Big Data Analysis of Individual Investors Behavioural Biases in China

by

HAN, DUN

A Thesis Submitted in accordance with the requirements for the Degree of Doctor of Philosophy

Newcastle University

Newcastle University Business School

December 2021

Abstract

This thesis analyses biases on individual investors selling decisions in the stock market from a very large and unique dataset from China. The dataset contains 100,000 individuals with more than 56 million daily holding records from January 2007 to May 2009. This thesis introduces the Chinese stock market and the Chinese investors in Chapter 1, and summarize the previous and current literatures in Chapter 2, then perform empirical tests in the following chapters.

In Chapter 3, I find that investors hold a preference on realizing large gain while no preference on loss, while this result is affected by the holding period and the market condition. More specifically, for short-term holding positions, individuals prefer to sell both large gain and large loss. But this effect reverses in the medium- and long-term. Regarding the different market scenarios, Chinese investors are more likely to sell large gains in all conditions but only prefer to realize large loss under booming market. Therefore, the confidence level to the market could be reasons for individuals to realize loss.

In Chapter 4, I investigate the investors' selling decision within their portfolios, and find that individual investors in China are more likely to sell a position with extreme good (the best) performance, and followed by the 2nd best position, but reluctant to sell the salience of extreme bad position, which is different from result in the US (sell both best and worst positions). When lottery-like positions are held by young, male and new investors, Chinese investors also sell the worst rank positions in their portfolios. Thus, the willing of gamble could be a reason for individuals to realize relative underperforming positions. This result is robust under different modelling method, extreme portfolio situation, measurement of rank and limit-down limitation, etc., and consistent in different market condition and holding period positions.

In Chapter 5, this study further tests how geographic factors impact selling biases above. Metropolis investors suffer more from disposition effect. Investors from different regions in China do not have different degrees on these effects. Furthermore, individual investors in China do not have significant preference on local stocks. Local stocks cannot influence disposition effect and V-shaped disposition effect. However, local position can moderate the preference of selling relative overperforming positions (rank effect in China). Since rank effect is harmful to profit when the position is local, local stocks benefit to profit to some degree.

This thesis further indicates that some investor sophistications (e.g. experience, trading frequency) can moderate the individual's selling biases. All results are concluded in Chapter 6.

Declaration

This thesis has not been submitted in any part for any other degree or qualifying examination at this or any other university. Unless the manuscript indicates otherwise, this thesis is entirely my own work.

Statement of Copyright

Acknowledgements

The Ph.D. process is not only the way of solving academic questions. It is the path of overcoming yourself, building your thoughts and defeating your anxiety. And that is the so-called the Doctor of *'Philosophy'*. During this journey, I would like to appreciate many people who have supported me.

I would like to give my first thank to my principal supervisor, Dr. Pei (Jose) Liu for his great supervision and inspiration. He is an outstanding mentor providing me helpful advices both in academic and life. I would also like to thank my secondary supervisor, Prof. David Higgins, for his valuable feedbacks and suggestions.

I am grateful to my beloved parents, Liyan Han and Li Quan. Without their support, I would never be what I am.

I also appreciate my panel members, my colleagues, seminar participants and conference discussants for their brilliant ideas. The people are not limited to the name list: Simon Hussain, Aly Salama, Ke Tang, Jie Guo, Yu-Jane Liu, Jiangze Bian, Yanran Wu, Jonathan Batten, Biao Guo, Songtao Tan, Shunming Zhang, Yingguang Zhang, Shaojun Zhang, etc. Shiqi Ren and Siqi Yue also provide computer programming assistance to this thesis.

Dedicated to My Family

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

1.1 Introduction of the Thesis

In traditional economics and finance, the majority of researches have been built on the assumption that human beings are rational, which means they are unbiased and efficient processors of relevant information, and their decisions are consistent with utility maximization under risk aversion. Markowitz (1952) describes how to choose a portfolio with the minimum possible risk for the given expected return but assumes that all investors are rational and risk adverse. The Capital Asset Pricing Model (CAPM) and Efficient Market Hypothesis (EMH) essentially require that markets are efficient and stock prices instantaneously reflect all publicly available information. However, a large number of empirical studies over the last twenty years indicate that investors do not behave the way that is predicted. They suffer from many behavioral biases, for example, they fail to behave rationally in even quite simple situations (Elton, et al., 2004), using too simple diversification portfolio (Benartzi and Thaler, 2001; Goetzmann and Kumar, 2008), buying stocks they are familiar with (Massa and Simonov, 2006), influenced by limited attentions (Seasholes and Wu, 2007; Barber and Odean, 2008), using mental accounting to evaluate stocks (Thaler, 1985), trading too much due to overconfidence (Barber and Odean, 2000), keeping loser and selling winner – known as disposition effect (Shefrin and Statman, 1985; Odean, 1998). All these biases contribute to the over-performance or under-performance of investors in the real world than in the ideal model.

Among these biases, disposition effect is one of the most well-known and widely discussed topics. Prospect theory is widely used to explain this effect (Kahneman and Tversky, 1979; Odean, 1998). It suggests that due to the S-shaped utility curve, people have tendency to realize gains and reluctant to realize losses. However, Ben-David and Hirshleifer (2012) examine the V-shaped in the selling and profit function. They observe that investors are more likely to sell big gains and losses than the small ones. The effect is more significant in short-term positions. Meanwhile, using data from Finnish market, Kaustia (2010) observes similar result from gain positions but find the selling probability is approximately constant in the realm of losses. After that, An (2016) discovers V-shaped disposition effect cause damage to investor profit.

Due to the inconsistency of results in previous literatures, the present thesis first discusses the relation between the choice of selling of individual investor and the magnitude of gain and loss in China. By applying a unique dataset from Chinese stock market 2007 to 2009, Chapter 3 reveals that individual investors prefer to sell large gains, but the scale of loss does not influence the probability of selling significantly. Nonetheless, it is worth noting that these results are shown under the models controlling holding period. When removing holding period effect, the results entirely change. This thesis therefore emphasizes the importance of impact of holding period. Based on the separated holding period sub-samples, results in Chapter 3 indicate the V-shaped disposition effect appears only on short-term positions. For medium- and long-term positions, investors slightly prefer to sell positions that are close to zero. The trading strategy and loss tolerate of short-, medium- and long-term positions is different among individual investors. At the end of Chapter 3, this study also finds evidence that the preference of selling large gains cause damage to individual investors' profit in China, which is consistent with the result of An (2016) in US.

In Chapter 3, this thesis contributes on several aspects to the prior literature. First, to the best of my knowledge, this is the first study focuses on impact of returns on selling decision in China. Previous studies investigate disposition effect, which discuss the impact of sign of returns (gain or loss) on selling decision. This study is the first to discuss how the magnitude of gain or loss influence in China. Second, this is the first study of this topic during financial crisis. Since the data period is 2007-2009 which covers the financial crisis, the behavior of investors during financial crisis is worth noting. More details of the dataset are written in Chapter 1.3.

After the discovery of V-shaped disposition effect, Hartzmark (2015) further develops rank effect. It shows that investors compare the returns of stocks in their portfolio when consider selling and they are more willing to sell stocks with extreme winning (the best performance) and extreme losing (the worst) positions. The most crucial contribution of rank effect is that it considers the comparison within one's portfolio when making decision of selling. In previous studies, although most research successfully explain investors' behavior to some extent, most of them still suffer from a stock-by-stock bias, which they assume that investors consider stocks one-by-one and ignore the comparison between stocks in the portfolio.

Chapter 4 explores the rank effect in Chinese stock market, which is different from the US. Individuals in China are more likely to sell the best and the 2nd best performance positions and ignore the underperforming positions in some degree. To further analyze the difference between US (selling both best and worst positions) and Chinese (only sell good performing positions) investors in terms of rank effect, Chapter 4 introduces the factor of gambling trading to discuss the trading of underperforming positions in China. Kumar (2009) defines the positions with short holding period, low price per share and high volatility as lottery-like (gambling) positions. Furthermore, male and new investors are aggressive investors and more likely to gamble in Chinese market (Liao, Liang and Zhang, 2016). This study discovers that when male and new investors holding gambling positions in China, the selling probability of the worst performing position is higher than middle positions, which is the same as US investors. The gambling factor affect rank effect in China. Furthermore, Hartzmark (2015) only considers the comparison within the large size portfolio (contains at least 5 stocks). The results in chapter 4 also discuss rank effect in small size portfolios. For investors with small size portfolios, they have the preference of selling relatively good performance positions as well. Chapter 4 also indicates that rank effect in China causes damage to investors' profit and it is indeed a trading bias.

To the best of my knowledge, Chapter 4 sheds light on several aspect. First, this is the first study of rank effect in any market besides US market. Chinese market as an emerging market is entirely different from US market. Second, since the selling behaviours on under-performing positions of Chinese investors and US investors are different, Chapter 4 further discusses when investors sell the worst position, which is the first in literatures. Third, pervious literatures of rank effect all only include the portfolios with 5 or more positions in the theory. Chapter 4 includes all portfolios with 2 or more positions. By doing this, this study includes roughly 100% more portfolios than pervious literatures while still keep the focus on the comparison of positions within one's portfolio. At last, as similar to Chapter 3, due to the special dataset, Chapter 4 is the first research on rank effect during financial crisis.

Dhar and Zhu (2006) and Feng and Seasholes (2005) find investor sophistication can moderate disposition effect. This thesis further investigates to what extent do investor characteristics and sophistication affect the selling behaviors based on magnitude of return and rank effect. Chapter 3 documents that investor with more experiences and more trading frequencies can moderate the preference of selling large gains while age has no influence on the effect. In Chapter 4, female and senior investors are rank traders in China, while less experience and high trading frequency moderate rank effect. Since both V-shaped disposition effect and rank effect cause damage to profit and are indeed biases, as a conclusion, trading more helps on V-shaped disposition effect and rank effect and is a sign of sophisticated in Chinese market. This is an indirect evidence of the statement that investors can learn from their trades. Meanwhile, although age is generally seen as an indicator of investor sophistication, senior Chinese investors do not perform differently on selling large gain and loss with juniors. Since Chinese stock market developed late in the 1990s, senior people may be lack of professional stock knowledge during their education while juniors are more educated with the development of the stock market. Therefore, senior investors are not sophisticated in China at least to some extent. Meanwhile, to the best of my knowledge, both Chapter 3 and Chapter 4 are the first to analyse the influence of investor heterogeneity on their topics respectively.

The impact of geographic factors is another commonly discussed topic in behavioral finance. Developed from correlated trading in herding effect (Grinblatt, Titman and Wermer, 1995; Lakonishok, et al., 1992), previous studies state that investors in the same or near area and community tend to trade similarity by holding highly related portfolios (Hong, Kubik, and Stein, 2002; Feng and Seasholes, 2004; Brown, et.al, 2008) and have similar trading timings (Baltakys et al., 2019). Another topic in geographic behavioral finance is local bias, which demonstrates that there is a preference of individuals to trade local stocks, stocks from firms that register in the same place as the investor registers. Grinblatt and Keloharju (2001) document Finish investors tend to trade stocks that are headquartered close to their location. Ivkovic and Weisbenner (2005) and Seasholes and Zhu (2010) discover similar phenomenon in the US market.

In Chapter 5, this thesis investigates the impact of geographic factors on disposition effect, V-shaped disposition effect and rank effect. Metropolis investors suffer more from disposition effect. And under theories of V-shaped disposition effect and rank effect, their behaviors have no significant difference with other investors. Furthermore, investors in eastern, mid and western China do not have significant difference either on trading behavior of these three effects as well. Investors in different regions do not trade differently. Thus, investors in the same or close region do not trade relative similarly to some degree. Since eastern China is the well-developed region in China as well as metropolis, investors in well-developed regions cannot perform better under the theories of disposition effect, V-shaped disposition effect and rank effect. In addition, Chapter 5 further discusses the local effect in China and the impact of local positions on the three behavioral biases. Individual investors in China do not have local bias at a significant level. Meanwhile, although individual investors do not sell local positions in different levels of disposition effect and V-shaped disposition effect, local stock can moderate rank effect. Therefore, local stock plays a crucial role in the comparison system when individual investors want to choose a position to sell. Since Chapter 5 finds evidence that rank effect is bias that cause damage to investors' profit, moderating rank effect by local positions helps investor performance to some degree.

Chapter 5 contributes to the previous literatures in several aspects. To the best of my knowledge, Chapter 5 is the first study of the influence of region on V-shaped disposition effect and rank effect in China. It is also the first study of local effect in China. Meanwhile, it is the first study of the impact of local stocks on disposition effect and rank effect in Chinese market.

The data used in this thesis is rich and unique, which is collected from a large brokerage firm in China[1]. It contains more than 3 million accounts and 2 billion daily stock dealing records over the period of January 2007 to May 2009. Due to the consideration on the cost of computation, this study uses a randomly selected sample of 100,000 investors and more than 56 million records sub-data in this thesis. The dataset is formed with 4 sub-datasets that are customer file, account file, stock file and transaction file. Customer file contains the information of each customer. Account file contains balance information of customer's account on daily basis. Stock file contains information of each stock held by each customer on daily basis. Transaction file contains each deal's information. Customer ID is used to merge all files. A specific introduction of this dataset is in Section 1.3.

Both the market and the time period of the dataset are worth noting. China is an ideal laboratory to study behavioral finance among investors. Due to its successful economic transition in the last three decades, Chinese market has become the world's second largest stock market in value since 2014 and has been added to MSCI Emerging Markets Index since 2017, indicating its increasing importance in global economy. However, Chinese stock market starts later (only from 1990s) and has generally been viewed as under-developed market with high degree of asymmetric information, due among other things to its unsound financial system and its weak shareholders' protection, as well as its weak corporate governance system. The time

1 This thesis does not disclose the specific name of the brokerage for confidentiality reasons, same as most of the previous studies (Odean, 1998; Feng and Seasholes, 2005; Dhar and Zhu, 2006, etc.).

period of the dataset in this research is from 2007 to 2009, which covers the financial crisis period. In China, suffering from the world financial crisis, there was also a huge bubble in 2007 and experiencing significant stock price falling in 2008 then gradually recovering in 2009. The changes of the investor sentiment and behavior when they face large profits and losses along with risks are interesting to research. A specific introduction of Chines stock market and Chinese stock market bubble during 2007 to 2009 is given in section 1.2.

Since our data covers financial crisis (2007-2009), this thesis further discusses to what extent do the booming, crushing and recovering (namely before, during, and after) period of financial crisis affect individual behavioral biases. To the best of my knowledge, this is the first paper to discuss disposition effect, V-shaped disposition effect and rank effect during financial crisis. The results from Chapter 3 show that disposition effect is consistent in all market conditions. Chinese investors consistently prefer to sell large gains in all scenarios as well. Meanwhile, individuals are more likely to sell large losses than small losses under booming market but no significant response to the magnitude of loss in both crushing and recovering period. When individuals lose confidence to the market during and after financial crisis, they are more patient to their large loss positions and more willing to keep them. However, when discussing rank effect, in Chapter 4, the rank effect has no significant difference among booming, crushing and recovering period of financial crisis. Thus, under financial crisis, investors have no significant difference behavior to relative bad performance positions in their portfolios. Investors have different attitudes to absolute bad performance positions (losses) and relative bad performance positions.

The rest chapters of this thesis are organized as follows. Chapter 2 is a review of relevant literature. Chapter 3 discusses the impact of magnitude of gain and loss on selling decision (V-shaped disposition effect). Chapter 4 investigates the rank effect. Chapter 5 documents the impact of geographic factors on behavioral biases. Chapter 6 is the conclusion and discussion.

1.2 Introduction of Chinese Stock Market

This section presents a brief introduction to Chinese stock market. Section 1.2.1 illustrates the establishment and development of Chinese market. Section 1.2.2 focuses on the bubble and crisis happened in Chinese market during 2007-2009, which is also the period of the dataset of this thesis.

1.2.1 Brief Introduction to Chinese Stock Market

With the development of Chinese economic system reform to the industrial sector, Shanghai Feile Audio Corporation was established in November 1984 as the first corporation since 1949, and two months later, a collectively owned enterprise, Shanghai Yanzhong Industrial Limited Corporation issued its shares publicly. With gradually developing of corporation for two years, the first OTC market was set in Industrial & Commercial Bank of China Shanghai Branch in September 1986, which prepared well for the establishment of Chinese stock markets.

In November 1990 and April 1991, Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE) was established, respectively, with the deepening of economic system reform and opening-up policy. The two stock markets trade four hours in two sections on each working day, from 9:30 am to 11:30 am and from 1:00 pm to 3:00 pm, and adopt call auction to form opening prices half an hour before opening in the morning. The common stocks traded in the markets are priced in CNY and for domestic investors, named A-shares. The listed firms can also issue special shares, invoiced in US dollar in SSE or invoiced in HK dollar in SZSE, to be traded by foreign investors from 1992 to 2001, which are named B-shares. The B-shares was open to all investors since 2011. As an item in the commitment entering WTO in 2001, the Qualified Foreign Institutional Investor (in short QFII) program was established in 2002, allowing certain foreign institutions to invest in A-shares. For financing small-sized private firms with excellent operating performance and promoting new economy with high technology, SZSE opened the small firm board in June 2004, and the pioneering work board (or growth enterprise board) in October 2009. Those two boards in SZSE form the Chinese characteristic second-board markets. From 2004 up to now, SSE works as the main board of Chinese stock markets, while SZSE operates the second-board market like NASDAQ in US.

The Chinese stock markets developed steadily over 28 years. There were 53 listed firms by the end of 1992. The SSE and SZSE reached their first thousand listed firms in 2000 and achieved 3500 firms by January 2018. The number of listed firms grew 125 annually on average over 28 years. By the end of 2017, the total market value reached 56.75 trillion yuan. The trading volume went up to 11.5 trillion yuan in January 2018, with daily trading volume as 522 billion yuan on average. It has become the world's second largest stock market by value since 2014.

A stock index works as basic single directing stock investment. The system of Chinese stock indexes has been well improved with the developing of the markets. There are two categories of stock indexes. One is made by stock exchanges, such as Shanghai Composite Index, Shenzhen Component Index, CSI 300 Index, Shanghai 50 Index. Another group comes from the third-party index service agencies, such as China Securities Index Co., Ltd, SPCITIC, FTSE Xinhua, with index like CSI 500 Small-cap Index.

As the part of a transformation economy, the Chinese stock market has a very unique performance in both trading system and trading behavior. The first is the A-share and B-share setting and QFII rules mentioned above. The second, one investor can only open one trading account in agent until the year of 2015. Therefore, the personal trading data is quite clean before 2015. The Third is limit up and limit down restriction, in the magnitude of 10%. The fourth is the short sale constraint. The short sale has not been allowed for quite a long time until April 2010 when HS300 stock index futures was introduced into Chinese stock markets. At the same time of introducing stock index futures, the securities margin trading was launched in the markets. The margin trading is implemented with experiment step by step. The first step experiment involved 180 index stocks from HSS 180 Constitute Index, which were well-performed stock with big shares. The margin was strictly controlled, with the initial margin and the maintaining margin as 70% and 30%, respectively.

As a typical emerging market born from both a transformation combined with rapid growth economy, Chinese stock markets passed through a not ordinary developing process. The first stage went through from 1991 to 2005, titled as the share-split initial stage. At that time, most listed firms were controlled by state owned firms, in which roughly 30% shares of a stock could be freely traded in the secondary market. This character determined the basic pattern of the markets at that time, which will be explained in next section. By the end of 2004, the number of listed firms was 1377, the market share in total was 3.70 trillion yuan while the negotiable market capitalization was 1.17 trillion yuan, the average PE ratio was 24.29 in SSE and 25.64 in SZSE. The second stage covered the period from 2006 to 2009. After success in reform on share-split structure and rapid development of mutual funds, the markets past through the biggest bullish market with the biggest bubble and experienced the deep bearish market after burst of the babble. Investors can only earn money by pushing the bullish markets when short sale trading was absent. By the end of 2009, the number of listed firms was 1718, the market share in total was 2.44 trillion yuan while the negotiable market capitalization was 1.51 trillion yuan, and the average PE ratio was 28.73 in SSE and 46.01 in SZSE. With such a high PE ratio, one can feel the sentiment at that time. This is the period of the dataset used in this study, a detailed description will be given later on in this chapter. The third stage covers the period from 2010 up to now. During this period, the margin trading rules was established and managed gradually well. Meanwhile, a kind of equalization-reserve fund, run by China Securities Finance Corporation Limited, was introduced to the market after the short flash babble in 2015. This event was a world-seeing shock. The average daily trading volume was 638.3 billion yuan in January 2015, while it went up 1745.7 billion yuan in June 2015, almost triple than in January. It should be blamed for the flash babble that uncontrolled trust leverage which financed the speculators with reasonable banking products in commercial banks. Drawing the lesson, the equalization-reserve fund was founded to smooth the market big wave, which are owned by major security companies.

The share of individual investors in Chinese stock markets accounts for near 70%, so their behaviors affect the trading waves and price volatility quite clearly, especially in bullish and bearish markets. Due to the shortage of information, individual investors possess serious massive herding behavior (Shi, et.al. 2009). Chasing investment hotspots often drive turn over in great deal for relatively short period, from 5 to 10 trading days on average.

The institutional investors have been grown up with market developing, which are divided into two major sectors, public offering funds and privately offered funds. By the end of February 2018, the number of public offering funds increased up to 5000, while privately offered funds achieved more than 70000 in total number. The public offering funds are managed by only 116 fund management companies, while the privately offered funds are belonging to 23400 general partners. Under the guidance of the fund industry society, the public fund sector plays clear stabilization in the markets. For private funds, some mature funds can work as hedge funds with excellent performance, in contrast, while quit a lot invest like individual investors, performing as noise trades.

As the typical emerging market with being supported by a transition real economy, Chinese stock market has been endowed with some features compared with developed security markets in the world. First, China Securities Regulatory Commission controls firm listing. Second, like American market, there are two major stock exchanges, SSE for ordinary firm listing and SZSE for growth firm listing. Third, margin trading and short sales are constraint only a part of good performance could be traded on margin and shortly. Fourth, an equalization fund, owned by SSE, SZSE and China Securities Depository and Clearing Corporation Limited, plays crucial role to control excessive fluctuating, which is analogous to Hong Kong and Japan. Fifth, the stock derivatives run under developed. There are only two stock index futures and one stock index ETF option, which are restricted by high deposit for security. Sixth, China Securities Regulatory Commission implement QFII, qualified foreign institutional investors, for managing foreign capital in investing Chinese stocks.

1.2.2 Chinese Market in 2007-2009

The year of 2008 saw a sequence of adverse financial news in the world and triggered the US credit crunch and market crisis. And it soon became the worldwide financial crisis. This poor external financial environment must have had a great impact on Chinese stock markets. There was indeed an extreme volatility of stock prices that signified a market bubble and bursting in Chinese market.

(Insert Figure 1.1 here)

The entire year of 2007 is a crazy year of Chinese stock market (see Fig.1.1). The biggest bull market came to Chinese stock markets in the spring of 2007. The Shanghai Composite Index stayed a little higher than 1000 points for more than one year since 2005. The market began its biggest bullish luck since the spring of 2006. In the initial stage, the index grew quite gently, while it came to crazy burst when entering to the year of 2007. The index surged over 3500 points from 2715.72 at the very begging of the year to 6124 on October 16, with the rise of 140%. And then it plunged all the way, felling back to roughly 2000 points in September of 2008, with the loss near to 70%. In sharp contrast, during the same time period, the Chinese real economy grew at more than 10%-year one year. The Split-Share Structure Reform happened in 2005 was the pushing hand of this bubble.

1.2.3 The Split-Share Structure Reform

In 1990, Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE) was established, with the deepening of economic system reform and opening-up policy. They adopted a unique mixed business model in which roughly 30% shares of a stock could be freely traded in the secondary market, while the Chinese government kept control of the listed firm by owing more than two of third non-tradable shares. Within this systematic framework, a typical firm's shares were split into state shares, legal-entity shares and tradable shares, which are named the split-share structure. Based on the system setting, only tradable shares, which account for about 30% of all shares, could be traded in stock exchanges by investors during that time period. This unique dual share structure had once played a double-edged role in the development of Chinese stock markets in both positive and negative ways. On one hand, this compromise in system promoted the initial rapid growth and kept the sustainable liquidity needed for trading. On the other hand, the split-share structure caused serious corporate governance problems, which allowed the controlling shareholders seized interests of small shareholders and blocked mergers and acquisitions (Allen, et al, 2005; Liao, et al, 2014). It also twisted the rationality of investors and encouraged short-term speculation in the secondary market, resulting in overheating turnover and severe volatility (Xiong and Yu, 2011). In contrast to the average turnover ratios of the stock markets in the US, UK, and Japan, being 129%, 142%, and 119%, respectively, China even heated over 900 % during 2005 to 2007 (Liao, et al., 2014). For system transfer reasons, therefore, individual investors ignored the long-term performance of listed firms they held with shares, while only focused on speculation for quite short time, quickly in and quickly out, which pushed the high volatility on waves (Lehkonen, 2010). As a result, all investors were eager to release the market for liquidity in a long period of time.

On April 30, 2005, the Notice of the official document of China Securities Regulatory Commission, in title of "Piloting the Share-Trading Reform of Listed Companies", was issued. It marked the official start of the Split-Share Structure Reform, which is the milestone for the institutional building of Chinese stock market when the markets fulfilled the ownership reform that allowed all shares of listed companies to be freely traded. From then on, the crazy bubble began.

1.2.4 The Market Sentiment and New Investors

Chinese stock market provides an interesting and unique research environment in terms of the market emotion and investor behavior. At the start of the bubble, with the extreme bullish market, both domestic and foreign investors were enticed to buy whatever shares were on offer without carefully analyzing the real performance and growth potential of the listed firms. The whole market was under the emotion of over-confidence and over-optimistic. Then when the bubble burst, all investors were facing extraordinary loss and risk. The overall emotion turns to fear and lack of confidence.

(Insert Figure 1.2 here)

One remarkable thing is that many individuals with no experience witnessed investors in market made large profit during 2007, so they chose to join the market. In fact, the numbers of new accounts in Shanghai Stock Exchange hiked the steep growth rate above 1 million per month lasting from March to December in 2007, which accounted for two to three times than before, as shown in Fig.1.2. The beneficial from highly increased newly opening accounts was to promote the liquidity. However, most of these new investors were lack of basic financial knowledge and market experience. They are more likely to suffer from individual investor behavioral bias. How these new investors performed under the following bear market, whether they would learn the lesson and became sophistication or just quit, is an interesting topic.

1.3 Introduction of the Unique and Large Dataset

This research is based on a very large database collected from a large nationwide brokerage firm in China, with more than 3 million accounts and 2 billion daily dealing records over the period of January 2007 to September 2009. The dataset is formed with 4 sub-datasets that are customer file, account file, stock file and transaction file. Customer file contains the information of each customer. Account file contains balance information of customer's account on daily basis. Stock file contains information of each stocks held by each customer on daily basis. Transaction file contains each deal's information. Customer ID links all files. Due to the computational capacity limitations, a random sample of 100,000 investors and more than 56 million trading records sub-data is applied in all empirical chapters to build the model. As far as my knowledge, the size of this sample dataset is still much larger than the whole dataset of most empirical behavioural finance researches.

The first unique feature of this dataset is the time period. In the period of 2007 to 2009, the Shanghai Stock Exchange (SSE) composite index was roughly 2500 at the beginning of 2007 and reached the top 6000 in October, with an increasing rate roughly 140%. Then the index dropped all the way to 2000 at the third quarter of 2008. The huge raise at the first half of 2007 attracted a large number of new accounts. The whole market was in the atmosphere of over-confidence and over-optimistic. However, at the second half of 2007 and the whole year of 2008, both experienced and new investors faced the extreme loss and risk. The special time provides us a good opportunity to study investor decision and behavior under large pressure and risk.

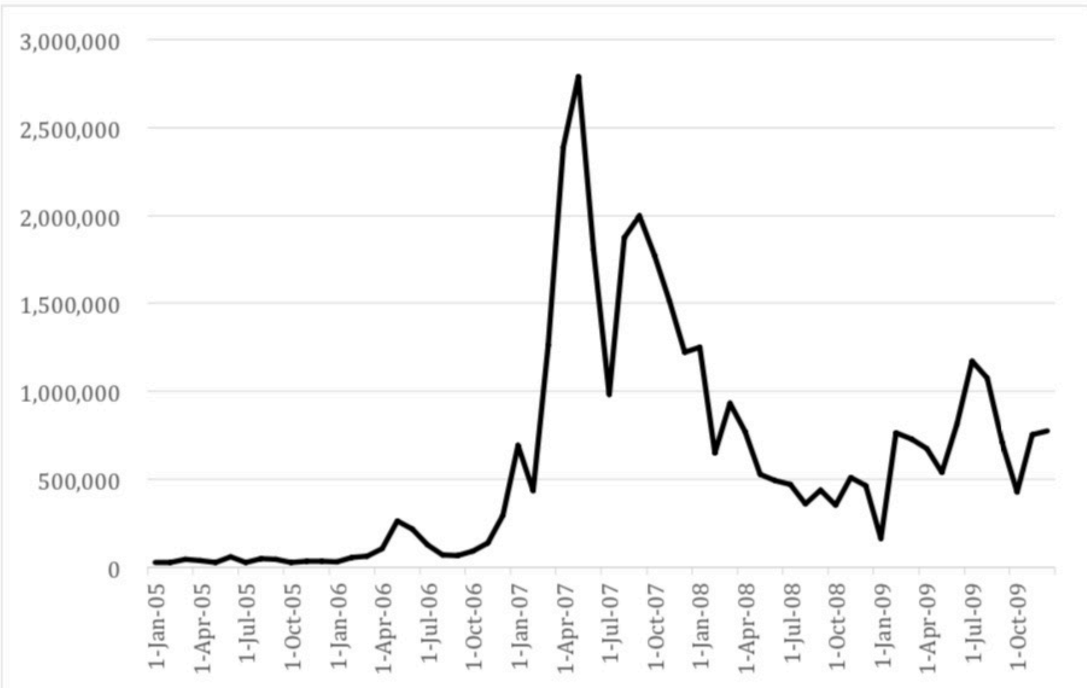
The second unique feature of this dataset is that it is from the Chinese market. To the best of my knowledge, only limited papers examine investor behavior in emerging markets especially on China. Chinese stock market was established late in 1990, but it has become the world's second largest stock market by value since 2014 and has been added to MSCI Emerging Markets Index since 2017, indicating its increasing importance in global economy. However, due to China's unsound financial system and heavy-handed government intervention, Chinese stock market still faces considerable challenges and exhibits many behavioral biases. It is also worth noting that due to the establishing time of Chinese market and the average education level as a developing country, the average experience and education level of investors in China, in other words the sophistication level, is lower than investors from developed countries. It may also cause the difference in behavior between Chinese investors and investors from other countries with both domestic professional investors and individual investors.

The third unique feature is the number of investors it contains. The dataset contains more than 3 million investors and 2 billion daily trading record. Most of dataset used in past studies only contains roughly no more than 100 thousand. The data used in this study brings several advantages to my research. It helps me to get a more comprehensive analysis of the entire market, and also significantly increases the performance in statistic level. As I will separate investors into different groups to exam the characteristics, the large number of total investors in the dataset provides me enough observations for even very subdivided group.

Figure 1.1: S&P 500 vs. SEE Composite Index

Source: S&P500 index and SSE composite index

Figure 1.2: Monthly Number of New Open Accounts in SSE, 2005-2009



Source: China Capital Market Development Report

Chapter Two: Literature Review

The literature review of this thesis is composed of three main perspectives. The first two sections introduce the development process of theories from disposition effect to V-shaped disposition effect and rank effect introduced in the previous chapter. The first section is based on disposition effect, as one of the most well-known investor behavior biases. The second section introduces the emerging of rank effect, including discovery of the V-shape disposition, the development of theory of attention-grabbing trading and rank effect. The rest three sections demonstrate factors that affect V-shaped disposition effect, rank effect and other aspects of individual investor behavior, which are discussed in the empirical chapters in this thesis. Considering investors' heterogeneity and sophistication, the third section investigates the effects of investor characteristics on behavioral finance.

2.1. The Development of Disposition Effect

2.1.1 The Early Stage of Disposition Effect

Since the first discovery of the prospect theory (Kahneman and Tversky, 1979), the biases of the investor behavior have been widely discussed. Among these, disposition effect (the tendency to hold losers too long and sell winners too soon), first pioneered by Shefrin and Statman (1985), is the most well-known one.

Weber and Camerer (1998) present an experimental investigation of the disposition effect. As the main result of the experiment, disposition effect does exist. Contrary to Bayesian optimization, subjects tend to sell winners and keep losers. When the shares are automatically sold after each period, the disposition effect is greatly reduced, which cannot be explained by mean-reversion. Beside the experimental result, Odean (1998) is the first to provide empirical result of disposition effect. To test the disposition effect, the trading records from 1987 through 1993 for 10,000 accounts at a large discount brokerage house are taken. And the result demonstrates that investors realize their gains more readily than their losses as well as many investors engage in tax-motivated selling, especially in December[2]. However, when the data are controlled for rebalancing and for share price, the disposition effect is still observed. And

2 Investors are reluctant to sell for a loss but recognize the tax benefits of doing so. They realize loss stocks to reduce the capital gains tax losses. The end of the year is the deadline for realizing these losses.

the winning investments that investors choose to sell continue in subsequent months to outperform the losers they keep.

The method used in Odean's paper is the PGR (Proportion of Gains Realized) PLR (Proportion of Losses Realized) measurement. For each day that a sale takes place in a portfolio of two or more stocks, the author compares the selling price for each stock sold to its average purchase price to determine whether that stock is sold for a realized gain or a loss. Each stock that is in that portfolio at the beginning of that day, but not sold, is considered to be a paper gain or loss. If both its daily high and low are above its average purchase price it is counted as a paper gain; otherwise, a paper loss; if its average purchase price lies between the high and the low, neither a gain nor loss is counted. Then two ratios are calculated as:

$$\frac{Realized\ Gains}{Realized\ Gains + Paper\ Gains} = Proportion\ of\ Gains\ Realized\ (PGR) \quad (1)$$

$$\frac{Realized\ Losses}{Realized\ Losses + Paper\ Losses} = Proportion\ of\ Losses\ Realized\ (PLR) \quad (2)$$

The difference between PGR and PLR (PGR-PLR) is then calculated. From January to November, PGR-PLR is statistically positive and significant (t-statistics greater than 35), which shows the existence of disposition effect. In December, the PGR is smaller than PLR (t-statistics equals to 16), which means a reverse disposition effect and can be explained by the tax-motivated selling. The PGR and PLR measurement becomes a common method used in disposition analysis since then. However, this method can only test disposition effect based on aggregate investor groups. It cannot exam the heterogeneity between individuals. Feng and Seasholes (2005) develop a new model to explore this question, which will be explained in later section.

2.1.2 Disposition Effect in Different Market

After the evidence found by Odean (1988) in U.S. market, disposition effect is widely examined in other markets, for example, Israel (Shapira and Venezia, 2001), Finland (Grinblatt and Keloharju, 2001a), China (Feng and Seasholes, 2005), Taiwan (Barber, et al., 2009), Germany (Lukas, et al., 2017), etc.

Shapira and Venezia (2001) analyze the trading of 4,330 investors with accounts at an Israeli brokerage in 1994. About 60% of these accounts are professionally managed, while for the rest of the accounts, investors make independent decisions. They measure the duration of round-trip trades conditional on whether the stock was sold for a gain or loss. A tendency to sell winners and hold losers would, ceteris paribus, yield shorter holding periods for winners versus losers. Both professionally managed accounts and independent accounts exhibit the disposition effect (the holding periods for winners is roughly half that of losers), and the effect is stronger for independent accounts.

By applying a shareholding data from Finnish Central Securities Depository (FCSD) from December 1994 through December 1996, the study from Grinblatt and Keleharju (2001a) analyzes the extent to which past returns determine the propensity to buy and sell. It finds that foreign investors tend to be momentum investors. Domestic investors, particularly households, tend to be contrarians and perform the disposition effect. Since both momentum behavior and performance (reverse disposition effect) appear to be associated with the level of sophistication of the investor, the portfolios of foreign investors seem to outperform the portfolios of households.

They study how investment behavior relates to past returns by examining whether the buy ratio of past winning stocks exceeds the buy ratio of past losing stocks. More specifically, investment style on day t for an investor category is measured as the difference between the average of the buy ratios of the four stocks with past returns that are in the top quartile (of the 16 stocks) less the average of the buy ratios of the four stocks with past returns that rank in the lowest quartile.

Finnish household investors exhibit negative buy ratio differences, buying losers and selling winners, and perform the disposition effect. Meanwhile, foreign investors tend to be momentum investors over all horizons. Foreign investors, who follow momentum strategies, have positive average performance, as exhibited by the abnormally high proportion of positive buy ratio differences. Household investors, who follow disposition strategies, have negative average performance.

Chinese market, as the second large market today and the market from a developing country, is also worth noting. Feng and Seasholes (2005) study the existence of disposition effect and the relation between disposition and sophistication in Chinese market. They also introduced the survival analysis model into the research of disposition effect, so that they can investigate behavior at the individual-investor level rather than at the group level as in earlier papers. In the result, sophisticated investors are 67% less prone to the disposition effect than the average investor in the sample. Trading experience on its own attenuates up to 72% of the disposition effect.

Their study contributes to the development of disposition theory in two main points. The first one is the discovery of the new methodology, survival analysis used in the paper. The second one is the result they found in Chinese market. In the paper, they regress a holding indicator at the stock position level (1 = Sell; 0 = Hold) on independent variables. The coefficient on the trading loss indicator indicates whether investors are reluctant to sell at a loss. This method allows researchers to test the heterogeneity in regarding of investor characteristic variables in individual level[3]. For each day t after a stock is bought, the authors calculate the conditional probability of the stock being sold. This conditional probability on any date t is called the baseline "hazard rate". Similar to Logit regressions, survival analysis regresses a sell/hold indicator variable on the baseline hazard function and other independent variables. We can think of a coefficient's hazard ratio as reporting a change in the hazard rate when the independent variable changes from zero to one.

The database used in Feng and Seasholes (2005) is an account-level data from a national brokerage firm in the People's Republic of China ("PRC"). The data are comprised of 1,511 investors (accounts). The data period is from January 1999 to December 2000. Since my data period is 2007 to 2009, it significantly updates the findings in Chinese market. The average age of an investor in our sample is 34.71 years old. The sample contains 51.42% males and 48.58% females. In the sample, the hazard rate of a sale decreases by 0.3679 (where $0.6321-1.0000 = -0.3679$) if a stock is trading below its reference price. This means the

3 Feng and Seasholes (2005) introduce trading experience, trading rights, indicator of initial diversification, gender and age as variables to measure the heterogeneity of investors. Due to the data availability, I will test all of these variables except trading rights. I will also add trading frequency, location, nationality and occupation as variables. Since their data period is from 1999 to 2000 when Internet and online trading were not popular in China, in my data period, 2007 to 2009, trading rights is not a crucial variable.

investors in China are less likely to sell losers and perform disposition behavior. In the study, the proxy variables of sophistication are the number of trading rights and an indicator variable of initial portfolio diversification. They further define experience as number of positions taken by investor i up until date t. Sophistication and trading experience reduce the disposition effect. Together, sophistication and trading experience eliminate the reluctance of investors to realize losses. Men are 30% more likely to realize a loss than women. Investors age 25–35 in the PRC are 20% more likely to realize losses than investors over 55.

Sumway and Wu (2005) also find disposition effect in Chinese investor. Their paper acquires a transaction-level dataset from a large brokerage firm in the city of Shanghai. They collect all transactions (purchases or sales) that originate from the brokerage firm from the beginning of 2001 to March 12, 2004. The dataset contains almost 17 million trades placed by 3.8 million different accounts. By using the similar cox-survival analysis model, they find that a large majority of Chinese investors exhibit the disposition effect. The average disposition coefficient for investors is 5.91 with a significant t-statistic of 22.71, indicating that disposition effect is a costly behavioral bias. More disposition-prone investors tend to hold more diversified portfolios, trade less frequently and in smaller sizes than other investors. And disposition does indeed drive momentum.

As also the research based on Chinese and other Asian market, Chen, et al. (2007) analyze around 50,000 Chinese investors using data from a Chinese brokerage firm over the period 1998 to 2002. They suggest that Chinese investors are 67% more likely to sell a winner than a loser. Choe and Eom (2009) find a disposition effect for investors in Korean stock index futures; the effect is strongest for individual investors.

In Taiwan market, by using a complete trading history of all investors in the market from 1995 to 1999, Barber, et al. (2007) document that there is a strong disposition effect for individual investors, who are nearly four times as likely to sell a winner rather than a loser. Corporate investors and dealers also are disposed to selling winners (though the effect is much weaker than that observed for individuals), but neither Taiwan mutual funds nor foreign investors in Taiwan are disposed to selling winners. In their later research, Barber, et al. (2009) find out that the aggregate portfolio of individuals suffers an annual performance penalty of 3.8 percentage points and it is caused by the aggressive trades and disposition effect.

Meanwhile, foreigners earn nearly half of all institutional profits. The profits of foreigners represent an unambiguous wealth transfer from Taiwanese individual investors to foreigners.

In a recent research, Lukas, et al. (2017) document the disposition effect in Germany social trading platform. By using the transaction data from a social trading platform website in Germany, they find that, on average, each trader realizes 38.4 (median 8) gains and only 23.6 (median 5) losses and around 7.94% of trading appear the disposition effect. They also suggest that disposition effect decrease significantly when the trades become visible to public.

These studies find evidence of disposition effect in a wide range of countries, both developed and developing continents. Since all of their data is earlier than 2004 and my dataset is from 2007 to 2009, my research will significantly update the findings. I will also add more evidence on the investor behavior bias in emerging markets.

2.1.2 Disposition Effect from Institutional Investors

Compared to individual traders, professional investors are thought to be more sophisticated and make better trading decisions. However, there are evidence that institutional investors are not such rational and sophisticated as well. Besides the findings among individual traders, disposition effect is also common in professional managers.

Frazzini (2006) study the disposition effect and its relationship with news based on mutual fund managers in the U.S. during the period of 1980 to 2002. He tests whether the disposition effect induces "underreaction" to news, leading to return predictability and shows that post-announcement price drift is most severe whenever capital gains and the news event have the same sign. The magnitude of the drift depends on the capital gains (losses) experienced by the stockholders on the event date. An event-driven strategy based on this effect yields monthly alphas of over 200 basis points. The results confirm the intuition: managers of underperforming funds appear reluctant to close their losing positions. Conversely, successful managers realize losses at higher rates than gains. Frazzini (2006) uses the time series of net purchases by mutual fund managers and their cost basis in a particular stock to compute a weighted average reference price. This is a measure of the cost basis based on portfolio holdings, and the author analyzes the transmission of information when firm-specific information is released in the form of public news.

The database is from several sources. Stock returns and accounting data are obtained from the CRSP/COMPUSTAT merged database. Quotes and trades are obtained from the New York Stock Exchange Trades and Quotations (TAQ) database. Analysts' stock recommendations are taken from the Institutional Brokers Estimates System (I/B/E/S). And Mutual fund holdings are obtained from the Thomson Financial CDA/Spectrum Mutual Funds database. The data contain end-of-quarter stock holdings for about 29,000 mutual funds between January 1980 and December 2003. The stock price at the report date is used as a proxy for the buying or selling price.

As the results in the paper, the magnitude of the aggregate difference (PGR − PLR) is around 3%, which is smaller than the average 5% reported by Odean (1998) for retail investors, but still of the same order of magnitude. This means that mutual fund managers also suffer from disposition effect. When facing a capital loss, disposition investors are reluctant to realize the loss, thereby generating underreaction to negative news. As a result, post- event risk-adjusted returns can be achieved by using a sort on the capital gains overhang, suggesting that such a variable predicts the gradual market response to new information.

Locke and Mann (2005) research the disposition effect in futures market. Contrary to other studies, they find that disposition effect does not cause damage to profit. The paper investigates the nature of trading discipline and whether professional traders are able to avoid the costly irrational behaviors found in retail populations. They discuss the relation between discipline and future success using two measures of trading discipline: trading speed and exposure, determined by the magnitude of paper losses per contract on trades held for a long time. Using high-frequency transactions data, they study the trading behavior of professional futures traders on the Chicago Mercantile Exchange (CME), where trades are typically offset in a matter of minutes. The full-time traders in the sample hold onto losses significantly longer than gains, but they find no evidence of costs associated with this behavior.

The paper construct trade sequences for each trader for each trading day during the entire 1995. For each minute of the trading day and for each contract, the authors determine the quantity of contracts that traders buy and sell. For each trade, they calculate the average position and mark-to-market across those potential exit minutes to complement the simple count of potential exit opportunity minutes. They also calculate the maximum and minimum marking to market over the trade's history.

They suggest that disciplined traders will offset trades more rapidly. Traders offsetting losses more quickly are more likely to be successful in the future, but speed in closing gains is equally useful as a success predictor. Thus, aversion to realizing losses is not driving the results. The evidence is strong that these traders hold losing trades longer than gains. However, no evidence is available of a costly disposition effect among professional futures traders, but that a relative lack of discipline in realizing both gains and losses promptly is harmful to the probability of success.

2.1.3 Disposition Theory in Other Aspects

In recent researches, Frydman, *et al.* (2014) document a study on how the neural in our brain related to disposition effect. They conduct a study in which subjects trade stocks in an experimental market while they measure their brain activity using functional magnetic resonance imaging. They use the neural data to test a "realization utility" explanation for their investing behavior, such as disposition effect. They find that activity in two areas of the brain that are important for economic decision-making exhibit activity consistent with the predictions of realization utility. These results provide support for the realization utility model.

Chang, *et al.* (2016) build a test that investors avoid realizing losses because they dislike admitting that past purchases were mistakes, but delegation reverses this effect by allowing the investor to blame the manager instead. By using the trading data, they show that disposition effect applies only to non-delegated assets like individual stocks; delegated assets, like mutual funds, exhibit a robust reverse-disposition effect. While in an experiment, the paper finds that increasing investors' cognitive dissonance results in both a larger disposition effect in stocks and a larger reverse-disposition effect in funds. Additionally, increasing the salience of delegation increases the reverse-disposition effect.

In the first, the paper examines the extent to which real world trading data are consistent with cognitive dissonance and other explanations of the disposition effect. The individual trader data used are the same as in Barber and Odean (2000). The data come from a large discount brokerage and include 128,829 accounts with monthly position information, comprising 73,558 households (out of 78,000 initially sampled) from January 1991 to November 1996. By building a regression model with Sale as dependent variable and Gain as one of the independent variables in monthly bias, they find a significant reverse-disposition effect in funds, even within the set of investors who simultaneously hold both assets.

To provide direct evidence of cognitive dissonance as a driver of the disposition effect, they run an experiment on 520 undergraduate students over 12 weeks. The students participated in a stock and mutual fund trading game. The game started on January 23, 2012 and ended on April 16, 2012 (12 weeks' duration). Although the students in business school may have the knowledge of disposition effect before, for the results, on days with a sale, students are 14.1% less likely to sell a fund that is at a gain. Students who are told they can blame the fund manager displayed a significantly larger reverse-disposition effect, consistent with the cognitive dissonance hypothesis. It is worth noting that even in trading games without real money, investors are also more likely to appear disposition effect, which is consistent with empirical results. This paper also leaves a question to further study that how to proof the result in experiment in real data empirical studies.

2.2 The Emerging of Rank Effect

This section will introduce how disposition theory develops into rank effect theory and the establishment of rank effect. In traditional disposition theories, prospect theory is applied to explain the disposition effect, which expects a S-shape curve in the probability of selling as a function of profit. However, Ben-David and Hirshleifer (2012) find that the curve is actually V-shape and they challenge the traditional theory. They also document that gains or losses is not the only issue when analyze disposition effect. How much is the gains or losses is a question as well. Attention trading provides a possible reason for the V-shape disposition since stocks with large gains or losses catch more attention of investors. This theory also suggests some other factors that may influence the attention. Rank effect further develop these findings into a new investor trading bias effect that individuals are more likely to sell the extreme winning and extreme losing positions in their portfolio.

2.2.1 V-shape in Disposition Effect

Prospect theory is most commonly applied to explain the disposition effect. Barberis and Xiong (2009) model the trading behavior of an investor with prospect theory preferences. The paper finds that, in realized gain/loss model, prospect theory preferences may lead to disposition effect. They further document that investors gain utility from realizing gains and dub this behavior "realization utility." They show that, if gains and losses are evaluated when they are realized, a disposition effect obtains.

However, prospect theory is challenged by some other studies. Ben-David and Hirshleifer (2012) examine how investor preferences and beliefs affect trading in relation to past gains and losses. In the model of Meng (2010), owing to loss aversion, prospect theory predicts that investors have a greater probability of selling risky positions when the returns are close to zero. However, by using the residuals method, Ben-David and Hirshleifer (2012) find no evidence of a jump for short-term prior holding periods (1 to 20 days since purchase). The probability of selling as a function of profit is V-shaped, and investors are more likely to sell big losers than small ones. These findings provide no clear indication that realization preference explains trading.

They use data on retail investor trading as used by Strahilevitz, et al. (2011), which is similar to the data used by Odean (1998). The dataset includes stock transactions from 77,037 unique accounts over the period from January 1990 through December 1996. By fitting the residual analysis based on regressions with different polynomial function and control variables, they find that across the polynomial specifications and the prior holding period, the jump around the zero point is never significant at the 5% level. In summary, there is no clear indication of a jump, and therefore no clear indication that realization preference is a contributor to the disposition effect. They also suggest that the probability of selling has an asymmetric V-shape around the origin: in the loss region, the probability of selling increases with the magnitude of losses, while in the gain region, selling increases even more sharply with the magnitude of gains. And the V-shape is strongest in short-term samples. They also perform cross-sectional tests, suggesting that the V-shape may be driven by speculative trading in the expectation of profits. They conclude that there is, at present, no general evidence that individual investors in U.S. stocks have an inherent preference or “disposition” to realize winner stocks or a direct reluctance to realize loser stocks. Since prospect theory expect a S-shape curve and a jump at the zero point, it is important to understand that in this study the disposition effect does not provide grounds for proving prospect theory.

Meanwhile, in Finnish market, Kaustia (2010b) get similar result. By using the Finnish stock data and run a logit regression, he shows that prospect theory is unlikely to explain the disposition effect. Trading data show that the propensity to sell is approximately constant over a wide range of losses and increasing or constant over a wide range of gains. The primary data source for his study is the registry of shareholdings and daily trades from the Finnish Central Securities Depository (FCSD) from December 27, 1994 through May 26, 2000. It covered 97% of the total market capitalization. An otherwise similar data set, but covering a shorter time span, is used in Grinblatt and Keloharju (2000, 2001). After cleaning, there are total 3,871,863 observations. In the paper, he uses logit regressions to estimate the propensity to sell versus to hold a stock in different profit intervals and holding period. He finds that the propensity to sell generally increases with the amount of gain as well as loss. But the overall likelihood of a sale is higher if a gain is realized, which indicate an asymmetric V-shape. And the evidence is stronger when the holding period is short.

These findings in V-shape challenge the traditional theories in disposition effect. They arise the question that what is the actual curve of return and trading and how to explain that.

2.2.2 The Effect of Attention in Trading

Attention-grabbing trading is one of the explanations to the V-shape. Since stocks with large returns or loss may catch more attention of investors, they are more likely to being traded.

In the buying side, Barber and Odean (2008) document that individual investors are net buyers of attention-grabbing stocks, e.g., stocks in the news, stocks experiencing high abnormal trading volume, and stocks with extreme one-day returns. When facing the problem of choice and searching, individuals only focus on stocks that catch their attention and buy them. While each investor does not buy every single stock that grabs his attention, individual investors are more likely to buy attention-grabbing stocks than to sell them. Thus, preferences determine choices after attention has determined the choice set.

The data period is mostly from 1991 to 1999, similar to Odean (1998). The paper sorts stocks on the basis of abnormal trading volume by calculating for each stock on each trading day the ratio of the stock's trading volume that day to its average trading volume over the previous one year (i.e., 252 trading days). For each partition and investor group combination, the authors construct a time series of daily buy-sell imbalances. The inferences are based on the mean and standard deviation of the time series. And how the buy-sell imbalance of a particular investor group changes with volume is the empirical question.

In the results of the study, individual investors display a great amount of attention-driven buying with the 30% buy-sell imbalance, which means that individuals are 30% more likely to buy attention-grabbing stocks than sell them. Institutional investors exhibit the opposite tendency of the individual investors: in general, their buy-sell imbalances are greater on low-attention stocks than high-attention stocks. The paper further argues that many individual investors solve this search problem by considering for purchase only those stocks that have recently caught their attention. Professional investors are less prone to indulge in attention-driven purchases.

Attention is measured not only by stocks in the news, stocks experiencing high abnormal trading volume, and stocks with extreme one-day returns, Jacobs and Hillert (2016) suggest that there is an advantage to those positioned in the beginning of an alphabetical listing. They find that US stocks that appear near the top of an alphabetical listing have about 5–15% higher trading activity and liquidity than stocks that appear toward the bottom. Meanwhile, by using data from 1985 to 2012 also in U.S. market, Itzkowitz, et al. (2016) find similar result. They document that investors are more likely to buy and sell stocks with early alphabet names. Consistent with this view, they find that early alphabet stocks are traded more frequently than later alphabet stocks and that alphabeticity also affects firm value. They also document how these effects have changed over time.

Using Google search frequency as a measure of investor attention, Da, et al. (2010) analyze whether investor attention can cause price pressure effects. Using data from 2004 to 2008, they document that increases in search frequency predict higher returns in the ensuing two weeks and an eventual reversal within the year.

2.2.3 Rank Effect

While early studies focus more on how the buying behavior affect by attention, Hartzmark (2015) documents a new stylized fact about how investors trade assets: individuals are more likely to sell the extreme winning and extreme losing positions in their portfolio ("the rank effect"). Using data from a large retail brokerage, the paper shows that on a day an investor sells a position in their portfolio, the investor has a 31% chance of selling the stock with the highest return in the portfolio and a 26% chance of selling the stock with the lowest return, after controlling for a number of factors discussed below. This effect is not driven by firm-specific information, holding period or the level of returns itself, but is associated with the salience of extreme portfolio positions. It is also worth noting that most early researches suffer from a stock-by-stock bias, analysis the trading behavior of stocks one by one instead of considering the comparison of the stocks within one's portfolio. The rank effect indicates that trades in a given stock depend on how the stock compares to other positions in an investor's portfolio. And this comparison is supported by a large literature in behavioral economics and psychology on the joint evaluation of decisions (Bazerman, *et al.* 1992; Hsee 1996; Hsee *et al.* 1999; Kahneman 2003; List 2002).

The analysis of Hartzmark (2015) is based on two main datasets. The first contains data on individual investors trading on their personal accounts. The analysis is also conducted on mutual funds. The individual investor data are the same as those used by Barber and Odean (2000), from 1991 to 1996 in US stock market. As the paper examines portfolio rank, investors must hold at least five stocks to be included in the analysis. This excludes about 19% of observations. The data include 10,619 unique accounts, 94,671 sell day by account observations with a sale, and 1,051,160 positions held on those days. The average portfolio is comprised of 11.1 stocks. The mutual fund analysis combines holdings data from Thompson-Reuters and fund price and volume information from CRSP, and stock return information from CRSP with a period of 1990 to 2010, quarterly basis.

Positions in one portfolio are sorted by the unit share return and divided into five ranks (best, 2nd best, middle, 2nd worst, worst). A best rank position is the position that have the highest return among the portfolio. Other four ranks have similar definition. A series of dummy variables are built to measure the rank of the position. The paper then fits Logit regression model used by Ben-David and Hirshleifer (2012) after adding the rank dummy variable. The best-ranked stock is 15.7% more likely to be sold, and the worst ranked stock (Worst) is 10.7% more likely to be sold, both with large t-statistics. After including the two dummy variables for rank, the Loss*Return and Gain*Return coefficients, which indicate the disposition effect, are insignificant and the Gain dummy coefficient decreases. This means that rank effect is at least as strong as disposition effect.

The salience of extreme outcomes offers one possible explanation for the rank effect for individual investors. Empirically, both rank extremeness (rank-dependent utility models predict that extreme returns receive the most attention (Tversky and Kahneman, 1992) and outlier extremeness (predict that a position is more salient when it is best- or worst-ranked (Hauser, 2014) are significant aspects of the rank effect. This is consistent with the theory of consideration sets, suggesting that the choice of paying attention to a particular stock is an important aspect of the trading decision. For fund managers, funds appear to exhibit the effect at least in part because of both window dressing and salience.

The author also argues that less sophisticated investors, when measured by a higher disposition effect or lower self-reported experience, display a rank effect that is skewed toward selling best-ranked positions, whereas more sophisticated investors display a more symmetric effect. Due to momentum, the rank effect displayed by more sophisticated investors is on average more profitable as a strategy.

In that paper, Hartzmark (2015) only analyzes the relation between investor sophistication and rank effect with limited sophistication measurement variables. Adding more demographic variables to find out which kinds of investors are more likely to appear rank effect is a question leaves to further study. And also, as mention by Barber and Odean (2008), extreme return is not the only thing that grabble attention. Return may not be the only thing investors consider when they compare and rank their stocks in their portfolio. Testing the rank effect with more measurement variables, such as extreme trading volume or volatility, is a job to be done.

After these studies, An (2016) finds asset pricing value based on V-shape disposition effect and rank effect. The study investigates the asset pricing implications of a newly documented effect, characterized by investors being more likely to sell a security when the magnitude of their gains or losses on it increases. By using 2.1 million stock-month combinations in US market during the period of 1963 to 2013, she finds that stocks with both large unrealized gains and large unrealized losses outperform others in the following month. This supports the conjecture that these stocks experience higher selling pressure, leading to lower current prices and higher future returns. Overall, the study adds value to the V-shape disposition effect and rank effect by the affect to asset pricing.

2.3 Investor Heterogeneity in Behavioral Finance

Why investors trade and why they trade differently is one of the key questions in behavioral finance. The difference demographic may influence investor behavior, for instance their age, gender, education level, occupation, nationality, location, culture background, etc. Investor sophistication, experience, past return and life cycle trading are also factors. As a large number of studies have discussed the influence of these factors in disposition effect and other investor behavior biases, I will try to exam the influence in rank effect in my research.

Goetzmann and Massa (2002) and Dhar and Kumar (2002), finds significant heterogeneity in investor beliefs and trading styles. They show that mean value of the aggregate group is not the whole story. Using a unique dataset from Finnish stock market, Grinblatt and Keloharju (2001) employ Logit regressions to identify the determinants of buying and selling activity over a period of 1994 to 1996. They find that tax-loss motivation, past returns and historical price patterns, such as being at a monthly high or low, affect trading. There is also modest evidence that life cycle trading[4] plays a role in the pattern of buys and sells. List (2003) uses data from sports card trading market to show that trading frequency is a key of investor heterogeneity. He also documents that the endowment effect and disposition effect is likely to be weaker for individuals who trade more. Krueger and Rouse (1998) and Bailey, et al. (2001)

4 The life-cycle hypothesis suggests that rational investors should smooth their consumption by appropriately investing and borrowing based on expectations about lifetime income.

find a link between the educational background and better decisions and performance. Hallahan, et al. (2004) investigates the impact of gender and age on financial risk assessment.

Dhar and Zhu (2006) try to identify differences in the disposition bias across individuals and explain this in terms of underlying investor characteristics. They argue that differences in investor literacy about financial markets and trading frequency are responsible in part for the variation in individual disposition effect. Using demographic and socioeconomic variables as proxies for investor literacy, they find empirical evidence that wealthier individuals and individuals employed in professional occupations exhibit a lower disposition effect. Consistent with experimental economics, trading frequency also tends to reduce the disposition effect. The data used in the research contains trading records of more than 50,000 individual investors from a large discount brokerage firm between 1991 and 1996. The median age was 48 and the median annual house- hold income was $50,000. Twenty-one percent of the investors are female and the remaining 79% are male. Using a Logit regression with PGR-PLR (the extent of disposition effect) as target variable and different investor characteristics as independent variable, they suggest that individuals who are low income and work in nonprofessional occupations show the highest disposition effect among all investors and trading frequency helps reduce the effect.

Cheng, et al. (2013) test on how gender and age, internal characteristics of retail futures traders and the security being traded and bull– bear market conditions are related to the disposition effect by separately tracking their trade-by-trade transaction histories over a period of close to six years on the Taiwan Futures Exchange. They show that women and mature traders, compared with their male and younger counterparts, exhibit a stronger disposition effect. The effect is also stronger among traders who trade financial-sector futures contracts than among those who trade electronic-sector futures contracts. The paper further demonstrates that a bear market sees a stronger disposition effect.

Also comparing with the result by Feng and Seasholes (2005) in Chinese market that female and old investors are more likely to exhibit disposition effect, it is interesting to find that both Dhar and Zhu (2006) and Feng and Seasholes (2005) document that man are more likely to perform as disposition effect, but Cheng, et al. (2013) argue that woman are more likely. Feng and Seasholes agree with Cheng, Lee and Lin that mature investors are disposition investors while Dhar and Zhu state the oppsite. Cheng, et al. argue for themselves and suggest that Dhar and Zhu's dataset is dominated by men observations. Since the Taiwan sample includes all individual future traders in the market, they represent very well the whole spectrum of the retail investors, rendering the results more generalizable.

Although the relation between gender and disposition effect is still unclear, it is probably sure that the structure of investors in Chinese and Taiwan, emerging market, and those in U.S., developed country market, is different. Since the late establishment of Chines and Taiwan market, it is reasonable that old investors in these markets do not have significantly experience advantage than young investors, thus should not be considered to be more sophistication. In my research, I will further analyze the relation between these characteristics and rank effect.

Chapter Three: Impact of Magnitude of Return on Selling Decision: Evidence from China

3.1 Introduction

How investors choose when they sell positions? Disposition effect indicate that investor prefer to sell gains than losses (Shefrin and Statman, 1985; Odean, 1998; Grinblatt and Keloharju, 2001; Feng and Seasholes, 2005). Ben-David and Hirshleifer (2012) further examine the relationship between the magnitude of gain or loss on the probability of selling, and find investors are more likely to sell a position as it becomes a larger gain or a larger loss (V-shaped). The effect is more significant in short-term positions. Meanwhile, using data from Finnish market, Kaustia (2010) observes similar result from gain positions but find the selling probability is approximately constant in the realm of losses. Bian et al (2018) document the selling probability increase with large gain intervals in a hazard model from Chinses market without the control of holding period, which is crucial in result of Ben-David and Hirshleifer (2012). The inconsistency of results in these papers raises the question that what is relationship between the magnitude of gain and loss and the probability of selling the position, and how holding period of the position affect this relationship.

In this chapter, I first try to analyze this question with a database from Chinese stock market. I find that individual investors are more willing to sell large gains, but the scale of loss does not significantly affect the probability of selling. This result is similar to Kaustia (2010) and Bian et al (2018). But it is worth noting that these results are shown under the models controlling holding period. When I remove holding period effect, the results entirely change. I therefore emphasize the importance of impact of holding period. Based on the separated holding period sub-samples, I indicate the V-shaped disposition effect appears only on short-term positions. This result is consistent with Ben-David and Hirshleifer (2012) on their finding in the US market. In regard to the medium- and long-term holding period, positions with extreme gains and losses are less likely to be sold, while investors prefer to sell the positions with returns close to zero. Therefore, I draw the conclusion that in general, individual investors have the preference of selling large gains, but the scale of loss does not significantly affect the probability of selling. However, this result is different in positions whit different holding period. For consider selling short-term positions, investors are more willing to sell both large gain and large loss. For medium- and long-term, investors choose to sell positions with returns close to zero. The trading strategy and loss tolerate of short-, medium- and long-term positions is different among individual investors.

Since the data covers financial crisis (2007-2009), I further discuss to what extent do the magnitude of gains and losses affect individual investors selling during the booming, crushing and recovering (namely before, during, and after) period of financial crisis. Under extreme market uncertainty, the behavior of individual investor is more worth noting. To the best of my knowledge, this is the first study to discuss the impact of return on selling during financial crisis. The results show that Chinese investors consistently prefer to sell large gains in all scenarios, however, they sell large losses under booming market but no significant response to the magnitude of loss in both crushing and recovering period. When individuals lose confidence to the market during and after financial crisis, they are more patient to their large loss positions and more willing to keep them.

Dhar and Zhu (2006) and Feng and Seasholes (2005) find investor sophistication can moderate disposition effect. I further investigate to what extent do investor characteristics and sophistication affect the selling behavior based on magnitude of return. I find that investor with more experience and more trading frequency can moderate V-shaped disposition effect. Since An (2016) discovers v-shaped disposition effect cause damage to investor profit, these investors are more sophisticated to deal with V-shaped disposition bias. Meanwhile, although age is generally seen as an indicator of investor sophistication, senior Chinese investors do not perform differently on selling large gain and loss with juniors. Since Chinese stock market developed late in the 1990s, senior people may be lack of professional stock knowledge during their education while juniors are more educated with the development of the stock market. Therefore, senior investors are not sophisticated in China at least in some extent.

The data used in this chapter is very large and unique, which is collected from a large brokerage firm in China. It contains more than 3 million accounts and 2 billion daily stock dealing records over the period of January 2007 to May 2009. Due to the consideration on the cost of computation, I use a randomly selected sample of 100,000 investors in this chapter. China is an ideal laboratory to study behavioral finance among investors. Due to its successful economic transition in the last three decades, Chinese market has become the world's second largest stock market in value since 2014 and has been added to MSCI Emerging Markets Index since 2017, indicating its increasing importance in global economy. Meanwhile, Chinese stock market has generally been viewed as under-developed market with high degree of asymmetric information, due among other things to its unsound financial system and its weak shareholders' protection, as well as its weak corporate governance system. The time period of the dataset in this research is from 2007 to 2009, which cover the financial crisis period. In China, suffering from the world financial crisis, there was also a huge bubble in 2007 and experiencing significant stock price falling in 2008 then gradually recovering in 2009.

The rest of this chapter is organized as follows. Section 3.2 reviews the relevant literatures. Section 3.3 introduces the dataset, background of Chinese stock market during this period and methodology. Section 3.4 discusses the main empirical results, while robustness tests are presented in Section 3.5. Section 3.6 concludes.

3.2 Literature Review

Among the investment biases in behavioral fiannce, disposition effect (the tendency to hold losers too long and sell winners too soon), pioneered by Shefrin and Statman (1985), is the most well-known one. Odean (1998) is the first to provide empirical result of disposition effect. The result demonstrates that investors realize their gains more readily than their losses. These investors demonstrate a strong preference for realizing winners rather than losers. After his findings in US market, disposition effect is widely examined in many countries and areas, for instance, Israel (Shapira and Venezia, 2001), Finland (Grinblatt and Keloharju, 2001), China (Feng and Seasholes, 2005; Sumway and Wu, 2005; Bian, et al., 2018), Taiwan (Barber, et al., 2009), Korea (Choe and Eom, 2009), etc. These papers show that investor's behavior bias occurs in a wide range of markets. Prospect theory is most commonly used to explain the disposition effect. Barberis and Xiong (2009) model the trading behavior of an investor with prospect theory preferences and show that, if gains and losses are evaluated when they are realized, a disposition effect obtains. Therefore, a S-shaped curve in the probability of selling as a function of profit is expected. However, Ben-David and Hirshleifer (2012) find that the curve is actually V-shaped in the short-term positions. They also document that gains or losses is not the only issue when analyze disposition effect. How much is the gains or losses is a question as well. Meanwhile, using the data from Finnish market, Kaustia (2010) supports this result on the gain side, while he finds the probability of selling on the loss side is constant. Since Ben-David and Hirshleifer (2012) suggest that V-shape appears in positions with short holding period and it perform differently in long-term positions, the impact of holding period on V-shaped disposition is also worth noting. Additionally, attention trading (Barber and

Odean, 2008) provides a possible reason for the V-shaped disposition since stocks with large gains or losses catch more attention of investors. An (2016) also demonstrates that V-shaped disposition trading cause damage to individual's profit.

(*Insert Figure 3.1 here*)

Bian, et al (2018) introduce a similar result to Kaustia (2010) in Chinses market based on a hazard model. However, the analysis is lack of control of holding period and other control variables. Since the result from short-term and long-term position is different (Ben-David and Hirshleifer, 2012), holding period could be crucial in the analysis. In this chapter, I will test the relation between scale of return and probability of selling under control of holding period and other control variables in China. I further analyze this question in subsample of different holding period positions. I then investigate selling decision under different market background. The booming, crushing and recovering period of financial crisis provide us a way to research the changes of the investor trading decision when they face large profits and losses along with risks. Finally, I also discuss investor heterogeneity on selling large gain and loss. I will discover how investor sophistication influence the selling decision on positions with different magnitude of return.

3.3 Background, Dataset and Methodology

3.3.1 Chinese Stock Market in 2007-2009

The year of 2008 saw a sequence of adverse financial news in the world and triggered the US credit crunch and market crisis. And it soon became the worldwide financial crisis. This poor external financial environment should have a great impact on Chinese stock markets. However, although there was indeed an extreme volatility of stock prices that signified a market bubble appearing and bursting in Chinese market, the story began at the start of 2007 before the financial crisis.

Due to the Split-Share Structure Reform[5] in China, the entire year of 2007 is a booming year of Chinese stock market. The biggest bull market arrives at the beginning of 2007. The Shanghai Composite Index surged over 3,500 points from 2715.72 at the start of the year to 6,124 on October 16, which reached the peak, with the rise of 140%. And then it plunged all the way, felling back to roughly 2000 points in November 2008, with the loss near to 70%. On 4th November 2008, it got the lowest point, which is 1706. Then the index got back to steady growth till the end of 2009. In sharp contrast, during the same time period, the Chinese real economy grew at average more than 10% per year.

Chinese stock market provides an interesting and unique research environment in terms of the market emotion and investor behavior. At the start of the bubble, with the extreme bullish market, both domestic and foreign investors were enticed to buy whatever shares were on offer without carefully analyzing the real performance and growth potential of the listed firms. The whole market was under the emotion of over-confidence and over-optimistic. Then when the bubble burst, all investors were facing extraordinary loss and risk. The overall emotion turns to fear and lack of confidence.

I also introduce some criteria related to my research in Chinese stock market during 2007 to 2009. In that time, one investor can only open one trading account in agent. Therefore, data for one investor is the entire trading behavior of this investor in the market which helps us to get a more comprehensive understanding of the investor. There is a limit up and limit down restriction, in the magnitude of 10% as well. In Chinese market, the short sale constraint existed until 2010. There are no short-selling records in the data.

3.3.2 Sample construction

This chapter is based on a very large database collected from a large nationwide brokerage firm in China, with more than 3 million accounts and 2 billion daily dealing records over the period of January 2007 to May 2009[6]. Due to the computational capacity limitations, I use a random sample of 100,000 investors and more than 56 million records sub-data to build the model. The

5 Liao, *et al* (2014) and Lehkonen (2010) introduce the reform in detail.

6 Although there are three missing months data, which are April 2007, May 2007 and March 2008, the sample size is large enough to conduct the research.

dataset is formed with 4 sub-datasets that are customer file, account file, stock file and transaction file. Customer file contains the information of each customer. Account file contains balance information of customer's account on daily basis. Stock file contains information of each stocks held by each customer on daily basis. Transaction file contains each deal's information. Customer ID is used to merge all files. I also get the stock price and volatility information from CSMAR (China Stock Market & Accounting Research Database).

Each row in the stock file indicates the holding record of one investor for one stock at the end of one trading day and it composes the main data table. The transaction file provides us the trading amount and price of each trades. There is also a column shows "selling" when the transaction record is a sell. The customer file shows the gender, account open date and birthday. All investors are individual investors and there is no foreign investor. Since all data is based on the ending data of each trading day, the trading sequence of multiple trades of one investor in one day cannot be observed. Day traders are no included as well.

After the data cleaning processes, there are 4,065,596 records remain. I show more details in Appendix 1. I then calculate the return as follows. The current price is the stock price at the end of a trading day when investor keeps this stock or sells part of his holding shares of that stock. And it is the stock price at the last trade when investor liquidated. Since the stock price in one day does not change too much, this setting is reasonable in the data. The cost price is calculated as the share weighted average buying price for multiple buying behavior for one stock. The return is current price minus cost price.

3.3.3 Methodology

To test the impact of on selling by the magnitude of the gain and loss in Chinese market during financial crisis, I follow the model used in Ben-David and Hirshleifer (2012) and Hartzmark (2015):

$$Sell = Constant + a_1(Gain) + a_2(Gain * Return) + a_3(Loss * Return) + a_4(Control\ Variables) + e$$

The model is on day-investor-stock level. Each observation is a position that one investor holds one stock in one day. The model is fitted as a Logit model by maximum likelihood. The dependent variable is a dummy variable, equals to 1 if the stock is sold that day by that investor and 0 otherwise. Both partial selling and liquidation are involved. Return is the unit share return of position which is calculated based on the buying price (trading cost involved and is weighted average price by shares in case of multiple purchase) and the current price of that stock at that day. Gain is a dummy variable that takes the value of 1 if the unit share return of the position is positive and 0 otherwise. Loss is the opposite of Gain. Including the interaction terms of Gain (Loss) and Return allows us to analyze the relationship between the probability of selling and the magnitude of gain and loss separately.

For control variables, I also follow the choice of these two papers. The effect due to holding period and volatility are controlled for. To control for the days a position is held from purchase to sell, the square root of the holding days (sqrt_holding_period) and interaction terms with gain dummy and return (sqrt_holding_period*return*gain) and loss dummy by return (sqrt_holding_period*return*loss) are included. To control for the stock variance, I also calculate the return variance of a stock over the last year (the variance of stock price over preceding 250 trading days, if there are at least 50 non-missing records). I include the interaction term Variance*gain and Variance*loss in the model.

In addition to the control variables choice in Ben-David and Hirshleifer (2012) and Hartzmark (2015), I further add investor characteristic control variables to control for the investor specific influence. Gender is a dummy variable that takes value of 1 for female and 0 for male. Sqrt_age is the square root of the investor's age[7]. To control the experience of investor, I introduce a dummy variable New_investor, which equals to one if the investor opened account in this brokerage after the start of the data period and zero otherwise. It is worth noting that at that time, in Chinese market, one individual can only open one account in the whole market. This makes experience more powerful. Sqrt_tradetimes is the square root of times of trading an investor made in the data period. It can indicate the activation of an investor in some degree. Portfolio_size is the number of stocks in one's portfolio that day.

7 I use the date difference between investor's birthday and May 31st 2009, which is the last day of the dataset.

Since the data cover the financial crisis in China, I further introduce two dummy variables to control the time and market condition. I divide the data period, Jan 2007 to May 2009, into three parts, from Jan 1st 2007 to Oct 16th 2007 as bull market, from Oct 17th 2007 to Nov 4th 2008 as bear market, from Nov 5th 2008 to May 31st 2009 as steady growth market. I define the three sub-time period by the value of Shanghai Composite Index, which has been discussed in detail in the early part of this chapter (section 3.3.1). I introduce dummy variable Bull_mkt, equals to 1 if the position is in bull market period, and 0 otherwise; Bear_mkt, equals to 1 if the position is in bear market period, and 0 otherwise. I set the steady period as benchmark.

(*Insert Table 3.1 here*)

In Table 3.1, I present the summary statistics of data I use in the model. After all cleaning, there are total 4,065,596 records (positions). For dummy variable (binomial variable), I present the number of 1. For dependent variable Sold, there are 1,586,158 positions that are sold in the end of the day. Since I only include positions that at least one stock in the portfolio is sold in that day. This number is reasonable. Since the data cover financial crisis, there are more loss positions than gains in the model. The number of gain position is 1,576,700. The number of positions held by new investors and old investors are balanced. There are 2,215,459 positions held by new investors. The data is also balance in gender. For numerical variables, due to financial crisis, the average return is negative, which is -1.1633. And the mean of stock variance is 0.8439. On average, the square root of days a position is held is 5.8116 and the portfolio size is 6.0896. The average sqrt_age is 7.0219. The average sqrt_tradetimes is 27.9682.

3.4 Main Result

3.4.1 Empirical Results of the Impact of Magnitude of Gain and Loss

I calculate the marginal effect of a Logit regression model to get empirical results of the impact of magnitude of gain and loss (V-shape disposition) in Chinese market. Since the observations are related to each other, I apply the clustered standard error of investor and date instead of the simple standard error.

(*Insert Table 3.2 here*)

Table 3.2 presents the marginal effect of a logit regression model to examine the impact of disposition effect and magnitude of gain and loss on selling. In column 1, a position is 12.73% more likely to be sold if it is a gain with a very large t-statistics value. It confirms the disposition effect in Chinses stock market.

In column 3, I show the result of the impact of magnitude of gain and loss with all controls. Since the marginal effect of Gain*Return is in positive sign and significant, the increase of magnitude of gain leads to an increase probability of selling in Chinses market, which is similar to the US market (Ben- David and Hirshleifer, 2012). The marginal effect of Loss*Return is negative and insignificant. Since all of the Loss*Return terms should be nonpositive, this means that in loss case, the probability of selling is insignificantly increasing when the magnitude of loss is large. The probability of selling a loss position is relatively constant. It is different with the result of Ben-David and Hirshleifer (2012) in US market. The V-shape disposition is only suitable on the gain side in Chinese market from 2007 to 2009.

However, in column 2, when not applying holding period control variable set, I surprisingly find out the sign of Gain*Return and Loss*Return are opposite. This indicate that investors are more likely to sell positions close to zero. Therefore, holding period is crucial to this model. I suggest to using these controls when researching selling decision. To further investigate this, I will discuss this in different subsample of holding periods in the next session.

This result is similar to Kaustia (2010)'s result from Finish market that the probability of selling is approximately constant over a wide range of losses and increasing over a wide range of gains. Furthermore, I add the sources of the disposition effect and test whether it reflects realization preference in the model. When comparing with the result in Bian, et al. (2018), I highlight the importance of holding period. By adding these variables, the result is entirely different.

To consider the performance of investors' characteristics control variables, all these variables except New_investor is significant. This indicate that the investors' heterogeneity plays an important role in individuals' decision of selling. These variables should be included in this model. Furthermore, since the dataset is large, I do not suffer from a luck of degree of freedom by adding these variables. Adding them benefits to the robustness level of the results.

3.4.2 The Impact of Holding Period on Selling Performance

Ben-David and Hirshleifer (2012) show that V-shape disposition is closely related to holding period. When the position is bought within 20 days, the V-shape disposition is very strong. While when the holding period comes to 21-250 days, V-shape disposition is still significant but much weaker than the short-term positions. And if holding days is larger than 250, the V-shape disposition is not significant. Following their model, I also build the model based on the sub-sample of holding period. I split the data into three sub-data: holding period from 1 to 20 days, from 21 to 250 days and more than 250 days. I fit the same model in table 3 on these holding periods separately to analyze the impact of magnitude of gain and loss on selling in Chinese market for different holding period positions. Due to multicollinearity, I do not include holding period control variables this time.

(*Insert table 3.3 here*)

Table 3.3 presents the V-shape disposition in different holding period. In all three columns, Gain is positive and significant, which indicates a strong disposition effect in all positions. When the position is a short-term position (holding days from 1 to 20), in column 1, individuals are 1.57% more likely to sell a gain stock if the return of stock increase one standard deviation (with t statistics 6.34). And they are 0.35% more likely to sell a loss when the magnitude of loss increases by one standard deviation (with t statistics -2.38). This shows a significant V-shape disposition in the short-term positions. While in column 2, for midrange positions that has its holding period 21 to 250 days, surprisingly, Gain*return is negative and significant, and Loss*return is positive and significant. This result is totally the opposite of the result from short-term positions. It appears a reverse V-shape disposition effect. And for long-term positions, the result is the same to it is for midrange positions with also a reverse V-shape disposition. Therefore, I find out that in Chines stock market, from 2007 to 2009, while for all positions, the selling probability is increasing for large gains and constant for losses (appears V-shape only on gain side), the relationship between selling and magnitude of gain or loss is closely related to the holding period. V-shape disposition only appears on short-term positions while midrange and long-term positions give us a revers V-shape.

This result is constant with Ben-David and Hirshleifer (2012), where they find V-shaped disposition on short holding period positions. I further indicate a revers V-shaped disposition on medium and long holding period positions.

This result supports that probability of selling is linked to holding period. Individual investors have different trading strategies and loss tolerate among positions with different holding period. For short holding period positions, the selling behaviour is highly induced by speculative motivation. Since the substantial gains and losses could catch attention of individual investors, the limited attention theory in capital markets can probably explain the V-shaped result (Seasholes and Wu 2007; Barber and Odean 2008; Ben-Davaid and Hirshleifer 2012). For midrange and long-term positions, the behaviour of individual investors could probably be more rational. One of the reasons to explain the preference of selling close-to-zero long-term positions is that investors are more likely to sell a reverse performing position. If a position was a gain (loss) over a long period, but it is a loss (gain) now, the position is more likely to be sold. Zero could be a common reference point of long-term positions for individual investors. This should be a good question for further studies.

3.4.3 The Impact of Magnitude of Gain and Loss in Different Market Condition

Since the data covers the financial crisis, analyzing how individual investors performance under financial crisis and extreme market condition adds more value to the data and research.

Chinese market did not suffer seriously from the worldwide financial crisis in 2008. However, there is an extreme volatility of stock prices that signified a market bubble appearing and bursting in Chinese market during 2007 and 2008. The Shanghai Composite Index was around 2,700 at the beginning of 2007. It surged and reached the peak at 6,124 on 16th October 2007, which is the highest point in 2007. And then it went all the down to 1,706 on 4th November 2008, which is the lowest point in 2008. After that the Shanghai Composite Index enter a steady growth period in 2009. More details are introduced in Chapter 1.2. Based on this I split the dataset into 3 subsets: 1st January 2007 to 16th October 2007, called bull market, 17th October 2007 to 4th November 2008, called bear market, and 5th November 2008 to 31st May 2009 (the end of the data period), called steady market. I fit the same model in Table 3.3 in these time periods separately to test the impact of magnitude of gain and loss on selling in Chinese market before, under and over financial crisis.

(*Insert Table 3.4 here*)

In Table 3.4, since the observations of these time periods are on the same level the data is balance to this split. In column 1, under bull market, Chinese individual investors are 11.93% (with t-statistics 14.1987) more likely to sell a gain position, which shows a significant disposition effect. Gain*Return has a significant and positive result. This indicates that if the position is a gain, investors are more likely to sell it if it is a large gain. This is consistent to the all period result in Table 3.3. On the other hand, the relationship between magnitude of loss and the probability of selling is negative since and significant. This shows a strong V-shape disposition on bull market which is different from result from all period. What is unexpected is that the bear market condition shares the same result with bull market condition on gain side. In column 2, the power of disposition effect is 11.31%. In the gain part, the impact of return on selling is also positive and significant. And in the loss part, the marginal effect of Loss*return is insignificant which shows a constant selling probability on the range of loss. This result is the same as it is on all period. In column 3, the result of disposition effect and magnitude of gain in steady growth period is also the same. But it is slightly different when it comes to a loss. Investors are slightly more willing to sell a stock with a small loss (only 10% significant level on t test). This tendency is different with all other results.

In all of the three periods and market conditions, Chinese investors all perform strong disposition effect and a strong willing to sell positions with a large magnitude of gain. When facing a loss, the performance is slightly different. In bull market, individuals are more likely to sell large loss and follow V-shape disposition effect. In bear market, the probability of selling is constant with the magnitude of loss. And in steady growth market, they are more likely to sell loss positions that are close to zero. As a conclusion, although on loss side the behaviors of investors are slightly different among all three market conditions, on gain side their behaviors are very similar.

In the paper from Hoffmann, et al. (2013) and Gerrans, et al. (2015), both of them state that although individual investors change their expectation of return and risk tolerance during financial crisis, their trading behavior do not change significantly. The result support their argument in Chinese market when the position is a gain but have some difference when the position is a loss. Chinese investors follow disposition effect in all three periods as well.

This result may also be explained by the macroeconomic situation in China from 2007 to 2009. The GDP of China grew on average 10% during this period. This can lead Chinese individual investors to have confidence on their local companies and stocks even if their performance on the stock market is not good.

3.4.4 Investor Heterogeneity and Impact of Magnitude of Gain and Loss

Several papers discover influence of investor characteristics on disposition effect (List (2003), Dhar and Zhu (2006) and Feng and Seasholes (2005)). However, as fat as I know, there is no paper discuss investor heterogeneity and the impact of magnitude of gain and loss. In this section, I examine the relation between investor characteristics and the impact of magnitude of gain and loss on selling. I do this by introducing interaction term of return and characteristics term into the main model. Characteristic variables I use here are age, gender, trading experience and trading frequency. I measure age as both continuous variable (square root age) and categorical variable (young, middle age and old). Young, middle age and old are investors are separated by age under 35, between 35 and 55 and over 55. Gender is a dummy variable equals to 1 when the investor is a female. Experience is a dummy variable which takes number 1 when the investor opens account during the data period and 0 otherwise. In China that time, one investor can only open one account in the entire of investment life. Therefore, my measurement of trading experience appropriate. Trading frequency variable is the square root of trade times of one investor in data period.

(*Insert table 3.5 here*)

Table 3.5 presents result of investor heterogeneity and impact of magnitude of gain and loss on selling decision. The difference between column 1 and column 2 is in column 1 I use continuous age variable and in column 2 I apply age group. In both column, age is not significant which means age has no influence on selling large or small gain and loss. The interaction term of gender and gain is positive and significant indicates female is more willing to sell a large gain. However, there is no significant difference among male and female on selling large loss. On gain side, investors with both more experience and trades more are more probably to sell a small gain. Meanwhile, on loss side, these investors are more willing to sell a small loss. This means, more experience and more trading frequency can moderate the phenomena of selling large gain and loss, which is the V-shaped disposition effect. Senior, rich experience and high trading frequency are often seen as sign of investor sophistication. Since China is a developing country, senior people could be lack of professional training and education. If I ignore age as a sign of sophistication, I can indicate that sophisticated investor can moderate the effect of selling large magnitude of gain and loss in some extent.

Additionally, senior in age should not be a sign of sophistication in China. Feng and Seasholes (2005) and Cheng, et al. (2013) state that senior in age cannot help investors prevent disposition effect. This study further indicates that senior in age cannot protect investor from V-shaped disposition effect. There are two reasons that can probably explain why age is not a proper indicator of sophistication in China. First, China is a developing country, senior people may be luck of education and professional training on finance. Also, Chinese stock market was established in 1990 and the data period of this research is 2007 to 2009. Therefore, the most trading experience of investor is less than 20 yeas. Senior in age cannot leads to rich in trading experience. Thus, trading experience could be a more effective indicator of sophistication than age in China. More details are discussed in the conclusion chapter, Chapter 6.

3.5 Robustness and Discussion

3.5.1 Re-examination by Probit regression and fixed effect

To further control for the influence of methodology, I run the same model as in Table 3.2 but based on a probit regression method and fixed effect. Due to the extreme large size of the data and computation limitation, I can only apply day and stock fixed effect separately. Table 3.6 presents the marginal effect of the probit model and fixed effect. The results are similar to the results in logit model. In column 1, for probit model, if the position is a gain, the probability of selling this position by Chinese individual investor is increased by 12.58%. The *Gain*Return* term is also positive and significant, which is the same as it is in logit model in Table 3.3. The large magnitude of gain improves the probability of selling. In the loss side, *Loss*Return* is negative and insignificant, which means that when a position is a loss, how large the loss is does not influence the selling decision significantly. In column 2 and 3, the result is similar when controlling for day and stock fixed effect. The V-shape disposition is only suitable in Chinese market from 2007 to 2009 on the gain side. This result is consistent with the result in logit regression. Result in this chapter is robust with different modeling methodology.

(*Insert Table 3.6 here*)

3.5.2 Portfolios without limit-down stocks and portfolios with liquidation positions

The influence of government policy is strong in Chinese stock market. The limit-down policy was established 1996. A limit-down stock is a stock that decrease more than 10% in one day, which means (today's price – yesterday's price) / yesterday's price is less than -10%. If a stock becomes a limit-down stock in a particular day, the policy limits the lower bound of its price by -10% rate of return, so it cannot be traded in a lower price. Therefore, the possibility of selling limit-down positions is limited and it may influence the probability of selling bad-performance positions in the model. Thus, it can affect how investors sell their portfolios.

(*Insert Table 3.7 here*)

If there is a limit-down stock in one portfolio, it can influence the selling choice of the whole portfolio. Therefore, I delete the whole portfolio. In this part, only portfolios without limit-down stocks are included. I run the same logit model in Table 3.2 using these portfolios. The result of logit model is shown in Table 3.7 column 1. Individuals are 11.98% more likely to sell a position if it is a gain. When facing a gain position, the probability of selling increases 1.50% when the magnitude of gain increases one standard deviation. Investors are also slightly more likely to sell large loss positions, but the significant level is only at 10%. To sum up, disposition effect is significant in Chines stock market and V-shape disposition effect is only significant on gain side. This result is robust to the limit-down trading policy.

The strategy of liquidating a position and partial selling a position might be different. To further add robustness power to the result, I use dummy of liquidation as dependent variable in the model rather than selling in this section. The result of liquidation is shown in Table 3.7 column 2. Individual investors are more likely to liquidate a large gain and in loss case, the impact of magnitude of loss on liquidation is insignificant. These results are robust to the main result.

3.5.3 Decision of V-shape selling and future return

To add asset pricing value to my result, I calculate the average future return separately to positions with different magnitude of gain and loss. This method is a simple way to calculate the profit of selling large gain and loss, which is also similar to method in Odean (1998). An (2016) documents that V-shape selling is harmful to profit since positions with large unrealized gain or loss are under more selling pressure based on US market data. I try to examine it in China. I calculate three future returns for different time period, one week later, one month later and one year later. Then I apply correlations with future returns and magnitude of gain and loss. This indicate whether the position with large magnitude of gain or loss is more likely to lead to positive return or the small one does.

(*Insert Table 3.8 here*)

Table 3.8 present the result of the correlations. For all positions, in column 1 and 2, positions with large gain are more likely to lead to more future returns. Since magnitudes of loss are all nonpositive, large in number of magnitudes of loss is more likely to causes future less return indicates that loss positions close to zero have a larger probability of having a large return. When only consider positions that are sold, in column 3 and 4, the results are similar. These results lead to the conclusion that the one side V-shape selling in this chapter, selling large gains, causes damage to investor profit and it is indeed a bias.

3.6 Conclusion

In this chapter, I study the impact of magnitude of gain and loss on selling decision in Chinese stock market. Under the control of holding period and other control variables, I find that in general individual investors hold a preference of realizing a large gain, but their preference of selling among different magnitude of loss is constant. However, when studying positions with different holding period in sub-groups, for short-term positions, the V-shaped disposition effect appears, and investors are more willing to sell large gain and large loss. However, for other positions, the result is entirely opposite.

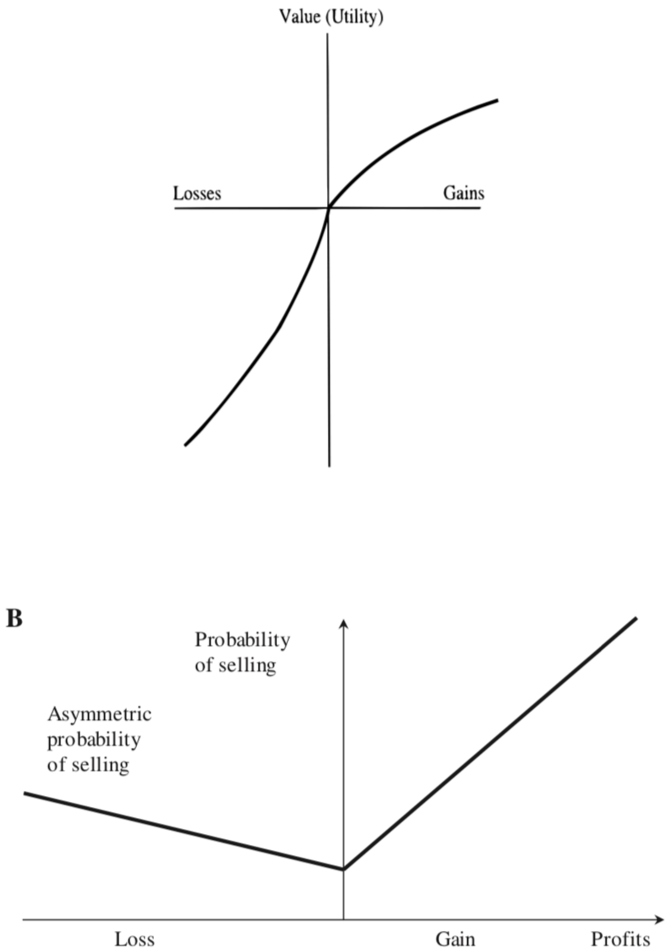
My findings indicate that holding period is a key control when analyzing investor behavior. By removing the holding period control variable in the model, I have totally different result. I strongly suggest studies on investor behavior should control the holding period of position.

My findings also document that the views of investors to positions with different returns and holding periods are different. For short-term positions, if the position performs exceed the expected interval of investor on either side, which means the position has a large gain or loss, investor is more likely to realize it. If the position survives the first month and become a midrange or long-term position, investors believe position with large gain will follow momentum theory and continue rising while position with large loss will follow reverse theory. They are more willing to realize returns close to zero. The difference in trading strategy to short-, medium- and long-term positions raises a worth noting question to further studies.

I then discover that during all booming, crushing and recovering period of financial crisis, individual investors hold the preference of realizing a large gain. Meanwhile, investors are only willing to realize large losses when the whole market is booming, when they have confidence to the market. And when they lose confidence to the market during and after financial crisis, they are more patient to their large losses.

I also document that sophisticated investors can moderate the bias of more willing to sell large gain and loss in some extent. Investors with more experience and trading frequency are less likely to trade follow V-shaped disposition effect while senior in age does not help. Since senior people could be lack of professional training and education in China as a developing country, I infer that senior in age is probably not a sign of sophistication for investors in developing country.

Figure 3.1: S-shape Disposition and V-shape Disposition



Source: Odean (1998) and Ben-David and Hirshleifer (2012)

Table 3.1: Summary Statistics

Observations	4,065,596			
Dummy variables	Number		Percentage (%)	
Sold	1,586,158		39%	
Gain	1,576,700		38%	
Gender	1,868,103		45%	
New_investor	2,215,459		54%	
Bull_mkt	764,294		18%	
Bear_mkt	1,801,764		44%	
Steady_mkt	1,499,538		36%	
Numerical variable	Average	S.D.	Min	Max
Return	-1.1633	4.0722	-165	113.413
Root_holding_period	5.8116	5.9084	0	29.4788
Variance	0.8439	2.7307	0.0013	106.2864
Portfolio_size	6.0896	5.0929	1	73
Root_tradetimes	26.0512	16.7221	1	148.7683
Root_age	6.5642	0.8804	3.8542	9.0514

Note: This table presents the summary statistics of variables in the model. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. All variables are defined in section 3.3.3.

Table 3.2: The Impact of Magnitude of Gain and Loss on Selling

	Dependent Variable: Dummy of Selling the Position		
	(1)	(2)	(3)
Gain	0.1273***	0.1229***	0.1251***
(t-statistics)	(18.1053)	(17.0902)	(19.6698)
Gain*return		-0.0019**	0.0154***
		(-2.1126)	(8.1766)
Loss*return		0.0092	-0.0042
		(17.9778)	(-1.5897)
Sqrt_holding_period	0.0014		0.0019
	(1.1951)		(1.4763)
Sqrt_holding_period*return*gain	-0.0007***		-0.0022***
	(-5.6000)		(-11.7796)
Sqrt_holding_period*return*loss	0.0008***		0.0011***
	(8.3267)		(4.1766)
Variance*gain	-0.0010**	-0.0014**	-0.0020***
	(-2.1654)	(-2.5436)	(-4.3430)
Variance*loss	0.0036***	0.0045***	0.0030***
	(10.0013)	(14.4933)	(8.8587)
Portfolio_size	-0.0370***	-0.0369***	-0.0371***
	(-26.4871)	(-24.7096)	(-26.8021)
Bull_mkt	-0.0106***	-0.0084**	-0.0145***
	(-2.6286)	(-2.1806)	(-3.9052)
Bear_mkt	0.0303	0.0351	0.0270
	(1.1221)	(1.3408)	(1.0482)
Root_trade_times	0.0016***	0.0016***	0.0016***
	(3.2466)	(2.9235)	(3.3342)
Gender	-0.0102***	-0.0100***	-0.0102***
	(-4.4093)	(-3.8005)	(-4.4628)
Root_age	-0.0102***	-0.0099***	-0.0102***
	(-7.7746)	(-7.2108)	(-7.8144)
New_investor	0.0054	0.0058	0.0052
	(1.4723)	(1.2748)	(1.4007)
Observations	4,065,596	4,065,596	4,065,596
Pseudo R^2	0.0860	0.0851	0.0863

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 shows disposition effect. Column 2 shows V-shape disposition effect. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. The dependent variable is a dummy variable equal to one if a stock is sold. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share price return since purchase. Other variables are control variables and are defined in part 3.3.3 in this paper. The top number is the marginal effect. The lower number in parentheses is the t -statistic. Clustered standard error is applied by date and investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 3.3: The Impact of Magnitude of Gain and Loss on Selling with Different Holding Period

	Dependent Variable: Dummy of Selling the Position		
Holding Period (day):	1-20	21-250	250+
	(1)	(2)	(3)
Gain	0.1158***	0.1281***	0.1138***
(t-statistics)	(13.8492)	(10.4013)	(10.6448)
Gain*return	0.0157***	-0.0054***	-0.0055***
	(6.3421)	(-5.5653)	(-3.4580)
Loss*return	-0.0035**	0.0088***	0.0060***
	(-2.3765)	(7.1503)	(7.0868)
Variance*gain	-0.0031***	-0.0004	0.0075***
	(-6.9487)	(-0.5296)	(2.8665)
Variance*loss	-0.0002	0.0073***	0.0176***
	(-0.6792)	(7.7392)	(6.9786)
Portfolio_size	-0.0360***	-0.0372***	-0.0271***
	(-22.3640)	(-20.0070)	(-21.0107)
Bull_mkt	-0.0350***	-0.0215***	0.0405
	(-8.5882)	(-3.8905)	(1.2219)
Bear_mkt	-0.0016	0.0435	0.0788**
	(-0.1463)	(1.0775)	(2.1432)
Sqrt_trade_times	0.0022***	-0.0001	-0.0004
	(11.93671)	(0.0717)	(-0.7349)
Gender	-0.0108***	-0.0068**	-0.0043
	(-4.9612)	(-2.0544)	(-1.5282)
Sqrt_age	-0.0085***	-0.0104***	-0.0034*
	(-5.5682)	(-7.3090)	(-1.9202)
New_investor	0.0040	-0.0002	-0.0069
	(1.5937)	(-0.0342)	(1.1176)
Observations	2,309,116	1,368,357	388,123
Pseudo R^2	0.0544	0.1022	0.1165

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 is from positions that their holding period is less than 20 days. Column 2 is from positions with holding period 21-250 days. Column 3 is from positions that had been held for more than 250 days. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. The dependent variable is a dummy variable equal to one if a stock is sold. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in part 3.3.3 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. The standard error is clustered by date and investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 3.4: The Impact of Magnitude of Gain and Loss on Selling in Different Market Condition

	Dependent Variable: Dummy of Selling the Position		
Market Condition:	Bull	Bear	Steady
	(1)	(2)	(3)
Gain	0.1193***	0.1131***	0.1281***
(t-statistics)	(14.1987)	(9.9267)	(11.9441)
Gain*return	0.0131***	0.0205***	0.0300***
	(7.2751)	(4.3374)	(8.8019)
Loss*return	-0.0232***	-0.0055	0.0040*
	(-3.9470)	(-1.340)	(1.9357)
Sqrt_holding_period	0.0167***	0.0073*	-0.0019***
	(14.1715)	(1.9517)	(-6.4100)
Sqrt_holding_period*return*gain	-0.0321***	-0.0024***	-0.0032***
	(-11.8039)	(-6.0751)	(9.5944)
Sqrt_holding_period*return*loss	0.0068***	0.0015***	0.0004***
	(7.0338)	(2.6602)	(3.6407)
Variance*gain	-0.0042***	-0.0023***	-0.0012**
	(-6.0251)	(-4.4373)	(-2.3827)
Variance*loss	-0.0005	0.0034***	0.0053***
	(-0.4785)	(9.6414)	(8.5708)
Portfolio_size	-0.0428***	-0.0386***	-0.0335***
	(15.4309)	(-18.7700)	(-27.1814)
Sqrt_trade_times	0.0027***	0.0011	0.0019***
	(12.8981)	(1.2013)	(13.6915)
Gender	-0.0170***	-0.0084***	-0.0106***
	(-5.0338)	(-2.7030)	(-4.8690)
Sqrt_age	-0.0146***	-0.0103***	-0.0101
	(-4.5876)	(-7.7032)	(-7.7513)
New_investor	0.0103***	0.0044	0.0084***
	(3.0505)	(0.6878)	(3.3223)
Observations	764,294	1,801,764	1,499,538
Pseudo R^2	0.0731	0.0844	0.1028

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 is from bull market and the time period is from Jan 1st 2007 to Oct 16th 2007. Column 2 is from bear market condition and the time period is from Oct 17th 2007 to Nov 4th 2008. Column 3 is the steady market condition and the time period is from Nov 5th 2008 to May 31st 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. The dependent variable is a dummy variable equal to one if a stock is sold. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in part 3.3.3 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. The standard error is clustered by date and investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 3.5: The Impact of Magnitude of Return on Investor Characteristics

	Dependent Variable: Dummy of Selling the Position	
	Continuous Age	Age Group
	(1)	(2)
Gain	0.1241***	0.1242***
(t-statistics)	(19.1010)	(19.1125)
Gain*return	0.0163***	0.0248***
	(3.9700)	(7.4625)
Loss*return	-0.0166***	-0.0171***
	(-2.8792)	(-2.8768)
Gain*return*sqrt_age	0.0009	
	(1.5941)	
Loss*return*sqrt_age	0.0001	
	(0.3103)	
Gain*return*young		-0.0027
(age<35)		(-1.5304)
Gain*return*middleage		-0.0026*
(35<=age<55)		(-1.9406)
Loss*return*young		0.0004
(age<35)		(0.5291)
Loss*return*middleage		0.0015**
(35<=age<55)		(2.3392)
Gain*return*gender	0.0045***	0.0045***
	(5.3459)	(5.3079)
Loss*return*gender	0.0002	0.0001
	(0.3396)	(0.2611)
Gain*return*exp	-0.0032***	-0.0033***
	(-3.4741)	(-3.6364)
Loss*return*exp	0.0018***	0.0018***
	(3.0442)	(3.0189)
Gain*return*sqrt_tradetimes	-0.0002***	-0.0002***
	(-2.8342)	(-2.8767)
Loss*return*sqrt_tradetimes	0.0004***	0.0004***
	(3.0836)	(3.1055)
Control Variables	Yes	Yes
Observations	4,065,596	4,065,596
Pseudo R^2	0.0869	0.0869

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. In column 1, we use age as continuous variable. In column 2, we discuss age groups. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. The dependent variable is a dummy variable equal to one if a stock is sold. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in part 3.3.3 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. The standard error is clustered by date and investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 3.6: The Impact of Magnitude of Gain and Loss on Selling using Probit Model and Fixed Effect

	Dependent Variable: Dummy of Selling the Position		
	Probit model	Fixed effect	
	(1)	(2)	(3)
Gain	0.1258***	0.138***	0.120***
(t-statistics)	(20.978)	(4.929)	(3.871)
Gain*return	0.0135***	0.016**	0.014*
	(8.007)	(2.286)	(1.750)
Loss*return	-0.0022	0.003	-0.003
	(-0.958)	(0.600)	(-0.231)
Control Variables	Yes	Yes	Yes
Fixed Effect		Day	Stock
Observations	4,065,596	3,985,114	4,065,596
Pseudo R^2	0.0819	0.082	0.098

Note: This table presents the marginal effect from probit regression and logit regression with fixed effect. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 is for probit model. Column 2 and 3 is the result of logit model with day and stock fixed effect respectively. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. The dependent variable is a dummy variable equal to one if a stock is sold. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share price return since purchase. Other variables are control variables and are defined in part 3.3.3 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is applied by date and investor. The first column shows the results with all variables. The second column shows the results without holding period. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 3.7: The Impact of Magnitude of Gain and Loss on Selling without Limit-Down Stocks and the Impact on Liquidating

	Dependent Variable: Dummy of Selling the Position	Dependent Variable: Dummy of Liquidating the Position
	(1)	(2)
Gain	0.1198***	0.0889***
(t-statistics)	(20.024)	(14.4446)
Gain*return	0.0150***	0.0045**
	(8.498)	(2.3778)
Loss*return	-0.0057*	0.0015
	(-1.879)	(0.4633)
Control Variables	Yes	Yes
Observations	3,877,061	4,065,596
Pseudo R^2	0.0862	0.0856

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. In column 1, if the whole portfolio of one investor has no limit-down stock, the data is included. Column 2 include all data. In column 1, the dependent variable is a dummy variable equal to one if a stock is sold. In column 2, the dependent variable is a dummy variable equal to one if a position is liquidated. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share price return since purchase. Other variables are control variables and are defined in part 3.3.3 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is applied by date and investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 3.8: The Impact of Magnitude of Gain and Loss and Decision of V-shape Selling on Future Return

	All Positions		Positions be Sold	
	Magnitude of Gain	Magnitude of Loss	Magnitude of Gain	Magnitude of Loss
	(1)	(2)	(3)	(4)
One Week Later Return	0.6694	0.9340	0.6114	0.8943
One Month Later Return	0.3636	0.8049	0.2897	0.7413
One Year Later Return	-0.2646	0.3920	-0.2682	0.4102

Note: This table presents the result of correlations. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. Returns are calculated as the difference of the unit share price in future and unit share cost. In column 1, we present for all gain positions, the correlation of the magnitude of gain and future return. In column 2, we include all loss positions. In column 3 and 4, we show positions that are sold that day.

Appendix 3.1: Data Preparation

There are totally 56,561,915 holding records in stock file from the 100,000 customers sample. Since the total size of data in this sample is over 10G, it is still too large to be handled by general PC. We need to choose the step that can significantly decreases data size at the beginning of our data preparation process. Following the method of Ben-David and Hirshleifer (2012) and Hartzmark (2015), we only consider portfolios in days that investors do sell at least one stock that day (active selling day). Since we discuss the comparison of stocks in one's portfolio when he considers selling a position, if the investor does not sell any position one day, he is considered to be inactive that day. To execute this, we compare the holding amount of one position with the position being held by the same investor in the previous trading day. If the holding amount decreases, we consider that the investor sells this position that day. Also, since the data is based on the shares one investor hold at the end of one trading day, if investor sells all his shares that day (liquidation), there is no record for him in that day. Therefore, we add these records in our data and marked them by liquidation. Due to the feature of our code[8], this step also includes duplicated records deleting. After this step, only positions in which at least one stock is sold (liquidated) is included in our data. There are 7,554,613 observations remain (roughly 13%). And by having this as our first step, we significantly decrease the load of calculation.

Since our data also contains some records from September and October 2009 but June, July and August 2009 are missing. We delete records after May 2009. We roughly delete 5% observations. We drop positions with unclear buying price. This is because the buying behavior is before the start date of our data. We accomplished this by deleting positions that are held in the first day and positions with same investor and same stock with these positions. This drops roughly 8% of all records. We furthermore delete extreme records and mistake records. We drop positions with trading price smaller than 0, number of shares smaller than 100. The initial dataset includes some records from HK market and other market. We only keep records from Chinese stock market (Shanghai exchange and Shenzhen exchange). There are 42 institutional investors in these 100,000 investors. We only focus on individuals and delete positions held by these investors. There are no foreign investors. Some of the records have a very small numbers

8 In python, dictionary {} cannot include duplicated key. By using dictionary with key equals to "person + stock code + date", we automatically delete records with the same person, same stock code and same date.

of shares holding by one person one day. We think it is caused by mistake. We keep holding records with more than (equal to) 100 shares per investor per day. There are also some mistake data that has 0 turnover rate in customer file, we drop them as well. We also delete positions with extreme return (>200 or <-200)

When dealing with stock disclosure data, we delete data from the IPO day, trade suspension day and the first day of trading after continuously 20 suspension day. We believe trading in these days is abnormal. When matching stock disclosure data and our holding records data, we drop positions on these days (approximately 11%). Since we use the stock price variance over last year to control the volatility (see more details in session 2.3), we also delete positions with stock IPO within 50 trading days. This delete 2% data.

When calculating holding period, we match our holding data with transaction data to get the first buying date of every positions. However, there are approximately 12% of positions that cannot find their first buying date in transaction data or have their first buying date after the positions record date. This is probably because of some bugs in this data. To discuss this, while we still delete this part of data in our main data in order to calculate holding period, we also run a robust test that keep this part (see more details in appendix 3.2).

When we delete these positions above, we may delete positions that are sold that day. And if this is the only position that is sold that day in the portfolio, there are no position that is sold in this portfolio remaining in our data and that day is not active selling day of this investor any more. Since we only want to keep portfolios that at least one stock is sold (the day is an active selling day of that investor), we check this again and delete roughly 11% observations. There are 4,065,596 records remain and this is the data we use to exam V-shape disposition.

Appendix Table 3.1: Data Preparation Process

	Data Preparation Process	
Original	56,561,915	
(percentage remain)	(100%)	
Active selling-day (at least one stock	7,554,613	
is sold on that day in the portfolio)	(13.36%)	
Date before 2009-06-01	7,140,349	
	(94.52%)	
Delete positions form the first day	6,565,088	
	(91.94%)	
Delete trading price smaller or	6,271,589	
equals to 0	(95.53%)	
Delete number of shares in a	6,071,948	
position < 100	(96.82%)	
Delete positions not in Shanghai	6,028,922	
and Shenzhen exchange	(99.29%)	
Delete institutional investors	6,023,504	
	(99.91%)	
Delete positions with extreme	6,023,159	
return (>200 or <-200)	(99.99%)	
Delete positions with inappropriate	5,352,904	
stock disclosure information	(88.87%)	
Delete investors with 0 turnover	5,352,894	
rate in customer file	(99.99%)	
Delete positions which stock	5,220,908	
variance over last year is null value	(97.53%)	
	Data with holding period	Data without holding period
Delete positions which cannot	4,597,491	--
calculate holding period	(88.06%)	--
Keep active selling-day again	4,065,596	4,622,239
	(88.43%)	(88.53%)

Note: This table presents the process of data preparation. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. The details of each step are described in part 3.2 in this paper. The top number is the number of observations(positions) after executing the particular data cleaning step. The lower number in parentheses is percentage of observations remain after this step.

Appendix 3.2: The robust test of deleting data that cannot calculate holding period

In appendix 1, we document that when calculating holding period, there are 12% data cannot calculate holding period. In this chapter, we discuss the robustness of deleting these 12% observations. We run a model that is similar to our main model Table 3.3 but it does not contain holding period variables. We fit the model with our main data (deleting these 12%) and the data keep this 12% observations separately.

(*Insert Appendix Table 3.2 here*)

As shown in Appendix Table 3.2, in column 1, it is result from our main data, and in column 2, it is from data with these 12%. The results from these two datasets are very similar. They both have large positive and significant coefficient on Gain. At the same time, they also both have negative and significant on Gain*return and positive and significant on Loss*return. These results demonstrate that the deleting of these 12% data does not influence our result.

In Appendix Table 3.2, investors are more likely to sell positions that are close to zero. This indicate a reverse V-shape disposition. And it is the opposite of our result in Table 3.2. As we miss the control variables of holding period here in Appendix Table 3.2, we find that holding period variables are important in our test of V-shape disposition. This result support our point that V-shape disposition is closely linked to holding period of position.

Appendix Table 3.2: The robust test of deleting data that cannot calculate holding period

	Dependent Variable: Dummy of Selling the Position	
Data:	Data used in our model	Keep data that cannot calculate holding period
	(1)	(2)
Gain	0.1201***	0.1210***
(t-statistics)	(18.187)	(19.169)
Gain*return	-0.0019**	-0.0030***
	(-2.286)	(-4.418)
Loss*return	0.0092***	0.0074***
	(19.845)	(16.927)
Control Variables	Yes	Yes
Observations	4,065,596	4,622,237
Pseudo R^2	0.0851	0.0861

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. Column 1 uses data we use in our model (around 10% of data that their holding period cannot be calculated correctly). Column 2 uses data that include this 10% data. The dependent variable is a dummy variable equal to one if a stock is sold. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share price return since purchase. Other variables are control variables and are defined in part 3.1 in this paper. All control variables are included. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is applied by date. The first column shows the results with all variables. The second column shows the results without holding period. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Chapter Four: Selling the Best: Rank Effect in Chinese Individual Investors

4.1 Introduction

In the study of behavioral finance, although most research successfully explain investors' behavior to some extent, how investors treat their own portfolio, how investors compare positions in their portfolio and how they choose when they want to sell are still questions.

To answer this, Hartzmark (2015) develops these theories and find a new stylized fact about how investors trade assets, named rank effect. It shows that investors compare the returns of stocks in their portfolio when consider selling and they are more willing to sell stocks with extreme winning and extreme losing positions. The most crucial contribution of rank effect is that it considers the comparison within one's portfolio when he/she is making decision of selling. In previous studies, although most researches successfully explain investors' behavior to some extent, but most of them suffer from a stock-by-stock bias, which they assume that investors consider stocks one-by-one and ignore the comparison between stocks in the portfolio. However, it has been proven in psychology that people consider what they have as a whole in the decision-making process. Therefore, comparison in one's own portfolio should be added as a factor when analyzing the decision making of an investor. Furthermore, rank effect causes damage to the profit of investors since both stocks with large unrealized gains and losses outperform other stocks (An, 2017). I also find rank effect hart investor's profit in Chinese market.

However, there is still lack of empirical evidence of rank effect due to the limited availability of detailed account-level data on trading activities. In this chapter, I explore rank effect in Chinese stock market. Based on a very unique and large database, I find a different rank effect exists in Chinese stock market, which is individuals are more likely to sell best and 2nd best performance positions.

One main limitation of rank effect is that because it needs to split positions into 5 groups (best, 2nd best, middle, 2nd worst and worst), it requires the portfolio to have at least 5 stocks. This condition significantly reduces the number of portfolios that can apply rank effect. Furthermore, since investors with a large portfolio size are thought to be more sophisticated, this leads to a bias as well. In this chapter, I enhance the previous result by including more portfolio into the theory. I only require at least 2 positions in a portfolio and develop rank effect into top effect. By doing this, I include roughly double the number investors in the theory and these investors are more likely to be individual investors with less skill. In the results, I find that the top one stock is the choice of selling.

Furthermore, to discover the reasons of the difference between US investors (selling both best and worst positions) and Chinese investors (only selling best position), I try to find who and in which situations Chinses investors will also sell the worst position. When the position is under a gambling situation (short period, low price per share and high volatility), the probability of selling the worst is close to the probability of selling the 2nd best. And when I further limit portfolios from young, male and new investors, investors are willing to sell the worst in China. Young, male and new investors are over-confidence and have less patient to their loss. Investors have less patient to gambling positions as well. When an aggressive investor holds a gambling position and it performs bad, the probability of selling this position is significantly high.

The data is collected from a large brokerage firm in China[9]. It contains more than 3 million accounts and 2 billion daily stock dealing records over the period of January 2007 to May 2009. Due to the consideration on the cost of computation, I use a sample of 100,000 investors in this chapter. The time period of the dataset in this research covers the financial crisis period. In China, although not mainly caused by the world financial crisis, there was also a huge bubble in 2007 and experiencing significant stock price falling in 2008. The changes of the investor emotion and behavior when they face large profits and losses along with risks are interesting to research. This adds more value to the dataset and this research.

Surprisingly, investors do not trade very differently before, during or over financial crisis. The extreme risk market condition in financial crisis does not change the behaviour of individual investors a lot. My results are the same under different time periods and market conditions.

Ben-David and Hirshleifer (2012) and Kaustia (2010b) show that the impact of past profit on disposition is closely related to holding period. Therefore, holding period can also play an important role on rank effect. Short-term, midrange and long-term positions may indicate

9 Most of the previous studies (Odean, 1998; Feng and Seasholes, 2005; Dhar and Zhu, 2006, etc.) in this area do not disclose the specific name of the brokerage for confidentiality reasons, neither do my research.

different trading stagey as well. In my study, I find that holding period is not a key in rank effect. Furthermore, I investigate how investor characteristics impact rank effect. female and senior investors are rank traders in China, while less experience and high trading frequency moderate rank effect.

The rest of this chapter is organized as follows. Section 2 reviews the relevant literatures. Section 3 introduces the dataset, data processing and Chinese stock market during this period. Section 4 document the model. Section 5 discusses the main empirical results, while robustness tests are presented in Section 6. Section 7 concludes.

4.2 Literature Review

As it is discussed in the previous chapter, Ben-David and Hirshleifer (2012) find a V-shaped curve of the position return and the probability of selling. They also document that gains or losses is not the only issue when analyze disposition effect. How much is the gains or losses is a question as well. Meanwhile, Kaustia (2010b) supports this result on the gain side. Additionally, attention trading (Barber and Odean, 2008) provides a possible reason for the V-shape disposition since stocks with large gains or losses catch more attention of investors. Rank effect (Hartzmark, 2015) further develops these findings into a new investor trading bias effect that individuals are more likely to sell the extreme winning and extreme losing positions in their portfolio. He criticizes that the previous studies have considered investor trading preliminary on a stock-by-stock bias and ignore the portfolio problem in its entirety. Using data from a large retail brokerage in US stock market from 1991 to 1996, Hartzmark (2015) shows that on a day an investor sells a position in their portfolio, the investor has a 31% chance of selling the stock with the highest return in the portfolio and a 26% chance of selling the stock with the lowest return, after controlling for a number of factors. Hartzmark (2015) also fits Logit regression model used by Ben-David and Hirshleifer (2012) with adding the rank dummy variable. The best-ranked stock (Best) is 15.7% more likely to be sold, and the worst-ranked stock (Worst) is 10.7% more likely to be sold, both significant with large t-statistics. After including the two dummy variables the Loss*Return and Gain*Return, which indicate the disposition effect, their coefficients are becoming insignificant and the Gain dummy coefficient decreases. This means that rank effect is at least as strong as disposition effect. After these, An

(2016) finds asset pricing value based on V-shape disposition effect and rank effect by showing that stocks with both large unrealized gains and large unrealized losses outperform others in the following month.

In this chapter, I significantly update the time period of the data to 2007 to 2009. I also discuss the rank effect in a brand-new market, Chinese market, which is a large and emerging market in a developing country. Since the year 2007 to 2009 cover the financial crisis, I also analyze rank effect in the extreme market condition. In addition, I also discuss rank effect on positions with different holding period and rank effect with investor heterogeneity.

4.3 The Dataset

This chapter is based on a very large database collected from a large nationwide brokerage firm in China, with more than 3 million accounts and 2 billion daily dealing records over the period of January 2007 to May 2009[10]. Due to the computational capacity limitations, I use a random sample of 100,000 investors and more than 56 million records sub-data to build my model. The dataset is formed with 3 sub-datasets that are customer file, stock file and transaction file. Customer file contains the information of each customer. Stock file contains information of each stocks held by each customer on daily basis. Transaction file contains each deal's information. Customer ID, stock ID and date are used to merge all files. I also get the stock price and volatility information from CSMAR (China Stock Market & Accounting Research Database).

Each row in the stock file indicates the holding record of one investor for one stock at the end of one trading day and it composes my main data table. The transaction file provides me the trading amount and price of each trades. There is also a column shows "selling" when the transaction record is a sale. The customer file shows the gender, account open date and birthday. Most of investors are individual investors and there is no foreign investor. Since all data is based on the end of each trading day, the trading sequence of multiple trades of one investor in one day cannot be observed. Day traders are no included as well.

10 There are three missing months, which are April 2007, May 2007 and March 2008.

There are totally 56,561,915 holding records in stock file from the 100,000 customers sample. Following the method of Ben-David and Hirshleifer (2012) and Hartzmark (2015), I only consider portfolios in days that investors do sell at least one stock that day (active selling day). Since I discuss the comparison of stocks in one's portfolio when he considers selling a position, if the investor does not sell any position one day, he is considered to be inactive that day. To execute this, I compare the holding amount of one position with the position being held by the same investor in the previous trading day. If the holding amount decreases, I consider that the investor sells this position that day. Also, since the data is based on the shares one investor hold at the end of one trading day, if investor sells all his shares that day (liquidation), there is no record for him in that day. Therefore, I add these records in my data and marked them by liquidation. After duplication, only positions in which at least one stock is sold (liquidated) is included in my data. There are 7,554,613 observations remain (roughly 13%).

Since my data also contains some records from September and October 2009, but June, July and August 2009 are missing, I delete records after May 2009. I roughly delete 5% observations. I drop positions with unclear buying price. This is because the buying behavior is before the start date of my data. I accomplished this by deleting positions that are held in the first day and positions with same investor and same stock with these positions. This drop roughly 8% of all records. I furthermore delete extreme records and mistake records. I drop positions with trading price smaller than 0, number of shares smaller than 100. The initial dataset includes some records from HK market and other market. I only keep records from Chinese stock market (Shanghai exchange and Shenzhen exchange). There are 42 institutional investors in these 100,000 investors. I only focus on individuals and delete positions held by these investors. There are no foreign investors. Some of the records have a very small numbers of shares holding by one person one day. I think it is caused by mistake. I keep holding records with more than (equal to) 100 shares per investor per day. There are also some mistake data that has 0 turnover rate in customer file, I drop them as well. I also delete positions with extreme return (> 200 or < -200)

When dealing with stock disclosure data, I delete data from the IPO day, trade suspension day and the first day of trading after continuously 20 suspension day. I believe trading in these days is probably abnormal. When matching stock disclosure data with my holding records data, I drop positions on these days (approximately 11%). Since I use the stock price variance over last year to control the volatility[11], I also delete positions with stock IPO within 50 trading days. This delete 2% data.

When calculating holding period, I match my holding data with transaction data to get the first buying date of every positions. However, there are approximately 12% of positions that cannot find their first buying date in transaction data or have their first buying date after the positions record date. This is probably because of some bugs in this data. To discuss this, while I still delete this part of data in my main data in order to calculate holding period, I also employ a robust test that keep this part[12].

When I delete these positions above, I may delete positions that are sold that day. And if this is the only position that is sold that day in the portfolio, there are no position that is sold in this portfolio remaining in my data and that day is not active selling day of this investor any more. Since I only want to keep portfolios that at least one stock is sold (the day is an active selling day of that investor), I check this again and delete roughly 11% observations.

The current price is the stock price at the end of a trading day when investor keeps this stock or sells part of his holding shares of that stock. And it is the stock price at the last trade when investor liquidate. Since the stock price in one day does not change too much, this setting is reasonable in my data. The cost price is calculated as the share weighted average buying price for multiple buying behavior for one stock. The return is current price minus cost price.

In my theory of top effect, an investor is more likely to sell the top-performance stock when he wants to sell. He prefers to choose the top one when comparing stocks in his portfolio. If there is only one stock in the portfolio, the comparison is meaningless. Therefore, I include records with at least 2 stocks in one's portfolio. There are 3,881,318 observations and 42,453 investors remain for me to examine top effect.

In rank effect theory, investors are more likely to sell best-performance stock and worst-performance stock. For similar reason, I follow Hartzmark's (2015) method to keep records with at least 5 stocks in one's portfolio one day. Since investors with a larger portfolio size are thought to be more sophisticated, this process is bias to some extent. After all the cleaning

11 See more details in Section 4.4.
12 See more details in Appendix 4.1.

process, there are 2,130,356 records and 18,610 investors remain in my data to analyze rank effect. Comparing to top effect, rank effect has the rule that needs to drop more than a half investor than top effect. My top effect can be applied to much more investors. I significantly expand and improve the theory. More details of the data preparation process are shown in Table 4.1 below:

(*Insert Table 4.1 here*)

4.4 The Model

To test the rank effect and top effect in Chinese market during financial crisis, I estimate a similar model as Ben-David and Hirshleifer (2012) and Hartzmark (2015) and is listed below as Equation 1:

$$Sell = Constant + \ a_1(Rank\ Variables\ or\ Top\ Variables) + \ a_2(Gain) \ + \\ a_3(Gain * Return) + \ a_4(Loss * Return) \ + \ a_5(Control\ Variables) + \epsilon \quad (1)$$

The model is on day-investor-stock level. Each observation is a position that one investor holds one stock in one day. The model is fitted as a Logit model by maximum likelihood. The dependent variable is a dummy variable, equals to 1 if the stock is sold that day by that investor and 0 otherwise. Both partial selling and liquidation are involved. Return is the unit share return of position which is calculated based on the buying price (trading cost involved and is weighted average price by shares in case of multiple purchase) and the current price of that stock at that day. Gain is a dummy variable that takes the value of 1 if the unit return of the position is positive and 0 otherwise. This controls the disposition effect. Loss is the opposite of Gain. Including the interaction terms of Gain (Loss) and Return allows me to control the relationship between the probability of selling and the magnitude of gain and loss separately.

When analyzing rank effect, I add a set of 5 rank variables into the model. I rank the positions in one's portfolio by the unit share return as best, 2nd best, worst, 2nd worst and middle. Best is a dummy variable that equals to 1 if the position is in 1st rank in one's portfolio on a particular day. This means the best stock has the largest unit share return in the portfolio. 2nd Best, 2nd Worst and Worst are defined similarly to indicate the 2nd best, 2nd worst and worst position in one's portfolio on a particular day. If one position is not Best, 2nd Best, 2nd Worst nor Worst (not ranked in the top or bottom two), it is defined as middle, which has a value of 1 in dummy variable Middle. When examining top effect, I employ a set of 2 top variables instead of rank variables. A position is Top when it performs the best in one's portfolio, and Other otherwise.

For control variables, I also follow the choice of these two papers. The effect due to holding period and volatility are controlled for. To control for the days a position is held from purchase to sell, the square root of the holding days (Root_holding_period) and interaction terms with gain dummy and return (Root_holding_period*return*gain) and loss dummy by return (Root_holding_period*return*loss) are included. To control for the stock variance, I also calculate the return variance of a stock over the last year (the variance of stock price over preceding 250 trading days, if there are at least 50 non-missing records). I include the interaction term Variance*gain and Variance*loss in my model.

In addition of the control variables choice in Ben-David and Hirshleifer (2012) and Hartzmark (2015), I further add investor characteristic control variables to control the investor heterogeneity influence. Gender is a dummy variable that takes value of 1 for female and 0 for male. Root_age is the square root of the investor's age[13]. To control the experience of investor, I introduce a dummy variable New_investor, which equals to one if the investor opened account in this brokerage after the start of my data period and zero otherwise. It is worth noting that at that time, in Chinese market, one individual can only open one account in the whole market. This makes my experience more powerful. Root_tradetimes is the square root of times of trading an investor made in my data period. It can indicate the activation of an investor in some degree. Portfolio_size is the number of stocks in one's portfolio that day.

Since my data cover the financial crisis in China, I further introduce two dummy variables to control the time and market condition. I divide my data period, Jan 2007 to May 2009, into

13 I use the date difference between investor's birthday and May 31st 2009, which is the last day of my dataset.

three parts, from Jan 1st 2007 to Oct 16th 2007 as bull market, from Oct 17th 2007 to Nov 4th 2008 as bear market, from Nov 5th 2008 to May 31st 2009 as steady growth market. I define the three sub-time period by the value of Shanghai Composite Index, which has been discussed in detail in the early part of this chapter (part 2.1). I introduce dummy variable Bull_mar, equals to 1 if the position is in bull market period, and Bear_mar, equals to 1 if the position is in bear market period. I set the steady period as benchmark.

(Insert Table 4.2 here)

In Table 4.2, I present the summary statistics of data I use in my model. After all cleaning, there are total 3,881,318 records (positions). For dummy variable (binomial variable), I present the number of 1. For dependent variable Selling, there are 1,401,880 positions that are sold in the end of the day. Since I only include positions that at least one stock in the portfolio is sold in that day. This number is reasonable. Since my data cover financial crisis, there are more loss positions than gains in my model. The number of gain position is 1,492,565. The number of positions held by new investors and old investors are balanced. There are 2,113,723 positions held by new investors. My data is also balance in gender. For numerical variables, due to financial crisis, the average return is negative, which is -1.1941. And the mean of stock variance is 0.8453. On average, the square root of days a position is held is 5.8599 and the portfolio size is 6.3312. The average Root_age is 6.5727. The average Root_tradetimes is 26.4244.

4.5 Rank Effect and Top Effect

4.5.1 Empirical Study of Rank Effect

I employee the model in section 3 to examine rank effect in Chinese stock market. For investors, only days that at least one stock is sold are included as sell day (active day). An investor needs to hold at least 5 stocks to be included as well. The observations are at day-investor-stock level. All results are presented as marginal effects.

Hartzmark (2015) states a rank effect in US market, which is investors are more likely to sell best and worst positions rather than middle positions. Therefore, he uses Middle as his benchmark in his model. Following his method, I set Middle as benchmark as well. I explore the it into Worst and 2nd worst as well. In my model, I compare the result when Middle, Worst or 2nd worst is the benchmark.

(Insert Table 4.3 here)

Table 4.3 presents the results of rank effect from the logit model. In column 1, when using Middle as benchmark, a best position is 12.92% (t-statistics 33.8272) more likely to be sold than a middle one and a 2nd best position is 5.51% (t-statistics 25.5293) more likely to be sold than a middle one. These are the top 1 and 2 positions to be sold. 2nd worst position is 0.52% (t-statistics -2.4782) less likely to be sold than a middle one. The difference of probability of selling is insignificant between middle positions and worst positions. In column 2, when Worst is the benchmark, the results of best and 2nd best positions are similar to they are in column 1. The difference of probability of selling between middle and worst positions is insignificant. The probably of selling a 2nd worst position is significantly smaller then selling worst position. While in column 3, when 2nd worst is the benchmark, the results of best and 2nd best positions are still constant. The probability of selling a middle position is 0.52% (t-statistics 2.4699) larger than selling a 2nd worst one. And the probability of selling a worst position is 0.74% (t-statistics 3.8345) larger than selling a 2nd worst one. As a conclusion, Chinese investors during financial crisis are more likely to sell a position with better performance (best, 2nd best) and the probability of selling other positions (middle, 2nd worst and worst) do not have a large gap. In Hartzmark (2015), the US investors are more likely to sell best and worst positions than the middle one. My result in Chinese market is significantly different.

On the same time, since in column 3, 2nd worst positions appear to have the smallest probability of selling. I will use 2nd worst positions as benchmark in further results of this chapter. Meanwhile, Gain is positive and significant in all three columns. This result is constant with previous studies and indicates a strong disposition effect in China. However, by adding rank variables, the sign of Gain*return turns is negative while Loss*return is insignificant. The

V-shape disposition effect is insignificant after including rank variables. This shows the explanatory power of rank variables. Rank effect is at least as strong as disposition effect.

Based on these results, I can conclude that in Chinese market during 2007 to 2009, The best position, the position that has the largest return in one's portfolio, has the largest chance to be sold and the 2nd best position follows. The results of probability of selling of middle, 2nd worst and worst positions may need more evidence. But based on the result for now, I can infer that Chinese individual investors do not treat them in a large magnitude of difference. The rank of probability of selling in China should be: Best >> 2nd best >> worst >= middle >= 2nd worst. The performance of Chinese investors during 2007 to 2009 is difference to the performance of US investors as it is in result of Hartzmark (2015). Chinese investor does not share the same rank effect with US investors.

In addition, to consider the performance of investors' characteristics control variables, all these variables except New_investor is significant. This indicate that the investors' heterogeneity plays an important role in individuals' decision of selling. These variables should be included in this model. Furthermore, since my dataset is large, I do not suffer from a luck of degree of freedom by adding these variables. Adding them benefits to the robustness level of my results.

4.5.2 Gambling Positions and Aggressive Investors in Rank Effect

Since Chinese investors only sell the best-performance position while US investors sell both the best and the worst position, I want to discover what causes the difference. Based on a data driven logic, I build many different sub-groups to find in what situations investors in China sell both best and worst positions, that consistent to US investors.

(Insert Table 4.4 Here)

In Table 4.4, I examine the situations when Chinese investors also sell worst positions. In column 1, I include positions that is short period (holding days less than 20), low on price per share (price per share less than 30 percental of all stocks price that day) and high on volatility (return variance of last 250 days above 70 percental of all stocks that day). I find that when position is under these situations, the probability of selling a worst position rather than a 2nd worst one is almost 2%, and it is close to the probability of selling a 2nd best one. Since the criteria of short period (indicate high trading frequency and high turnover rate), low price and high volatility is close to the criteria of gambling stocks (Kumar, 2009 and Liao, Liang and Zhang, 2016) in both US and Chinese market, I define these positions as gambling positions. Furthermore, young, male and new investors are aggressive investors and more likely to gamble in Chinese market (Liao, Liang and Zhang, 2016), I apply these criteria in column 2.

However, I find aggressive investors perform similar to the whole population. They are more likely to sell best and 2nd best positions. Furthermore, I combine the criteria in column 1 and column 2 together. In column 3 the probability of selling worst position is greater than the probability of selling 2nd worst. When the position is a gambling position and it is held by an aggressive investor, Chinese investors also sell the worst position, which is the same as US investors.

Young, male and new investors are over-confidence and have less patient to their loss. Investors have less patient to gambling positions as well. When an aggressive investor holds a gambling position and it performs bad, the probability of selling this position is significantly high. Since the previous paper on rank effect is based on Odean's data in US (Hartzmark, 2015) and Odean's data is dominated by male investors, I doubt that rank effect is biased by investors' characteristic.

4.5.3 Developing Rank Effect into Top Effect

Due to the result of rank effect in China, best-performance position has the largest probability of being sold. Rank effect requires a portfolio to at least have 5 positions, but a portfolio needs only at least 2 positions to have a best-performance position. Therefore, the question raises: when the condition expands and becomes at least 2 positions in a portfolio instead of 5, do the result of the best (top) performance have the largest probability of being sold still exist? To explore this question, I employ portfolios with at least 2 stocks and name a dummy variable

Top (equals to 1 when the position has the largest unit share return in the portfolio) into my model. In this case, I am using positions that do not have largest unit share return as benchmark.

(Insert Table 4.5 Here)

Table 4.5 present the marginal effect of logit regression. In column 1, the portfolio size is larger or equal to 2. Individual investors are 13.53% more likely to sell a position when it is a top (best) performance one. And they are 9.00% more likely to sell a gain position as well. If the portfolio has at least 5 stocks, it is included in column 2, this is also the condition of rank effect. In column 2, investors are still more likely to sell top stock and gain stocks. These results indicate that no matter the portfolio size, my result that individual investors are more likely to sell the top 1 performance stock in their portfolio always exist. Disposition effect is strong in China as well.

These results show the choice of individual investors when they want to sell a position is the top 1 performance stock. I name this preference of investors the top effect. Due to the comparison of marginal effects of top effect and disposition effect, I can say in general, top effect is strong among Chinese investors and it is at least as strong as disposition effect.

4.5.4 Rank Effect under Different Market Conditions

Since the particularity of my data period, I test rank effect before, during and after financial crisis. The market condition changed dramatically during financial crisis. Under financial crisis, when all investors are facing extreme loss and risk, the mind and selling choice of investors may change. Similar to chapter 4.3, I depart data time period into three parts that represent three market conditions: bull market, bear market and steady market.

(Insert Table 4.6 here)

Table 4.6 shows the marginal effect of rank effect in different market conditions. In column 1, under bull market, individual investors are 11.94% more likely to realize best-ranked positions than 2nd worst-ranked (benchmark in this model) positions. They are also 7.09% more likely to realize 2nd best-ranked positions. The selling probability of middle positions is 1.79% larger than 2nd worst positions. And the probabilities of realizing worst and 2nd worst positions do not appear a significant difference. Results during financial crisis are shown in column 2. Best and 2nd best ranked positions still get the largest and second largest probability of being sold. However, different from bull market, there is no significant difference among 2nd worst and middle positions. And for worst positions, individual investors are 0.94% more likely to sell them than 2nd worst positions. In column 3, after financial crisis, investors in Chinese market are 12.77% more likely to sell best positions and 6.54% more likely to sell 2nd worst positions. They are also more likely to sell middle and worst positions than 2nd worst positions with percentage 0.83% and 1.08% respectively.

Based on these results, the behaviors of individual investors before, under and after financial crisis are similar. Rank effect in Hartzmark (2015) does not appear the same in Chinese market. Chinese investors are more likely to sell best and 2nd best positions while their behavior on other positions are still not clear enough. But in general, they do not discriminate middle, 2nd worst and worst positions when they consider selling a position. Their selling behavior before, during and after financial crisis do not change very much.

In Hoffmann, et al. (2013) and Gerrans, et al. (2015), both of them state that although individual investors change their expectation of return and risk tolerance during financial crisis, their trading behavior do not change significantly. My results support their argument in Chinese market. This result may also be explained by the macroeconomic situation in China from 2007 to 2009. The GDP of China grew on average 10% during this period. This can lead Chinese individual investors to have confidence on their local companies and stocks even if their performance on the stock market is not good.

4.5.5 Rank Effect by Different Holding Period

Ben-David and Hirshleifer (2012) show that V-shape disposition is closely related to holding period. Therefore, holding period can also play an important role on rank effect. Short-term, midrange and long-term positions may indicate different trading stagey as well. Because of that, I analyze rank effect on different holding period positions. I split my data into three sub-data: holding period from 1 to 20 days, from 21 to 250 days and more than 250 days. I fit the same model on three sub-datasets separately to analyze rank effect in Chinese market for different holding period positions. Due to multicollinearity, wo do not include holding period control variables this time.

(Insert Table 4.7 here)

In Table 4.7, column 1 shows results of short-term positions (holding period from 1 to 20 days). Individuals are 13.61% more likely to sell a best-ranked position than 2nd worst-ranked position with a very large t-statistics. They are also 5.66% more likely to sell a 2nd best position. The difference between middle positions and 2nd worst positions is not significant. And the preference of selling worst positions than 2nd worst is relatively small, only 1.56%. Meanwhile, for midrange positions (holding period from 21 to 250 days), in column 2, investors are also more likely to sell best and 2nd best positions. Both the preferences of selling middle and worst positions than 2nd worst positions are smaller than 1% and not significant on 1% level. In column 3, I present results from long-term positions (holding period larger than 250 days). The performance of best and 2nd best positions are similar; both have large preference of selling. The probability of selling middle positions than 2nd worst positions is 2.05%. There is no significantly difference between selling of worst positions and 2nd worst positions.

To sum up, the Chinese choices, best and 2nd best positions, perform constantly strong among all holding period positions. The results of middle, worst and 2nd worst positions are also similar. There is no large difference among these three positions. None of the results has a preference larger than 2.5%. Not as it is for V-shape disposition, rank effect is robust with holding periods of positions. This result can also improve the robustness power of my result on Chinese rank effect.

4.5.6 Impact of Investor Heterogeneity on Rank Effect

Goetzmann and Massa (2002) and Dhar and Kumar (2002) finds significant heterogeneity in investor beliefs and trading styles. List (2003), Dhar and Zhu (2006) and Feng and Seasholes (2005) discover influence of investor characteristics on disposition effect. In this section, I am going to discuss the impact of investor heterogeneity on rank effect. I do this by introducing interaction term of rank term and characteristics term into my main model.

(Insert table 4.8 here)

In my model, I apply two measurements of age, a continuous variable square root of age and age groups. In Table 4.8 column 1, I display the first measurement of age as well as other investor characteristics. Although aged investors are more likely to best positions, I can hardly fine strong influence of age on other rank variables. Since gender is 1 for female, female investors have larger probability of selling a best or a 2nd best position than male. Thus, they perform follow rank effect. Comparing with new investors, experienced investors are more likely to sell best, 2nd best and middle positions. They are rank effect traders. Investors with less trading frequency are more likely to behave follow rank effect as well. On the other side, in column 2, while other characteristics stay the same, I use age group to test effect of age. Old investors (age >= 55) like to sell best position and do not like sell a worst one. Old investors are rank traders in China.

Since trading follow rank effect causes damage to investor's profit (An, 2017), I treat rank effect as a trading bias together with disposition effect. Combining with results in Chinese market from Feng and Seasholes (2005), female and old investors are more likely to trade follow rank effect as well as disposition effect. I also find that years since one open an account cannot improve their performance on defending rank effect. But trading more helps.

4.6 Robustness Test

4.6.1 Rank Effect with Small Size Portfolios

One main limitation of rank effect is that it requires investors to have at least 5 stocks in their portfolio. I enhance it by developing rank effect into top effect that only requires at least 2 stocks. To further test rank effect in small size portfolios, I run a model similar to my main model in three sub-samples, portfolio contains 2, 3 or 4 positions. When portfolio size is 2, I employ Best variable and use Worst as benchmark, this is also the same model as top effect. When portfolio size is 3, I employ Best and Middle variables and use Worst as benchmark. And when portfolio size is 4, I employ Best, 2nd best, Worst variables and use 2nd worst as benchmark, which follow the benchmark choice of rank effect model in my main result.

(Insert table 4.9 here)

Table 4.9 present the results of rank effect in small size portfolios. All of the three sub-samples follow my result in rank effect that investors are most likely to sell best position and also like to sell 2nd best position (middle is also 2nd best in a 3 stocks portfolio). My result is robust in small portfolios.

4.6.2 Rank Effect by Probit Model

For rank effect, in order to control the influence of modelling methodology, I apply a test similar with my model in rank effect by probit model. Table 4.10 presents the result from probit model. In Chinese market, a best rank position is 12.65% more likely to be sold than a 2nd worst position (benchmark). A 2nd best position is 5.95% more likely to be sold than a 2nd worst position. With preference smaller than 1%, a middle position is 0.44% more likely to be sold and for worst position, the preference is 0.68%. Chinese individual investors have larger preference on selling best and 2nd best positions and the preferences on selling middle, worst and 2nd worst positions are very small. The results from porbit model is the same as results from logit model. My results are robust with the choice of model.

(Insert Table 4.10 here)

4.6.3 Rank Effect in All Gain/ Loss Portfolios

In order to research rank effect in extreme condition and the relation between rank effect and disposition effect, I estimate a similar logit model that restricts the portfolios into all gain portfolios and all loss portfolios. By doing this, I also consider the robustness of the overall performance of the portfolio. In all gain portfolios, all positions in this portfolio at that day are gains. These positions are at very good situation and may lead investors to overconfidence. In all loss portfolios, everything is the opposite. This test provides a more precise control for the disposition effect.

(Insert Table 4.11 here)

Table 4.11 shows the results of rank effect in all gain/loss portfolios in logit model. When all positions in a portfolio is gains, in column 1, investors are 11.00% (with significant t value 3.732) more likely to sell a best position than a 2nd worst one and 4.78% more likely to sell a 2nd best position. The probabilities of selling middle and worst positions are not larger than selling 2nd worst positions significantly. The differences among middle, 2nd worst and worst positions are not strong. When all positions come to a loss, in column 2, the probabilities of realizing best and 2nd best positions are still significantly larger than the probability of 2nd worst positions. Both of the preferences on selling middle and worst positions rather than 2nd worst positions are smaller than 1%. My result that Chinese investors are more likely to sell best and 2nd best positions and treat other positions similarly is found in both all gain and all loss portfolios.

4.6.4 Rank Effect with Liquidation

In my main model, I test the probabilities of selling of different rank positions. Selling here contains both partial sale and liquidation. Since liquidation can indicate a different strategy rather than selling, I use dummy of liquidation as dependent variable in my model rather than selling in this section.

(Insert table 4.12 here)

The result of liquidation is shown in Table 4.12. Individuals are 6.77% more likely to liquidate a best position than a 2nd worst one and 2.59% more likely to liquidate a 2nd best one. These results are similar to my main result. The only different is that when it comes to worst position, an investor is more likely to liquidate a worst position as well. However, the preference of liquidating a worst position is no more than half of a 2nd best one and also significantly smaller than a best one. My main result that investors are more likely to sell best and 2nd best positions than others is generally robust when I use liquidation.

4.6.5 Rank Effect in No Limit-Down Portfolios

The influence of government policy is strong in Chinese stock market. The limit-down policy was established 1996. A limit-down stock is a stock that decrease more than 10% in one day, which means (today's price – yesterday's price) / yesterday's price is less than -10%. If a stock becomes a limit-down stock in a particular day, the policy limits the lower bound of its price by -10% rate of return, so it cannot be traded in a lower price. Therefore, the possibility of selling limit-down positions is limited and it may influence the probability of selling bad-performance positions in my model. Thus, it can affect how investors sell their portfolios.

(Insert Table 4.13 here)

If there is a limit-down stock in one portfolio, it can influence the selling choice of the whole portfolio. Therefore, I delete the whole portfolio. In this part, only portfolios without limit-down stocks are included. I employ the same logit model in Table 4.3 using these portfolios. In Table 4.13, when a portfolio has no limit-down position, a best-ranked position is 12.18% more likely to be sold than a 2nd worst one, with a very large t-value (26.191). The probability of selling a 2nd best position is also significantly larger than selling a 2nd worst position. The preferences of selling middle and worst positions than 2nd worst positions are very tiny, smaller than 1%. This result is similar to my result in rank effect model. My result is robustness to the policy of limit-down stock.

4.6.6 Rank Effect Measured by Rate of Return

In this chapter, I use return of a position per share to measure the rank effect. In the real world, there are many factors that may draw investors' attention. Thus, these factors can lead investors to rank positions in their portfolio by other measurements. To test this, I change the measurement of rank into rate of return, which equals to (current price per share – purchase price per share) / purchase price per share. This measurement is commonly used in stock analysis.

(Insert Table 4.14 here)

When using rate of return as the measurement of rank, in Table 4.14, I find that the result is robust to my main result. Best rank stock has the largest probability to be sold, which is 12.18% more likely than 2nd worst with t-value 26.191. 2nd best stock follows with a 5.80% more likely to be sold than worst stock and also large t-value. The preferences of selling middle and worst positions than 2nd worst positions are very tiny, smaller than 1%. This indicate that my result is robust when I change the measurement of rank from return to rate of return.

4.6.7 Rank Effect by Day Fixed Effect

To further control the time and market condition, I apply day fixed effect to rank effect and top effect. The results are robust to the fixed effect method.

(Insert Table 4.15 here)

4.6.8 Future Returns of Rank Effect

To investigate the asset pricing value of rank effect in China, I use a very simple method (similar to Odean, 1998) to briefly test if rank effect benefit or causes damage to investors' profit. I first calculate average future returns of every position ranks. Then I calculate the differences of future returns among each rank and the Bonferroni-adjusted significance. In Table 4.16, for one week and one month later returns, good-rank positions today still have larger returns in the future and all the differences are significant. For one year later returns, although 2nd best positions will outperform best positions, these two rank positions are still significantly better than other rank positions. Therefore, the best rank position and 2nd best position in the portfolio have a large probability to have better returns in the future than other positions. The rank effect in China, willing to sell best and 2nd best positions, causes damage to investors' profit. And it is indeed a behavioral bias.

(Insert Table 4.16 Here)

4.7 Discussions and Conclusions

In this chapter, I test rank effect among individual investors in Chines stock market before, during and after financial crisis. I further develop rank effect into top effect and discuss in what situation investor sell bad-performance positions. I also support that disposition effect is significant in China.

The rank effect in Chinese market is different from US market. The positions with best performance in one's portfolio have the largest probability to be sold. The 2nd best on follows. However, the probabilities of selling middle, 2nd worst and worst positions are not significantly different. In general, investors want to sell positions with better return. Furthermore, I find that when the positions are lottery like and the investors are young and male, investors in China have the preference of selling the worst rank position as well. Thus, selling the worst position can be explained by the willing of gamble to some extent. This result is also similar in different market conditions and on different holding period positions. I also document that senior, female investors with long trading experience and less trading frequency are more likely to sell good rank positions.

Since one main limitation of rank effect is that it requires at least 5 stocks in the portfolio, I introduce top effect and significantly improve this number to 2. When there are more than 1 stocks in investors' portfolio which means they can choose when selling, investors have a significant preference of choosing the top one performance position. The top one is choices of selling.

In my study, I try to discover how investor choose which position to sell when they want to sell their stocks. I find investors are more willing to sell positions with better performance, especially the top one position. In further studies, I aim to employ all the models I run on rank effect to test top effect. I will try to find more evidences and reasons of top effect. In addition, since investors do not behave significant differently before, during and after financial crisis, how to explain this and how to explain the tiny difference in different market conditions are also good questions.

Table 4.1: Data Preparation Process

	Data Preparation Process
Original	56,561,915
(percentage remain)	(100%)
Active selling-day (at least one stock	7,554,613
is sold on that day in the portfolio)	(13.36%)
Date before 2009-06-01	7,140,349
	(94.52%)
Delete positions form the first day	6,565,088
	(91.94%)
Delete trading price smaller or	6,271,589
equals to 0	(95.53%)
Delete number of shares in a	6,071,948
position < 100	(96.82%)
Delete positions not in Shanghai	6,028,922
and Shenzhen exchange	(99.29%)
Delete institutional investors	6,023,504
	(99.91%)
Delete positions with extreme	6,023,159
return (>200 or <-200)	(99.99%)
Delete positions with inappropriate	5,352,904
stock disclosure information	(88.87%)
Delete investors with 0 turnover	5,352,894
rate in customer file	(99.99%)
Delete positions which stock	5,220,908
variance over last year is null value	(97.53%)
Delete positions which cannot	4,597,491
calculate holding period	(88.06%)
Keep active selling-day again	4,065,596
	(88.43%)
Portfolio contains at least 2 stocks	3,881,318
(to test top effect)	(95.47%)
Portfolio contains at least 5 stocks	2,130,356
(to test rank effect)	(54.89%)

Note: This table presents the process of data preparation. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. The details of each step are described in section 3.2 in this paper. The top number is the number of observations(positions) after executing the particular data cleaning step. The lower number in parentheses is percentage of observations remain after this step.

Table 4.2: Summary Statistics

	Data after Preparation	
Observation	3,881,318	
Dummy variable	Number of 1s	
Selling	1,401,880	
Top	943,135	
Gain	1,492,565	
Gender	1,792.979	
New_investor	2,113,723	
Bull_mar	728,198	
Bear_mar	1,720,882	
Steady_mar	1,432,238	
Numerical variable	Average	S.D.
Return	-1.1941	4.1177
Root_holding_period	5.8599	5.9649
Variance	0.8453	2.7330
Portfolio_size	6.3312	5.0873
Root_tradetimes	26.4244	16.8250
Root_age	6.5727	0.8792

Note: This table presents the summary statistics of variables in my model. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. An investor must hold at least two stocks to be included in the data. All variables are defined in section 3.1 in this paper.

Table 4.3: The Test of Rank Effect

	Dependent Variable: Dummy of Selling the Position		
Benchmark:	Middle	Worst	2^{nd} worst
	(1)	(2)	(3)
Best	0.1292***	0.1266***	0.1353***
(t-statistics)	(33.8272)	(21.4500)	(27.8662)
2^{nd} best	0.0551***	0.0527***	0.0607***
	(25.5293)	(11.4488)	(17.4442)
Middle		-0.0022	0.0052**
		(-0.6700)	(2.4699)
2^{nd} worst	-0.0052**	-0.0074***	
	(-2.4782)	(-3.8849)	
Worst	0.0022		0.0074***
	(0.6687)		(3.8345)
Gain	0.0866***	0.0866***	0.0866***
	(17.4740)	(17.4740)	(17.4740)
Gain*return	-0.0044**	-0.0044**	-0.0044**
	(-2.5478)	(-2.5478)	(-2.5478)
Loss*return	-0.0001	-0.0001	-0.0001
	(-0.0431)	(-0.0431)	(-0.0431)
Control variables	Yes	Yes	Yes
Observations	2,130,354	2,130,354	2,130,354
Pseudo R^2	0.0575	0.0575	0.0575

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 uses rank middle as benchmark. Column 2 uses worst. Column 3 uses 2nd worst. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.4: Rank Effect with Gambling Positions and Aggressive Investors

	Dependent Variable: Dummy of Selling the Position		
	Gambling Positions	Aggressive Investors	Gambling Positions & Aggressive Investors
	(1)	(2)	(2)
Best	0.0900***	0.1277***	0.1461**
(t-statistics)	(3.8176)	(15.9853)	(2.2086)
2nd best	0.0260	0.0559***	0.0728
	(1.2768)	(9.1048)	(1.1526)
Middle	-0.0232	0.0017	0.0676
	(-1.3479)	(0.4291)	(1.2468)
Worst	0.0191	0.0180***	0.1154
	(0.7499)	(3.6865)	(1.6227)
Gain	0.1292***	0.0796***	0.0694
	(5.1520)	(11.2871)	(0.7418)
Control variables	Yes	Yes	Yes
Observations	7,145	196,225	853
Pseudo R^2	0.0732	0.0441	0.0741

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. A portfolio needs to have at least 2 positions to be included. If a position is short period (holding days less than 20), low on unit share price (unit share price less than 30 percental of all stocks price that day) and high on volatility (return variance of last 250 days above 70 percental of all stocks that day), it is defined as a gambling position and is included in column 1. In column 2, I only include aggressive traders (<35, male and new investor). In column 3, based on requirements in column1 and 2, I combine them together. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. The dependent variable is a dummy variable equal to one if a stock is sold. Rank variables are dummy variables equal to 1 if the position is that particular rank in the portfolio. Gain is a dummy variable indicating a positive return. Other variables are control variables and are defined in section 3.1 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.5: The Test of Top Effect

	Dependent Variable: Dummy of Selling the Position	
Portfolio size:	2+	5+
	(1)	(2)
Top	0.1353***	0.1083***
(t-statistics)	(35.3203)	(32.4736)
Gain	0.0900***	0.0980***
	(15.5655)	(19.1316)
Gain*return	0.0040**	-0.0030*
	(2.4118)	(-1.7687)
Loss*return	-0.0063**	0.0010
	(-2.3354)	(0.3658)
Control variables	Yes	Yes
Observations	3,881,318	2,130,354
Pseudo R^2	0.0747	0.0560

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. If the portfolio size of an investor is larger or equal to 2, it is included in column 1. If the portfolio size of an investor is larger or equal to 2, it is included in column 1. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. The dependent variable is a dummy variable equal to one if a stock is sold. If the position has the largest unit share return in the portfolio, the dummy variable top takes value of 1. Otherwise, it is classified into other and is applied as benchmark here. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.6: Rank Effect in Different Market Conditions

	Dependent Variable: Dummy of Selling the Position		
Market Condition:	Bull	Bear	Steady
	(1)	(2)	(3)
Best	0.1194***	0.1252***	0.1277***
(t-statistics)	(20.5903)	(14.6289)	(30.1280)
2nd best	0.0709***	0.0500***	0.0654***
	(14.6079)	(8.7292)	(17.4247)
Middle	0.0179***	-0.0023	0.0083***
	(3.7878)	(-0.8308)	(2.7749)
Worst	0.0042	0.0094***	0.0108***
	(1.2062)	(3.8770)	(3.8090)
Gain	0.0755***	0.0768***	0.0858***
	(12.5693)	(8.6587)	(13.5603)
Gain*return	-0.0031*	-0.0022	-0.0095***
	(-1.6450)	(-0.5831)	(-3.1459)
Loss*return	-0.0093*	-0.0009	0.0139***
	(-1.9437)	(-0.2232)	(6.2582)
Control variables	Yes	Yes	Yes
Observations	367,635	965,389	797,330
Pseudo R^2	0.0352	0.0486	0.0846

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 is the test of rank effect using middle positions as benchmark. Column 2 uses worst positions as benchmark. Column 3 uses 2nd worst positions as benchmark. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. All control variables except market condition variables (bull_market, bear_market) are included. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.7: Rank Effect by Different Holding Period

	Dependent Variable: Dummy of Selling the Position		
Holding Period (day):	1-20	21-250	250+
	(1)	(2)	(3)
Best	0.1361***	0.1453***	0.1436***
(t-statistics)	(22.3917)	(21.6307)	(13.4621)
2nd best	0.0566***	0.0657***	0.0843***
	(11.4117)	(17.5677)	(12.5418)
Middle	0.0001	0.0046**	0.0205***
	(0.0363)	(2.1082)	(6.2769)
Worst	0.0156***	0.0049**	-0.0046
	(4.9930)	(2.2855)	(-1.3273)
Gain	0.0920***	0.0800***	0.0547***
	(13.6712)	(7.9029)	(4.8731)
Gain*return	-0.0083***	-0.0116***	-0.0123***
	(-4.2908)	(-8.4789)	(-6.3757)
Loss*return	-0.0039**	0.0061***	0.0031***
	(2.2781)	(4.8816)	(3.7467)
Control variables	Yes	Yes	Yes
Observations	1,052,895	803,872	273,587
Pseudo R^2	0.0351	0.0504	0.0685

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 is from positions that their holding period is less than 20 days. Column 2 is from positions with holding period 21-250 days. Column 3 is from positions that had been held for more than 250 days. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. All control variables except holding period variables (root_holding_period, root_holding_period*return*gain, root_holding_period*return*loss) are included. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.8: Rank Effect and Investor Characteristics

	Dependent Variable: Dummy of Selling the Position	
	Continuous age	Age group
	(1)	(2)
Best	0.1337***	0.2096***
(t-statistics)	(6.7547)	(18.8585)
2^{nd} best	0.0563***	0.0859***
	(3.7228)	(11.3848)
Middle	0.0142	-0.0033
	(1.2953)	(-0.6279)
Worst	0.0334***	-0.0042
	(2.7184)	(-0.8083)
Gain	0.0872***	0.0872***
	(17.6649)	(17.6642)
Gain*return	-0.0035**	-0.0036**
	(-2.0299)	(-2.0465)
Loss*return	-0.0001	-0.0001
	(-0.0457)	(-0.0440)
Best*root_age	0.0083***	
	(3.3344)	
2^{nd} best*root_age	0.0035*	
	(1.7299)	
Middle*root_age	-0.0014	
	(-0.8351)	
Worst*root_age	-0.0041**	
	(-2.5194)	
Best*young		-0.0159***
(age<35)		(-2.8671)
2^{nd} best*young		0.0066
(age<35)		(-1.3256)
Middle*young		0.0101*
(age<35)		(1.8679)
Worst*young		0.0168***
(age<35)		(2.7254)
Best*middle-age		-0.0075*
(35<= age<55)		(-1.9162)
2^{nd} best*middle-age		-0.0035
(35<= age<55)		(-1.0082)
Middle*middle-age		0.0102***
(35<= age<55)		(2.6155)
Worst*middle-age		0.0102**
(35<= age<55)		(2.0566)
Best*gender	0.0292***	0.0293***
	(6.9775)	(6.9803)

2^{nd} best*gender	0.0192***	0.0192***
	(5.6920)	(5.6907)
Middle*gender	0.0042*	0.0044*
	(1.6707)	(1.7819)
Worst*gender	-0.0031	-0.0031
	(-1.2063)	(-1.1953)
Best*new_investor	-0.0281***	-0.0297***
	(-6.6580)	(-7.0213)
2^{nd} best*new_investor	-0.0234***	-0.0241***
	(-6.3741)	(-6.5712)
Middle*new_investor	-0.0105***	-0.0113***
	(-3.6177)	(-3.8356)
Worst*new_investor	0.0029	0.0023
	(1.0193)	(0.7922)
Best*root_trade_times	-0.0018***	-0.0018***
	(-7.4812)	(-7.5189)
2^{nd} best*root_trade_times	-0.0005***	-0.0005***
	(-2.8404)	(-2.8352)
Middle*root_trade_times	0.0001	0.0001
	(0.9667)	(1.1206)
Worst*root_trade_times	0.0000	0.0000
	(0.4878)	(0.0658)
Root_age	-0.0104***	-0.0086***
	(-5.9640)	(-5.4330)
Gender	-0.0175***	-0.0176***
	(-5.2383)	(-5.2418)
New_investor	0.0257***	0.0266**
	(4.8767)	(4.9075)
Root_trade_times	0.0020***	0.0020***
	(4.6741)	(4.6744)
Control Variables	Yes	Yes
Observations	2,130,354	2,130,354
Pseudo R^2	0.0586	0.0586

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. I employ a continuous variable the square root of age to measure age in column 1. I use age group in column 2. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by date and investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.9: Rank Effect with Small Size Portfolios

	Dependent Variable: Dummy of Selling the Position		
Portfolio Size:	2	3	4
	(1)	(2)	(3)
Best	0.1305***	0.1524***	0.1339***
(t-statistics)	(23.346)	(30.220)	(30.142)
2^{nd} best			0.0429***
			(16.694)
Middle		0.0466***	
		(16.035)	
Worst			-0.0196***
			(-6.941)
Gain	0.0799***	0.0813***	0.0850***
	(11.767)	(12.510)	(13.270)
Gain*return	0.0383***	0.0147***	0.0055***
	(10.934)	(6.630)	(2.660)
Loss*return	-0.0267***	-0.0162****	-0.0107***
	(-9.237)	(-5.165)	(-3.528)
Control variables	Yes	Yes	Yes
Observations	526,458	635,012	589,494
Pseudo R^2	0.0394	0.0277	0.0282

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. If a portfolio contains 2 stocks, it is included in column 1. If a portfolio contains 3 stocks, it is included in column 2. If a portfolio contains 4 stocks, it is included in column 3. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. The dependent variable is a dummy variable equal to one if a stock is sold. Rank dummy variables are included to test rank effect. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.10: Rank Effect by Probit Model

	Dependent Variable: Dummy of Selling the Position
	(1)
Best	0.1265***
(t-statistics)	(31.602)
2^{nd} best	0.0595***
	(19.539)
Middle	0.0044**
	(2.539)
Worst	0.0068***
	(4.152)
Gain	0.0846***
	(18.749)
Gain*return	-0.0052***
	(-3.349)
Loss*return	0.0010
	(0.474)
Control variables	Yes
Observations	2,130,354
Pseudo R^2	0.0575

Note: This table presents the marginal effect from probit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.11: Rank Effect in All Gain/Loss Portfolios

	Dependent Variable: Dummy of Selling the Position	
Portfolios:	All gain	All loss
	(1)	(2)
Best	0.1100***	0.1142***
(t-statistics)	(3.732)	(31.757)
2^{nd} best	0.0478***	0.0555***
	(3.543)	(17.997)
Middle	0.0094	0.0083***
	(0.838)	(3.624)
Worst	-0.0096	-0.0024
	(-1.147)	(-0.887)
Return	0.0034	0.0003
	(0.619)	(0.201)
Control Variables	Yes	Yes
Observations	78,302	378,843
Pseudo R^2	0.0239	0.0447

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. If the whole portfolio of one investor in one day are gain positions, the data is included in column 1. If the whole portfolio of one investor in one day are loss positions, the data is included in column 2. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. Control variables in this table includes root_holding_period, root_holding_period*return, stock variance, market condition variables and demographic variables. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.12: Rank Effect Examining Liquidation Rather than Selling

	Dependent Variable: Dummy of Liquidation
	(1)
Best	0.0677***
(t-statistics)	(13.361)
2^{nd} best	0.0259***
	(7.050)
Middle	-0.0032
	(-1.635)
Worst	0.0097***
	(4.843)
Gain	0.0588***
	(13.661)
Gain*return	-0.0090***
	(-4.632)
Loss*return	0.0029
	(0.946)
Control variables	Yes
Observations	2,130,354
Pseudo R^2	0.0696

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. The dependent variable is a dummy variable equal to one if a stock is liquidated. 5 dummy rank variables are included to test rank effect. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.13: Rank Effect from Portfolios without Limit-Down Stocks

	Dependent Variable: Dummy of Selling the Position
	(1)
Best	0.1218***
(t-statistics)	(26.191)
2^{nd} best	0.0580***
	(16.709)
Middle	0.0055***
	(2.795)
Worst	0.0077***
	(3.972)
Gain	0.0817***
	(17.545)
Gain*return	-0.0050***
	(-3.085)
Loss*return	-0.0012
	(-0.384)
Control variables	Yes
Observations	2,011,191
Pseudo R^2	0.0562

Note: This table presents the marginal effect from logit regression. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor- stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. If the whole portfolio of one investor has no limit-down stock, the data is included. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by date. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.14: Rank Effect Measured by Rate of Return

	Dependent Variable: Dummy of Selling the Position
	(1)
Rate_best	0.1313***
(t-statistics)	(33.141)
Rate_2nd best	0.0660***
	(22.046)
Rate_middle	0.0080***
	(4.424)
Rate_worst	0.0066***
	(3.463)
Gain	0.0774***
	(17.392)
Gain*return	-0.0027*
	(-1.807)
Loss*return	-0.0002
	(-0.076)
Control variables	Yes
Observations	2,130.354
Pseudo R^2	0.0587

Note: This table presents the marginal effect from logit regression. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor- stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rate_rank variables (measured by rate of return instead of return) are included to test rank effect. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.15: Rank Effect and Top Effect by Fixed Effect

	Dependent Variable: Dummy of Selling the Position	
	Rank Effect	Top Effect
	(1)	(2)
Best	0.1200***	
(t-statistics)	(28.5009)	
2^{nd} best	0.0506***	
	(16.5505)	
Middle	0.0009	
	(0.4339)	
Worst	0.0110***	
	(6.3896)	
Top		0.1253***
		(37.8873)
Gain	0.0979***	0.1032**
	(20.7448)	(18.1111)
Control variables	Yes	Yes
Day FE	Yes	Yes
Observations	2,098,291	3,809,888
Pseudo R^2	0.0670	0.0814

Note: This table presents the marginal effect from logit regressions controlled by day fixed effect. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 examines rank effect. Column 2 examines top effect. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. If the position has the largest unit share return in the portfolio, the dummy variable top takes value of 1. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share return since purchase. Other variables are control variables and are defined in section 3.1 in this paper. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 4.16: Rank Effect and Future Returns

	Best	2nd best	Middle	2nd worst	Worst
	(1)	(2)	(3)	(4)	(5)
One Week Later Return	1.1696	0.0331	-1.5682	-3.1147	-5.2134
Row – Col					
(Bonferroni)					
2nd best	-1.1365				
	(0.000)				
Middle	-2.7379	-1.6014			
	(0.000)	(0.000)			
2nd worst	-4.2842	-3.1478	-1.5464		
	(0.000)	(0.000)	(0.000)		
Worst	-6.3830	-5.2466	-3.6452	-2.0988	
	(0.00)	(0.000)	(0.000)	(0.000)	
One Month Later Return	0.6672	-0.2820	-1.8410	-3.4831	-5.6691
Row – Col					
(Bonferroni)					
2nd best	-0.9492				
	(0.000)				
Middle	-2.5082	-1.5591			
	(0.000)	(0.000)			
2nd worst	-4.1503	-3.2012	-1.6421		
	(0.000)	(0.000)	(0.000)		
Worst	-6.3363	-5.3872	-3.8281	-2.1860	
	(0.000)	(0.000)	(0.000)	(0.000)	
One Year Later Return	-2.9559	-2.5218	-3.4244	-5.6993	-7.7755
Row – Col					
(Bonferroni)					
2nd best	0.4341				
	(0.000)				
Middle	-0.4685	-0.9026			
	(0.000)	(0.000)			
2nd worst	-2.7434	-3.1775	-2.2749		
	(0.000)	(0.000)	(0.000)		
Worst	-4.8196	-5.2537	-4.3511	-2.0762	
	(0.000)	(0.000)	(0.00)	(0.000)	

Note: This table presents the result of average unit share returns, differences and Bonferroni-adjusted significance. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. Returns are calculated as the difference of the unit share price in future and unit share cost. Column 1 to 5 shows the 5 ranks of positions respectively. The future return is presented followed by the matrixes of differences among different ranks and the Bonferroni-adjusted significance.

Appendix 4.1: The Univariate T-test Result of Rank Effect

By a similar method with Hartzmark (2015), I run a similar t-test to test the rank effect based on a random sample of 5,000 investors. I rank the positions in one's portfolio by the unit share return as best, 2nd best, worst, 2nd worst and middle. A position is ranked best if it has the highest unit share return in the portfolio of that particular investor in that particular day. 2nd best, worst and 2nd worst are defined in a similar way. Middle includes all positions not ranked in the top or bottom two positions. For investors, only days that at least one stock is sold are included as sell day (active day). The observations are at day-investor-stock level. I define *Best%* as:

$$Best\% = \frac{\#Best\ Sold}{\#Best\ Sold + \#Best\ Not\ Sold}$$

#Best Sold is the number of best stocks that had their number of shares decreased. *#Best Not Sold* is the number of best stocks that had their number of shares increased or remained the same. *2nd Best%, Middle%, 2nd Worst%* and *Worst%* are defined in a similar way. To calculate the t-statistics, I cluster the data by investor and date and calculate the average.

(*Insert Appendix Table 4.1 here*)

In Appendix Table 4.1, I present the result. In Chinese stock market, the order of probability of selling from large to small is the same order of the rank of return. A best position with a 32% probability of selling is 7.05% more likely to be sold than a 2nd best position, with a very large t-statistics. In a similar way, a 2nd best position is more likely to be sold than a middle one, a middle position is more likely to be sold than a 2nd worst one and the worst position has the lowest probability to be sold. However, when I compare the difference between ranks, although all differences pass statistic test, the difference between best and 2nd best is 7.05% and the difference between 2nd best and middle is 6.89%. These two differences are more than 2 times larger than the rest two differences. In further results, when control variables are added, the differences among middle, 2nd worst and worst become insignificant. The best stock is the one that is most likely to be sold by Chinese investors, and the 2nd best stock follows. The rest stocks, middle, 2nd worst and worst, are treated similarly. As a conclusion, Chinese investors during financial crisis are more likely to sell a position with better performance. In Hartzmark (2015), the US investors are more likely to sell best and worst positions than the middle one. My result in Chinese market is significantly different.

To compare with results from Hartzmark (2015), the average probability of selling for all stocks is 12.1%. My result is 19.67%. All other selling percentage is larger as well. This is not because that Chinese investors are more likely to trade. This is because both of the papers only include selling day position, which means on each day that is included, at least one position is sold. Since the average portfolio size of Chinese investor is significant smaller than US investor and at least one stock is sold in one portfolio one day, the average selling probability in Chinese market is of course larger than it is in US.

Appendix Table 4.1: The Univariate Test of Rank Effect

	(1)
Best%	0.3189
2nd Best%	0.2484
Middle%	0.1795
2nd Worst%	0.1404
Worst%	0.1194
All	0.1967
Best% - 2nd Best%	0.0705***
	(10.3105)
2nd Best% - Middle%	0.0689***
	(12.4040)
Middle% - 2nd Worst%	0.0391***
	(8.1051)
2nd Worst% - Worst%	0.0210***
	(4.1094)
Observations	8,625

This table presents the t-test result of rank effect. The data contains daily holding records of 5,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Best% is calculated as the ratio of best positions that are sold divided by all best positions. Others are defined in a similar method. The data is clustered by investor and date. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included. In the second half of this table, the top number is the difference in average, and the lower number in parentheses is the t -statistic. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Chapter Five: The Impact of Location and Local Stock on Individual Investors Biases

5.1 Introduction

The individual investors behavioral biases have been widely discussed by a number of literatures. Among all the biases, the most well-known one is the disposition effect. Odean (1998) is the first to provide empirical result of disposition effect. The result demonstrates that investors realize their gains more readily than their losses. In the theory, it divides positions into gains and losses, but it fails to discuss the magnitude of gain and loss. Ben-David and Hirshleifer (2012) further discover that US investors prefer to realize large gains and large losses. This phenomenon is name after V-shaped disposition effect. Meanwhile, using the data from Finnish market, Kaustia (2010b) supports this result on the gain side, while he finds the probability of selling on the loss side is constant. These studies consider the return of the positions separately and ignore the comparison among positions in one's portfolio. Hartzmark (2015) develops the rank effect to investigate this comparison. He finds that when sorting positions from one portfolio by return from large to small, US individuals are more likely to sell extreme wining and extreme losing positions, which are also the best and worst performance positions in the portfolio. In chapter 3 and 4 of this thesis, I discuss V-shaped disposition effect and rank effect in China. In this chapter, I will investigate these theories from a new perspective.

The impact of geographic factors on behavioral finance has been commonly studied by literatures as well. Investors in the same or near area and community trend to trade similarity. They hold highly related portfolios (Hong, Kubik, and Stein, 2002; Feng and Seasholes, 2004; Brown, et.al, 2008) and have similar trading timings (Baltakys et al., 2019). And when dealing with local stocks, individuals prefer to trade local stocks. Grinblatt and Keloharju (2001) document Finish investors tend to trade stocks that are headquartered close to their location. Ivkovic and Weisbenner (2005) and Seasholes and Zhu (2010) discover similar phenomenon in US market. This preference is called local bias.

In this chapter, I will combine these two effects together. Among what I know, I am the first to do this. By using a very large and unique dataset from China, I discuss how the geographic factors impact on disposition effect, V-shaped disposition effect and rank effect. Metropolis investors suffer more from disposition effect. And under theories of V-shaped disposition effect and rank effect, their behaviors have no significant difference with other investors. Furthermore, investors in eastern, mid and western China do not have significant difference on trading behavior of these three effects as well. Investors in different regions do not trade differently. Thus, investors in the same or close region do not trade relative similarly to some degree. My results cannot fully support previous literatures on this.

For local stock trading, individual investors do not sell local positions in different levels of disposition effect and V-shaped disposition effect. However, local stock can moderate rank effect. When the position is local, investors have less probability to suffer in rank effect when they make choice of selling among positions in their portfolio. Therefore, I indicate that local stock plays a crucial role in the comparison system when individual investors want to choose a position to sell. Furthermore, the local effect is also the first effect in literature which can moderate rank effect with empirical evidence. I also find evidence that rank effect is bias that cause damage to investors' profit when the position is local. Thus, moderating rank effect by local positions helps investor performance to some degree.

The rest of this chapter is organized as follows. Section 2 is a brief introduction of previous related literatures. Section 3 presents the data and methodology. Section 4 shows the mainly empirical results. Section 5 is the conclusion and discussion and also provide thoughts for further studies.

5.2 Literature Review

The theory of location and investor behavior comes from the studies of correlated trading in herding effect of finance. Grinblatt, Titman, and Wermers (1995) discover causal-link to tradings among fund managers. They find managers tend to buy and sell stocks at the same time. Lakonishok et al. (1992) provides evidences of more correlated trading in small stocks than in the average stock. They indicate stocks that herd buy perform better than stocks that herd sell. Hong, Kubik, and Stein (2002) present that trades from investors within a region are highly correlated. They suggest that the similar information diffused throughout a region cause the effect.

Brown, et.al (2008) further establish a causal community effect and find evidence on the impact of social interaction on investors trading decision. By testing the similarity of average portfolio ownership of an individual's community with lagged average ownership of one's neighbors, the authors discover that investors in the same community are more likely to trade similar positions.

In a study more related to this chapter, Feng and Seasholes (2004) discovers investors in one region of the country tend to trade in a similar way. Data using in Feng and Seasholes (2004) is individual investor trading records data from a Chinese brokerage in 1999 to 2000. They consider investors from 4 branches of this brokerage from Guangdong and three branches from Shanghai. They also only include top 25 highest-volume stocks as measured by total trading value which have their headquarters in Guangdong. Thus, for Guangdong investors, these stocks considered in the paper are all local stocks. In existing literatures, investors have different trade preference on local and non-local stocks. This sample bias could cause bias in results. Their draw the conclusion, isolated groups of investors in one region of the country tend to buy and sell together, and different region investor trade differently. However, this conclusion could be caused by the fact that Guangdong investors are trading local stocks in their model, while Shanghai investors do not. In addition to this, both Shanghai and Guangdong are metropolis with high income level in China. The trading behaviors of investors in towns and rural regions in China is the question remain. In this chapter, the data I applied is from the entire China and entire Chinses stock market. I overcome the biases in Feng and Seasholes (2004) and continue to discover more on this topic.

In a recent study, Baltakys et al. (2019) investigate trade timing of investors in a short geographical distance. By using a unique data from Finland, the authors find neighboring investors have similar trading behavior and trade timing. In their analysis, they put this effect down to the sharing of information between investors in the same region and discover the information transfer channels are used in trading decision making.

Since previous studies indicate similar buying/selling and trade timing behavior in the same region. In this chapter, I first discuss the relationship between region in China and the disposition effect, the V-shaped disposition effect and the rank effect.

Question 1: Is region factor affect the magnitude of investors behave in disposition effect and V-shaped disposition effect in China?

Question 2: Is region factor affect the magnitude of investors behave in rank effect in China?

Beside the discussion of trading behavior and region, there are a few papers in geographic behavior finance discuss how investors trade stocks that are located close to them. Grinblatt and Keloharju (2001) document Finish investors tend to trade stocks that are headquartered close to their location, that use their first language to communicate and that have CEO is of similar cultural origin. This effect is moderated when the stock is from Helsinki (the capital city of Finland, thus the stock is more nationally known) and the investor is sophisticated. The authors also indicate the reason of this local effect is that investors can get more useful information from familiar firms, so they are more willing to trade these stocks.

Ivkovic and Weisbenner (2005) find similar result in US. By analyzing US individual investors from 1991 to 1996, they discover that households exhibit a strong preference for local stocks. Individual's local investments outperform their nonlocal investments, which indicates that individuals have ability to get asymmetric information from their local companies. This effect is stronger among stocks not in S&P 500 index. These stocks are thought to have more asymmetric information.

However, Seasholes and Zhu (2010) present a different result, which is portfolios of local holdings do not generate abnormal performance. By using calendar-time portfolio method, they overcome fmy pitfalls from previous studies: 1) cross- sectional correlation of portfolio returns, 2) small stock effects in individual portfolios, 3) geographic selection biases, and 4) time-series selection biases. This result indicates that individual investors do not have asymmetric information of the local stocks they are more willing to hold and trade.

From previous studies, there is no doubt that individual investors prefer local stocks. However, the question of relation of local stock investment and return remains. In this chapter, beside this question, I combine the local effect with disposition effect, V-shaped disposition effect and rank effect. I discuss the difference of the magnitude of these behavior biases among local and nonlocal stocks. I am willing to figure out if the familiarity of local stocks could moderate these biases. Furthermore, I will also discuss the impact of this issue on investor profit.

Question 3: Do the magnitude of disposition effect and V-shaped disposition effect differ among local and nonlocal stocks in China?

Question 4: Does the magnitude of rank effect differs among local and nonlocal stocks in China?

Question 5: Are these behavior biases (disposition effect, V-shaped disposition effect and rank effect) cause damage on investor profit from local stocks?

5.3 Data and Methodology

5.3.1 Database and the definition of regions in China

The data applied in this chapter is similar to the database in last two chapter, which is a very large database collected from a large nationwide brokerage firm in China, with more than 3 million accounts and 2 billion daily dealing records over the period of January 2007 to May 2009. Due to the computational capacity limitations, I use a random sample of 100,000 investors and more than 56 million records sub-data to build my model. I also get the stock price and volatility information from CSMAR (China Stock Market & Accounting Research Database).

I acquire the location information of individual investors from their ID that are contained in the customer file of my dataset. In the ID system in China, the first 6 numbers are codes for the administrative divisions, which indicate the location information of an individual. The first and second numbers are code for provincial-level administrative regions (including provinces, municipalities and special administration regions, using province for short in the rest of this chapter), the third and fourth numbers are code for prefecture-level administrative regions (using city for short in the rest of this chapter), and the fifth and sixth numbers are for county-level administrative region (including districts, countries, etc.). I use the table of this code and location from Chines government to match with the ID in my dataset in order to get location information of the individual investor. This location is the place where the individual register for his/her ID card. This is the place of birth for most people. Therefore, if one birth at one place then migrates, lives and invests at another, I cannot get this information. In China, the number of migrant people reached 221 million in 2010, accounting for 16.5% of the total population[14]. For these 16.5% migrant people, I argue that people have more interest and pay more attention on stocks from their hometown. Using location information from investor's ID is appropriate in this research.

After the data preparation process of the last two chapter, I further delete investors that do not have correct ID card information as well as investors not in mainland China. This process deletes roughly 11% records. There are 3,604,961 trading records remain, which is applied to analysis v-shaped disposition effect and geographic factors. In the theory of rank effect, it requires the investor to have at least 5 stocks in their portfolio. Therefore, I only keep records that fulfill this requirement when analyzing geographic factors and rank effect. There are 1,891,112 trading records remain for rank effect analysis.

I further use the location information to split investors from different region. I apply two different method to divide region into groups. First, I define metropolis (large cities) and other regions. I state that Beijing, Shanghai, Tianjin, Chongqing, Guangzhou and Shenzhen are metropolis in China in this chapter. Beijing, Shanghai, Tianjin and Chongqing are the only fmy municipalities directly under the Central Government in China. Furthermore, during my data period, 2007 to 2009, these six cities are top six in GDP, balance of saving deposit and total volume of retail sales. These cities also take 6 places in the top 8 rank of the total output value of high and new tech enterprises. This indicates the relatively high education level of people who live in these cities. Beijing, Shanghai, Tianjin, Guangzhou and Shenzhen are also top 5 cities on residents' average incomes[15]. Furthermore, the only two stock exchange market in China located in Shanghai and Shenzhen. So, I indicate that investors in these cities have more potential information smyces. In conclusion, investors in these six metropolises have the top rank savings, incomes, consumption level, relatively high education level, more investment knowledge and skills and more potential information smyces. I will investigate the investment difference between investors in metropolises and other region under behavioral finance biases in this chapter. Second, I further divide regions into east, middle and west region in China. I

14 Reported by China's Migrant Population Development 2011, The Department of Services and Management of Migrant Population of National Population and Family Planning Commission of China
15 Data in this paragraph are from *China Statistical Yearbook* 2007, 2008 and 2009.

dividend all province into three groups east, middle and west[16]. This method of grouping is from National Development and Reform Commission in China. It considers not only the geographic location but also the economics and development level of the region. In China, the east region has the best level of economics and development level, followed by the middle region and the west region has the worst level. Therefore, I investigate the difference of investment behavior among investors from well-developed region, medium region and less-developed region.

In order to analysis how investors trade local stocks, I acquire the stock's location information from Chinese stock database, CSMAR. The data from CSMAR contains the stock ID, firm's name, the city that the firm registered in and the area code of the city. The city here is the prefecture-level city by the government administrative divisions. I use the stock ID to be the key to match the stock information table with investors' holding data in my dataset. For a particular investor holds a particular stock one day, which is one raw of data in my holding dataset, if the city code of the company registered in and the city code of the investor's ID card registered in are identical, I define a dummy variable local equals to 1. Therefore, in this chapter, a local stock is a position held by one investor, which the registered prefecture-level city of this investor and the registered prefecture-level city of this stock are identical.

Brown, et.al (2008) use Metropolitan Statistical Areas (MSA) as their definition of local communities in their research of investor behavior and local communities in US. This definition is also commonly used in other empirical literatures in economics and finance. The prefecture-level administrative region in China is similar to MSA in US, since it is also an administrative region contains a city and areas surrounding the city defined by the government. Grinblatt and Keloharju (2001) find investors in Finland have the preference of holding and trading stocks located near them than those far from them. And the relationship between the distance of firm and investor and the preference of holding and trading is piecewise linear with a break point at 100 kilometers. The average area of prefecture-level administrative regions is approximately 9,000 square kilometers; thus, it is a radius of approximately 96 kilometers.

16 East: Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong and Hainan. Middle: Shanxi, Neimenggu, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei and Hunan. West: Sichuan, Chongqing, Guizhou, Yunnan, Xizang, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang and Guangxi.

Therefore, my definition of local stock using the prefecture-level administrative region is supported by previous literatures.

The brokerage that offers database used in this chapter is headquarter in Nanjing (provincial capital city of Jiangsu province). There are more customers from Nanjing and Jiangsu province than other parts of China. 51.5% records in the dataset is from Jiangsu. Since I discuss how investors from different regions of China impact the selling behavior, I use region categories in this study. Jiangsu is in the east region of China and all of cities in Jiangsu are not metropolis. Therefore, by using the whole dataset, there are much more investors in east China and not in metropolis than other region categories. To avoid the data selection bias and check the robustness, in every statistical model, I applied the same model twice, once with the whole dataset and all records, and then with investors that do not live in Jiangsu. I show details in empirical result sections.

5.3.2 Empirical Modeling

In this chapter, I estimate an empirical model similar to Ben-David and Hirshleifer (2012) and my model in the last two chapter. For the research question of region factor and individual bias (research question 1 and 2), the Logit model is listed below as Equation 1 and 2:

$$Sell = Constant + a_1(Gain) + a_2(Gain * Return) + a_3(Loss * Return) + a_4(Region\ Variables) + a_5(Gain * Region\ Variables) + a_6(Gain * Return * Region\ Variables) + a_7(Loss * Return * Region\ Variables) + a_8(Control\ Variables) + \epsilon \quad (1)$$

$$Sell = Constant + a_1(Rank\ Variables) + a_2(Region\ Variables) + a_3(Region\ Variables * Rank\ Variables) + a_4(Gain) + a_5(Gain * Return) + a_6(Loss * Return) + a_7(Control\ Variables) + \epsilon \quad (2)$$

The model is on day-investor-stock level. Each observation is a position that one investor holds one stock in one day. The model is fitted as a Logit model by maximum likelihood, and also controlled by marginal effect and clustered standard error. The dependent variable, Sell, is a dummy variable equals to 1 if the position is sold that day by that investor and 0 otherwise. Both partial selling and liquidation are included. Return is the unit share return of position which is calculated based on the buying price (trading cost involved and is weighted average price by shares in case of multiple purchase) and the current price of that stock at that day. Gain is a dummy variable that takes the value of 1 if the unit share return of the position is positive and 0 otherwise. This controls the disposition effect. Loss is the opposite of Gain. The interaction terms of Gain (Loss) and Return indicate the impact of magnitude of gain or loss on individual selling decision (V-shaped disposition effect). Rank Variables are a set of dummy variables. I sort the positions in one portfolio of one investor one day by unit share returns from large to small. If the position has the largest return in that portfolio, the dummy variable Best is 1, otherwise 0. The other Rank Variables, 2nd Best, Middle, 2nd Worst and Worst are defined in similar way. By adding these Rank Variables, I control for the rank effect.

When discussing the topic of metropolis regions, Region Variables is a dummy variable equals to 1 if the investor is from a big city and 0 otherwise. When discussing the east, middle and west regions in China, Region Variables is a set of dummy variables equals to 1 if the investor is from east, middle, west of China respectively. I define big cities and east, middle and west regions in China in last session (session 3.1). Since I want to examine how the region factors influence disposition effect, V-shaped disposition effect and rank effect, I involve the interaction terms of the Region Variables and others. If the interaction term is positive and significant, investors in this region prefer to trade in a larger level of the particular trading bias than the baseline region. If the interaction term is negative and significant, the region factor in this region can moderate the trading bias. And if the interaction term is insignificant, it indicates that investors in this region have no different degrees of trading follow the particular trading bias than the baseline region.

For control variables, I also follow the choice from last two chapters. The effect due to holding period and volatility are controlled for. To control for the days a position is held from purchase to sell, the square root of the holding days (Root_holding_period) and interaction terms with gain dummy and return (Root_holding_period*return*gain) and loss dummy by return (Root_holding_period*return*loss) are included. To control for the stock variance, I also calculate the return variance of a stock over the last year (the variance of stock price over preceding 250 trading days, if there are at least 50 non-missing records). I include the interaction term Variance*gain and Variance*loss in my model. I further add investor characteristic control variables to control the investor heterogeneity influence. Gender is a dummy variable that takes value of 1 for female and 0 for male. Root_age is the square root of the investor's age[17]. To control the experience of investor, I introduce a dummy variable New_investor, which equals to one if the investor opened account in this brokerage after the start of my data period and zero otherwise. It is worth noting that at that time, in Chinese market, one individual can only open one account in the whole market. This makes my measurement of investor experience more powerful. Root_tradetimes is the square root of times of trading an investor made during my data period. It can indicate the activation of an investor in some degree. Portfolio_size is the number of stocks in one's portfolio that day.

In order to study the trading of local positions, I replace Region Variables into Local. Local is a dummy variable equals to 1 if the position is from a company which headquarter in the investor's city and 0 otherwise. The interaction term between Local and financial behavior bias variables indicates how local stocks have an impact on the investor trading decisions following the behavior biases (disposition effect, V-shaped disposition effect and rank effect). If the interaction term is positive and significant, investors suffer more in the particular trading bias when the trade local positions. If the interaction term is negative and significant, local positions can moderate the bias. And if the term is insignificant, local positions do not influence the trading bias in some degree.

(*Insert Table 5.1 here*)

Table 5.1 presents the summary statistics of my data and variables. After data preparation process, there are total 3,604,961 records. Each record is a record of one investor hold one position one day. There are 10.8% records from investors registered in metropolis. Meanwhile, 81.8%, 13.7% and 4.5% records are from investors from eastern, middle and western China respectively. Only 2.1% records are local positions. This percentage is much smaller than it is

17 I use the date difference between investor's birthday and May 31st 2009, which is the last day of my dataset.

in US. The local effect in China could be different. I will discuss it in detail in session 4.5. All other control variables are balanced and reasonable as well.

5.4. Main Result

5.4.1 Empirical Study of Metropolis Region and V-shaped Disposition Effect

In this session, I apply the model in session 3 to study how investor's region influences disposition effect and V-shaped disposition effect. In my model, I introduce the interaction terms of region variables and dummy variable Gain to analyze the impact of region on disposition effect. I also use the interaction terms of region variables and Gain (Loss)*return to study the impact of region on V-shaped disposition effect. The observations are at day-investor-stock level. All results are presented as marginal effects.

(*Insert Table 5.2 here*)

I first test whether investors in metropolis or not trade differently by following disposition effect and V-shaped disposition effect. Table 5.2 presents the empirical result from the logit model. In column 1, I include all investors. When facing the decision of selling a gain position, investors in large cities have 1.27% more probability to sell it (t-statistics 3.1395), which indicates that investors live in large cities suffer more in disposition effect. Meanwhile, the influence of the magnitude of gain or loss on selling has no difference between investors from large cities or not. The statistics of both Bigcity*gain*return and Bigcity*loss*return is insignificant. Since my data is from a nationwide brokerage firm which is headquartered in Nanjing (provincial capital city of Jiangsu province), 51% percent of my records is from Jiangsu. In order to check the robustness of this data bias, I delete all records from Jiangsu and apply the same model. In column 2, I show the empirical results. Investors in large cities have 1.73% (t-statistics 4.1123) preference to sell the position if it is a gain than investors not in large cities. The magnitude of gain and loss does not influence the probability of selling significantly differently among investors in large cities or not. These results in column 2 is similar to column 1. Thus, my result is robust to the data selection.

My result cannot support the theory in previous literatures. These literatures believe that investors in the same or near region are more likely to trade similarly (Brown, et.al, 2008; Feng and Seasholes, 2004). I find investors in metropolis suffer more from disposition effect. But the magnitude of gain and loss do not play different role in selling decisions among investors in metropolis or not (V-shaped disposition effect). Investors in metropolis or not trade similar under the theory of V-shaped disposition effect. Since metropolis investors have higher income and consumption level, better investment skills, higher education level, and more potential information smyces, they are thought to be more sophisticated. However, these virtues cannot help them from suffering in investment biases. I cannot support that investors in similar region, like metropolis, trades similarly.

5.4.2 Empirical Study of Metropolis Region and Rank Effect

I investigate how the region influence rank effect in this session. The theory of rank effect considers the choice of selling and the comparison of positions in one's portfolio. Therefore, it requires the portfolio to have at least 5 positions. I follow this requirement and build 5 dummy variables to test the rank effect. I also use the interaction terms of metropolis dummy variable and rank dummy variable to analyze the impact of region on rank effect.

(*Insert Table 5.3 here*)

Table 5.3 shows the marginal effect from the logit regression. Same to the last table, in column 1, the result is tested on all investors. Investors in the large cites group are 2.21% more likely to sell the best performance position in their portfolio than other investors with t-statistics 3.2429. The 2nd best performance are also 1.19% (t-statistics 2.2353) likely to be sold for metropolis investors than others. The difference on selling preference of middle rank positions, 2nd worst position and worst position from metropolis investors and other investors are all insignificant due to the t-statistics. In column 2, when I remove all records from Jiangsu to check the robustness of my result, I find similar result to column 1. The preference of selling the best and 2nd best position in their portfolio from investors in metropolis are significantly heavier than investors not in metropolis.

I find rank effect in China is that investors trend to sell their best and 2nd best performance positions, and I prove that it causes damage to their portfolio profit in the previous chapter. Therefore, investors in metropolis cannot perform better than other investors on avoiding the rank bias. Furthermore, here I support the previous studies on herding and geographic finance that investors in the same or similar region trade similarly. Under the theory or rank effect, investors in metropolis have similar selling choice that they are more likely to sell well-performing positions than investors not in metropolis.

5.4.3 Empirical Study of Eastern, Mid and Western China and V-shaped Disposition Effect

I divide all province in China into three region groups in session 3.1: east, middle and west, which indicate well-developed, medium-developed and less-developed regions in China respectively. In this session, I applied method similar to session 4.1 to discuss how investors in these region groups trade under the theory of disposition and V-shaped disposition effect. I use the interaction terms to test the influence of these region factors to disposition effect and V-shaped disposition effect.

(*Insert Table 5.4 here*)

Table 5.4 presents the result from logit model. In column 1, I applied the model by all investors. The interaction terms East*gain is very weakly significant in statistics and Mid*gain is not significant. This result indicates that the preference of selling the gain positions has very weakly significant or no significant difference among investors in east, middle and west part of China. In addition, investors in east and middle region group have very weakly or no significant difference on the preference of trading large gain position or loss position than west region investors. In column 2, I remove all records from Jiangsu since investors in Jiangsu are nearly a half in my dataset. The results are similar. Therefore, generally, investors in east, middle and west region group of China trade similarly under the theories of disposition effect and V-shaped disposition effect.

In summary, living in eastern, mid or western China do not cause significant difference in trading behavior under the theories of disposition effect and V-shaped disposition effect. Since eastern, mid and western China indicate well-developed, mid-developed and less-developed region respectively, the difference on level of development does not result in difference on investors' trading behavior and avowing behavioral biases.

5.4.4 Empirical Study of Eastern, Mid and Western China and Rank Effect

Following previous sections, in this session, I discuss the rank effect among east, middle and west region groups in China. Since the rank effect theory require investor's portfolio to have at least 5 positions, I follow this in this session. I use the model similar to session 4.2. I apply the interaction term to test the impact of region factors on rank effect as well.

(*Insert Table 5.5 here*)

I present the result of marginal effect in Table 5.5. In this model, I use the west region as the benchmark. In column 1, I include all investors. In all interaction variables of region dummy variables and rank dummy variables, only East*best and Mid*best are weakly significant, all others are not. Investors' preference on selling relatively good performance positions in their portfolio have no significant difference among different regions. In column 2, for the robust test of data selection, I remove all investors from Jiangsu. The results are similar. Under the theory of rank effect, investors from eastern, mid and western China trades similarly.

The level of preference of selling the good performance positions in one's portfolio has no significant difference among investors from east, middle and west region groups in China. Although I cannot support the previous literatures that discover investors from the same or near regions trade similarly, I argue that one reason could be that in this session, each my region group contains several provinces and cover a very large area. The previous literatures use cities or even communities instead. My focus here is the difference of development level among region groups. I find that born or living in a well-developed region cannot moderate the rank effect, which cause damage to investors' profit and is indeed a bias.

5.4.5 Empirical Study of Local Stocks and Disposition Effect, V-shaped Disposition Effect

In this section, I discuss how local stock influence disposition effect and V-shaped disposition effect. Local is a dummy variable equals to 1 if the position is a local stock. And local stock is the position that the registered city of the stock's firm and the registered city of the investor's ID card are identical. I apply a logit model similar to the previous sessions. I also add variable Local to test the local effect, whether investors are more willing to sell or hold local stocks. I further add interaction terms to analyze if local stock influence investor's selling decision under the theory of disposition effect and v-shaped disposition effect.

(*Insert Table 5.6 here*)

Table 5.6 presents the marginal effect of the logit regression model. In column 1, I test local effect and disposition effect. When I take the factor of local effect into consideration, investors are 12,90% more likely to sell a gain position with t-statistics 18.6477. This indicates the existence of disposition effect. The coefficient of Local is insignificant. Individual investors in China have no preference on selling or not selling (holding) local positions when they make trading decisions. My result goes against local effect. The interaction term Local*gain is negative and insignificant, which indicate that trading local stock can moderate disposition effect slightly. In column 2, I analyze how local effect affect V-shaped disposition effect. After adding the local effect terms, the asymmetric V-shaped disposition effect at gain side still exist. Investors prefer to sell large gains significantly. Meanwhile, local stock has no significant influence on V-shaped disposition effect, because both of the interaction terms are insignificant.

My result cannot support previous literatures on the preference of trading local stocks in Finland and US (Grinblatt and Keloharju, 2001; Ivkovic and Weisbenner, 2005; Seasholes and

Zhu, 2010). Individual investors in China have no preference on local stocks. The probability of trading *local* stocks and nonlocal stocks are even. This balance on trading choice have no significant influence on disposition effect and V-shaped disposition effect.

5.4.6 Empirical Study of Local Stocks and Rank effect

Individual investors suffer in rank effect (preference of selling the best performance and other good performance stocks in their portfolio rather than relatively bad performance stocks in their portfolio). In this section, I discuss whether investors trade local stocks also following ran effect or trading local stock can moderate rank effect. The interaction terms of dummy variable Local and a series of dummy variables Rank are applied in the model. I further fit the model with all nonlocal positions and all local positions.

(*Insert Table 5.7 here*)

In Table 5.7, I present the marginal effect of the logit model. I include all position records and applied interaction terms in column 1. After adding local variables, the preferences of selling best performance and 2nd best performance positions in the portfolio are still significant. Meanwhile, investors are 2.48% (t-statistics -2.0617) less likely to sell the best performance position in their portfolio if the best position is a local position than the benchmark position (2nd worst position). The preferences of selling the 2nd best position, middle position and the worst position decrease by 3.14% (t-statistics -2.9170), 3.14% (t-statistics -2.7933), 2.72% (t-statistics -2.9709) respectively as well if the position is a local one. Local stock can moderate rank effect significantly. In order to check the robustness of my result and discuss the rank effect in nonlocal stocks and local stocks, I involve only nonlocal positions and local positions in column 2 and 3 respectively. In column 2, the rank effect is still the same and significant if all positions are nonlocal. However, in column 3, when all positions are local position, the rank effect is different. After several unpresented tests of the choice of different benchmark variables, I choose the worst position as the benchmark of the series of rank dummy variables. Investors are still preferring to sell the best and the 2nd best performance positions in their portfolio when the position is local. However, the selling probably of the middle position and the worst position have no significantly difference. A 2nd worst position is 2.12% slightly more likely to be sold than a worst position with a t-statistics 2.1022. Therefore, when considering the comparation within the portfolio, the choice of selling positions is Best>>2nd best>>worst>middle>2nd worst when the position is nonlocal. And when the position is local, the choice is Best>>2nd best>>2nd worst>middle≈worst.

Local stocks can moderate rank effect. The differences among selling preference of different rank positions decrease when the position is local. Although the best performance and 2nd best performance positions still have larger probabilities to be sold, individual investors are more willing to keep the worst position if the position is local. Investors have more patient and loss tolerance for the relative bad-performing local stocks in their portfolio. Since investors could gain more asymmetric information from local firms (Ivkovic and Weisbenner, 2005), this could be the reason of this patient and tolerance. The favor of the local firms or the more confidence to local firms could also be the reason.

5.4.7 Empirical Study of Local Stocks and the Impact of Investor Biases on Position Future Return

Previous literatures have different opinions on the question that local stocks benefit or cause damage to investors' profit (Seasholes and Zhu, 2010; Ivkovic and Weisbenner, 2005). In this section, I first briefly discuss the profit of local stocks in Chinese stock market. Furthermore, since I investigate the impact of local positions on disposition effect, V-shaped disposition effect and rank effect, I test the profit of local positions with these effects.

(*Insert Table 5.8 here*)

Table 5.8 presents the results of several statistics tests. In panel A, I demonstrate the future returns of local positions and nonlocal positions. The one week later potential returns of local positions are slightly smaller than nonlocal positions (p-value 0.150). And one month later, the returns of local stocks are significantly smaller than nonlocal stocks. However, it is opposite when considering one-year long-term returns. Local stocks preform significantly better than nonlocal stocks. In panel B, I discuss the disposition effect profit based on local positions. On all short, medium and long terms, when the position is local, gain positions have better potential returns than loss positions, which indicates that dispositions causes damage to investor profit. In panel C, the correlation coefficients of magnitude of gain and future returns for local positions are positive when the time period is one week and one month. The correlation coefficient of magnitude of gain and one-year later return is negative, but the absolute value is relatively small. Therefore, in general, positions with large gain have large potential return in the future. The V-shaped disposition effect on the gain side that investors prefer to sell large gains hurts the profit of investors when the position is local. Meanwhile, on the loss side, positions that are close to zero have larger potential returns in the future. In panel D, I discuss rank effect profit on local stocks. For short-term (one week) and medium-term (one month) returns, the order of future returns is identical to the rank of return, and all the return difference among different ranks are statistically significant. Meanwhile, when investigating long-term (one year) return, the best, 2nd best and middle rank positions have no significant difference among their long-term future returns, while the underperforming positions are still underperforming.

In previous sections, I find evidences that local stock preference is not significant in China. Trading local positions cannot moderate disposition effect nor V-shaped disposition effect. But it can decrease the degree of rank effect. In this section, I prove that local stocks do not have more future potential returns in short-term and mid-term, while local stocks can over performance nonlocal stocks in long-term. Furthermore, when the position is local, disposition effect and one-side V-shaped disposition effect cause damage to investors' profit, while rank effect is harmful to profit in short-term and mid-term. These effects are investment biases indeed. Thus, when the local stock moderate rank effect, it helps the profit of the portfolio to some degrees.

5.5. Conclusion and Discussion

In this chapter, I discover how geographic region factors impact investors' behavior on disposition effect, V-shaped disposition effect and rank effect. I find investors in metropolis suffer more from disposition effect. Meanwhile, born and living in metropolis cannot moderate

V-shaped disposition effect and rank effect. When I divide China into east, middle and west region groups, the preference of selling a position is similar among different regions. There is no significant difference in the theory of disposition effect, V-shaped disposition effect and rank effect among investors in eastern, mid and western China.

Investors in metropolis have high income and consumption level, relatively high education level, more investment knowledge and more potential information smyces. Therefore, these investors are thought to be more sophisticated. Meanwhile, east, middle and west regions is well-developed, mid-developed and less-developed region respectively. Investors in regions with higher development level have more potential resmyces. However, in this chapter, I find that born and living in metropolis or high development level region cannot moderate investors to suffer from investment biases. Since in previous chapters, I already prove that disposition effect, V-shaped disposition effect and rank effect cause damage to investors' profit, investors in metropolis or well-developed regions cannot perform better avoiding these biases. Born and living in metropolis or well-developed regions cannot on average improve investors' trading behavior. Thus, these investors are not more sophisticated than others on average. The performance of investment depends more on an individual level, not a region level.

My result cannot support previous literatures well on herding effect and the similarity of investment from investors that are close on geographic. Investors in different regions tend to trade similarly under theory of disposition effect, V-shaped disposition effect and rank effect. However, I argue that the regions in this chapter are much larger than regions in previous papers. In this chapter, I use metropolis and general regions divided by directions (eastern, mid, western). In previous paper, specific cities, districts or block communities are applied. In addition, in this chapter, I focus more on the econometric distance (development level, econometric index, etc.) within one region group instead of geographic distance. Therefore, the difference on definition of regions could cause the difference on results. The financial behavioral biases in more specific and small region groups could be a topic for further studies.

I further discuss how local effect influence selling decisions in China and the impact of local effect on disposition effect, V-shaped disposition effect and rank effect. Investors in China have no preference on selling local stocks, which is different from US and Finland

(Grinblatt and Keloharju, 2001; Ivkovic and Weisbenner, 2005; Seasholes and Zhu, 2010). Local stocks cannot influence disposition effect and V-shaped disposition effect as well. Meanwhile, local stocks moderate rank effect. The preference of selling relative good-performance positions is weaker when the position is from a local firm. Since rank effect focus on the comparison among positions within one's portfolio while disposition and V-shaped disposition effect consider more on the return of one specific position, the local effect probably influence more on the situation when investor need to make a selling choice among positions. And when that happens, if at least some of the selected selling potential positions are local, the local positions can prevent investors from suffering from rank effect in some degree. Since rank effect causes damage to investors' profit on both local and nonlocal positions, local stock can help the investment in some degree. Thus, this result indirectly reveals that investors have asymmetric information on local stocks.

Table 5.1: Summary Statistics

	Data after Preparation	
Observation	3,604,961	
Dummy variable	Number of 1s	
Selling	1,401,768	
Gain	1,406,745	
Gender	1,667,631	
New_investor	1,871,472	
Bigcity	389,639	
East	2,950,399	
Mid	495,681	
West	158,881	
Local	76,530	
Numerical variable	Average	S.D.
Return	-1.1554	4.1046
Root_holding_period	5.8567	5.9296
Variance	0.8355	2,7066
Portfolio_size	6.1381	5.2184
Root_tradetimes	26.1502	16.6603
Root_age	6.5907	0.8915

Note: This table presents the summary statistics of variables in my models. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. All variables are defined in session 3.2 in this chapter.

Table 5.2: Analysis of Metropolis Investor and V-shaped Disposition Effect

	Dependent Variable: Dummy of Selling the Position		
Investor:	All	Not in Jiangsu	Jiangsu
	(1)	(2)	(3)
Gain	0.1253***	0.1213***	0.1283***
(t-statistics)	(19.8658)	(17.7550)	(20.6356)
Gain*return	0.0144***	0.0177***	0.0122***
	(7.9193)	(7.4027)	(0.6790)
Loss*return	-0.0038	-0.0046*	-0.0031
	(-1.4931)	(-1.6646)	(-1.2810)
Bigcity	-0.0045	-0.0061	
	(-1.1258)	(-1.5662)	
Bigcity*gain	0.0127***	0.0173***	
	(3.1395)	(4.1123)	
Bigcity*gain*return	0.0001	-0.0007	
	(0.0466)	(-0.4084)	
Bigcity*loss*return	0.0006	0.0003	
	(0.7768)	(0.3338)	
Control variables	Yes	Yes	Yes
Observations	3,604,961	1,746,498	1,858463
Pseudo R^2	0.0869	0.0906	0.0836

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 includes all investors. Column 2 includes investors with their ID card not registered in Jiangsu. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. The dependent variable is a dummy variable equal to one if a stock is sold. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share price return since purchase. Big city is a dummy variable that equals to 1 when the investor's ID card is registered in big city in China. Other variables are control variables and are defined in section 3.1 in this chapter. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 5.3: Analysis of Metropolis Investor and Rank Effect

	Dependent Variable: Dummy of Selling the Position		
Investor:	All	Not in Jiangsu	Jiangsu
	(1)	(2)	(3)
Best	0.1321***	0.1342***	0.1298***
(t-statistics)	(26.9778)	(22.1465)	(24.7982)
2^{nd} best	0.0596***	0.0593***	0.0593***
	(16.9069)	(13.2219)	(15.8032)
Middle	0.0058***	0.0033	0.0068**
	(2.7110)	(1.1572)	(2.4529)
Worst	0.0079***	0.0077***	0.0081***
	(3.9770)	(2.8960)	(3.3867)
Bigcity	-0.0092*	-0.0122**	
	(-1.8423)	(-2.2888)	
Bigcity*best	0.0221***	0.0231***	
	(3.2429)	(3.2208)	
Bigcity*2^{nd} best	0.0119**	0.0138**	
	(2.2353)	(2.4828)	
Bigcity*middle	0.0014	0.0076*	
	(0.3328)	(1.6513)	
Bigcity*worst	0.0007	0.0007	
	(0.1486)	(0.1420)	
Control variables	Yes	Yes	Yes
Observations	1,891,122	917,035	974,087
Pseudo R^2	0.0576	0.0616	0.0541

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 includes all investors. Column 2 includes investors with their ID card not registered in Jiangsu. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Big city is a dummy variable that equals to 1 when the investor's ID card is registered in big city in China. Other variables are control variables and are defined in section 3.1 in this chapter. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 5.4: Analysis of Different Region Investor and V-shaped Disposition Effect

	Dependent Variable: Dummy of Selling the Position	
Investor:	All	Not in Jiangsu
	(1)	(2)
Gain	0.1191***	0.1194***
(t-statistics)	(13.5401)	(13.4427)
Gain*return	0.0165***	0.0185***
	(7.0749)	(7.2500)
Loss*return	-0.0032	-0.0043
	(-1.1347)	(-1.3983)
East	-0.0075*	-0.0065
	(-1.8419)	(-1.4560)
Mid	0.0008	0.0009
	(0.1822)	(0.1929)
East*gain	0.0091*	0.0089
	(1.7130)	(1.6202)
East*gain*return	-0.0025*	-0.0018
	(-1.5646)	(-1.0985)
East*loss*return	-0.0005	0.0002
	(-0.1599)	(0.2064)
Mid*gain	-0.0008	-0.0005
	(-0.1599)	(-0.1001)
Mid*gain*return	0.0005	0.0004
	(0.2240)	(0.1997)
Mid*loss*return	-0.0015*	-0.0014
	(-1.5316)	(-1.4325)
Control variables	Yes	Yes
Observations	3,604,961	1,746,498
Pseudo R^2	0.0869	0.0907

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 includes all investors. Column 2 includes investors with their ID card not registered in Jiangsu. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. The dependent variable is a dummy variable equal to one if a stock is sold. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share price return since purchase. East, mid, west is a set of dummy variables indicate different regions in China, and west is the baseline. Other variables are control variables and are defined in section 3.1 in this chapter. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 5.5: Analysis of Different Region Investor and Rank Effect

	Dependent Variable: Dummy of Selling the Position	
Investor:	All	Not in Jiangsu
	(1)	(2)
Best	0.1546***	0.1566***
(t-statistics)	(13.1621)	(13.3813)
2^{nd} best	0.0591***	0.0601***
	(6.7113)	(6.8555)
Middle	-0.0005	0.0048
	(-0.0811)	(0.7263)
Worst	0.0047	-0.0043
	(0.7261)	(-0.6452)
East	-0.0044	-0.0122
	(-0.6989)	(-0.6452)
Mid	0.0033	0.0031
	(0.4516)	(0.4179)
East*best	-0.0170*	-0.0138
	(-1.8730)	(-1.4428)
East*2^{nd} best	0.0225	0.0037
	(0.2978)	(0.4909)
East*middle	0.0064	0.0024
	(1.1028)	(0.3894)
East*worst	0.0031	0.0028
	(0.4748)	(0.4140)
Mid*best	-0.0177*	-0.0174*
	(-1.7019)	(-1.6709)
Mid*2^{nd} best	-0.0004	-0.0002
	(-0.0477)	(-0.0247)
Mid*middle	0.0077	0.0080
	(1.1331)	(1.1761)
Mid*worst	0.0051	0.0050
	(0.6811)	(0.6637)
Control variables	Yes	Yes
Observations	1,891,122	917,035
Pseudo R^2	0.0577	0.0616

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 includes all investors. Column 2 includes investors with their ID card not registered in Jiangsu. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Big city is a dummy variable that equals to 1 when the investor's ID card is registered in big city in China. Other variables are control variables and are defined in section 3.1 in this chapter. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 5.6: Analysis of Impact of Local Stock on Disposition Effect and V-shaped Disposition Effect

	Dependent Variable: Dummy of Selling the Position	
Investor:	All	All
	(1)	(2)
Gain	0.1290***	0.1268***
(t-statistics)	(18.6477)	(20.1843)
Gain*return		0.0144***
		(7.8824)
Loss*return		-0.0038
		(-1.4734)
Local	0.0115	0.0102
	(1.6215)	(1.2233)
Local *gain	-0.0060	-0.0048
	(-0.9737)	(-0.7753)
Local *gain*return		0.0005
		(0.2756)
Local *loss*return		-0.0005
		(-0.3405)
Control variables	Yes	Yes
Observations	3,604,961	3,604961
Pseudo R^2	0.0866	0.0869

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 discusses disposition effect. Column 2 discusses V-shaped disposition effect. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. The dependent variable is a dummy variable equal to one if a stock is sold. Gain (Loss) is a dummy variable indicating a positive (non-positive) return. Return is the unit share price return since purchase. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Local is a dummy variable that equals to 1 when the region investor's ID card registered in and the region stock's company registered in are in the same city. Other variables are control variables and are defined in section 3.1 in this chapter. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 5.7: Analysis of Local Stock and Rank Effect

	Dependent Variable: Dummy of Selling the Position		
Positions:	All	Nonlocal Positions	Local Positions
	(1)	(2)	(3)
Best	0.1357***	0.1353***	0.1463***
(t-statistics)	(28.0179)	(27.9603)	(8.9582)
2nd best	0.0617***	0.0615***	0.0581***
	(17.8840)	(17.8298)	(4.0022)
Middle	0.0063***	0.0062***	0.0068
	(2.9461)	(2.9115)	(0.4996)
2nd worst			0.0212**
			(2.1022)
Worst	0.0086***	0.0086***	
	(4.4624)	(4.4826)	
Local	0.0319***		
	(-1.8423)		
Local*best	-0.0248**		
	(-2.0617)		
Local *2nd best	-0.0314***		
	(-2.9170)		
Local *middle	-0.0314***		
	(-2.7933)		
Local *worst	-0.0272***		
	(-2.9709)		
Control variables	Yes	Yes	Yes
Observations	1,891,122	1,853,364	37,758
Pseudo R^2	0.0576	0.0576	0.0596

Note: This table presents the marginal effect from logit regressions. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Column 1 includes all positions. Column 2 includes nonlocal positions. Column 3 includes local positions. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included, and an investor must hold at least five stocks to be included in the model. The dependent variable is a dummy variable equal to one if a stock is sold. 5 dummy rank variables are included to test rank effect. Local is a dummy variable that equals to 1 when the region investor's ID card registered in and the region stock's company registered in are in the same city. Other variables are control variables and are defined in section 3.1 in this chapter. The top number is the marginal effect, and the lower number in parentheses is the t -statistic. Clustered standard error is by investor. ***, **, * indicate statistical significance at the 1% level, 5% level and 10% level respectively.

Table 5.8: The Impact of Investor Behavior Biases on Future Returns from Local Positions

Panel A: (All Positions)	Local	Nonlocal	Local – Nonlocal (p-value)
	(1)	(2)	(3)
One Week Later Return	-1.2186	-1.1948	-0.0238 (0.150)
One Month Later Return	-1.6405	-1.4957	-0.1448 (0.000)
One Year Later Return	-2.8408	-3.5824	0.7418 (0.000)

Panel B: (Local Positions)	Gain	Loss	Gain – Loss (p-value)
	(1)	(2)	(3)
One Week Later Return	0.7012	-2.3946	3.0957 (0.000)
One Month Later Return	-1.6125	-3.5933	1.9808 (0.000)
One Year Later Return	0.2684	-2.8099	3.0783 (0.000)

Panel C: (Local Positions)	Magnitude of Gain	Magnitude of Loss
	(1)	(2)
One Week Later Return	0.3699	0.8533
One Month Later Return	0.2557	0.7014
One Year Later Return	-0.1195	0.3157

Panel D: (Local Positions)	Best	2^{nd} best	Middle	2^{nd} worst	Worst
	(1)	(2)	(3)	(4)	(5)
One Week Later Return	0.7402	-0.2362	-1.6244	-2.9219	-4.8959
Row – Col (Bonferroni)					
2^{nd} best	-0.9765 (0.000)				
Middle	-2.3646 (0.000)	-1.3882 (0.000)			
2^{nd} worst	-3.6621 (0.000)	-2.6856 (0.000)	1.2975 (0.000)		

Worst	-5.6361	-4.6596	-3.2715	-1.9740	
	(0.00)	(0.000)	(0.000)	(0.000)	
One Month Later Return	0.0909	-0.7397	-2.0116	-3.4670	-5.4886
Row – Col					
(Bonferroni)					
2nd best	-0.8306				
	(0.000)				
Middle	-2.1026	-1,2720			
	(0.000)	(0.000)			
2nd worst	-3.5579	-2.7273	-1.4553		
	(0.000)	(0.000)	(0.000)		
Worst	-5.5796	-4.7490	-3.4770	-2.0217	
	(0.000)	(0.000)	(0.000)	(0.000)	
One Year Later Return	-2.6207	-2.4950	-2.4227	-4.7738	-6.0735
Row – Col					
(Bonferroni)					
2nd best	0.1257				
	(1.000)				
Middle	0.1979	0.0722			
	(1.000)	(1.000)			
2nd worst	-2.1531	-2.2789	-2.3511		
	(0.000)	(0.000)	(0.000)		
Worst	-3.4528	-3.5785	-3.6507	-1.2997	
	(0.000)	(0.000)	(0.00)	(0.000)	

Note: This table presents the result of average unit share returns, differences, t-statistics, correlation coefficients and Bonferroni-adjusted significance. The data contains daily holding records of 100,000 investors from a large nationwide brokerage in the period from January 2007 to May 2009. Each position is an observation which is at investor-stock-day level. Only days in which a stock is sold are included. Returns are calculated as the difference of the unit share price in future and unit share cost. Panel A includes all positions. Column 1 presents future returns of local positions. Column 2 presents future returns of nonlocal positions. Column 3 presents the difference and p-value. In panel B, only local positions are involved. Column 1 presents for gain positions, column 2 presents for loss positions and column3 presents the difference and p-value. Panel C demonstrates the correlation coefficients of marginal of gain and loss and future returns of local positions. Panel D only includes positions if the position is in a portfolio with at least 5 positions and the position is local. Column 1 to 5 shows the 5 ranks of positions respectively. The future return is presented followed by the matrixes of differences among different ranks and the Bonferroni-adjusted significance.

Chapter Six: Conclusion

This thesis discusses several individual investors behavioral biases in Chinese stock market from January 2007 to May 2009. Based on a very large and unique dataset with individual investor daily trading and holding records, this thesis discovers that investors prefer to sell large gains. For relative performance of the position, when comparing within the portfolio and sorting the positions by return, best rank position has the largest probability to be sold followed by the 2nd best one. Furthermore, investors from different regions in China do not have significant different degrees of the previous phenomenon in general. Meanwhile, when the position is from a local firm, the preference of selling good rank positions in the portfolio is moderated. This study also finds evidences that the previous two types of selling preference cause damage to profits of investors and are indeed biases.

In Chapter 3, this thesis investigates the impact of magnitude of gain and loss on selling decision. Applying the control variable of holding period and others, individual investors hold a preference of realizing a large gain rather than a small one, but their preference of selling among different magnitude of loss is constant (one side V-shaped). Meanwhile, when removing the control of holding period or applying the model with sub-groups of different holding period positions, the result is different. Therefore, this finding emphasizes the importance of holding period in behavioral finance and related studies. Investors hold different views and have different strategies for short-term and long-term positions. Chapter 3 also illustrates that during all booming, crushing and recovering period of financial crisis, individual investors hold the preference of realizing a large gain. Meanwhile, investors are only willing to realize large losses when the whole market is booming, when they are confidence to the market. And when they lose confidence to the market during and after financial crisis, they are more patient to their large losses. Furthermore, sophisticated investors can moderate the bias of more willing to sell large gain and loss to some extent. Investors with more experiences and higher trading frequencies are less likely to trade follow V-shaped disposition effect while senior in age does not help. Since V-shaped disposition effect causes damage to investor profit in China, the feature of more experiences and higher trading frequencies can probably be the signs of sophistication in China. In addition, because senior people could be lack of professional training and education in China as a developing country, I infer that senior in age is probably not a sign of sophistication for investors in developing country.

Chapter 4 discusses rank effect in China. The positions with best performance in one's portfolio have the largest probability to be sold. The 2nd best on follows. However, the probabilities of selling middle, 2nd worst and worst positions are not significantly different. Meanwhile, in US market, investors sell both best and worst positions. The rank effect in Chinese market is different from US market. Chapter 4 further illustrates that when the positions are lottery like and the investors are young and male, investors in China have the preference of selling the worst rank position as well. Thus, selling the worst position can be explained by the willing of gamble to some extent. These results are also robust in different market conditions and on different holding period positions. This chapter also documents that senior, female investors with long trading experience and less trading frequency are more likely to sell good rank positions. In addition, Chapter 4 provides evidence that in small size portfolios, investors are more willing to sell the best performance position in their portfolio.

In chapter 5, this thesis discovers how geographic region factors impact investor behavioral biases discussed in Chapter 4 and 5. Investors in metropolis suffer more from disposition effect. Meanwhile, born and living in metropolis cannot moderate V-shaped disposition effect and rank effect. When dividing China into east, middle and west region groups, there is no significant difference among these regions in terms of the previous effects. Since metropolis and eastern region are well-developed regions in China, investors from these regions have high income and consumption level, relatively high education level and more potential information sources. However, this thesis cannot find evidence that investors from well-developed regions are more sophisticated than others. The performance of investment depends more on an individual level, not a region level. Furthermore, Chapter 5 analyzes the local effect and the impact of local effect on the effects discussed above. When the firm of the position is local to the investor, Investors in China have no preference on selling local stocks, which is different from US and Finland. While local stocks cannot influence disposition effect and V-shaped disposition effect as well, local stocks moderate the rank effect. Since rank effect causes damage to investors' profit on both local and nonlocal positions, local stock can help the investment in some degree. Thus, this result indirectly reveals that investors have asymmetric information on local stocks.

For the policy and industry implication, this thesis suggests that since the preference of selling well-performing positions causes damage to investor's profit, individual investors should try to prevent this preference. The well-performing positions probably catch more attention of investors. Investors should try to avoid the attention driven trading and do more analysis on their investment. From the perspective of policy maker, China Securities Regulatory Commission (CSRC) should continue promoting the stock market efficiency and reducing the information asymmetry in different areas, especially supervise and stabilize the stock market during the financial crisis. In addition, media from different channels (including social media, newspapers, magazines, televisions, etc.) should introduce knowledge on financial markets and strengthen education on stock investment to the public to rich their investment experience. From the perspective of security companies and banks, they should regularly organize the investment training to the individual investors and help them aware the investment bias.

As a final conclusion of this thesis, individual investors hold the preference of selling both absolute well-performing (gains and large gains) and relative well-performing (large return positions in the portfolio) positions. This preference causes damage to the profit and it is indeed a bias. While the preference of selling large gains is closely related to the holding period of the position and the market condition, the preference of selling relative well-performing positions is consistent. However, when the position is local, this preference is moderated. Furthermore, this thesis suggests that the preference of selling underperforming positions in US market is not consistent in China. And it is related to the investor confidence to the market and the willing of gamble.

References

An, L., 2016, Asset Pricing When Traders Sell Extreme Winners and Losers, *The Review of Financial Studies* 29:823-861

Bailey, T., Berg, P., and Sandy, C. 2001. The effect of high-performance work practices on employee earnings in the steel, apparel, and medical electronics and imaging industries. *Industrial and Labor Relations Rev.* 54-2A 525–543.

Baltakys, K., M. Baltakiene, H. Karkkainen and J. Kanniainen. 2019. Neighbors matter: Geographical distance and trade timing in the stock market, *Finance Research Letters,* 31:250-257

Barber, B.M., and T. Odean. 2000. Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors. *Journal of Finance* 55:773- 806.

Barber, B.M., and T. Odean. 2008. All that Glitters: The Effect of Attention on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies* 21: 785-818.

Barber, B.M., Y. Lee, Y. Liu, and T. Odean. 2007. Is the Aggregate Investor Reluctant to Realize Losses? Evidence from Taiwan. *European Financial Management* 13:423- 447.

Barber, B.M., Y. Lee, Y. Liu, and T. Odean. 2009. Just How Much Do Individual Investors Lose by Trading? *Review of Financial Studies* 22:609-632.

Barberis, N., and W. Xiong. 2009. What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-based Explanation. *Journal of Finance* 64:751-784.

Bazerman, M. H., G. F. Loewenstein, and S. B. White. 1992. Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives. *Administrative Science Quarterly* 37:220–40.

Ben-David, I., and D. Hirshleifer. 2012. Beyond the disposition effect: Do investors really like realizing gains more than losses. *Review of Financial Studies* 25:2485–532

Bernartzi, S., and R. Thaler. 2001. Naïve diversification strategies in defined contribution savings plans. *American Economic Review* 91:79–98.

Bian, J., K. Chan, D. Shi and H. Zhou. 2018. Do behavioral biases affect order aggressiveness? *Review of Finance* 1121-1151

Brown, J., Z. Ivkovic, P. Smith and S. Weisbenner. 2008. Neighbors Matter: Causal Community Effects and Stock Market Participation, *The Journal of Finance* 3:1509-1531

Chang, T. Y., D. H. Solomon, and M. M. Westerfield. 2016. Looking for Someone to Blame: Delegation, Cognitive Dissonance, and the Disposition Effect. *Journal of Finance*, 71, 267-302.

Chen, G., K. Kim, J. Nofsinger, and O. Rui. 2007. Trading Performance, Disposition Effect, Overconfidence, Represenativeness Bias, and Experience of Emerging Market Investors. *Journal of Behavioral Decision Making* 20:425-451.

Cheng, T., C. Lee, and C. Lin. 2013. An examination of the relationship between the disposition effect and gender, age, the traded security, and bull–bear market conditions. *Journal of Empirical Finance* 21: 195-213.

Choe, H. and Y. Eom. 2009. The Disposition Effect and Investment Performance in the Futures Market. *Journal of Futures Markets* 29:496-522.

Coval, J. D., and T. J. Moskowitz. 2001. The geography of investment: Informed trading and asset prices, *Journal of Political Economy*, 109:811–841.

Da Z., J. Engelberg, and P. Gao 2010. In Search of Attention. *Journal of Finance*, 5:1461-1499

Dhar, R., and A. Kumar. 2002. A non-random walk down the main street: Impact of price trends on trading behavior of individual investors. Working paper, International Center for Finance, Yale University, New Haven, CT.

Dhar, R., and N. Zhu. 2006. Up close and personal: investor sophistication and the disposition effect. *Management Science* 52: 726–740.

Elton, E., M. Gruber, and J. Busse. 2004. Are Investors Rational? Choices among Index Funds. *Journal of Finance* 59:261–288.

Feng, L. and M. Seasholes (2004), Correlated Trading and Location, *The Journal of Finance*, 5:2117-2144

Feng, L. and M. Seasholes. 2005. Do Investor Sophistication and Trading Experience Eliminate Behavioral Biases in Financial Markets? *Review of Finance* 9:305-351.

Frazzini, A. 2006. The Disposition Effect and Underreaction to News. *Journal of Finance* 61:2017:2046.

Frydman, C., N. Barberis, C. Camerer, P. Bossaerts, and A. Rangel. 2014. Using neural data to test a theory of investor behavior: An application to realization utility. *Journal of Finance* 69:907–46.

Gerrans, P., R. Faff, and N. Hartnett. 2015. Individual financial risk tolerance and the global financial crisis. *Accounting and Finance* 55:165-185

Goetzmann, W. N., and A. Kumar. 2008. Equity portfolio diversification. *Review of Finance* 12:433–63.

Goetzmann, W., and M. Massa. 2002. Daily momentum and contrarian behavior of index fund investors. *Journal of Financial and Quantitative Analysis* 37:375–390.

Grinblatt, M., and M. Keloharju. 2000. The Investment Behavior and Performance of Various Investor Types: A Study of Finland's Unique Data Set. *Journal of Financial Economics* 55:43-67.

Grinblatt, M., and M. Keloharju. 2001a. What Makes Investors Trade? *Journal of Finance* 56:589-616.

Grinblatt, M. and M. Keloharju. 2001b. How Distance, Language, and Culture Influence Stockholdings and Trades, *The Journal of Finance*, 3:1053-1073

Grinblatt, M., S. Titman, and R. Wermers. 1995. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior, *American Economic Review* 85:1088–1105.

Hallahan, T., R. Faff, and M. McKenzie. 2004. An empirical investigation of personal financial risk tolerance. *Financial Service Review* 13:57–78.

Hartzmark, S. M. 2015. The worst, the best, ignoring all the rest: The rank effect and trading behavior. *Review of Financial Studies* 28:1024–59.

Hauser, J. R. 2014. Consideration-set heuristics. *Journal of Business Research* 67:1688-1699.

Hoffmann, A., T. Post, and J. Pennings. 2013. Individual investor perceptions and behavior during the financial crisis. *Journal of Banking and Finance* 37:60-74

Hong, H., Jeffrey D. Kubik, and Jeremy C. Stein. 2002. Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, Working paper, Stanford Business School.

Hsee, C. K. 1996. The evaluability hypothesis: An explanation of preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes* 66:122–9.

Hsee, C. K., G. F. Loewenstein, S. Blount, and M. H. Bazerman. 1999. Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin* 125:576–90.

Itzkowitz, J., J. Itzkowitz and Rothbort, S. 2016. ABCs of Trading: Behavioral Biases affect Stock Turnover and Value. *Review of Finance* 20:663–692

Ivkovic, Z. and S. Weisbenner. 2005. Local Does as Local Is: Information Content of the Geography of Individual Investors' Common Stock Investments, *The Journal of Finance*, 1:267-306

Jacobs, H., and A. Hillert. 2016. Alphabetic Bias, Investor Recognition, and Trading Behavior. *Review of Finance* 20:693–723

Kahneman, D. 2003. Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review* 93:1449–75.

Kahneman, D. and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, 47:263-291.

Kaustia, M. 2010b. Prospect Theory and the Disposition Effect. *Journal of Financial and Quantitative Analysis* 45:791–812.

Krueger, A., and C. Rouse. 1998. The effect of workplace education on earnings, turnover, and job performance. *Journal Labor Economics* 16(1):61–94.

Kumar, A. 2009. Who Gambles in the Stock Market? *Journal of Finance* 4:1889-1933.

Lakonishok, J., A. Shleifer, and R. W. Vishny. 1992. The impact of institutional trading on stock prices, *Journal of Financial Economics* 32:23–43

Lehkonen, H. 2010. Bubbles in China. *International Review of Financial Analysis* 19: 113-117

Liao, L., B. Liu and H. Wang. 2014. China's secondary privatization: Perspectives from the Split-Share Structure Reform. *Journal of Financial Economics* 113: 500-518.

Liao, L., Y. Liang and W. Zhang. 2016. Who gambles in the Chinese Stock Market? (in Chinese). *Journal of Tsinghua University (Sci & Technol)* 6: 677-684.

List, J. A. 2002. Preference reversals of a different kind: The ‘more is less’ phenomenon. *American Economic Review* 92:1636–43.

List, J. A. (2003). Does market experience eliminate market anomalies? *The Quarterly Journal of Economics* 118(1):41–71.

Locke, P., and S. C. Mann. 2005. Professional trader discipline and trade disposition. *Journal of Financial Economics* 76:401–44.

Lukas, M., A. Eshraghi, and J. Danbolt. 2017. The Disposition Effect in Social Trading: Influence of Transparency on Investment Decision. working paper.

Markowitz, H. 1952. Portfolio selection. *Journal of Finance* 7:77–91.

Odean, T. 1998. Are investors reluctant to realize their losses? *Journal of Finance* 53:1775–98.

Meng, J. 2010. The Disposition Effect and Expectations as Reference Point. Working Paper, Guanghua School of Management, Peking University.

Seasholes, M.S., and G. Wu. 2007. Predictable Behavior, Profits, and Attention. *Journal of Empirical Finance* 15:590-610.

Seasholes, M. and N. Zhu. 2010. Individual Investors and Local Bias, *The Journal of Finance*, 5:1987-2010

Shapira, Z., and I. Venezia 2001. Patterns of Behavior of Professionally Managed and Independent Investors. *Journal of Banking and Finance* 25:1573-1587.

Shefrin, H.M., and M.S. Statman 1985. The Disposition to Sell Winners too Early and Ride Losers too Long: Theory and Evidence. *Journal of Finance* 40:777-790.

Shi, Y., Z. Li, and W. Chen. 2009. An empirical study of trading behavior in Chinese security markets. *Financial Research* 11: 129-142, in Chinese.

Shumway, T., and Wu. G. Does disposition drive momentum? AFA 2006 Boston Meetings Paper.

Strahilevitz, M., T. Odean, and B. M. Barber. 2011. Once Burned, Twice Shy: How Naïve Learning, Counterfactuals, and Regret Affect the Repurchase of Stocks Previously Sold. *Journal of Marketing Research* 48:102–20.

Thaler, R. 1985. Mental Accounting and Consumer Choice. *Marketing Science* 3:199-214.

Tversky, A., and D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5:297–323.

Weber, M. and C.F. Camerer 1998. The Disposition Effect in Securities Trading: An Experimental Analysis. *Journal of Economic Behavior and Organization* 33:167-184.

Yao, S., and D. Luo. 2009. The Economic Psychology of Stock Market Bubbles in China. *The World Economy* 32 (8): 667-690