10:00	70.0%	100%	60.0%
negative	10:25	10:30	70.0%	100%	80.0%
negative	10:45	10:50	60.0%	50%	60.0%
negative	10:55	11:00	60.0%	50%	60.0%
positive	13:20	13:45	90.0%	100%	60.0%
positive	13:50	13:55	50.0%	100%	100.0%
negative	14:00	14:05	70.0%	100%	100.0%
positive	14:15	14:25	70.0%	50%	60.0%
positive	14:30	14:45	60%	0%	60.0%
positive	14:50	14:55	70%	50%	40%
negative	14:55	15:00	80%	100%	60%
the total success rat			69.2%	83.3%	66.7%
profit (yuan)			41951.54	7628.28	15424.53
the biggest profit (yuan)			4284.81	1521.907	3121.601
the biggest loss (yuan)			−2877.28	−1799.34	−1799.34

4. Conclusions

In this paper, we construct different virtual variable regression models for high-frequency yield, volume change rate, position change rate, price fluctuation

range and price difference series by means of stationarity test, autocorrelation test and heteroscedasticity test. Then, from the sampling point and the distance of the contract, the intraday effect form of CSI 500 stock index futures is analyzed and summarized. And the statistical arbitrage strategy is constructed based on the manifestation of the intraday effect of the spread. When the coefficient of the significant variable in the intraday effect test model is positive, the intraday effect period corresponding to the variable is positively arbitrarily; the starting point is opened, and the ending time is closed. Secondly, the arbitrage strategy is revised according to the success rate of each arbitrage interval in the sample, and the modified arbitrage strategy is used to simulate the arbitrage of the sample closing price. The main conclusions are as follows:

Intraday effect of high frequency yield: The CSI 500 stock index futures yield does have an intraday effect, and the intraday effect is different under different samples. Overall, the high-frequency yields have changed significantly in at least four of the 48 5-minute periods of the day (20 minutes), and the intraday effect in the near-month contract is more pronounced than the far-month contract.

Intraday effect of volume change rate: The trading rate change rate of the recent month contract is significant in the two periods of 13:00 - 13:05 and 14:25 - 14:30, and the corresponding dummy variable coefficient is negative, indicating that the trading volume has decreased. There are 27 zero-volume cases in the 5-minute data of IC1712 stock index futures within 10 trading days, mainly distributed in the morning at 10:35 - 10:40, 11:20 - 11:25, 11:25 - 11:30, 13:00 - 13:05, 13:15 - 13:20 and 14:20 - 14:25.

Intraday effect of the rate of change in positions: The rate of change in positions has a significant intraday effect in both overall and near-term contracts. During the period of 9:30 - 9:55, the volume of positions changed significantly within 25 minutes of opening. The change rate of position holding in the near month contract mainly has intraday effect within 40 minutes after the opening of the morning and 5 minutes after the opening of the afternoon, while the intraday effect in the morning of the trading day in the far month contract is within 25 minutes after the opening of the morning, and the significant period in the afternoon is concentrated within 14:20-14:25.

Intraday effect of price volatility: The CSI 500 stock index has obvious intraday price fluctuations within 15 minutes of the opening period. The near-month contract is more significant than the far-month contract. The dummy variable coefficient is significant every 5 minutes during the 25-minute opening period.

Intraday effect of spread between near and far month contracts: The spread is most affected by the intraday effect among the five test variables. In general, the time period during which the daytime changes are mainly distributed is 13:20 - 13:25, 13:35 - 13:45, 13:50 - 13:55, 14:00 - 14:05 and 14:15 - 14:25.

Based on the intraday effect of the spread within 10 trading days in the sample, after the strategy correction of merging adjacent arbitrage intervals, the total success rate of the new arbitrage strategy in the sample is equal to 69.2%, and the

profit after deducting the commission fee is 41951.54 yuan. The arbitrage total output power of the two trading days is 83.3%, the profit is 7629.28 yuan, and the maximum arbitrage gain is 1521.907 yuan. At 9:30 - 9:40, 9:55 - 10:00, 10:25 - 10:30, 13:20 - 13:45, 13:50 - 13:55, 14:00 - 14:05 and 14:55 - 15:00, this is 7 arbitrage period; the success rate is 100%. To sum up, statistical arbitrage strategy based on the intraday effect of spread has better arbitrage effect. This paper also has some shortcomings, which need to be further studied and improved. Therefore, in future studies, intraday effect and statistical arbitrage can be further discussed from the perspective of sample size.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Wu, G.-H. (2014) Empirical Study on the Liquidity of Domestic Stock Index Futures. Fudan University, Shanghai.

[2] Wang, W. and Zhai, H. (2015) Research on the Adjustment Method of Intra-Day Effect of UHF Data. *Chinese Management Science*, **23**, 49-56.

[3] Wang, S. (2014) Modeling of Jump Behavior and Volatility. Xiamen University, Xiamen.

[4] Lin, X.-Y., Dai, H.-X. and Gan, Y.-X. (2015) Study on the Calendar Effect of Securities Market Friday and Stock Index Futures Due Date. *Journal of Guizhou University of Finance and Economics*, No. 5, 48-57.

[5] Zhao, X.-J. (2015) Analysis of the Arbitrage of Shanghai and Shenzhen 300 Stock Index Futures Based on the Intraday Effect. *Journal of Management Sciences*, **18**, 73-86.

[6] Feng, G.-Y. (2013) An Empirical Study on the Function of Stock Index Futures in China. Southwest Jiaotong University, Chengdu.

[7] Yang, H.-Q. (2015) Research on Calendar Effect of Stock Index Futures in China. Hunan University, Changsha.

[8] Zhou, S.-J. (2016) Research on Dynamic Optimal Hedging Ratio of Stock Index Futures Based on Copula-Realized GARCH Model. Nanjing University, Nanjing.

[9] Yang, Y. (2016) An Empirical Study on the Impact of China's Stock Index Futures Trading on the Volatility of Related Spot Markets. Xinjiang University of Finance and Economics, Urumqi.

[10] Hao, N. (2016) Research on Arbitrage Strategy Based on CSI 500 Index. *Modern Economic Information*, No. 15, 291.

[11] Qin, L.-W., Tang, G.-Q. and Lin, J. (2016) Study on Optimal Threshold Statistical Arbitrage Based on Co-Integration-GARCH Model. *Journal of Guilin University of Technology*, **36**, 625-631.

[12] Zhang, Z. and Xu, W. (2017) Study on the Influence of Data Frequency of Stock Index Futures on Statistical Arbitrage Performance—Time-Based Trading Mechanism Based on Dynamic Prediction Interval. *Scientific Decision*, No. 2, 61-75.

[13] Xu, B.-J. (2016) Analysis of the Impact of Stock Index Futures on the Index Volatility—Based on the CSI 500 Stock Index Futures. *Financial Economy*, **10**, 87-89.

[14] Bai, W. (2017) Analysis of the Impact of CSI 500 Stock Index Futures on Stock Market Volatility. Inner Mongolia University, Hohhot.

[15] Wu, X.-H. (2016) Study on the Weekday Effect of China's Stock Index Futures Market. *Friends of Accounting*, No. 10, 86-89.

[16] Zhao, Y. (2012) Study on Risk Management of Soybean Futures Price Fluctuation. Beijing Institute of Technology Press, Beijing, 62-63.

[17] Liang, W. (2016) Empirical Analysis of the Calendar Effect of SSE 50 ETF. Hebei Finance Institute, Baoding.

[18] Lu, X.-G. (2011) An Empirical Study on the Weekly Calendar Effect of China's Shanghai and Shenzhen Index. Qingdao University, Qingdao.