The impact of high-frequency trading and its strategies on China’s market
高频交易及其策略对中国市场上的影响

a thesis

submitted to School of Mathematics & Physics

of Xi’an Jiaotong-Liverpool University

in partial fulfilment for the award of the degree of

BSc Financial Mathematics

By

Yulin Huang 2033597

Supervisor: Dr. Arodh Lal Karn

May 6, 2024

Abstract

High-frequency trading is a special category of algorithmic trading, which is based on a certain trading strategy, the use of high-speed computers at a very high frequency of the relevant information for centralised processing, and issue trading instructions to automatically complete the purchase and sale transactions. With the existence of etf (exchange-traded funds) and underlying stocks, investors can directly or disguisedly use credit trading and other means to realise T+0 trading, so as to obtain the lucrative profits from intra-day trading in the securities market. As a result, high-frequency trading has been used in the trading of commodity futures, ETFs and warrants in China. This report focuses on the impact of HFT in the Chinese securities market and the performance of HFT strategies in the Chinese market. During the experiment, Matlab was successfully used to perform time series analysis, frequency domain analysis on intraday trading data of CSI 300 stock index futures. In the return analysis, volatility optimisation empirical evidence has been successfully implemented using. At the end, an empirical simulation analysis of high-frequency trading strategies is conducted. In general, high-frequency trading activities have developed more rapidly in the Chinese securities market in recent years. In future research, it may focus more on the in-depth analysis of high-frequency trading strategies and their feasibility.

高频交易是一类特殊的算法交易，它是根据一定的交易策略，利用高速计算机以极高的频率对相关信息进行集中处理，并发出交易指令，自动完成买卖交易。随着 etf(交易所交易基金) 和标的股票的存在，投资者可以直接或变相利用信用交易等手段实现 T+0 交易，从而获得证券市场日内交易的丰厚利润。因此，高频交易已经在中国的商品期货、ETF 和权证交易中被使用。本报告主要研究 HFT 在中国证券市场的影响以及 HFT 策略在中国市场的表现。在实验过程中，成功用 Matlab 对 CSI 300 股指期货的日内交易数据进行时间序列分析，频域分析。在回报率分析中，成功利用已实现波动率优化实证。在最后，对高频交易策略进行了实证模拟分析。总的来说，高频交易活动近几年在中国证券市场发展较为迅猛。在以后的研究中，可能会更专注于高频交易策略的深度分析及其可行性。

KEY WORDS: High frequency trading(HFT), Fast Fourier Transform Algorithm(FFT), Stock index futures, HFT strategies.

Acknowledgements

I will take this opportunity to thank my supervisor Dr. Arodh Lal Karn and Dr. Hong Yi.

In my final year project, I am very grateful to met Dr. Karn, who gave me plenty of choice in selecting a topic that interested me and allowed me to carry out the research, and who provided me with a lot of useful technical support and encouragement during the research process, giving me the confidence to carry out my own research on a relatively complex topic. He gave me the confidence to work on a relatively complex topic on my own, which is a quality that a student graduating from an undergraduate programme needs to have. I would also like to thank Dr. Hong Yi for his generous assistance in obtaining high frequency trading data, which helped my research go smoothly.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, high-frequency trading has gradually expanded from the United States to major financial markets such as Europe and Asia. The proliferation of high-frequency trading and the continuous innovation of trading strategies have become a major focus in the global field (Hasbrouck and Saar, 2013). In Europe, the market share of high-frequency trading (HFT), after peaking at 40 percent in 2010, accounted for about 35 percent of total stock trading volume. In the United States, HFT's market share is even higher, having reached around 50 % of total equity trading after peaking at around 60 % in 2009 (Zaharudin et al., 2022).

First of all, High-frequency trading is a special class of algorithmic trading, which is based on a certain trading strategy, the use of high-speed computers to focus on the relevant information at a very high frequency, and issue trading orders to automatically complete the purchase and sale of transactions.However, there is no uniform authoritative definition in the international academic and industry circles with regard to high-frequency trading.Different authorities hold their own views on the definition of HFT, but a great deal of overlap and coincidence occurs based on comparisons and summaries(Zaharudin et al., 2022).For instance, The Arthur et al. (2010) defines high-frequency trading as a form of automated trading based on mathematical algorithms that utilize advanced technology to implement certain short-term trading strategies, as opposed to the individual trading strategies themselves.ASIC (2010) describes high-frequency trading as a special form of trading that employs a high-speed, low-latency technological infrastructure.The predominant feature of high-frequency trading involves employing high-speed, intricate computer algorithms or strategies for order generation and execution. Additionally, high-frequency trading utilizes sophisticated systems and efficient infrastructure to minimize network latency, including hosting services, individual or direct market data feeds, and proximity hosting.

At first, due to the restriction of T+1 trading system, most investors paid little attention to day trading opportunities. With the launch of innovative products such as commodity futures and stock index futures that allow T+0 trading and the possibility of redemption through the primary market, most investors paid little attention to day trading opportunities. With the existence of ETFs (exchange-traded funds) and underlying stocks that achieve T+0 trading directly or in disguise utilizing credit trading and other means, investors can capture the rich profits of intra-day trading in the securities market. Therefore, high-frequency trading has been used in commodity futures, ETF and warrant trading in China.

In fact, the number of domestic quantitative investment funds is increasing, such as Guangfa, Daimo Huaxin, Everbright Prudential and other fund companies launched a large number of quantitative investment strategy funds. In May 2012, like Nanhua Futures Company, CITIC Securities selected the Progress Apama algorithmic trading platform, and the algorithmic trading system has since been expanded into a multi-product program trading system, which is an important step in the development of high-frequency trading in China(Markets, 2018). With the deepening of the reform of the securities market, high-frequency trading has become one of the most important trading means in China's financial market. Since a large number of experiments are based on foreign data, it is difficult to find research on high-frequency trading in China, this paper decided to study the impact of HFT on China's securities market to have a deeper understanding of it.

Chapter 2

Literature Review

There has been considerable interest in recent years in the study of the impact of high-frequency trading and its strategies on markets. Usually, liquidity, volatility, transaction cost, and transparency are used as four indicators to reflect market quality (Menkveld, 2013).

According to Hasbrouck and Saar (2013), market quality has been improved by the introduction of algorithmic trading in the New York Stock Exchange Algorithmic trading results in a lower bid-ask spread. The lower bid-ask spread allows market buyers and sellers to trade at more advantageous prices, thereby reducing transaction costs and increasing market liquidity.Zhang (2010) utilized the company sample in CRSP and Thomson Reuters Institutional Holding Database from 1985 to 2009 in examining the impact of high-frequency trading on stock price volatility and price discovery. He found that after controlling for the company's fundamental volatility and other exogenous volatility factors, high-frequency trading is positively correlated with stock price volatility. This positive relationship is more pronounced in stocks with larger market values. HFT reduces the impact cost of the market. Impact cost refers to the impact on the price when a certain number of commissions (orders) are sold quickly. Brogaard (2010) Research shows that before the emergence of HFT traders, the market caused a price move of $0.013 per 100 shares traded while trading 1,000 shares caused a price move of $0.056 . Cumming, Hendershott and Riordan (2012) examined the frequency and severity of after-market price manipulation at 22 stock exchanges around the world from February 2003 to June 2011 and found that the emergence of high-frequency trading in some markets significantly alleviated the frequency and severity of after-market manipulation.

Chapter 3

Methodology

3.1 Research question

- Does High frequency trading have an impact on China's securities market?
- How are HFT strategies performing in the Chinese market?

3.2 HFT Strategies

Strategy model based on market microstructure

Market microstructure is a discipline that studies the process of price formation, and trading based on market microstructure is the core of high-frequency trading. Trading based on the microstructure of the market is to obtain information through the flow of quotations and gain information advantages through trading.

Li (2014) pointed out that there are two basic models in market microstructure: Inventory Models and Information Models. Both explain the process of market price formation at the microscopic level. The difference is that the information model explains the process of information being reflected in prices after market news is announced, and its basic theory is that the flow of orders containing information leads to price changes; The inventory model explains the mechanism of short-term price formation in the absence of information disclosure. The inventory model no longer focuses on the information contained in the order flow, but on the changes in the order flow.

- Basic inventory model:

Garman (1976) was the first to propose using the uncertainty of inventory and cash positions of market makers to study the optimal market making environment. Garman's research process is no longer limited to merely the order flow of individual investors but focuses on the total order flow of the whole market, which makes the total order flow of this market follow the Poisson distribution in a statistical sense. Therefore, Garman first proposed the "microstructure of the securities market" model (Shukla, 2015). Garman assumes that there is only one monopolistic market maker, whose job is to clear all orders in the market by setting buy and sell prices (the buy and sell price order flow follows an independent Poisson stochastic process). At the same time, market makers can profit from market making, and their goal is to maximize their interests without going bankrupt. Based on Gambler' s Ruin Problem,

$$P(InitialWealth) = P(Gain)P(InitialWealth + Gain) + P(Loss)P(InitialWealth - Loss)$$

By the skills of series,

$$P_{failure} = \frac{P(Loss) \times Loss}{P(Gain) \times Gain}^{InitialWealth}$$

Garma proposed two conditions based on the Gambler model and applied them to the study of work environments,

- If the cash in the hands of the market maker runs out, the market making fails;
- If the inventory in the hands of the market maker is used up, the market making fails.

From a financial perspective:

$$\lim_{t\to\infty} P_{failure}(t) \approx (\frac{\lambda_a}{\lambda_b})^{InitialWealth}/(E_0(p_a, p_b)), if \lambda_b > \lambda_a = 1;$$

and from a capital perspective:

$$\lim_{t\to\infty} P_{failure}(t) \approx (\frac{\lambda_b p_b}{\lambda_a p_a})^{InitialWealth}, if \lambda_a p_a > \lambda_b p_b = 1;$$

then it is introduced that $p_a > p_b$ is satisfied at any time, which is the source of bid ask spread. The bid ask spread allows market makers to make profits on the basis of having enough inventory.

- Basic information model:

Copeland (1983) proved Bagehot (1971)'s assertion that "price difference is explained by information cost" from a mathematical point of view. The author introduced the concept of information cost into the model and established a single trading cycle market maker pricing model. The information model explains how information is reflected in prices after the market publishes information. The information model uses Game theory model to reverse engineer and analyze the quotation and quotation flow, so as to find the information owned by market makers. The information model is based on the phenomenon of information asymmetry in asset trading. Different markets have different transparency, and the flow rate of information is also different. The transparency of OTC markets such as foreign exchange market, bond and derivatives market is low, and the flow rate of information is relatively slow; Most stock markets and markets that complete transactions through electronic trading systems are more transparent, and their information flows faster. An important conclusion drawn from the information model is that even if market makers have unlimited inventory to maintain liquidity, the bid ask spread still exists, which is brought about by information.

Strategy model based on market making (liquidity provision)

In general terms, market making means placing limit orders on either side of the market, i.e. placing limit orders for buying (selling) slightly under (over) the market. By continuously supplying the market with standing limit orders, market making essentially provides liquidity to the market. The spreads between the buy and sell limit orders represent the market maker's profit.

Rather than seeking to make directed bets, high-frequency trading market makers seek to maximize their inventory turns by taking positions on both sides of the order book. This differs from high-frequency trading, which employs a directed trading strategy. High-frequency trading market makers typically turn over their inventory more than five separate times per day, which explains their high share of the market's trading volume(Shukla, 2015). In order to safeguard their investment, they either maintain a minimum inventory position at the end of the trading day or even a zero inventory position.

Benos and Sagade (2016) identify that high-frequency traders engage in liquidity-taking and liquidity-making behaviours, which are consistent with market-making activity. This involves trading in the opposite direction of recent price changes (i.e., reverse trades), issuing limit orders, and using aggressive trades to quickly adjust inventories. In any case, the authors also find that passive high-frequency traders have high informational volume ratios, which suggests that high-frequency traders may employ a variety of market-making strategies, in addition to aggressive orders, to facilitate market making.

Based on the Markovian queuing model, where Xn is the state of the order book after n transitions, and X_0 is the initial state. An agent wishes to buy a unit at price $j_0 < j^A(X_0)$. If the price is below the pre-determined stop-loss price $J > j^A(X_0)$, the agent has the option of keeping the limit order or canceling it and placing a market buy order at the best ask level $j^A(Xn)$. In the event that j^a reaches J before the order is executed, the bid order is cancelled and a market order is placed at J to fulfil the trade. It is assumed that there is always a sufficiently high number of limit orders at price level J. The position of the limit order of the agent after n transitions of the order book is represented by Y_n. The Markov chain S_n=(X, Y) remains valid within $\mathbb{S} \subset \mathbb{N}^{2K} \times \{0, 1, 2, ...\}$ where $Y_n \leq X_n^{j_0}$.

The extended Markov process is defined by$(X_n, Y_n^0, Y_n^1, j_n^0, j_n^1)$, where $Y_n^0(Y_n^1)$ represents the position of the market maker's bid (ask) order in the corresponding bid (ask) price level, denoted by $j_n^0(j_n^1)$. It can be demonstrated that both Y_n^0 and Y_n^1 are non-increasing. Furthermore, $X_n^{j_n^0} \geq Y_n^0 \geq 0$ and $X_n^{jn^1} \geq Y_n^1 \geq 0$.

The market maker has predetermined a best buy level, $J^{B^0} < j^A(X_0)$, a worst level, $J^{B^1} > j^A(X_0)$, a best sell level, $J^{A^1} > j^B(X_0)$, and a worst sell level, $J^{A^0} < j^B(X_0)$. Analogously to the 'buy one' strategy, an order is cancelled and executed at the stop-loss price if the correlating best limit price equals or exceeds the worst price level. It is postulated that the execution at the stop-loss price is always available. The state space is defined as $\mathbb{S} \subset \mathbb{Z}^d \times \{0,1,2,...\} \times \{0,1,2,...\} \times \{J^{B^0},...,J^{B^1}-1\} \times \{J^{A^0}+1,...,J^{A^1}\}$.

Potential courses of action for this strategy are as follows:

- The market maker may choose to wait for the next market switch and cancel both orders before any order is executed.
- Alternatively, when one order is executed, the market maker has an order waiting to be executed on the opposite side. The market maker adheres to the strategy of purchasing (selling) one unit(Abergel et al., 2020).

The optimal (minimal) expected buy price (cost rather than value) in the state (x,y,j) for buying one unit is denoted by $V_\infty^B(x,y,j)$. This is the price at which the best buy level, J_B^0, and the worst level, J_B^1, are achieved. In a similar manner, $V_\infty^A(x,y,j)$ denotes the optimal (maximum) expected sell price in a given state (x,y,j) for the sale of a unit. It should be noted that the optimal expected value can be determined by

$$V_\infty(s) = \begin{cases} max(\Sigma_{s'} \in \mathbb{S} P_{ss'} V_\infty(s'),0), & y^0 > 0, y^1 > 0 \\ \pi^{j^1} - V_\infty^B(x,y^0,j^0), & y^0 > 0, y^1 = 0 \\ V_\infty^A(x,y^1,j^1) - \pi^{j^0}, & y^0 = 0, y^1 = 0 \end{cases} \tag{3.1}$$

- Basic return measure:

There are various trading strategies available in high-frequency finance to generate profits. Returns can be measured in different ways to measure time frequency, such as seconds, minutes, hours, days, years, etc. There are various performance measurement techniques to evaluate the basic return characteristics of these strategies. This article will analyze trading strategies using basic regression characteristics such as annual average return rate, standard deviation, volatility, maximum decline, skewness and kurtosis.

- Comparative Ratios:

The foundation of high-frequency trading is its strategy model, and an important issue is how to evaluate the performance of different high-frequency trading strategies. Traditional methods such as average returns can describe the performance of a single trading strategy, but cannot compare two or more strategies. Sharp (1966) proposed the Sharpe ratio, which later became one of the most commonly used comparison methods. The comparison ratios used in this article include Sharp, Omega, Sottino, Kappa, VAR, and CVAR which are listed below(Aldridge,2009).

1. Sharpe Ratio:

$$SR = \frac{E[r] - r_f}{\sigma[r]}$$

A comparative method is provided, which is sufficient when the returns follow a normal distribution. For high-frequency trading strategies, there is no risk-free term rf in the Sharpe ratio definition. The Sharpe ratio integrates mean, variance, and other factors to measure the performance of different trading strategies, covering three standard mean returns, standard deviations, and cost of funds. As there is no return on investment in risk-free assets in high-frequency trading, there is no cost of funds.

1. Omega Ratio:

Shadwick and Keating (2002) propose the use of a ratio called Omega, which provides insights into the higher moments of a return distribution. Omega, also referred to as the gain-loss ratio, takes into account both skewness and kurtosis, effectively considering the potential for both positive and negative returns. The calculation formula for Omega Ratio is as follows:

$$\Omega_i = \frac{E[r_i] - \tau}{LPM_{1i}(\tau)} + 1$$

1. Sortino Ratio:

The Sortino ratio is a risk adjusted performance indicator used to evaluate the relationship between investment returns and downside risk. It is a well-known extension of the Sharpe ratio, but unlike the Sharpe ratio that considers total volatility (up and down), the Sortino ratio only focuses on downward volatility, which is usually associated with negative returns.The principle behind the Sortino ratio is that investors are usually more concerned about the risk of losses rather than the potential for higher returns. This ratio is named after the renowned financial researcher Frank A. Sortino(1991), who used it as a substitute for the Sharpe ratio(Sapkota, 2014). The calculation formula for Sortino Ratio is as follows:

$$Sornito_i = \frac{E[r_i] - \tau}{LPM_{2i}(\tau)^{1/2}}$$

1. Kappa Ratio:

Kappa Ratio is an indicator used to evaluate portfolio performance, which considers the balance between excess returns and systemic risk. This indicator was initially proposed by David F. Babbel(1995) and has been widely used for measuring and comparing risk adjusted returns.The calculation of Kappa Ratio is based on the performance of the excess return of the investment portfolio relative to the overall risk of the investment portfolio. It can be calculated by the following formula:

$$Sornito_i = \frac{E[r_i] - \tau}{LPM_{3i}(\tau)^{1/3}}$$

1. VaR:

VaR is a measure of financial portfolio risk, which is widely used in financial risk management. VaR measures the maximum loss that a portfolio may suffer in the future at a given level of confidence. VaR is calculated based on the probability distribution function of portfolio value and gives the amount of loss at a specific confidence level. Generally, higher confidence levels indicate more conservative risk measurements. VaR can be calculated by many methods, including historical simulation method, Monte Carlo simulation method and parameter method. The core principle of VaR is based on probability statistics and risk distribution measurements (Jorion, 2000).

1. CVaR:

CVaR is an extension of VaR (value at risk), which not only measures the maximum possible loss of the portfolio at a given confidence level but also provides the average value of the loss when VaR is touched. CVaR is based on the integral of the loss distribution at VaR level, reflecting the additional loss when the loss exceeds var. CVaR is a relatively conservative risk indicator, which can measure portfolio losses more comprehensively. The calculation method of CVaR can be estimated by using the loss distribution after var. Typically, CVaR can be estimated by direct calculation or by optimization methods for a given loss distribution (Rockafellar and Uryasev, 1999).

Chapter 4

Empirical Analysis

The development of high-frequency trading in China's securities market is restricted by costs, systems and regulations. The objective of this study is to observe the trading frequency distribution curve of the CSI 300 stock index futures after the launch of stock index futures from the perspective of trading volume. This will be achieved by analysing the possible trading conditions of high-frequency traders in the market. Subsequently, an empirical analysis will be conducted from the perspective of trading price to ascertain the potential for high-frequency trading in CSI 300 stock index futures.

4.1 CSI 300 stock index futures trading volume time series

The research on the trading habits of traders, especially the trading frequency, can learn from signal processing In the field of time-frequency domain conversion method, a more accurate transaction frequency distribution curve can be obtained. FFT The algorithm (Fast Fourier Transform) is an efficient algorithm to realize the conversion from time domain to frequency domain.

To study the trading frequency of high-frequency traders in China's A-share market, we study the trading volume time domain of CSI 300 stock index futures distribution and frequency domain distribution, to examine the introduction of stock index futures from the perspective of trading volume. The market is affected by the frequency with which traders trade. The selected data were taken from Minute trading volume of CIFX CSI 300 Index futures from July 17, 2017 to July 17, 2020. During the research process, transactions are studied by year.

Based on MATLAB software, we draw the intraday trading volume distribution of CSI 300 index from 2017 to 2020. The horizontal coordinate in Figure 11 is the trading time from 9:30 to 11:30 and 13:00 to 15:00, which is the continuous bidding trading phase with a 1-minute interval; the vertical coordinate is the annual average of the minute trading volume, which is taken as the arithmetic average of the trading volume corresponding to the trading time of each trading day within the year.

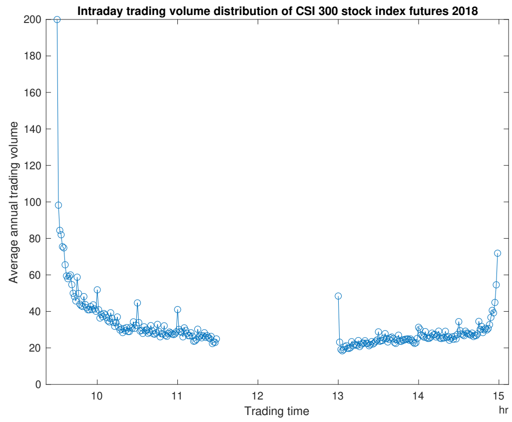


Figure 4.1: Intraday volume distribution(2018)

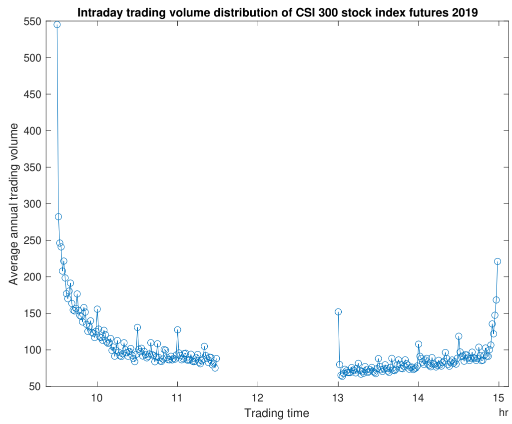


Figure 4.2: Intraday volume distribution(2019)

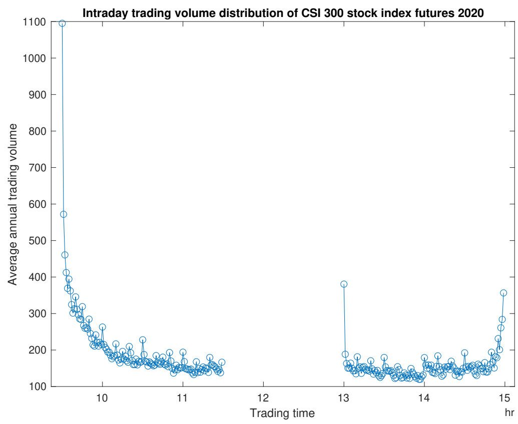


Figure 4.3: Intraday volume distribution(2020)

From Figures 4.1, 4.2, and 4.3, we can see that from 2018 to 2020, the trading volume shows jagged fluctuations with more regular fluctuation intervals. Each year compared to the previous year the minute trading volume has improved, and in 2020 the minute trading volume is above 100 lots per minute and the highest trading volume is over 1,000, reflecting the increase in market participation heat in stock index futures after 2020.

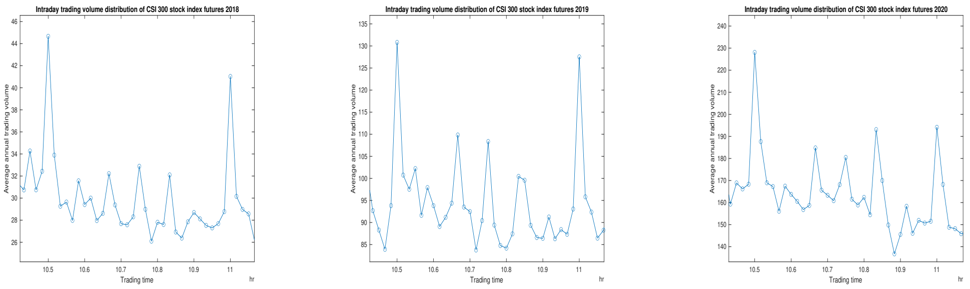
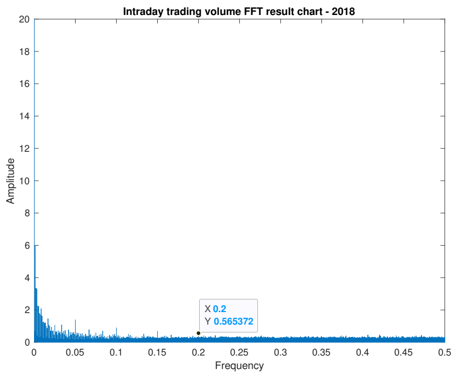


Figure 4.4: Intraday trading volume distribution of CSI 300 stock index futures

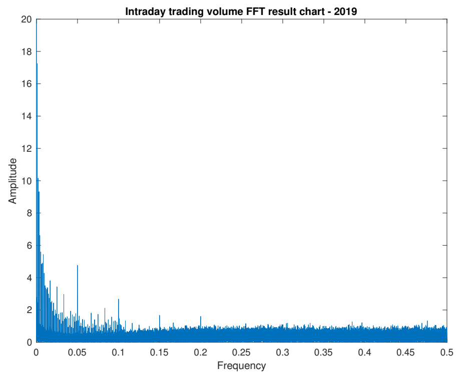
Intercepting the time window from 10:30 to 11:00, we can see the change of trading volume fluctuation pattern from 2018 to 2020: there are obvious peaks of trading volume in all three years, the maximum difference of trading volume in different moments is nearly double, and it shows cyclical movement, the fluctuation period is five minutes, and the peaks appear in 10:30, 10:35, 10:40, 10:45, 10:50, 10:55, and 11:00, it is hypothesized that market traders will make larger trading moves after the appearance of the five-minute K-line. Observation of the 2020 volume distribution chart also shows that there is a large increase in the peak trading volume compared to 2019, and the peak trading volume is higher at 190 lots or more at 10:30,10:45,11:00 minutes, which are the moments when the five-minute and fifteen-minute K-lines overlap and it is speculated that there are more traders trading on the basis of the K-line pattern.

4.2 FFT analysis of CSI 300 stock index futures trading volume

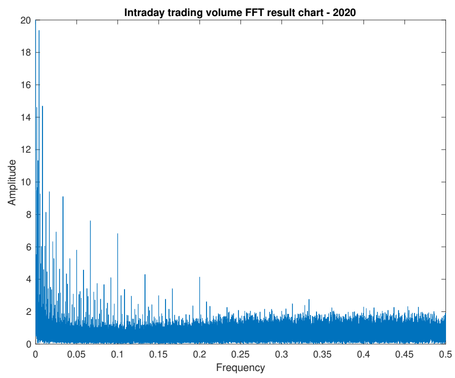
Based on the minute trading volume data, we use MATLAB software to perform FFT transformations in an attempt to observe the number of traders at different trading frequencies in the stock index futures market and analyze the trader activities.



(a) Guinea



(b) Liberia

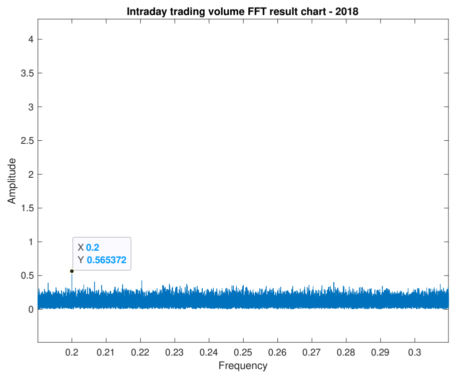


(c) Sierra Leone

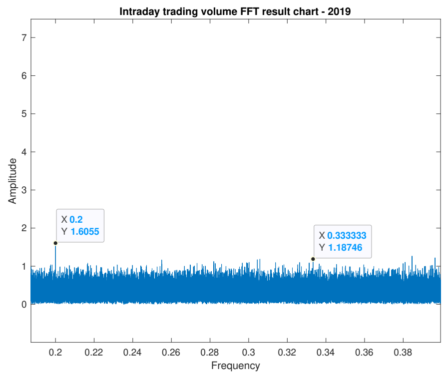
Figure 4.5: FFT analysis of trading volume of CSI 300 stock index futures from 2018 to 2020

Figure 4.5 shows the frequency domain analysis of stock index futures trading volume from 2018 to 2020. The horizontal coordinate represents the trading frequency as the number of minute trades, its reciprocal represents the trading period, and the vertical coordinate represents the volume intensity. Given that the sampling frequency of the raw data is 1 minute, the maximum trades with a frequency of 0.5 trades per minute can be observed here.

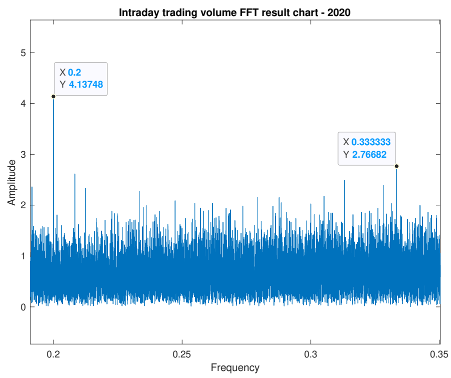
Zooming in on Figure 4.5 reveals clearer results, and from Figure 4.6 we can see that there is a clear peak at the three-year 5-minute trading cycle (at 0.2 on the horizontal coordinate), and at the 3-minute trading cycle (at 0.33 on the horizontal coordinate), with peaks in 2019 and 2020. It can be seen that more traders in the stock index futures market take 5-minute trading frequency to operate, which may be related to the fact that most of the market participants refer to the 5-minute K-chart to trade, and the results of the frequency domain analysis also confirm the inference of the previous time domain analysis.



(a) Guinea



(b) Liberia



(c) Sierra Leone

Figure 4.6: FFT analysis of trading volume of CSI 300 stock index futures in 3 years(zoom in

From the analysis results, we can qualitatively see that from the official trading of stock index futures from 2018 to the present, the trading volume of 5-minute cycle is significantly more than the trading volume of other frequencies, and in 2019 and 2020 compared with 2018 is more significant, reflecting the trading pattern of more participants in the futures market to enter and exit quickly.The larger trading volume of 5-minute cycle may also be related to the futures market participants' trading strategies, where buy and sell signals are based on 5-minute K-line patterns, especially as programmatic trading mostly triggers orders based on K-line patterns.

4.3 The return rate analysis of CSI 300 stock index futures under different frequency

Next, we conduct an empirical simulation analysis of possible high-frequency trading in CSI 300 stock index futures from the perspective of trading prices. The selected data are taken from the minute trading prices of the main contract from July 17, 2017 to July 17, 2020 for CSI 300 index futures of CICC. The empirical analysis tool is MATLAB software.

High-frequency data behaves in certain statistical ways that are quite different from traditional low-frequency data, and it opens up a new direction for investors to explore and research. Generally speaking, price returns are the main object of research and analysis. Returns are usually obtained by dividing the change in two consecutive prices by the first price, but academic researchers believe that logarithmic returns better reflect the distributional characteristics of price returns, so we take the logarithmic return approach for the return calculation here.

Time intervals	1 minute	5 minutes	10 minutes
Minimum	-0.028063	-0.031952	-0.031952
Median	0	0	0
Mean	3.4525e-06	2.7637e-07	4.3258e-07
Maximum	0.028115	0.079145	0.079145
Standard Deviation	0.0065533	0.0065474	0.0065461
Skewness	-0.022548	-0.019251	-0.019182
Kurtosis	4.73	4.7803	4.7824

Table 4.1: The return statistics of CSI 300 stock index futures from 2018 to 2020

Table 4.1 shows the sub-pen data in the sample and the data in 1 minute, 5 minutes and 10 minutes Logarithmic return statistics for observation intervals. From Table 4.1, we can see several characteristics of price return:

① When the observation frequency is high, both the mean and the median of the return distribution tend to be Zero, that is, the mean of the short-term volatility distribution of the index price yield can be approximately zero.

② When the observation frequency is higher, the skewness of the return distribution is smaller, and the positive and negative extreme values are absolute. The closer the logarithms are to each other, the symmetry of the returns is satisfied in the higher frequency data.

③ When the observation frequency is higher, the kurtosis of the return distribution is larger, that is, the frequency is higher, the greater the chance of a thick tail.

④ As an estimate of price volatility, variance does not exist with observation time interval Linear growth. It is shown that when the observation interval increases, there is an autocorrelation of price changes in the observation period.

While realized volatility makes effective use of the informativeness of high-frequency data, the price we observe in the market can be decomposed into the true intrinsic price plus a noise term due to market microstructural noise such as discontinuities in price changes and delays in responding to information. When the amount of intraday HF data is very large and the interval between price observations is very small, market noise becomes the main factor affecting price changes, at which point realized volatility (RV) is no longer an estimate of true price volatility, but instead becomes an estimate of noise volatility. According to theoretical research, the realized volatility is approximately equal to the noise volatility of 2 times the sample size, i.e $RV \approx 2n\sigma$ Then there are $\log(RV) \approx \ln(n) + ln(2\sigma^2)$ where n is the sample size and 2 noise is the noise volatility.

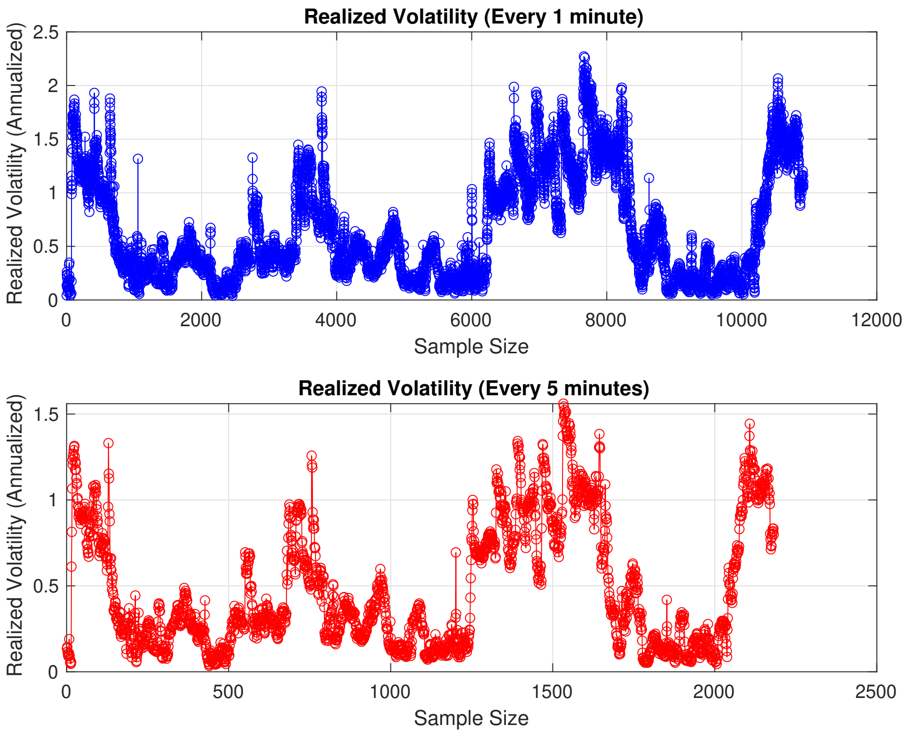


Figure 4.7: Realized volatility (annualized) in relation to sample size

Figures 4.7 show the realized volatility (annualized) versus sample size at one-minute intervals and the realized volatility (annualized) versus sample size at five-minute intervals. It can be seen that when the observation interval is very small, the realized volatility (RV) is large and becomes an estimate of the noise volatility; when the observation interval grows, the realized volatility (RV) slowly decreases.

4.4 Analysis of HFT opportunities

Firstly, we determine the triggering direction of each transaction based on the relationship between the price and the pre transaction quotation, that is, whether it is triggered by the buyer or the seller. If the price is close to buy one, then this order is designated as a sell order trigger; If the price is close to the selling price, it is set as a buy order trigger. If the price is equal to the distance between buying and selling, the triggering direction of this order is not confirmed.

Secondly, we compare the relationship between transaction volume and quotation volume. When the volume of transaction orders is greater than the volume of both buy one and sell one quotation, we believe that the preliminary signal is established, and the futures market may follow the triggering direction of this transaction to change.

Finally, we compare the current one-minute transaction price with the arithmetic mean of the past 10 minute transaction price. If the transaction price exceeds the average by 0.2 points and is triggered by the buyer and meets the preliminary signal conditions, the buy signal is confirmed; If the transaction price is 0.2 points lower than the average value and triggered by the seller and meets the preliminary signal conditions, the sell signal is confirmed.

According to the above strategy, the 1-minute trading prices of the main contract of CSI 300 stock index futures for 749 trading days from July 17, 2017 to July 17, 2020 are used for simulation, and the results of trading signal profitability statistics are shown in Table 4.2.

	Number of trades	**Number of earnings**	**PW**
Long position	21874	13007	0.596
Short position	21809	15914	0.727
Total	43683	28920	0.662
Total return	10016700		

Table 4.2: CSI 300 stock index futures high-frequency trading signal profit

Table 4.2 shows that based on the aforementioned strategy, the probability of profit for long position is 59.6% and the probability of profit for short position is 72.7% in the above trading range, with an average of about 58 trading signals occurring on a daily basis, and the total return reaches 100,167,000 points

4.5 Restrict factor analysis

- T+1 transaction mechanism

China's stock market adopts the T+1 trading system, and the operation of high-frequency trading requires a large number of intra-day transactions. T+1 trading mechanism will undoubtedly bring great obstacles to high-frequency traders, and the purpose of high-frequency trading is to capture the instantaneous price differences of many stocks or options and arbitrage from them. If the T+1 system can only be complied with, high-frequency traders can only judge the price movements of individual stocks or options and choose to trade on the front end of the market.

- Lack of market maker system

A market maker system does not exist in China, and it is unlikely that high-frequency traders will be able to obtain the privilege of early access to information from market makers or sufficient liquidity from market makers. As a result, investors are currently unable to conduct high-frequency trading based on inventory models, information models and other market microstructure strategy models.

- Frequent pending orders are illegal

High-frequency trading is typically based on the rapid submission of a large number of orders with the objective of detecting price information in the market. Once this information has been obtained, the orders are either executed or cancelled in order to carry out operations that can benefit the trader. These operations are usually executed at a price that is only 24% of the quoted price. In the United States, such trading strategies are permitted as long as they are within the reasonable range of regulatory requirements. However, in China, the current regulatory framework identifies the act of frequently placing trading orders and then cancelling them as illegal, disrupting the securities market and false trading, which limits high-frequency trading.

- Trading in stock index futures is restricted

As the most important derivative of stock market, stock index futures has attracted much attention since its launch on April 16, 2010. In order to limit excessive speculation in the market and limit investors' unilateral holdings and trading times, the realization of high-frequency trading will be subject to greater resistance before the reform of the trading system.

Chapter 5

Conclusion

This thesis focuses on the essence of high-frequency trading (HFT) and its methodology as well as its development in financial markets in recent years, and successfully tests the performance of the Chinese HFT market in an empirical analysis. In the introduction of HFT strategies, market microstructure strategies and market making strategies are highlighted, including the principles involved. Next, this paper examines several statistical models of data for assessing the performance of HFT strategies.

In addition, this study demonstrates the activity of the Chinese HFT market by analysing the CSI 300 stock index futures. the study shows that the trading volume on a 5-minute cycle is larger compared to other frequencies, with significant peaks. In the return analysis, the calculation of volatility is optimised according to the characteristics of high-frequency trading. Afterwards, the report analyses high-frequency trading opportunities in CSI 300 stock index futures, which is proved successfully with signal profitability statistics. Finally, This paper lists several factors that limit HFT activity in China.

The final results show that China's high-frequency trading market is more active in 2018-2010, especially with a five-minute cycle, which proved to have an impact on China's securities market.Furthermore, with one-minute, five-minute, and ten-minute nodes, the high-frequency trading returns are valid. In the simulation proof, the strategy applied in this study is also fully proved to be profitable.

The main limitation of this thesis is that there is limitation of ways to obtain more accurate LOB activity data from NASDAQ for the Chinese market, so it is difficult to accurately validate the feasibility of all the strategies in the Chinese market, which would be relatively accessible if we choose to obtain data from other countries.

Bibliography

Abergel, Hur, Pham, and Huyn (2020). Algorithmic trading in a microstructural limit order book model. *Post-Print*.

Arthur, W. B., Iii, A. R. B., Borrill, P., Brown, M. W., and Weinstein, E. R. (2010). Comment letter on sec's "concept release on equity market structure" file no s7-02-10.

Benos, E. and Sagade, S. (2016). Price discovery and the cross-section of high-frequency trading. *Journal of Financial Markets*, 30(sep.):54–77.

Brogaard, J. (2010). High frequency trading and volatility. *SSRN Electronic Journal*.

Copeland, B. J. (1983). Pure semantics and applied semantics. *Topoi*.

Hasbrouck, J. and Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, 16(4):646–679.

Hendershott, T. and Riordan, R. (2012). High frequency trading and price discovery. *SSRN Electronic Journal*.

Jorion, P. (2000). Value at risk: The new benchmark for managing financial risk. *risk management*.

Li, W. (2014). Essays on market microstructure and high frequency trading. *Business*.

Markets, E. (2018). High frequency trading(hft) curbs in china.

Menkveld, A. J. (2013). High frequency trading and the new market makers. *Journal of Financial Markets*, 16.

Rockafellar, R. T. and Uryasev, S. (1999). Optimization of conditional valueat-risk. *Journal of Risk*, 2(3).

Sapkota, N. (2014). Evaluating performance capacity of high frequency trading strategies, based on comparative ratios and market inefficiency at helsinki stock exchange.

Shukla, P. (2015). High frequency market microstructure. *CFA Digest*, 45(11).

Zaharudin, K. Z., Young, M. R., and Hsu, W.-H. (2022). High-frequency trading: Definition, implications, and controversies. *Journal of Economic Surveys*, 36(1):75–107.

Zhang, F. (2010). High-frequency trading, stock volatility, and price discovery. *Social Science Electronic Publishing*, 54(2):91–101.