이학석사 학위논문

Frequency and Resonance in Financial Data

(금융 데이터에서의 진동수와 공명 현상에 대한 관찰)

2020년 8월

서울대학교 대학원
수리과학부
이 루 다

이학석사 학위논문

Frequency and Resonance in Financial Data

(금융 데이터에서의 진동수와 공명 현상에 대한 관찰)

2020년 8월

서울대학교 대학원
수 리 과 학 부
이 루 다

Frequency and Resonance in Financial Data

(금융 데이터에서의 진동수와 공명 현상에 대한 관찰)

지도교수 Otto van Koert

이 논문을 이학석사 학위논문으로 제출함

2020년 4월

서울대학교 대학원

수 리 과 학 부

이 루 다

이루다의 이학석사 학위 논문을 인준함

2020년 7월

위 원 장: ____________________________

부위원장: ____________________________

위 원: ____________________________

Abstract

This paper investigates the volatility of the stock market in two ways. We decompose the one minute scale trading data of the stock market into the frequency and spectrum part using Discrete Fourier Transform(DFT).

The first methodology is to observe the average value of the high-frequency part of the spectrum using DFT with sliding windows of trading data. We are able to find a characteristic pattern of trading activity during a normal trading day. In addition, we detected an unusually high value of this spectrum information before a flash crash happens, suggesting a possibility for predicting flash crashes. Moreover, we apply this methodology to other days when there was a steep rise or drop in price to show the method explains the market well.

The second methodology is to directly analyze the spectrum using DFT of sliding windows of trading data to find resonance in the stock market. This has an implication that we can infer the trading pattern of the automated trading algorithm without additional data.

keywords: Discrete Fourier Transform, Time-series, Spectrum, Resonance, Stock transaction data

student number: 2018-29001

v

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

More than 70% of the US stock equities are executed by High Frequency Trading (HFT) [1], which uses a powerful computer algorithm for stock trading. If the algorithms try to buy or sell large amounts of shares at the same time, the market is likely to fluctuate. As an example, about the flash crash that happened on May 6, 2010, experts estimate that HFT exacerbated the crash [2]. Thus, one might think that trading frequency and spectrum of HFT are related to the volatility of the stock market. There are numerous studies to quantify the volatility of the stock market suggesting that can be a key to predict flash crashes [3]. In this paper, we are able to find a preliminary signal of unusual market crashes by observing the frequency of trading systems in the stock market. Instead of looking at the frequency of HFT directly, which happens at timescales of milliseconds, we look at minute scale frequencies.

In Chapter 2, we apply the Discrete Fourier Transform (DFT) to decompose the time-series into frequency and spectrum components. Since it is nearly impossible or inconvenient to compute the entire real-time data at once, we select a suitable size of the length of M of the sliding window for DFT, which is used in [4]. The entire time-series is processed shifting the window by time-delay parameter τ.

In Chapter 3, we observe the average value of the high frequency part of the spectrum using DFT with sliding windows of trading data, which is detailed in Definition

3.2.2. Unless otherwise stated, we call the value briefly as *the high-frequency part of the spectrum*. We use transaction data from the US futures in this chapter. First, we observe the distribution of *the high-frequency part of the spectrum* of each data segment along trading time and over a period $\mathcal{D}$. Second, we investigate how trading activity varies typically during the day by looking at this spectrum information that might reflect trading pattern of trader or HFT system. Then, we are able to find a characteristic pattern of trading activity during normal trading day. Third, we observe how *the high-frequency part of the spectrum* varies on flash crash days and suggest a possibility for predicting flash crashes. Finally, we repeat the same process to predict unusual crash in the market.

In Chapter 4, we give a definition of resonance, which is a phenomenon that the amplitude of a wave increases when several waves are synchronized. Then, we observe resonance in transaction data. By searching for a certain pattern in the frequency of the data, we might surmise trading patterns of trader and HFT algorithms.

Chapter 2

Preliminaries

2.1 Definitions and Motivation

We begin this section by introducing the classical Fourier Transform. The Fourier transform decomposes a signal into sum of sines, and corresponding amplitude.

Definition 2.1.1. *Let f be an L_1 integrable function $f : \mathbb{R} \to \mathbb{C}$. Then, the Fourier transform of f at ξ is*

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\xi}dx$$

for any $\xi \in \mathbb{R}$. At $\xi = \xi_0$, $\hat{f}(\xi_0)$ represents a spectrum component of the sinusoidal signal $e^{2\pi i\xi_0 x}$.

Definition 2.1.2. *Let x be an independent variable, and ξ be transform variable that represents frequency. f is determined by $\hat{f}$ by via the inverse transform*

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{2\pi i x\xi}d\xi$$

for any $x \in \mathbb{R}$

Definition 2.1.3. *Time-series f is a map $f : T \to \mathbb{R}$ where $T = \{t_i | i = 1, \ldots, N\}$ is an ordered set of time, which are successively equally spaced.*

The Discrete Fourier Transform (DFT) acts on a finite time-series to decompose it into the frequency and the spectrum part on the assumption that the sequence is generated by a continuous signal.

Definition 2.1.4. *Let N data points be denoted $f[0], f[1], \ldots, f[N-1]$. The discrete Fourier Transform of time-series $f[t]$ is defined by*

$$\hat{f}[n] = \sum_{k=0}^{N-1} f[k] e^{-\frac{2\pi i}{N} nk}$$

The $|\hat{f}[n]|$ is called the nth spectrum of f for $n = 0, 1, \ldots, N-1$.

Next, we introduce some mathematical tools compute DFT in computer as referred from [5].

Definition 2.1.5. *Let n be a positive integer. Then, complex number ω is called an nth root of unity if $\omega^n = 1$.*

Remark 2.1.1. *Let $\omega_0, \omega_1, \ldots, \omega_{n-1}$ be nth root of unity. Then, ω_k is explicitly represented by $exp\left(\frac{2k\pi i}{n}\right) = cos\frac{2k\pi}{n + isin\frac{2k\pi}{n,}}$ for $k = 0, 1, \ldots, n-1$. Thus, ω_k and ω_{n-k} are complex conjugate each other.*

Definition 2.1.6. *Vandermonde matrix for nth root of unity ω, which is denoted by*

$$V_\omega = VDM(1, \omega, \ldots, \omega^{n-1}) = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^{-1\cdot 1} & \omega^{-1\cdot 2} & \cdots & \omega^{-(n-1)} \\ 1 & \omega^{-1\cdot 2} & \omega^{-1\cdot 4} & \cdots & \omega^{-2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{-1\cdot(n-1)} & \omega^{-2\cdot(n-1)} & \cdots & \omega^{-(n-1)^2} \end{pmatrix}$$

$$= \left(\omega^{jk}\right)_{0 \le j,k < n} \in \mathbb{C}^{n \times n}$$

We may write Discrete Fourier Transform of signal $f[t]$ using Vandermonde matrix as following

$$\begin{pmatrix} \hat{f}[0] \\ \hat{f}[1] \\ \hat{f}[2] \\ \vdots \\ \hat{f}[N-1] \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega^{-1\cdot 1} & \omega^{-1\cdot 2} & \dots & \omega^{-(N-1)} \\ 1 & \omega^{-1\cdot 2} & \omega^{-1\cdot 4} & \dots & \omega^{-2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{-1\cdot(N-1)} & \omega^{-2\cdot(N-1)} & \dots & \omega^{-(N-1)^2} \end{pmatrix} \cdot \begin{pmatrix} f[0] \\ f[1] \\ f[2] \\ \vdots \\ f[N-1] \end{pmatrix}$$

Remark 2.1.2. *Consider*

$$\begin{aligned} \hat{f}[N-n] &= \sum_{k=0}^{N-1} f[k]e^{-\frac{2\pi i}{N}(N-n)k} = \sum_{k=0}^{N-1} f[k]e^{\frac{2\pi i}{N}nk} \cdot e^{2\pi ik} \\ &= \sum_{k=0}^{N-1} f[k]e^{\frac{2\pi i}{N}nk} = \hat{f}^*[n] \end{aligned} \tag{2.1}$$

where $\hat{f}^*$ represents complex conjugate of $\hat{f}$. Thus, $\hat{f}[n] = \hat{f}^*[N-n]$, and therefore $|\hat{f}[n]| = |\hat{f}^*[N-n]|$. Thus, for odd N, $\hat{f}(0)$ to $\hat{f}\left(\left\lfloor \frac{N}{2} \right\rfloor - 1\right)$ have the same spectrum from $\hat{f}(N-1)$ to $\hat{f}\left(\left\lfloor \frac{N}{2} \right\rfloor + 1\right)$. For even integer N, $\hat{f}(0)$ to $\hat{f}\left(\left\lfloor \frac{N}{2} \right\rfloor - 1\right)$ have the same spectrum from $\hat{f}(N-1)$ to $\hat{f}\left(\left\lfloor \frac{N}{2} \right\rfloor\right)$.

In this paper, we deal with the spectrum from $\hat{f}(0)$ to $\hat{f}\left(\left\lfloor \frac{N}{2} \right\rfloor - 1\right)$.

Theorem 2.1.1. *Let ω be a primitive nth root of unity. Then ω^{-1} is a primitive nth root of unity and $V_\omega \cdot V_{\omega^{-1}} = nI$ where I is the $n \times n$ identity matrix.*

Definition 2.1.7. *The inverse transform of $\hat{f}[n]$ is defined by*

$$f[k] = \frac{1}{N} \sum_{n=0}^{N-1} \hat{f}[n]e^{\frac{2\pi i}{N}nk}$$

The inverse Discrete Fourier Transform decomposes a signal $f[k]$ into sinuous functions of frequency n/N of amplitude $\hat{f}[n]$.

We now introduce the Fast Fourier Transform(FFT) algorithm and its computational complexity.

Definition 2.1.8. *Let f be a real or complex valued function and g be a real valued function. Let both functions be defined on some unbounded subset of real positive numbers, and $g(x)$ be strictly positive for all large enough values of x. Then, $f(x) = \mathcal{O}(g(x))$ implies that there exists a positive real number M and a real number x_0 such that $|f(x)| \leq Mg(x)$ for all $x \geq x_0$*

Then, for DFT, we need N^2 times of multiplication for N data points. With Fast Discrete Fourier Transform algorithm, we can reduce it to $\mathcal{O}(Nlog(N))$ [5]

Step 1. If $N = 1$, return $f[0]$.

Step 2. Let $f[0], f[1], \ldots, f[N-1]$ are real-valued data points.Separate this signal into two sequences of length $N/2$. One is with odd indices, say $f_{odd} = [f[1], f_[3], \ldots]$ and $f_{even} = [f[0], f[2], \ldots]$.

Step 3. Let $r_0(x) = \sum_{0 \leq j \leq n/2}(f[j] + f[j+n/2])x^j$ and let $r_1(x) = \sum_{0 \leq j \leq n/2}(f[j] - f[j+n/2])\omega^j x^j$.

Step 4. Call the algorithm recursively to evaluate r_0 and r_1^* at the powers of ω^2.

Step 5. Return $= \big(r_0(1), r_1^*(1), r_0(\omega^2), r_1^*(\omega^2), \ldots, r_0(\omega^{n-2}), r_1^*(\omega^{n-2})\big)$.

Theorem 2.1.2. *Let N be a power of 2 and $\omega \in \mathbb{C}$ be a primitive Nth root of unity. Then the algorithm correctly computes DFT_ω using $N \log N$ additions in $\mathbb{C}$ and $(N/2)\log N$ multiplications by powers of ω, in total $\frac{3}{2}N \log N$ additions in $\mathbb{C}$ and $(N/2)\log N$ multiplications by powers of ω, in total $\frac{3}{2}N \log N$ operations.*

Chapter 3

Frequency in Financial Data

3.1 Data

In this chapter, we study historical data of Dow Jones Industrial(DJIA) Futures, Nasdaq(mini) Futures, and S&P500 Futures from April 2, 2007 to April 2, 2019. The range of periods may vary for each section. These datasets are consisted of trading information of OHLC(opening, high, low, closing) price, and trading volume for every minute from opening to closing of the stock market. Unless there is no missing data or break of the market, there are 401 minutes of trading data from $8:31$ am to $3:11$pm. We refer to days having 401 minutes of trading history as *full-trading day* for the above three futures indices. We may say *normal trading day* when there are more than 390 minutes of trading history for a day. We re-sampled the data by taking the average price of the OHLC price by every minute. We represent this average price of data with time-series. We use the following terminologies.

- Let $\mathcal{D}$ be a set of trading days that satisfies certain condition for a given period. Then, we say ith trading day to indicate ith date in $\mathcal{D}$.

- Time sequence T is a sequence of time by minute scale from opening to closing of the market.

- jth trading minute indicates jth time in a time sequence T.
- To indicate the time-series of volume of ith trading day of a period $\mathcal{D}$, we use $f_i[T]$ where T is a minute-scale time-series while the market is open. Depending on the context, it may refer time-series of the average price of OHLC (open-high-low-close) prices.
- The set of f_i is written as $\mathcal{P}(\mathcal{D})$
- When we need to refer real date(time)of ith trading day(jth trading minute),we say corresponding date(minute) of ith(jth) trading day(minute).

3.2 The high-frequency part of the spectrum

We begin this section by introducing the concept of sliding windows for time-series. Sliding window is widely used in computer science to use data efficiently.

Definition 3.2.1. *Let $T = \{t_i\}_{i=0}^{N-1}$ be a sequence of time and M be a positive integer. Then, j-th sliding window of size M of T is a set of M consecutive elements up to t_j in T, which is denoted by $W_{M,j}$ for $j = M, \dots, N$. i.e.,*

$$W_{M,j} = \{t_{j-k}\}_{k=0}^{M-1} \tag{3.1}$$

for $k = 0, \dots, M-1$.

Similarly, j-th sliding window for time-series $f[T] = \{f[t_i]\}_{i=0}^{N-1}$ of size M is a set of M consecutive elements up to $f[t_j]$ in $f[T]$. i.e.,

$$f[W_{M,j}] = \{f[t_{j-k}]\}_{k=0}^{M-1} \tag{3.2}$$

for $j = M, \dots, N$ where $W_{M,j}$ is a jth sliding window of size M of T.

Remark 3.2.1. *To avoid using future data in predicting future, when we use jth sliding window we need data up to jth trading day(or minute), not over jth trading day(or minute).*

Remark 3.2.2. *Otherwise stated, we assume that sequence of time or dates are sorted in ascending order.*

Definition 3.2.2. *Let $f[T]$ be a times-series of length N and k be a quarter of N. Then, the high-frequency part of the spectrum of $f[T]$ is defined by the average value of spectrum of the k-highest frequency of discrete Fourier transform of $f[T]$, and is denoted by*

$$hS\left(f[T]\right) = \frac{1}{k}\sum_{i=0}^{k-1} \hat{f}\left[\left\lfloor \frac{N}{2} \right\rfloor - i\right].$$

In this chapter, we observe this value for every sliding window. When f is a time-series of trading price or volume, *the high-frequency part of the spectrum $hS\left(f[T]\right)$* may represent the trading activity or volatility of the market in time T.

Correlation between volume and the spectrum of high-frequency part

Now, we deal with transaction data from the US futures: Dow Jones Industrial Average(DJIA) futures, Nasdaq(mini) futures, S&P 500 futures. In this section, we define $\mathcal{D}$ and f_i as follows.

- Period $\mathcal{D}$ is $\{$normal trading days of trading history from Jan 01, 2010 to Jan 01, 2019$\}$.
- $f_i[T]$ is a time-series of price on ith trading day of a period $\mathcal{D}$ where T is a minute scale time sequence during the market opens.

Correlation shows relation between variables. In financial data, trading volume is one of important indicators to analyze the trend of the stock market. To observe the impact of frequency in stock market, we study correlation between trading volume and the high frequency part of the spectrum of discrete Fourier Transform.

Definition 3.2.3. *Let X, Y be random variables with expected values μ_X and μ_Y and standard deviation σ_X and σ_Y. Then, correlation between X and Y is defined as*

$$corr(X,Y) = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}$$

where E is the expected value operator.

In this chapter, we focus on *the high-frequency part of the spectrum* of time-series of price to observe high frequency trading or volatility of the stock market.

In Figure 3.1(left), each point represents $hS\left(f_i[T]\right)$ on ith trading day of $\mathcal{D}$ to detect unusual day. The average and standard deviation of $hS\left(f_i[T]\right)$ is 54.623 and 50.925, respectively. We investigated the days when $hS\left(f_i[T]\right) \geq 258.322$, which value is 4 standard deviation further from the average. There were five days: Jan 23,2008, Oct 9, 2008, August, 2015, Feb 6, 2018, and December 26, 2018.

Next, we calculated correlation between the trading volume and $hS\left(f_i[T]\right)$, and it comes out to be 0.642 which can be considered high in psychology. Figure 3.1(right) shows trading volume in horizontal axis and $hS\left(f_i[T]\right)$ in vertical axis. Figure 3.1 are plotted with giotto-tda([6])

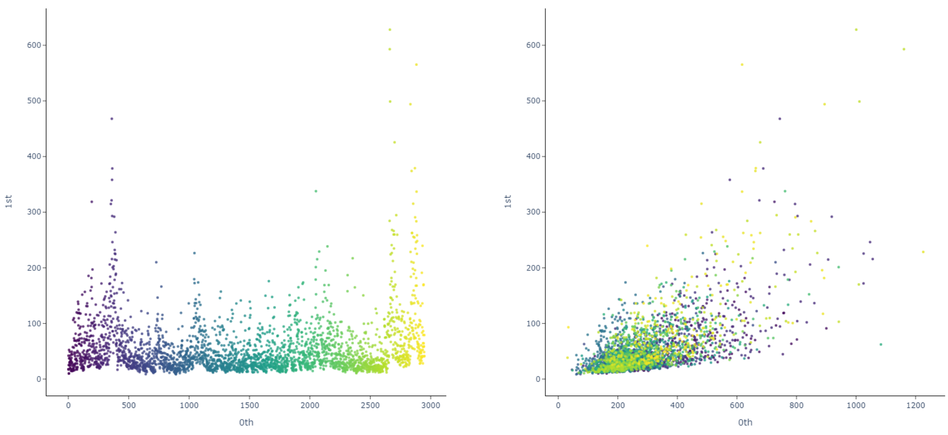


Figure 3.1: (a) *The high-frequency part of the spectrum* of $f_i[T]$ over a period $\mathcal{D}$. (b) Correlation between volume and *the high-frequency part of the spectrum* for period $\mathcal{D}$.

In Figure 3.1, we would say that the average value of *the high-frequency part of the spectrum* is a measure for the activity level of the market. Thus, a high average value of *the high-frequency part of the spectrum* may suggest that the market is turbulent, often occurring high volume of trade.

3.3 Distribution of daily trading activity via spectral analysis

In this section, we find a characteristic pattern of trading activity during a normal trading day. We investigate how trading activity varies typically during the day by looking at *the high-frequency part of the spectrum* of sliding windows $f[W_{M,j}]$. In this section, we define f_i as follows.

- $f_i[T]$ is a time-series of price on ith trading day of a period $\mathcal{D}$ where T is a minute scale time sequence during the market opens.

We observe *the high-frequency part of the spectrum* using DFT with sliding windows. To check whether *the high-frequency part of the spectrum* has a characteristic pattern according to trading time, we observe mean and standard deviation for each sliding window $W_{M,M+15j}$ of size $M = 32$ over a period $\mathcal{D}$.

Remark 3.3.1. *We regard the set of $hS(f_i[W_{M,j}])$ for $f_i \in \mathcal{P}(\mathcal{D})$ as random variables that are normally distributed.*

Definition 3.3.1. *Let $\mathcal{D}$ be a period $f_i[T] \in \mathcal{P}(\mathcal{D})$ where T is a minute scale time sequence while the market opens. We denote the average of the high-frequency part of the spectrum of sliding window $f[W_{M,j}]$ over a period $\mathcal{D}$ by $\mu\left(\mathcal{P}(\mathcal{D})_{M,j}\right)$, i.e.,*

$$\mu\left(\mathcal{P}(\mathcal{D})_{M,j}\right) = \frac{1}{|\mathcal{P}(\mathcal{D})|}\sum_{f_i \in \mathcal{D}} hS\left(f_i[W_{M,j}]\right)$$

Similarly, standard deviation of the high-frequency part of the spectrum of sliding window $f_i[W_{M,j}]$ over a period $\mathcal{D}$ is denoted by

$$\sigma\left(\mathcal{P}(\mathcal{D})_{M,j}\right) = \sqrt{\mu\left(hS\left(f_i[W_{M,j}]\right) - \mu\left(\mathcal{P}(\mathcal{D})_{M,j}\right)^2\right)}$$

Remark 3.3.2. *In the rest of this chapter, we use the high-frequency part of the spectrum to refer the high-frequency part of the spectrum of sliding window* $f[W_{M,j}]$ *when sliding windows* $W_{M,j}$ *are defined in the context.*

In Figure 3.2 and 3.3, we slide windows by 15 minutes. Thus, sliding windows are represented by $W_{M,M+15j}$. We observe $\left(j, \mu\left(\mathcal{P}(\mathcal{D})_{M,j}\right)\right)$ in the plane where $k = M + 15j$ for $j = 0, \ldots, 24$.

- In Figure 3.2, each green, red, orange, and blue graph corresponds to the graphs of $\mu\left(\mathcal{P}(\mathcal{D})_{M,j}\right)$ where $\mathcal{D}$ is 1st, 2nd, 3rd, and 4th quarter of the year, respectively, and grey plot is for a period $\mathcal{D}$ of the year. The horizontal axis corresponds to the jth sliding window in time.

- For instance, green graph of Figure 3.2a is plotted with connected points of $\left(j, \mu\left(\mathcal{P}(\mathcal{D})_{M,j}\right)\right)$ over a period $\mathcal{D}$ where $\mathcal{D}$ is a set of full-trading days in the 1st quarter of 2013's, that is, from Jan-01-2013 to March-31-2013.

- Red graph of Figure 3.2a is plotted with a period $\mathcal{D}$, where $\mathcal{D}$ is a set of full-trading days in the 2nd quarter of 2013's, that is, from April-01-2013 to June-30-2013

- In Figure 3.3, the graphs are plotted over period

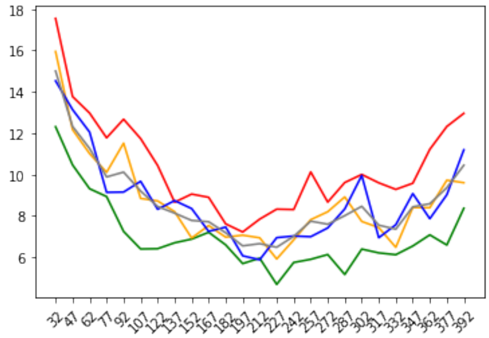
  $$\mathcal{D} = \{\text{full-trading days from April-02-2007 to April-02-2019}\}$$

  , and the whole value is divided by their minimum value to normalize the graphs.

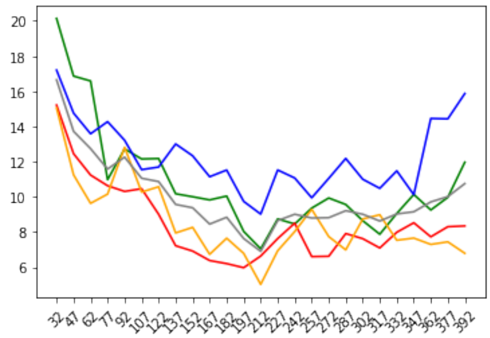
- In Figure 3.3, blue graph used price of DJIA futures to obtain the average of *the high-frequency part of the spectrum* over a period $\mathcal{D}$.

- In Figure 3.3, green graph used price of Nasdaq futures to obtain the average of *the high-frequency part of the spectrum* over a period $\mathcal{D}$.

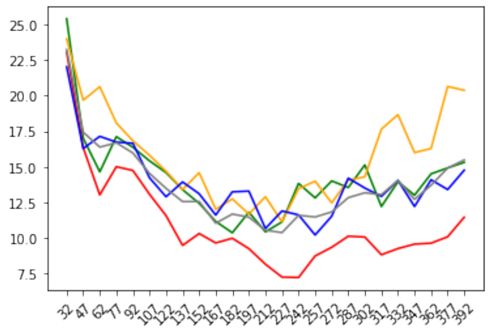
- In Figure 3.3, red graph used price of S&P 500 futures to obtain the average of *the high-frequency part of the spectrum* over a period $\mathcal{D}$.



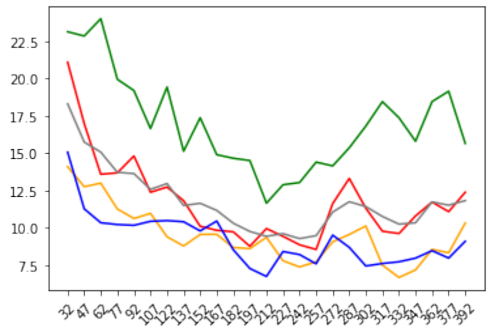
(a) Quarterly graph for 2013's



(b) Quarterly graph for 2014's



(c) Quarterly graph for 2015's



(d) Quarterly graph for 2016's

Figure 3.2: Average of *the high-frequency part of the spectrum* over a period $\mathcal{D}$.

Results

In Figure 3.2, we can observe that there are no quarterly characteristics in *the high-frequency part of the spectrum*, instead, in any quarter in any year, we can find a pattern with wide U-shape. This is more clear in Figure 3.3.

In Figure 3.3, We can observe that the graphs have the highest value when the market opens. we can observe that the highest value is almost twice the minimum value which appears around noon for all three futures. We may interpret this result as follows. First, stock trading is most active in the morning during 32 trading minutes right after the market opens. Traders or algorithms are busy to make decisions based on the information generated in the last night. Second, after 200 minutes of opening

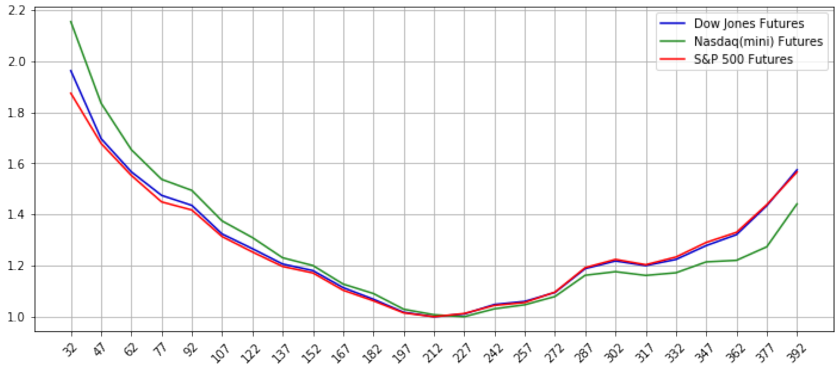


Figure 3.3: Characteristic pattern of the spectrum information during typical trading day for 13 years.

of market, around 12 O'clock, trading is the least active, it may be related to lunch time or they had trade enough in the morning. Last, at the end of the market, trading gets gradually active, even if it is not active as in the morning. They are likely to make orders to hold shares that are likely to rise in price in the night. This pattern appears regardless of season and year.

Next, we observe the relation between the number of shares and *the high-frequency part of the spectrum*. The number of *full-trading days* were 2832, 2867, and 2880 for DJIA futures, Nasdaq(mini) Futures, and S&P500 Futures, respectively. All of them has maximum value of *the high-frequency part of the spectrum* at the first window that contains trading information from 8:30am to 9:02 am. DJIA futures and Nasdaq(mini) futures have their minimum values at 212th sliding window, which contains 32minutes of trading information up to 12:02 pm. Similarly, S&P 500 futures has its minimum at 227th sliding window that contains 32minutes of trading information up to 12:17 pm. Then, transaction becomes gradually activated until the market is closed.
Table 3.1 shows values before dividing minimum of *the high-frequency part of the spectrum*. Over all, the average of *the high-frequency part of the spectrum* of Dow

	Dow Jones		Nasdaq(mini)		S&P500
# of full-trading days	2832		2867		2880
maximum	26.587	$>$	8.093	$>$	2.973
minimum	13.551	$>$	3.758	$>$	1.589
# of stocks	30	$<$	100	$<$	500

Table 3.1: Comparison between three futures

Jones Futures has the highest value, Nasdaq(mini) futures was the second highest, and S&P500 has the lowest spectrum. Considering that the number of stock they include has opposite relation, high frequency of trading is more sensitive of collection of smaller number of stocks.

3.4 Predicting Flash Crash

A flash crash in stock market is a phenomenon in which a plunge and a surge occur in a very short time. Experts estimate that one of the causes of this phenomenon is High Frequency Trading(HFT). HFT is literally an extremely fast stock trading. Their transaction is concluded within milliseconds using powerful algorithms. The computer analyzes multiple markets, and places orders according to the algorithms. A huge scale of transactions is done in a very short time, and it can have a significant impact on the entire stock market. For example, on May 6, 2010, due to the concern of the Greeks' economic crisis, the Dow Jones Industrial Average had a severe drop of 998.5 point in its price in the day and 600 point loss of which took only 5 minutes. Though it recovered most of its loss in 20 minutes, the market had almost a trillion-dollar loss, see Figure . In this section, we define f_i as follows.

- $f_i[T]$ is a time-series of price on ith trading day of a period $\mathcal{D}$ where T is a minute scale time sequence during the market opens.

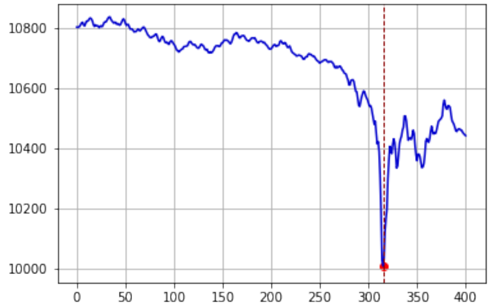
In Figure 3.3, we observed that there is a similar pattern of *the high-frequency part* *of the spectrum* among DJIA futures Index, Nasdaq(mini) futures Index, and S&P 500 futures Index. Thus, in the rest of this chapter, we focus on DJIA futures based on an assumption that same strategy can be applied on the other futures. Sometimes market breaks for minutes or data are missing, even in trading days. In this section, we deal with *normal-trading days* that have more than 390 minutes of trading history for one day. Thus, size of time sequence for a day satisfies $391 \leq |T| \leq 401$.

We use sliding windows of time sequence T of size $M = 32$, and shift window for every minute for real-time analysis of trends of *the high-frequency part of the spectrum*. There are windows $W_{M,M+j}$ for $j = M, M+1, \ldots, N-M$ where $N = |T|$. We can find that there was an extremely high value in *the high-frequency part of the spectrum* more than 30 minutes before flash crash on May 6, 2010. Note that the red plot is $\mu\left(\mathcal{P}(\mathcal{D})_{M,j}\right) + 2\sigma\left(\mathcal{P}(\mathcal{D})_{M,j}\right)$ with a period $\mathcal{D}$ that is a whole year that includes the flash crash day. This plot plays a role of baseline to show clearly how high $hS\left(f_i[W_{M,j}]\right)$ is on flash crash days, for each window $W_{M,M+j}$.

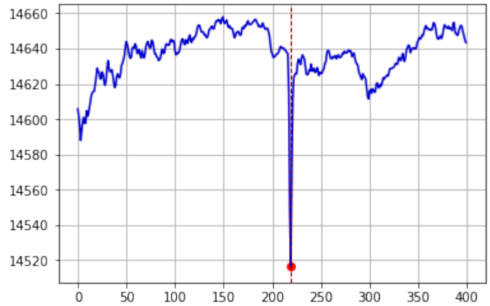
Results

Each column of Figure 3.4 shows charts of the index and the high-frequency part of the spectrum of a day when a flash crash happened. First column is for May 6, 2010, and second column is for April 23rd, 2013. First row is index chart of DJIA futures, and second row is *the high-frequency part of the spectrum* of jth sliding window size of M shifting by one minute. Red line is a baseline for unusual volatility of the market. To closely observe the change in *the high-frequency part of the spectrum* value before the flash crash occurs, the graphs of the second row are zoomed in the third row.

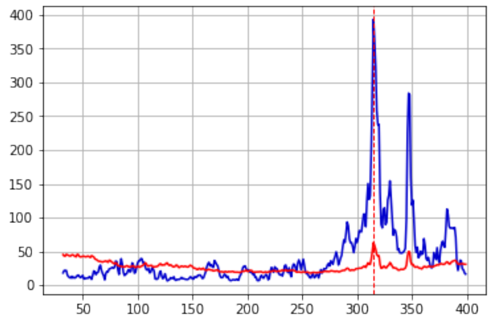
On May 6, 2010, DJIA futures had its lowest index at 13:45pm, that is, at 315th trading minute. However, *the high-frequency part of the spectrum* had already exceeded its baseline at 268 trading minute (See Fig. 3.4c. That is, there were preliminary sign at 268 trading minute, which is 47 minutes before the actual crash. From on economic point of view, on the day, the world economy was unstable due to Greek crisis.



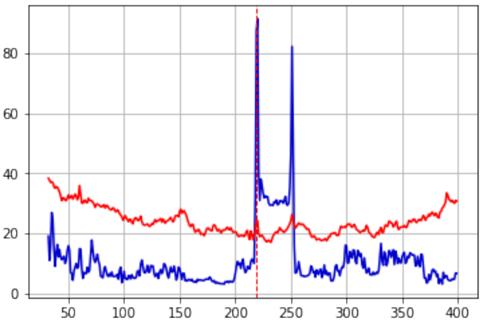
(a) DJIA futures chart on May 6, 2010.Crash happened at 315 trading minute



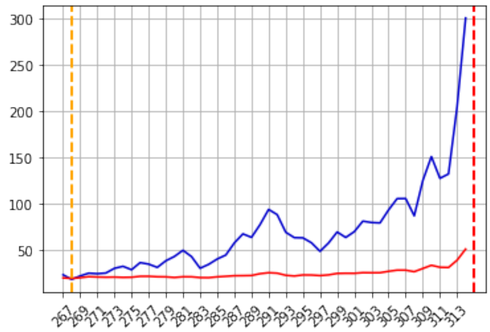
(b) DJIA futures chart on April 23rd, 2013. Crash happened at 219 trading minute



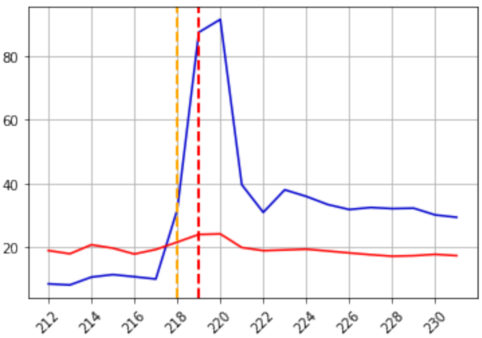
(c)



(d)



(e)



(f)

Figure 3.4: **(a)** DJIA futures chart on May 6, 2010. Crash happened at 219 trading minute. **(b)**DJIA futures chart on April 23rd, 2013. Crash happened at 315 trading minute. **(c)**,**(d)** *the high-frequency part of the spectrum* of trading price in sliding windows of size $M = 32$ of the chart (a),(b), respectively, shifting by one minute. Red line is baseline for defining usual day. **(e)** Zoomed figure of (c) from 267 to 315 trading minute. **(f)** Zoomed figure of (d) from 212 to 230 trading minute.

Concerns might be reflected in HFT algorithms, making the market in a panic.

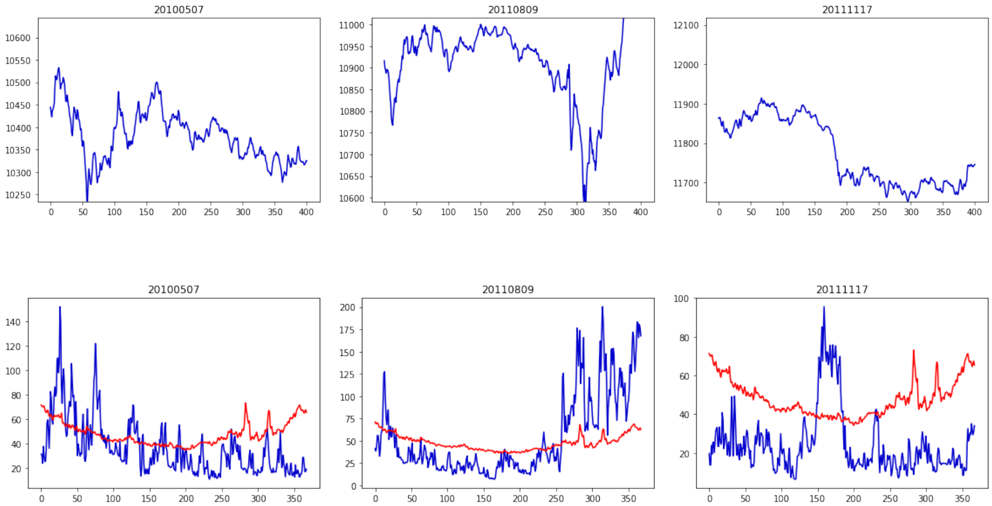
On the other hand, on April 23, 2013, *the high-frequency part of the spectrum* had exceeded its baseline one minute before the actual crash. Crash happened at 12:09 pm which is 219 trading minute.*the high-frequency part of the spectrum* does not have strange value until 218 trading minute. That is, we cannot find meaningful signal of the crash. In society on the day, Associated Press news(AP) twitter account was hacked and made false claim of explosions at White House. It took only two minutes to get 1% drop of Dow Jones futures Index at 12:09pm(219th trading minutes). In this case, people may have rarely known the news in advance, and neither it did not have effect on the stock market. In this case, a preliminary signal cannot present in the data before 12:09. Thus, it is more natural that the flash crash cannot be detected.

3.5 Spectral analysis of some other crashes

This methodology can be applied to other days not severe drop as flash crashes as. We begin this section by defining abnormal trading time, and days with *the high-frequency part of the spectrum* for each sliding window of size $M = 32$.In this section, we define $\mathcal{D}$ and f_i as follows.

- Period $\mathcal{D}$ is $\{$normal trading days from Jan 01, 2010 to Jan 01, 2014$\}$.
- $f_i[T]$ is a time-series of price on ith trading day of a period $\mathcal{D}$ where T is a minute scale time sequence during the market opens.

In this section, baseline is $\mu\left(\mathcal{P}(\mathcal{D})_{M,j}\right) + 4\sigma\left(\mathcal{P}(\mathcal{D})_{M,j}\right)$ with the period $\mathcal{D}$ for $j = M, M+1, \ldots, 391$. This baseline is more strict than that of previous section which is plotted with red in Figure 3.5. We call jth trading minute as an unusual trading minute when *the high-frequency part of the spectrum* exceeds the baseline in the time window $W_{M,j}$, and a trading day is called as an unusual trading day when there are more than 15 minutes of unusual trading minutes.



(a) May 7, 2010 (b) August 9, 2011 (c) November 17, 2011

Figure 3.5: top: index chart of DJIA futures. bottom: *the high-frequency part of the spectrum*(blue) and baseline(red) for each sliding window.

In the Figure 3.5, we choose three unusual days that are randomly selected to explain how *the high-frequency part of the spectrum* behaves on days of the fluctuating market: May 7, 2010, and August 9, 2011, and November 17, 2011. Figures in the first row are index charts of DJIA futures on these days. The height of the boxes is limited up to 1.04% minimum index of the day. Each figure in the second row has two plots, red plot is baseline, and blue plot is *the high-frequency part of the spectrum*. We can observe a pattern similar to that analyzed on days of flash crashes. For example, on November 17, 2011, crash happened around trading minute 190, and $hS\left(f[W_{M,j}]\right)$ exceeds the baseline at trading minute 150.

Chapter 4

Resonance in Financial Data

4.1 Resonance

Resonance is a phenomenon that the amplitude of a signal is increased when the signals are synchronized. The collapse of Takoma bridge is a famous example and it shows how resonance can have a powerful impact. In Topological Data Analysis(TDA) point of view, resonance can be detected using persistence homology and its barcode [7]. However, since DFT is a very effective method in computing frequency, we keep using DFT to find resonance. When $\hat{f}[k]$ is outstanding compared to its neighborhood, then it means that the original function f has many compositions of frequency k.

Definition 4.1.1. *A function f is $T-periodic$ if $f(t+T) = f(t)$ for $t \in [0, 2\pi]$.*

Definition 4.1.2. *Let f_1, f_2 be $T_1, T_2 - periodic$ functions, respectively. Then, we say f_1, f_2 are in resonance, if $T_1/T_2 \in \mathbb{Q}$.*

Since we deal data of finite points, $\hat{f}[k]$ has period in $\mathbb{Q}$. Thus, any pair of $\hat{f}[k]e^{\frac{2\pi i}{N}nk}$ and $\hat{f}[k']e^{\frac{2\pi i}{N}nk'}$ resonance, mathematically. To observe strong features in data, we define strong resonance.

Definition 4.1.3. *Let $r = n/m$ be a rational number for $n, m \in \mathbb{Z}$. Then, r is called a nice rational number if $m, n \in \{1, 2, 3, 4\}$.*

Then, for two periodic functions f_1, f_2 of period T_1, T_2, respectively, we say f_1, f_2 are in strong resonance if T_1/T_2 are nice rational number.

Results

We observe *strong resonance* in stocks listed on KOSPI(Korea Composite Stock Price Index) as shown in the Figure 4.1.

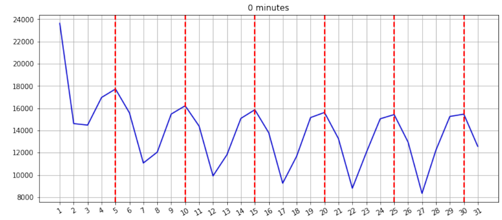
- Period $\mathcal{D}$ is whole trading days in June,2016.
- Index is KR7000210005, DAELIM INDUSTRIAL CO.,LTD.

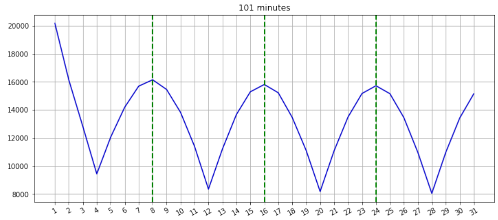
In Figure 4.2a and 4.2b, average spectrum of DFT of $f_i[W_{M,j}]$ for $f_i \in \mathcal{P}(\mathcal{D})$ for fixed $M = 64$ and $j = M, \ldots, 391$ where period $\mathcal{D}$ is a set of trading days during June, 2016. Pattern in 4.2a lasts while j shifts from M to $M + 27$, and 4.2b lasts for 55 minutes while j shifts from $M + 52$ to $M + 107$.

We can observe that these *strong resonances* are dominated by one day, respectively. The first resonance that has peaked for every 5 frequency seems to be dominated by June 24, 2016 (See 4.1c). On June 23, 2016, a referendum was conducted in the U.K for Brexit agenda. In the midnight, voting and counting had been broadcast worldwide. The impact of the vote had hit the World economy, including South Korea. We may guess the result is reflected in frequency.

On the other hand, the second resonance that has peaked for every 8 frequency seems to be dominated by June 21, 2016 (See 4.1d)

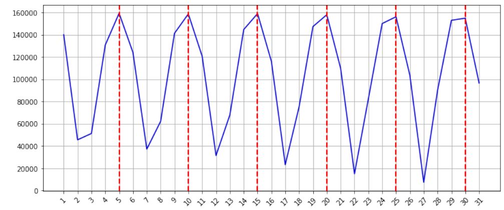
Also, we can find a strong resonance in some companies. We selected two representative examples of ISIN KR7078930005 and KR7007570005 as shown in 4.2. This phenomenon might happen because a large number of trading systems are synchronized. We might infer the trading pattern of the automated trading algorithm without additional data.

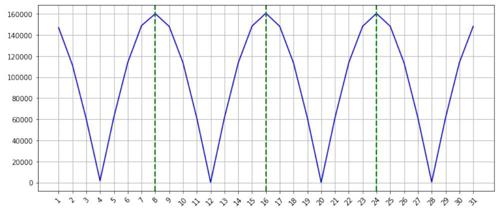




(a) Average of spectrum of $f_i[W_{M,j}]$ for $f_i \in \mathcal{P}(\mathcal{D})$ for fixed $M = 64$, $j = 0$

(b) Average of spectrum of $f_i[W_{M,j}]$ for $f_i \in \mathcal{P}(\mathcal{D})$ for fixed $M = 64$, $j = 101$

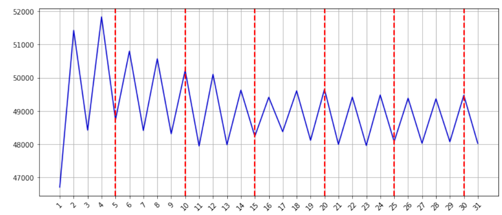


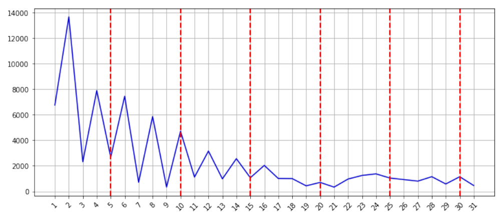


(c) DFT of $f_i[W_{M,j}]$ for $j = M, \ldots, M+27$ where f_i is a time-series of price on June 24, 2016

(d) DFT of $f_i[W_{M,j}]$ for $j = M+52, \ldots, M+107$ where f_i is a time-series of price on June 21, 2016

Figure 4.1: Resonance in DAELIM INDUSTRIAL CO.,LTD.





(a) DFT of time-series of price of KR7078930005 on June 24, 2016

(b) DFT of time-series of price of KR7007570005 on June 24, 2016

Figure 4.2: Strong resonance

Bibliography

[1] A. Haldane, "Patience and finance," in *Speech to Oxford China Business Forum, Beijing, Bank of England*, vol. 2, 2010.

[2] G. W. Shorter and R. S. Miller, *High-frequency trading: background, concerns, and regulatory developments*. Congressional Research Service Washington, DC, 2014, vol. 29.

[3] M. Gidea and Y. Katz, "Topological data analysis of financial time series: Landscapes of crashes," *Physica A: Statistical Mechanics and its Applications*, vol. 491, pp. 820–834, 2018.

[4] J. A. Perea and J. Harer, "Sliding windows and persistence: An application of topological methods to signal analysis," *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 799–838, 2015.

[5] J. Von Zur Gathen and J. Gerhard, *Modern computer algebra*. Cambridge university press, 2013.

[6] G. Tauzin, U. Lupo, L. Tunstall, J. B. Pérez, M. Caorsi, A. Medina-Mardones, A. Dassatti, and K. Hess, "giotto-tda: A topological data analysis toolkit for machine learning and data exploration," *arXiv preprint arXiv:2004.02551*, 2020.

[7] M. Coccolo, B. Zhu, M. A. Sanjuán, and J. M. Sanz-Serna, "Bogdanov–takens resonance in time-delayed systems," *Nonlinear Dynamics*, vol. 91, no. 3, pp. 1939–1947, 2018.

초 록

본 논문은 주식 시장의 변동성을 관측하고, 플래시 크래시를 예측하기 위해 진동수를 두 가지 방법으로 분석한다. 우리는 주식 거래의 분봉 데이터셋을 진동수와 스펙트럼으로 분해하는 이산 푸리에 변환을 이용하여 분석한다.

첫 번째 방법은 가격 데이터의 슬라이딩 윈도우를 이산 푸리에 변환을 취하여 고진동수 부분의 평균 스펙트럼을 관찰한다. 우리는 이 값을 통해 거래 활동의 경향성을 발견할 수 있다. 또한, 플래시 크래시 전에 고진동수 부분의 평균 스펙트럼이 비정상적으로 높아짐을 감지하였으며, 이를 통해 플래시 크래시를 예측 가능성을 제시한다. 또한, 우리는 이 방법론을 주가의 급락 및 급증이 있는 날에 적용하여 우리가 제시한 모델이 실제 시장을 잘 설명함을 보인다.

두 번째 방법은 가격 데이터의 슬라이딩 위도우의 이산 푸리에 변환을 있는 그대로 분석하여 주식 시장에서의 공명 현상을 관찰한다. 이는, 자동 매매 알고리즘에 대한 정보를 추가 데이터 없이 거래 데이터로부터 추론한 것에 의의를 가진다.

주요어: 이산 푸리에 변환, 시계열 분석, 스펙트럼, 공명 현상, 주식 거래 데이터

학번:2018-29001

감사의 글

이 논문이 완성되기까지 도움을 주신 분들께 이 자리를 빌려 감사의 말씀을 전합니다. 우선, 석사 과정 동안 성심성의껏 지도해주신 Otto van Koert 교수님께 감사의 말씀 드립니다. 교수님의 지속적인 관심과 걱정 덕분에 여기까지 올 수 있었습니다. 그리고 바쁜 일정 중에도 논문 심사를 맡아주시고, 소중한 조언과 충고를 해주신 임선희 교수님과 국웅 교수님께 감사드립니다.

늘 응원해주시고 격려해주신 사랑하는 부모님께 감사드립니다. 믿고 지지해주신 덕분에 한 번 더 문턱을 넘을 수 있었습니다. 또한, 같이 함께 지도를 받은 연구실 분들께 감사드립니다. 함께여서 헤쳐나갈 수 있었고, 힘을 받을 수 있었습니다. 이 외에도 대학원에서 같이 공부하고 도움을 준 선후배 및 동기들, 소중한 인연이 되어주신 모든 분께 감사의 말씀 드립니다.