VSB – TECHNICAL UNIVERSITY OF OSTRAVA
FACULTY OF ECONOMICS

DEPARTMENT OF FINANCE

Empirical Analysis of Herding Behavior in the Chinese Stock Market
Empirická analýza stádního chování na čínském akciovém trhu

Student: Bc. Xinran Chen

Supervisor of the Diploma thesis: Ing. Martina Novotná, Ph.D

Ostrava 2019

VŠB - Technical University of Ostrava
Faculty of Economics
Department of Finance

Diploma Thesis Assignment

Student: **Bc. Xinran Chen**

Study Programme: N6202 Economic Policy and Administration

Study Branch: 6202T010 Finance

Title: Empirical Analysis of Herding Behavior in the Chinese Stock Market
Empirická analýza stádního chování na čínském akciovém trhu

The thesis language: English

Description:

1. Introduction
2. Overview of Herding Behavior
3. Description of Methodology
4. Empirical Analysis of Herding Behavior in the Chinese Stock Market
5. Conclusion

Bibliography
List of Abbreviations
Declaration of Utilisation of Results from the Diploma Thesis
List of Annexes
Annexes

References:

ANDRIKOPOULOS, P., G. N. GREGORIOU and V. KALLINTERAKIS. *Handbook of frontier markets*. Boston: Academic Press, 2016. ISBN 978-012-8037-768.
BURTON, T. Edwin and Sunit, N. SHAH. *Behavioral finance: Understanding the social, cognitive, and economic debates*. Hoboken: Wiley, 2013. ISBN 978-1-118-30019-0.
RICHARDS, Tim. *Investing psychology: the effects of behavioral finance on investment choice and bias*. Hoboken: Wiley, 2014. ISBN 978-1-118-72219-0.

Extent and terms of a thesis are specified in directions for its elaboration that are opened to the public on the web sites of the faculty.

Supervisor: **Ing. Martina Novotná, Ph.D.**

Date of issue: 23.11.2018
Date of submission: 26.04.2019

VYSOKÁ ŠKOLA BÁŇSKÁ-TECHNICKÁ UNIVERZITA * OSTRAVA * Ekonomická fakulta 1

Ing. Iveta Ratmanová, Ph.D.
Head of Department

prof. Dr. Ing. Zdeněk Zmeškal
Dean

The declaration

"I hereby declare that I have elaborated the entire thesis including annexes myself."

Ostrava date 2019. 4. 20 ....

Xinran Chen 陈欣冉 ....

Student's name and surname

Contents

1 Introduction ... 7

2 Overview of Herding Behavior ... 8

2.1 Efficient Market Hypothesis ... 8

2.2 Behavioral Finance ... 9

2.2.1 Herding Behaviour ... 9

2.2.2 Theoretical Research of Herding Effect ... 11

3 Description of Methodology ... 17

3.1 Herd Behavior Model ... 17

3.1.1 LSV Model ... 17

3.1.2 PCM model ... 18

3.1.3 CSSD Model ... 19

3.1.4 CSAD Model ... 20

3.1.5 Adjusted CSAD Model ... 23

3.2 Linear Regression Model ... 25

3.2.1 Multiple Linear Regression Model ... 25

3.2.2 Stationary ... 25

3.2.3 Autocorrelation ... 26

3.2.4 Heteroscedasticity ... 27

3.2.5 Multicollinearity ... 29

3.2.5 Model Specification ... 29

3.2.6 Normality ... 30

4 Empirical Analysis of Herding Behavior in Chinese Stock Market ... 32

4.1 Description of Selected Stock Market ... 32

4.2 Data Collection ... 33

4.3 Empirical Results ... 34

4.4.1 Shanghai A-Share Market ... 34

4.4.2 Shanghai B-Share Market 44

4.4.3 Shenzhen A-Share Market 46

4.4.4 Shenzhen B-Share Market 47

4.4.5 Shenzhen Second Board Market 53

4.4.6 Results Overview and Comparation 54

5 Conclusion 57

Bibliography 59

List of Abbreviations 61

Declaration of Utilisation of Results from the Diploma Thesis

List of Annexes

Annexes

1 Introduction

Behavioural economics is different from traditional finance, which mainly explains the phenomena in the current financial field from the aspects of market participants' behaviour habits, psychological and emotional thinking. The herding effect is one of the research hotspots in the field of behavioural finance. The herding effect describes the behaviour of social subjects that ignore their own information and adopt the same strategy like most other subjects in society, having convergence characteristic nature.

In this thesis, the objective is to test the presence of herding behavior in the Chinese stock market, including five submarkets, Shanghai A-share, Shanghai B-share, Shenzhen A-share, Shenzhen B-share and Shenzhen second-board stock markets. We use data for the last four years during the period from January 2, 2014 to October 31, 2018 to run the regression models. According to the coefficient we get from the regression model, we can conclude is there any herd behavior in the Chinese stock market and in which submarket exists the most serious herd effect. There are five chapters in this thesis. This chapter is the general structure and organization of this thesis.

Chapter two mainly about the literature background of this work. It starts with traditional economic theory. Here, we focus more on detail assumption about efficient market hypotheses. Then we will introduce behavioural finance, the main part is herding behaviour and its causes.

In Chapter three, we first describe the five economic models and one model will be used to measure the herd behavior. Then we introduce the linear regression model and its assumption. All assumption mentioned in this part will be exam in chapter four step by step.

In Chapter four, we introduce the background of the selected stock market, then we describe the data collection. We apply adjusted CSAD model in the empirical market analysis. We will show the results calculated by STATA software and make comparison of five markets.

In the last chapter, we will make a summarize and comparison of previous results. We will also provide some suggestions on future herd behavior measurement and future equity market development.

2 Overview of Herding Behavior

In this chapter, we first introduce the traditional economic theory, mainly focus on the efficient market hypothesis, which based on the rational man assumption. Then we will introduce behaviour finance, which believes the ration of man is limited. Then we focus on heading effect.

2.1 Efficient Market Hypothesis

The study of the "Efficient Market Hypothesis" originated from Bachelier (1900), who studied the randomness of Brownian motion and stock price changes from a stochastic process perspective. And he recognizes the effectiveness of the market in terms of information: past, present events, and even discounted values for future events are reflected in market prices. His "basic principle" is that stock prices follow a “fair game” model. Samuelson (1965), Mandelbrot (1966) clarified the relationship between fair game models and random walks through mathematical proofs and theoretically discussed the correspondence between effective markets and fair game models. It also made a theoretical foundation for the efficient market hypothesis. Based on summarizing the theory and empirical evidence of predecessors, Fama (1970) proposed an efficient market hypothesis by using the analytical methods of Samuelson (1965) and the three effective forms proposed by Roberts (1967), Fama (1970) proposed an efficient market hypothesis.

There are three variations of the hypothesis – the weak, semi-strong, and strong forms – which represent three different assumed levels of market efficiency. In the case of the weak-form efficiency, the market price has fully reflected all the past historical securities price information, including the stock transaction price, trading volume, short selling amount, financing finance, etc. If the weak efficient market hypothesis is established, the technical analysis of stock prices will be ineffective, and the fundamental analysis may also help investors to obtain excess profits. In a semi-strong effective market, prices have fully reflected all publicly available information about the company's operating prospects. This information includes transaction price, volume, profit data, profit forecast, company management status and other publicly disclosed financial information. If investors can get this information quickly, the stock price should respond quickly. If the semi-strong effective hypothesis is established, the use of technical analysis and fundamental analysis in the market will be lost, and insider information may gain excess profits. The strong efficient market hypothesis states that prices have adequately reflected all information about the company's operations, including publicly available or undisclosed information. In a strong and efficient market, there is no way to help investors get excess profits, even for insiders.

The three different forms of markets have some same assumptions. There are a large number of both buyers and sellers in the market. And everyone in the market is a rational economic man. So, price movements always occur efficiently, which means stocks are always trading at their current fair market value. The hypothesis of rational man require people acts rationally and with complete knowledge, who seeks to maximize personal utility or satisfaction. The assumptions create an ideal world and everything go as our tentative plan. But in reality, the rationality of human is limited, and it has been observed that investors may exhibit apparently irrational and predictable biases mainly attributable to psychological factors, which are excluded from traditional economic theory.

2.2 Behavioral Finance

Due to the limitation of traditional finance, the behavioural finance theory is established. Behavioural economics is practical economics that combines economics with behavioural analysis theory, psychology and sociology, to find errors or omissions in today's economic models. Furthermore, behavioural finance revises the shortcomings of the basic assumptions of mainstream economics, such as human rationality, self-interest, complete information, utility maximization and preference consistency.

Nicholas Barberis and Richard Thaler (2003) believed the behavioural finance has two building blocks: limit to arbitrage, which argues that it can be difficult for rational traders to undo the dislocations caused by less rational traders; and psychology, which catalogues the kinds of deviations from full rationality we might expect to see.

All these factors will lead to irrational decision making for market participants. And the wrong decisions will cause market anomalies, which can be observed and measured by financial models. Here we will introduce herding behaviour.

2.2.1 Herding Behavior

The herding behaviour at first is used to describe the behave of the sheep flock. If a wooden stick is placed in front of a group of sheep, the first sheep jump over, then the second and third will follow. At this time, the stick was removed, and the following sheep came here, still, jumping up like the sheep in front, even though the stick of the road was gone. The sheep flock will follow the front sheep of the flock wherever it goes and whatever it does. This is the so-called "herd effect", also known as "constrained psychology."

Then herding behaviour gradually cited to describe human social behaviour, which refers to the decision-making approach of taking the same actions as most people while ignoring valuable private information. In the financial market, the behaviour of the herd is that in a certain period of time, a large number of investment entities abandon their original investment strategies after obtaining the investment decision information of others and adopt an investment strategy similar to others. Due to the influence of information transmission among a large number of investors, the herd effect has a greater impact on market resource allocation and market stability. At the same time, historical experience shows that herding behaviour is also closely related to the occurrence of the financial crisis, which also affects the normal order of financial markets. Therefore, the herding effect has become the focus of scholars in various countries and the focus of the regulatory authorities.

Regarding herd behaviour, Keynes pointed out in 1934: In the day-to-day fluctuations of investment income, there is some inexplicable group bias in the county, and even a ridiculous emotion affects the entire market. The prevalence of herd behaviour has led to two outcomes of market efficiency: rational and irrational results.

Suppose there are 100 investors in the market, and there will be different assessments of investment opportunities in emerging markets. Among them, 20 people think the investment is profitable, while the other 80 people hold the opposite view. In this way, if the information held by 100 investors and their assessments are concentrated, they will generally think that this investment is unwise. In reality, however, they are unable to communicate information or make investment decisions at the same time. If the initial decision-making investor comes from the 20 people, then they will invest, then the 80 people will change their minds and invest. This creates a Snowballing Effect that causes most investors to invest. Finally, it is too late when investors find out that this is indeed an unprofitable investment.

From this example, we derive some of the characteristics of “herd behaviour” or “Information Cascades”:

1) The decisions are not made at the same time, there should have the early sheep and the followers.

2) The earlier investors decisions could be seen or observed. Thus, the early decision will affect the overall final decision;
3) When investors find themselves adopting failed decisions, as new news arrives, they eventually follow up in groups in the opposite direction, which in turn exacerbates market volatility.

2.2.2 Theoretical Research of Herding Effect

In the financial market, there are many potential reasons for herd behaviour. The most important ones can be classified into the following three categories: Imperfect Information, Concern for Reputation and Compensation Structure.

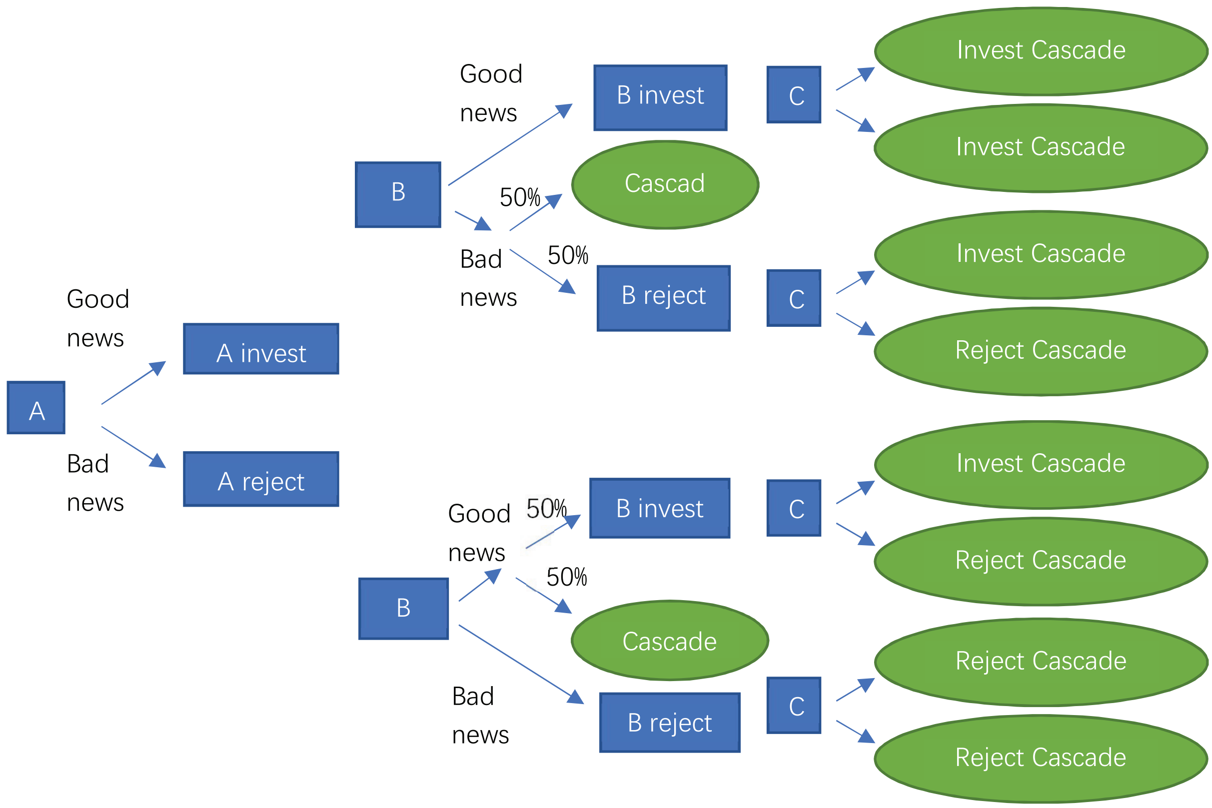
Imperfect Information

We assume that every investor is uncertain when facing the same investment decisions. Through hard research, they each obtained a portion of the correct information about the investment. Therefore, individuals can only evaluate the investment through some limited information they have. Because investors can't exchange information, they can only judge each other's information through mutual investment behaviour (of course, even if they can exchange information, the actual action is more persuasive), so under such conditions, the herding behaviour is very likely to happen.

Imagine that there are several investors in the above market, who take turns to make decisions on whether to invest in specific stocks, and the order in which decisions are made is externally determined. The entire investment decision process is shown in Fig. 2.1. The first investor A will make decisions based on the information he gets: invest if he gets good news or gives up. For the second investor B, if he gets the news in the case of A investment, he will invest without hesitation; but if he gets bad news, then the probability of his investment is 50% if B also invests at this time. Then the third investor C will think that the first two investors have received good news, even if they get bad news, they will insist on investment decisions. As the fourth investor D, he knows that the investment behaviour from C does not reveal the real information that C has mastered, so he can only observe the investment decisions of A and B. The following investors E, F... are all so, thus forming an invest cascade.

If the first investor A invests in the good news he gets, but the second investor B gets the bad message and chooses not to invest (the probability is 50%), then the second investor C will use the information he gets by himself to choose whether to invest or not. Because the prior probabilities generated by the behaviour of the first two investors offset other out, then the fourth investor D faces the same choice as B... This forms another chain of relationships.

Figure 2.1 Invest decision-making process



Source: Author's work

So, we get the following conclusion: If and only before an investor, the number of investors who decide to invest is at least two more than the investor who decides to reject, the investor will be in the spread of investment behaviour. That is, when an investor finds out that among all former investors, the number of people choosing one decision-making method is two more than the one who chooses another decision-making method, starting from this investor, all subsequent investors will behave in the same way as many decision-making behaviours. These people who have the same decision-making behaviour, which we call information cascade. Investment cascade usually occurs in the first investor who finds more people choose to invest rather than refuse to invest. Like all the following investors, he is in a state of rational consideration, preferring to follow previous investments without considering the real information he has.

The spread of investment information and the order of investor investment decision are closely related, in other words, they are closely related to the order of information arrive. For example, if the order of information arrives is: good news --- good news --- bad news --- bad news - bad news..., then this is an information that everyone is going to invest, because the first two good news has made the former two investors choose to invest, so the latter news does not matter any longer. If the order of the arrival of the message is good news --- bad news --- good news --- bad news ..., this is a way of information that may not cascade. Of course, when the second bad news arrives, investors choose to invest, then it is the invest cascade. It can be seen that the small difference in the order of arrival of the news may lead to huge differences in investment behaviour across the investment market. In addition, decision-making or behavioural cascade usually appear very quickly and disappear quickly.

Reputation-Based Herding

In the financial market, information about the capabilities of specific investment managers is undetermined. Because investment managers and their customers are not sure about their investment ability, herd behaviour can occur when investment managers are in the same environment. The so-called reputation-based herd behaviour means that for an economic person, if he doubts his ability to choose stocks correctly, then it will be a better choice to be consistent with other investment experts. Because this can at least maintain average performance without damaging your reputation (Scharfstein & Stein, 1990).

Considering the behaviour of two investment managers U and V in the market, they may have either high or low capability. Firstly, U decides to invest according to the information he has, and V needs to take into account his own information and U's investment behaviour. Because V is not sure about his ability, he dares not take the risk of making decisions that are contrary to U's. Therefore, even if V gets bad news, he will stick to the investment decision. U is also satisfied with this arrangement, because he is not sure of his ability, and V's follow-up has strengthened his investment confidence.

If this behaviour is promoted, it is obviously inefficient for the whole society. Because if the investment decision chosen by the first manager is inefficient, then all the managers behind will choose the wrong decision. However, as mentioned above, the post-investment managers reduce the risk through this behaviour, so that even if there is a bad result, they will not be blamed by the boss or cause other losses. In this way, the manager will have a higher scare for follow-up decision than their own judgment. For them, the herding effect is good and effective. For the manager who invests first, if someone follows the trend, then they will be more confident. And if something goes wrong, it can be attributed to luck or other factors. So, the manager who invested first also hopes to have such a result. The outcome of this game is valid for every manager. Reputation-based herding behaviour is also known as "blame-sharing effect" (Scharfstein & Stein, 1990).

Return-Based Herding

If the compensation for an investment manager depends on performance compared to other managers, the incentives of these agents will be distorted and their investment behaviour will eventually end in a highly similar portfolio. This situation also causes herd behaviour.

The fund managers compensation increases with individual performance improvement and decreases with the performance of the benchmark improvement. The agent and his reference investors have some information about the stock returns and other agents' investment portfolio. Under the risk-averse assumption, they will choose the follow-up strategy to maintain their performance around the benchmark instead of taking the risk to do their own decision. For this kind of compensation programs, investors will pay close attention to the behaviour of the reference investors. However, for the investment manager's customers, it is the best choice to sign such a contract with the agent's performance in the face of moral hazard or adverse selection.

The three main sources of herding behaviour are mentioned above. At the same time, economics also classify the kinds of herding behaviour.

True or Spurious Herd Behavior

True herding behaviour refers to investors giving up their own information and simply imitating and following others' decisions. This kind of follow-up behaviour shows irrational of investors, which violates the assumptions of rational people. Spurious herd behaviour means that investors take actions for their own interests after obtaining enough useful information, and their behaviours are just similar. For example, when the regulator announced that the IPO was restarted, investors chose to sell the stocks in consideration of the dilution function of the new stocks, causing the stock price to fall. This is not a simple follow-up, but a rational behaviour in the face of bad news. This is a rational behaviour taken by a rational investor in the face of the same information, rather than herd behaviour abandoning the information then followed.

This behaviour is called a spurious herding behaviour. The true herd behaviour is the irrational behaviour of investors, while the false herding behaviour shows the perfect transaction of information and information symmetry. It is difficult for researchers to distinguish between true herd behaviour and spurious herd behaviour. Because there is a lot of information on the market and there is a great correlation, it is difficult to determine whether investors have given up private information and choose to follow others.

Rational or Irrational Herd behaviour

Rational herding behaviour means that the investor's behaviour in the market is based on the maximization of their own interests. In the securities market, the basis of such behaviour is the change of the intrinsic value of stocks, and the rational herd behaviour can achieve the effectiveness of resource allocation. Irrational herd behaviour means that market investors do not start from the maximization of their own interests, but only a random behaviour, which has nothing to do with the actual value of the stock, which will lead to a decline in market efficiency. In the actual investment process, investors are more likely to show limited rationality. That is, investors make investment decisions based on objective factors in most time, but sometimes they will violate the principle of maximizing profits due to subjective factors such as psychological emotions, then abandon private information and choose to follow other investors to buy and sell.

Sequence or Non-sequence Herd Behavior

The theory of sequential herd behaviour suggests that investors have a certain order in their investment decisions. Assuming that investors are rational, investors will observe the decisions made by the decision makers in front of them. Since investors think that the first mover may have information that he does not own, he will refer to the decision of the first mover, and then choose to abandon private information when making decisions.

Another theory is the non-sequence herd behaviour. In reality, it is very difficult to distinguish the order of investors' decision-making. Therefore, the theory of non-sequence herd behaviour suggests that the decisions between investors do not show a clear sequence. If the market participants have the same tendency when the behaviour of mutual imitation between investors is weak, the market performance is that the investment return is subject to Gaussian distribution; when there is strong imitation behaviour among investors, the market is in a state of collapse. According to the non-sequence herd theory, the above situation will not cause the stock market to have zero-point symmetry in the modern financial theory and the thick tail distribution of the single mode.

3 Description of Methodology

In this chapter, we first introduce the economic models we can use to measure the herd behaviour, the most important method is the CSAD model, which will be used in chapter four. Then we will introduce the statistical method used in data analysis, linear regression method and its assumptions.

3.1 Herd Behavior Model

At present, the empirical test model for measuring the herd effect is mainly divided into two categories according to the research direction. The first direction is to analyze the investment behavior of institutional investors such as funds to test whether the institutional investors have a herding effect, including LSV method and the PCM method. The other direction is to analyze the dispersion of the rate of return to study whether there is a herding effect in the market as a whole, including the CSSD method and the CSAD method.

3.1.1 LSV Model

The LSV model (Lakonishok, Shleifer & Vishny,1992) is an empirical model for studying the group effect of institutional investors.

They use the average trend of the fund managers to buy or sell specific assets at the same time, that is, the long and short numbers of the two sides of the transaction as indicators of herd behaviour to measure whether the investors make investment decisions independently. They proposed a following indicator of measuring herd behaviour,

$$HM_{(i,t)} = \left|P_{(i,t)} - E\left(P_{(i,t)}\right)\right| - AF, \tag{3.1.}$$

where $P_{(i,t)}$ is the fund's net purchase ratio of stocks i during the period t. To be more specific, the equation is below,

$$P_{(i,t)} = \frac{B_{(i,t)}}{B_{(i,t)} + S_{(i,t)}}, \tag{3.2.}$$

where $B_{(i,t)}$ represents the number of fund companies that buy stocks i during the period t , $S_{(i,t)}$ represents the number of fund companies that sell stocks i during the period t. For $E(P_{(i,t)})$, the LSV method uses the arithmetic mean of all stocks over the period instead. Since $|P_{(i,t)} - E(P_{(i,t)})|$ may not be 0 under the assumption, in order to have a theoretical zero point, the adjustment factor AF is introduced,

$$AF = E\left[\left|P_{(i,t)} - E\left(P_{(i,t)}\right)\right|\right]. \qquad (3.3.)$$

When the trading amount of the stock i increases, the proportion of the investor who buys the stock $P_{(i,t)}$ in the period t will approach to the expected value, and the value of the adjustment factor will also approach to zero. LSV method use the HM as a coefficient to measure the herd effect, if the HM value is equal to 0, it can be considered that the investor does not have a herd effect in the transaction process. If the HM value is not significantly zero, then the herd effect can be considered to exist, and the larger the value, the more obvious the herding effect. HM is calculated as below,

$$\mathrm{HM} = \frac{1}{N}\sum_{i=1}^{N} HM_{(i,t)}\,. \qquad (3.4.)$$

The LSV test method has the advantages of easy data and easy operation, however, in the actual test, there are some problems in the use of this method. First of all, the factors considered are too singular. This method only considers the number of investors in the market, but the transaction volume is not taken into account. Thus, when the number of buyers and sellers is similar, but the number of shares bought and sold by the two parties is very different, it is impossible to effectively test the possible herding effect. Second, the method does not effectively distinguish the difference in investment horizon strategies. If a stock has both long-term investment and short-term investment, LSV will not be able to get accurate results. Investors are constantly changing their portfolios and investment horizons. It is difficult to match portfolios and maturities. If these variables are not consistent, it will have an impact on the accuracy of the test results of the herd effect.

3.1.2 PCM model

In response to the shortcomings of the LSV model, Wermers (1995) proposed a portfolio-change measurement (PCM) method. This method takes into account the investor's investment weight and investment direction and effectively overcomes the shortcomings of the LSV method that only considering the number of investors in the market.

The PCM model considers the cross-section correlation coefficients between portfolio portfolios, which are expressed as:

$$\rho = \frac{1/_{N_t} \sum_{n=1}^{N_t} (\Delta\omega_{n,t})(\Delta\omega_{n,t-1})}{\sigma(\tau)}, \tag{3.5.}$$

where $\Delta\omega_{n,t}$ is the weight change of each stock in the portfolio during period $t-1$, $\Delta\omega_{n,t-1}$ is the weight change of each stock in the portfolio during the period $t-\tau-1$. N_t is the amount of stock available for trading in each portfolio, $\sigma(\tau)$ is the standard deviation of the cross-section. This method has been used in the detection of herding effects in US mutual funds, and the results show that there is a significant herd effect in the fund. Stocks are traded by more funds, and the greater the volume of trading, the more likely the herd effect will occur. Therefore, it can be speculated that the large-cap stocks are more likely to show the herd effect, which is also consistent with the reality.

The defects of the PCM method are also obvious. The first is the stock with a larger market value, which is more likely to have an impact on the measurement of the herd effect because the stocks with large market capitalization will have a larger weight. Secondly, the daily fluctuations of stocks make the market value constantly changing, which will affect the weight of stocks, which will lead to deviations in the measurement of herd behaviour.

3.1.3 CSSD Model

Christie and Huang (1995) proposed using the cross-sectional standard deviation of returns (CSSD) to check the overall herd effect of the market. The main idea of the method is to detect the return dispersion of the investment, that is, whether the return of individual stocks and the return of the whole market tend to be similar. According to the asset pricing model, individual stocks are not exactly the same as the overall market performance. When the market returns become larger, the income deviation of individual stocks will also increase. When herd behaviour exists in the market, investors will abandon the information they own and choose to follow others' actions, resulting in individual stock returns that will not deviate from market returns or present a relatively independent trend. Therefore, the herd behaviour can be measured by the dispersion index of the individual stock returns relative to the standard deviation of the market average return. The standard deviation of the section can be expressed as:

$$\text{CSSD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(R_{it} - R_{mt})^2}, \tag{3.6.}$$

where N is the number of stocks included in the market, and R_{it} is the yield of stock i during the period t. R_{mt} is the average rate of return for all stocks during the period t. It can be seen from this formula that the dispersion index will increase with the deviation of individual stock returns from the market yield, but when the herd behavior occurs, the individual stock returns are close to the market yield, and the dispersion index will become smaller. Thus, we can test whether the herd effect exists by checking whether there is a significant difference between the CSSD indicators in the market stable period and the market volatility. Its regression model is:

$$\mathrm{CSSD} = \alpha + \beta_1 D_t^L + \beta_2 D_t^U + \varepsilon_t, \quad (3.7.)$$

where D_t^L and D_t^U are dummy variables. When the total return of the market portfolio is lower than the extreme tail of the income distribution in the period t, D_t^L is equal to 1, otherwise D_t^L is equal to 0. When the total return of the market portfolio is higher than the extreme tail of the income distribution in the period t, D_t^U is equal to 1, otherwise D_t^U is equal to 0. α is the sample dispersion after the extreme tail has been removed. β_1 and β_2 are parameters used to determine whether there is a herd effect. If both are significantly negative, it indicates a herd effect and if both are positive, it does not exist. The advantage of the law is that the data is easy to get, but it also has shortcomings. Only when β_1 and β_2 are significantly negative can we confirm the existence of herd effect in the market. Therefore, the conclusions drawn by this method are conservative, and only a very severe herding effect can be found, and the sensitivity to the measurement of herd effect is insufficient.

3.1.4 CSAD Model

Because the accuracy of the CSSD method is not high enough to capture the herd effect sensitively, Chang, Cheng and Ajay (2000) proposed a cross-sectional absolute dispersion (CSAD) model to measure the herd effect. Based on the cross-section absolute dispersion CSSD method proposed by Christie and Huang (1990), based on the rational capital asset pricing model (CAPM), the absolute deviation of stock returns is used as a measure of market return deviation. We can measure the herd effect of the stock market, by analyzing the relationship between the degree of deviation and the market rate of return.

Compared with the CSSD model, the CSAD model has further developed the effectiveness of the herding effect test. First, the CSAD model uses market rate of return and deviation as variables rather than dummy variables in the process of variable regression, so it provides more abundant data support in the empirical analysis of herding effect. Second, the quadratic term of the variable is set in the regression equation of the CSAD extended model, and the nonlinear regression form can describe the herding effect more specifically. That is to say, when the yield of the market is increased, the deviation of the stock return rate is reduced, or the increase is reduced, indicating that the market has a certain herding effect. Third, the CSAD model is based on the rational capital asset pricing model. Therefore, in the process of empirical analysis, once the empirical results reject the rational hypothesis, it indicates that there are some irrational factors in the market, and these irrational factors may lead to the emergence of the herd effect and provided a theoretical basis for further deep research on the herd effect.

In the CSAD model, the absolute cross-sectional dispersion (CSAD) is used as a fillowing indicator of the degree of deviation as a test for the herd effect,

$$CSAD_t = \frac{1}{N}\sum_{i=1}^{N}|R_{it} - R_{mt}|, \qquad (3.8.)$$

where, R_{it} represents the yield of stock i during the t period, R_{mt} represents the market return rate of N stocks at time t after calculation, and the deviation of stock return rate in the form of absolute dispersion. The original regression equation of the CSAD model uses linear regression, as shown below:

$$CSAD_t = \beta_1 + \beta_2|R_{mt}| + \varepsilon_1. \qquad (3.9.)$$

According to the CSAD model, if there is a herding effect in the stock market, the regression coefficient β_2 should be negative. Because if there is a herding effect in the stock market, as the market yield increases, the effect of the herd effect will prompt investors to chase the market's development trend, thus narrowing the deviation between stock returns and market returns. If all investors in the stocks market are rational, there is no herding effect, then the fluctuation of stock return rate should eventually stabilize. Therefore, there is a linear relationship between the deviation of the stock return rate and the market return rate, and the coefficient of the dependent variable is positive, which proves as follows.

According to the rational capital asset pricing model,

$$\mathrm{E}(R_i) = R_f + \beta_i E(R_m - R_f), \qquad (3.10.)$$

where R_f is the risk-free rate, and β_i is the risk coefficient of the stock i. Under the premise of equal weight, the market risk β_m can be expressed as

$$\beta_m = \frac{1}{N}\sum_{i=1}^{N}\beta_i . \tag{3.11.}$$

Therefore, the absolute value of deviation (AVD) between stock i and market yields is calculated by

$$\begin{aligned} AVD_{it} &= |E(R_{it}) - E(R_{mt})| \\ &= \left|R_f + \beta_i E\left(R_{mt} - R_f\right) - R_f - \beta_m E\left(R_{mt} - R_f\right)\right| \\ &= |\beta_i - \beta_m|\left|E\left(R_{mt} - R_f\right)\right|. \end{aligned} \tag{3.12.}$$

The absolute deviation of the cross-sectional yield of the entire market is shown below,

$$\begin{aligned} \mathrm{E}(CSAD_t) &= \frac{1}{N}\sum_{i=1}^{N} AVD_{it} \\ &= \frac{1}{N}\sum_{i=1}^{N}|\beta_i - \beta_m|\left|E\left(R_{mt} - R_f\right)\right| . \end{aligned} \tag{3.13.}$$

Using the absolute dispersion of the cross section to derive the market yield, you can get a linear relationship between the two factors as below,

$$\frac{\partial \mathrm{E}(CSAD_t)}{\partial |E(R_{mt})|} = \frac{1}{N}\sum_{i=1}^{N}|\beta_i - \beta_m| > 0, \tag{3.14.}$$

$$\frac{\partial \mathrm{E}(CSAD_t)}{\partial |E(R_{mt})|^2} = 0. \tag{3.15.}$$

Theoretically, under the premise of the rational capital asset pricing model, there is a positive correlation between the stock yield deviation and the market rate in the CSAD model, and it shows a linear relationship. It can be seen that the rational capital asset pricing model and the herding effect are the opposite. This shows that once the phenomenon of herding is indicated in the market, the hypothesis of the rational market can be rejected, and there are irrational factors in the market. From the perspective of the Chinese stock market, due to its low degree of marketization, the premise of the rational CAPM model cannot be fully satisfied. Thus, the interpretation ability of a single linear regression equation is not strong. Therefore, even if the regression coefficient in the regression equation is positive, it does not indicate that there is no herding effect. On this basis, the extended CSAD model further proposes the use of polynomial regression equations as below to test for the existence of herding effects:

$$CSAD_t = \beta_1 + \beta_2|R_{mt}| + \beta_3|R_{mt}|^2 + \varepsilon_1. \qquad (3.16.)$$

If there is a herding effect in the stock market, investors' investment decisions will tend to be market-oriented, that is, the return rate of investment stocks will gradually converge to the market return rate. Therefore, in the regression equation, the stock return deviation will decrease with the increase of the market yield (β2 is negative), or it can be expressed as the decrease of the deviation degree (β1 is negative). That is to say, the two exhibit a negative correlation or a decreasing nonlinear relationship. And if β_1 and β_2 are negative at the same time, then this indicates that there is a very significant herding effect in the market.

3.1.5 Adjusted CSAD Model

In this paper, when the empirical test of the existence of the herd effect in the Chinese stock market is carried out, the regression model adopted is a further improved CSAD model, that is, a nonlinear polynomial regression equation (3.16) is basis.

Using equation (3.16) to regress, on the one hand, can make full use of existing data to make a clearer and more effective empirical study of the herd effect in the stock market. Since the quadratic term of the independent variable is added to the regression equation, we can judge whether there is a herding effect in the market, and we can judge the strength of the herd effect and its changing trend by the size of the regression coefficient. On the other hand, because the CSAD model is based on the rational capital asset pricing model, once the empirical results reject the original hypothesis, it indicates that there are irrational factors in the real market.

Because the study of the herding effect of the stock market adopts the CSAD model, the most critical indicator is the absolute dispersion of the cross-section. Therefore, the cross-sectional absolute dispersion (CSAD) index needs to be calculated. The basic calculation formula is equation (3.8). However, in the study of this thesis, we use market capitalization value as weight to calculate the rate of market return R_{cmt}, the process is shown below,

$$R_{cmt} = \sum_{i=1}^{N} \alpha_i R_{it}\, , where\ \alpha_i = \frac{CV_i}{\sum_{i=1}^{N} CV_i}, \qquad \sum_{i=1}^{N} \alpha_1 = 1. \qquad (3.17.)$$

Therefore, the absolute deviation of the cross-section calculated on the basis of the weighted market rate of return is

$$CSAD_{ct} = \frac{1}{N}\sum_{i=1}^{N}|R_{it} - R_{cmt}|. \quad (3.18..)$$

The reason why we use the weighted market rate of return is that there is a large difference among the constituent stocks' capitalization. However, due to the imperfect development of China's stock market, there is often a certain "small company effect", that is, the stocks with lower market capitalization tend to have higher yields. Specifically, Banz (1981) proposed the concept of "Small Firm Effect". He found that there is a negative correlation between the listed company's circulation market and its stock return rate. Further research by Roll (1981) and Basu (1983) shows that the risk factors of smaller companies in the CAPM model are underestimated. The rate of return should be higher considering the higher risk. He and Rao (2013) conducted a correlation analysis of stock excess return rate by using the market value of circulation as a standard indicator of company size. It is found that there is a staged small company effect in China's A-share market. According to this theory, if an equal market rate of return is adopted, it is impossible to accurately measure the relationship between the market yield and the influence of stocks of different sizes. Therefore, in the process of measuring the market rate of return, we should take small company effect into consideration, which is caused by the difference level of market capitalization among China's stock market.

Although calculated using a weighted market rate of return, this calculation does not violate the preconditions of the CSAD model based on rational CAPM theory. It is also an empirical analysis of the relationship between the CSAD indicator and the weighted market rate of return to test the herding effect in the stock market. At the same time, according to the development characteristics of China's stock market, the active degree of stock market trading is currently a data indicator that investors pay more attention to. The active degree of trading reflects the investment direction of market investors, market investment sentiment, etc. Thus, when ordinary stock investors make investment strategies, they often use this as important data information for their investment decisions. Since these data information is transparent and open, it is the public information enjoyed by investors, so the public factors affecting the market are used as model explanatory variables. Therefore, this thesis selects daily turnover (V_m) as a variable to measure market oscillations and market trading activity. Therefore, the regression equation of this thesis is as follows,

$$CSAD_{ct} = \beta_0 + \beta_1|R_{cmt}| + \beta_2|R_{cmt}|^2 + V_m + \varepsilon_1. \quad (3.19..)$$

In the subsequent empirical analysis, this paper will use this adjusted CSAD model as the main model to develop an empirical analysis of the herd effect.

3.2 Linear Regression Model

In this thesis, we use linear regression to get the coefficient β_1 and β_2 to if there is a herd effect in the Chinese stock market. To achieve this aim, we need to carry out the linear regression model for five times in five different market to test each of them separately. In this part, the definition of the linear regression model and its assumption that needs to be satisfied in regression processes are introduced, which will help you understand the empirical part better.

3.2.1 Multiple Linear Regression Model

In statistics, linear regression is a linear approach to modelling the relationship between a dependent variable and one or more explanatory variables. The case of one explanatory variable is called simple linear regression. Explanatory variable, the process is called multiple linear regression. In this paper, we use multiple linear regression. The equation is shown below,

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + \beta_k X_{kt} + \mu_t, \ \mathrm{t} = 1,2,3 \ldots., \mathrm{k}, \qquad (3.20.)$$

where Y_t is the dependent variable, β_1 is the constant, $\beta_2, \beta_3, \beta_3 \ldots and\ \beta_k$ are the coefficient for each X_t, μ_t is the error term.

3.2.2 Stationary

A time series Y_t said to be strictly stationary if the joint distribution of $(\mathrm{Y}_{t1}, \mathrm{Y}_{t2}, \ldots, \mathrm{Y}_{tk})$ is identical to that of $(\mathrm{Y}_{t1+s}, \mathrm{Y}_{t2+s}, \ldots, \mathrm{Y}_{tk+s})$ for all integers s. Strict stationarity requires that the joint distribution of the subsequence $(\mathrm{Y}_{t1}, \mathrm{Y}_{t2}, \ldots, \mathrm{Y}_{tk})$ does not change when it is shifted by an arbitrary amount s. If we consider that stationarity requires that all moments of the joint distribution are invariant to time shifts, we can easily understand that the distribution that generate most economic time series are not strictly stationary.

A time series Y_t is said to be weakly or covariance stationary if the three condition hold true. The mean of the process is constant and equal to a μ; the variance of the process is time invariant and equal to finite constant σ^2; the covariance of the process should not be time dependent, it can be affected just by the distance between the two-time stick considered.

3.2.3 Autocorrelation

Autocorrelation characterizes the dependency between Y_t and its past values $\mathrm{Y}_{t-k}, k > 0$. The main idea shown in below,

$$\mu_t = \rho_1\mu_{t-1} + \rho_2\mu_{t-2} + \cdots + \rho_k\mu_{k-1} + \epsilon_t. \quad (3.21.)$$

The causes can be:

- Inertia time series,
- Nonstationarity,
- Manipulation of data,
- Neglet the lagged term for explanatory variables, and
- Inappropriate specification of the mathematical form of model.

If there is autocorrelation in the model, the estimation of β_i still linear unbiased as well as consistent and asymptotically normally distributed, but it is no longer asymptotically efficient. If there is autocorrelation, we should remove it away. We get two method to estimate the autocorrelation: graphical tests and DW tests.

We can use Ljun-Box test:

$H_0: \rho_1 = \cdots = \rho_m = 0\ againts\ H_1: \rho_k \neq 0, for\ k - 1, \ldots, m$

Calculated statistic $\mathrm{Q(m)} = \mathrm{n(n+2)}\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k} \sim \chi_{1-\alpha,m}^2$

For selected α if $\mathrm{Q(m)} > \chi_{1-\alpha,m}^2$, we reject H_0, it means that the time series Y_t exhibits statistically significant dependence structure up to lags m.

We can also use the DW test because it is better to see the result.

$H_0: \rho_1 = \cdots = \rho_m = 0$

$H_1: \rho_k \neq 0, for\ k - 1, \ldots, m$

In large samples d-statistics: $d = 2(1 - \hat{\rho})$ the formula will be as below,

$$\mathrm{d} = \frac{\sum_{t=2}^{n}(\hat{\mu}_t - \hat{\mu}_{t-1})^2}{\sum_{t=2}^{n}\hat{\mu}_t}. \quad (3.22.)$$

If there is no autocorrelation, ρ is 0 and d should be distributed randomly around 2.

If there is severe positive autocorrelation, ρ close to 1 and d close to 0.

If there is severe positive autocorrelation, ρ close to –1 and d close to 4.

There are various methods for elimination of serial autocorrelation：

1. To include lagged explained variable and use h-statistic
2. To use the Cochrane-Orcutt iterative procedure
3. To include lagged explanatory variables
4. To include trend variable if it is in residuals

Here we will explain the second method Cochrane-Orcutt iterative procedure in detail because this method is suited for our model improvement.

The original model is shown below,

$$Y_t = \beta_1 + \beta_2 X_{2_t} + \beta_3 X_{3_t} + \cdots + \beta_\text{k} X_{k_t} + u_t \, . \qquad (3.23.)$$

We multiply both sides of the equation by ρ, we get transformation model as below,

$$\rho Y_t = \rho\beta_1 + \rho\beta_2 X_{2_{t-1}} + \rho\beta_3 X_{3_{t-1}} + \cdots + \rho\beta_\text{k} X_{k_{t-1}} + u_{t-1}. \qquad (3.24.)$$

Equation (3.19) and equation (3.20) need to be subtracted and after the transformation we get

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_{2_t} - \rho X_{2_{t-1}}) + \beta_3(X_{3_t} - \rho X_{3_{t-1}})$$
$$+ \cdots + \beta_k(X_{k_t} - \rho X_{k_{t-1}}) + u_t - u_{t-1} \, . \qquad (3.25.)$$

Rewrite the equation (3.21), we get our new model:

$$Y_t{}^* = \beta_1{}^* + \beta_2 X_{2_t}{}^* + \beta_3 X_{3_t}{}^* + \cdots + \beta_\text{k} X_{k_t}{}^* + \epsilon_t, \qquad (3.26.)$$

where $Y_t{}^* = Y_t - \rho Y_{t-1}$, $\beta_1{}^* = \beta_1(1-\rho)$, $X_{i_t}{}^* = X_{i_t} - \rho X_{i_{t-1}}, \forall \text{i} = 2, \ldots, \text{k}$

3.2.4 Heteroscedasticity

One of the important assumptions of the classical linear regression model is that the variance of disturbances term, conditional on the chose values of the explanatory variables, is some constant number. The main idea is shown below,

$$\text{E}(\mu_t) = \sigma^2, t = 1,2, \ldots, n. \qquad (3.27.)$$

The causes can be:

- Data of outliers,
- Incorrect specified regression model,

- Incorrect data transformation, and
- Incorrect functional form.

We get two methods to detect the heteroscedasticity: graphical methods and statistic methods.

Here we only introduce one of the statistic methods, white test.

Residuals are determined from the estimate of the original regression model as below,

$$Y_t = \beta_1 + \beta_2 X_{t2} + \beta_3 X_{t3} + u_t. \quad (3.28.)$$

Estimation of a new regression model is shown as follows,

$$\hat{u}_t^{\ 2} = \alpha_1 + \alpha_2 X_{t2} + \alpha_3 X_{t3} + \alpha_4 X_{t2}^{\ 2} + \alpha_5 X_{t3}^{\ 2} + \alpha_6 X_{t2} X_{t3} + \varepsilon_t. \quad (3.29.)$$

Hypotheses formulation is shown as below,

$$H_0: \alpha_2 = 0 \wedge \alpha_3 = 0 \wedge \alpha_4 = 0 \wedge \alpha_5 = 0 \wedge \alpha_6 = 0,$$

$$H_1: \alpha_2 \neq 0 \vee \alpha_3 \neq 0 \vee \alpha_4 \neq 0 \vee \alpha_5 \neq 0 \vee \alpha_6 \neq 0.$$

We use following methods to get calculated statistics,

$\chi^2_{cal} = nR^2_{new} \sim \chi^2_{df}, df = k,$

where k is the number of explanatory variables in the new model.

The decision for the selected level of significance is 0.05, expressed in math language is $\alpha = 0.05$

If $\chi^2_{cal} = nR^2_{new} > \chi^2_{1-\alpha}(df) = CHINV(\alpha; df)$, then we reject H_0. It means that the variance of residuals depends on at least one explanatory variable in the new model at alfa level of significance. If there is heteroscedasticity, we have some approaches to remedial the problem. We investigate model specification (omitting a relevant explanatory variable, including an unnecessary or irrelevant explanatory, wrong functional form) variable.

There are several kinds of heteroscedasticity, linear dependence, quadratic dependence and quadratic dependence on estimated Y. The type can be shown by the graph test. The main method of reducing heteroscedasticity is using weighted least squares method. Different types of heteroscedasticity will use different weight. Here we take quadratic dependence on estimated Y as an example. The new model will be transformed as the equation below,

$$\frac{Y_t}{\hat{Y}_t} = \beta_1 \frac{1}{\hat{Y}_t} + \beta_2 \frac{X_{2_t}}{\hat{Y}_t} + \beta_3 \frac{X_{3_t}}{\hat{Y}_t} + \cdots + \beta_k \frac{X_{k_t}}{\hat{Y}_t} + \frac{u_t}{\hat{Y}_t}. \quad (3.30.)$$

If we transfer our model into the form of equation (3.27), the heteroscedasticity is removed for certain and no needs to test again.

3.2.5 Multicollinearity

The assumption of the classical regression model is that among the regressors included in the regression model is no multicollinearity. It means that there does not exist a "perfect" linear relationship among some or all explanatory variables of the regression model.

The causes can be:

- The data collection method employed,
- Constraints on the model or in the population being sampled,
- Model specification, and
- An overdetermined model.

If there is multicollinearity, we can still get estimates unbiased and consistent, but the variances and standard errors increase, confidence intervals are very large, estimates are less reliable and unstable estimates, which means that a slight change in dependent variables will cause a significant change in estimates.

Here are some ways to detect multicollinearity. We can test the pair-wise correlation among regressors. If detection rule meet $|\text{correlation}(X_{i1}, X_{i2})| < 0.8$, there is no multicollinearity.

Or we can use the variance inflation factor (VIF)

Calculated for each variable individually then regress one explanatory (X_i) the variable on the remaining set of variables

$$\text{VIF} = \frac{1}{1 - R_{xi}^2} \qquad (3.31.)$$

For VIF between 1 and 10, there is no multicollinearity.

3.2.5 Model Specification

Model specification is made if all important variables are in the model and linear dependence is the good one for the model. In this part, we should focus on predicted values and residuals. In the graphical test, we are looking at the development of standardized residuals. The development has to be in the confidential interval 95% with [-1.96;1.96].

One statistic method is linktest which works with predicted values. The main idea is to generate a new model as equation (3.32).

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \beta_{k+1}\hat{Y}_t + \beta_{k+2}\hat{Y}_t^2 + \mu_t, \ \ t = 1,2,3 \ldots., k \qquad (3.32.)$$

Compared new model equation (3.32) and the old model in equation (3.20), the difference is that in the new model we add two variables, the estimated Y and the squares of estimated Y.

Hypotheses:

H_0 : regression model is correctly specified

H_1 : regression model is not correctly specified

Decision rule:

If the squares of estimated Y is significant, we disapprove H_0 at significance level of α, which means the model is not correctly specified.

If the squares of estimated Y is not significant, we accept H_0 at significance level of α and our model is correctly specified.

Another condition is that the estimated Y should be significant considering it is an estimated value.

Another is ovtest. We use ovtest to exam is there any omitted variables in our model specification. The idea behind ovtest is very similar to linkest. It also creates new variables based on the predictors and refits the model those new variables to see if any of them be significant. The ovtest command indicates that there are omitted variables. So we have tried both the linktest and ovtest, and both of them tell us that we have a specification error at 5 % level of significance.

3.2.6 Normality

One assumption for the regression model is that the residuals have a normal distribution. The residuals are expressed as u_t in equation (3.17). After we regress the model, we can predict the u_t. Then use stat to draw the distribution of u_t, and compare it with normal distribution. From distribution picture, we can know the skewness and kurtosis of the residuals. We can also use statistic test to identify the characteristic of normality. Here we use Jarquethe -Bera test. In Jarque-Bera test, we first need to compute the skewness and kurtosis of the regression model.

Hypotheses:

H_0 : residuals are normal distributed

H_1 : residual are NOT normal distributed

Then we compute the Jarque-Bera statistic (for large samples).

$$\mathrm{JB} = \mathrm{n}\left[\frac{S^2}{6} + \frac{(k-3)^2}{24}\right] \approx \chi^2_{\alpha,df=2} \tag{3.33.}$$

If decision rule for alpha=0.5 significant level, we reject H_0.

$$\chi^2_{vyp} > \chi^2_{\alpha,df=2} = \mathrm{CHIINV}(\alpha = 0.05; \mathrm{df} = 2) \tag{3.34.}$$

In STATA, we use “sktest” to run above procedure, if Prob>chi2 is lower than 0.05, we reject H_0, which means residuals are not normal distributed.

4 Empirical Analysis of Herding Behavior in Chinese Stock Market

In this chapter, we focus on the empirical analysis of herding behavior in Chinese stock market. First, a brief introduction about the selected stock market will be introduced. Then we start the analysis part. Before the data analysis, we first prepare the data we used in regression analysis. Then we carry out adjusted SCAD model to test herd behavior in five different markets. All final model is shown in this chapter. For more detailed process, we take Shanghai A-share market and Shenzhen B-share market as examples. Finally, we show the final results for all five markets and make comparison.

4.1 Description of Selected Stock Market

The Shanghai Stock Exchange (SSE) came into existence on November 26, 1990, and on December 19 of the same year, it started formal operations. After 28 years of rapid growth, SSE has developed into a comprehensive exchange with stocks, bonds, funds and derivatives products, world-class exchange system and communication infrastructure, and effective self-regulatory system. With these advantages, the scale of the Shanghai securities market and the number of investors is growing fast. The Shanghai Stock Exchange is the world's 4th largest stock market by market capitalization at US$6 trillion as of December 2018.

Shenzhen Stock Exchange (SZSE), established on 1st December 1990, is a self-regulated legal entity under the supervision of China Securities Regulatory Commission (CSRC). It also organizes, supervises securities trading and performs duties prescribed by laws, regulations, rules and policies. The market capitalization of this exchange is smaller than the Shanghai Stock Exchange. SZSE is committed to developing China's multi-tiered capital market system, serving national economic development and transformation and supporting the national strategy of independent innovation. The SME Board was launched in May 2004. The ChiNext market was inaugurated in October 2009.

The official name of the A share is the ordinary stock of the RMB. It is issued by a company in China for domestic institutions, organizations or individuals (from April 1, 2013, domestic, Hong Kong, Macao, and Taiwan residents can open A-share accounts) to subscribe and trade common stocks in RMB. A-shares are not physical stocks. They are booked electronically without paper, and the "T+1" delivery system is implemented. There are restrictions on price increase (10%). Participating investors are institutions or individuals in mainland China.

The official name of the B-shares is the RMB special stock. It is a foreign share listed and traded on the stock exchanges in China (Shanghai, Shenzhen) in the form of RMB denominations, foreign currency subscriptions and purchases. The registration and listing of B-share companies are in the territory. Before 2001, investors were restricted to foreigners. After 2001, domestic private residents were allowed to invest in B-shares. Shanghai B-shares are quoted in U.S. dollars and Shenzhen B-share is quoted in HK dollars.

There are two types of stocks issued in SSE which are A-shares and B-shares. A-shares are quoted in RMB currency, while Unlike the Hong Kong Stock Exchange, the Shanghai Stock Exchange is still not entirely open to foreign investors. Foreign investors are now allowed (with limitations) to trade in A-shares under the Qualified Foreign Institutional Investor (QFII) program which was officially launched in 2003, while B-shares are available to both domestic (since 2001) and foreign investors.

The GEM, also known as the Second-board Market, is the second stock exchange market. It is a different type of securities market from the Main-Board Market. It is designed for entrepreneurial companies and small and medium-sized enterprises that cannot be listed on the Main Board for the time being. Enterprises and high-tech industrial enterprises that need financing and development provide financing channels and growth space for the securities trading market. The GEM is an important complement to the main board market and has an important position in the capital market. Compared with the mainboard market, the GEM is often more lenient, mainly reflected in the requirements of establishment time, capital scale, and medium and long-term performance. On October 30, 2009, China GEM (ChiNext) was officially run.

4.2Data Collection

Since the herd effect is reflected in the stock market through the longer-term effect, the sample time span selected in this paper is from January 2, 2014 to October 31, 2018, using 1178 daily data for empirical testing. In addition, the data in this paper is from the CSMAR database, and the Stata 15.1 software is used for empirical testing. The main data are downloaded included, daily return of each stock in five market (R_{it}), capital weighted daily market return (R_{cmt}), and daily trading volume in currency RMB (V_m).

We first calculate $CSAD_{ct}$ according to the equation (3.18) to get the controlled variable. The absolute value of capital weighted daily market return ($|R_{cmt}|$) and the square of absolute value of capital weighted daily market return ($|R_{cmt}|^2$) are also computed as independent variables. Combined with daily trading volume (V_m), all of them are input data for regression model. To make the regression process clearer, the data expression in Stata are shown in Tab. 4.1.

Table 4. 1 Data expression in Stata

Stata	Meaning		Stata	Meaning
csadcw	Original $CSAD_{ct}$		ny	$CSAD_{ct}$ after remove autocorrelation
absrm	Original $\|R_{cmt}\|$		nx1	$\|R_{cmt}\|$ after remove autocorrelation
rm2	Original $\|R_{cmt}\|^2$		nx2	$\|R_{cmt}\|^2$ after remove autocorrelation
vm	Original V_m		nx3	V_m after remove autocorrelation

Source: Authors' work

In Annexes 1, downloaded original data are available, while in Annexes 2 final input data are provided. Because of the large volume of data collection and calculation, the annexes only provide a small sample. The complete data and calculation process are available in attached CD. And all input data used in regression are prepared and stored in each market attached Stata file.

4.3 Empirical Results

In this thesis, we use adjusted SCAD model as equation (3.19) to test herd behavior in five different markets. It means that we carry out the linear regression for five times. Here we will take the Shanghai A-share market and Shenzhen B-share market as examples to show the detailed process of how we get the results and show the final results for all five markets and make comparison.

4.4.1 Shanghai A-Share Market

First of all, we need to test the stationary of the dependent and independent variables. We use the ADF test as we mentioned above, the results are shown below in Fig.4.1.

Figure 4. 1 ADF test for variables

```
. dfuller csadcw

Dickey-Fuller test for unit root                     Number of obs   =       929

                               ─────────── Interpolated Dickey-Fuller ──────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
────────────────────────────────────────────────────────────────────────────────
 Z(t)            -18.068            -3.430            -2.860            -2.570
────────────────────────────────────────────────────────────────────────────────
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller absrm

Dickey-Fuller test for unit root                     Number of obs   =       929

                               ─────────── Interpolated Dickey-Fuller ──────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
────────────────────────────────────────────────────────────────────────────────
 Z(t)            -23.550            -3.430            -2.860            -2.570
────────────────────────────────────────────────────────────────────────────────
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller rm2

Dickey-Fuller test for unit root                     Number of obs   =       929

                               ─────────── Interpolated Dickey-Fuller ──────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
────────────────────────────────────────────────────────────────────────────────
 Z(t)            -26.742            -3.430            -2.860            -2.570
────────────────────────────────────────────────────────────────────────────────
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller vm

Dickey-Fuller test for unit root                     Number of obs   =       929

                               ─────────── Interpolated Dickey-Fuller ──────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
────────────────────────────────────────────────────────────────────────────────
 Z(t)             -6.669            -3.430            -2.860            -2.570
────────────────────────────────────────────────────────────────────────────────
MacKinnon approximate p-value for Z(t) = 0.0000
```

Source: Authors' calculation

From the Fig. 4.1, we can find that MacKinnon approximate p-value for Z(t) is equal to zero, which is lower than 0.05, which means that for 95% confidence level that all variables are stationary.

Then we carry out the regression of our original model, and test for its autocorrelation.

Here we use the DW test to show the results. The details are provided in Fig 4.2 and Fig 4.3.

Figure 4. 2 Regression results of original model

```
. regress csadcw absrm rm2 vm

      Source |       SS           df       MS      Number of obs   =     1,178
-------------+----------------------------------   F(3, 1174)      =    306.17
       Model |  .037804309         3  .012601436   Prob > F        =    0.0000
    Residual |  .048320478     1,174  .000041159   R-squared       =    0.4389
-------------+----------------------------------   Adj R-squared   =    0.4375
       Total |  .086124787     1,177  .000073173   Root MSE        =    .00642

------------------------------------------------------------------------------
      csadcw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       absrm |   .2937347   .0367764     7.99   0.000     .2215799    .3658896
         rm2 |  -2.997931   .5899398    -5.08   0.000    -4.155385   -1.840477
          vm |   2.36e-14   1.03e-15    22.91   0.000     2.16e-14    2.56e-14
       _cons |   .0097091   .0003476    27.93   0.000     .0090271    .0103911
------------------------------------------------------------------------------
```

Source: Authors' calculation

Figure 4. 3 DW test of original model

```
. dwstat

Number of gaps in sample:  248

Durbin-Watson d-statistic(  4,  1178) =  .8041685
```

Source: Authors' calculation

From Fig. 4.3, the DW test indicates that the model has autocorrelation, because the Durbin-Watson d-statistic is 0.8, which is closer to 0 instead of 2. So, we need to find a method to remove the autocorrelation. There are four different methods are introduced in section 3.2.3, we test them all and only the Cochrane-Orcutt iterative procedure worked well in this model, which can be seen in Fig 4.4.

Figure 4. 4 Cochrane-Orcutt iterative procedure

```
. prais csadcw absrm rm2 vm, corc

Number of gaps in sample:  248
(note: computations for rho restarted at each gap)

Iteration 0:  rho = 0.0000
Iteration 1:  rho = 0.3985
Iteration 2:  rho = 0.4503
Iteration 3:  rho = 0.4565
Iteration 4:  rho = 0.4572
Iteration 5:  rho = 0.4572
Iteration 6:  rho = 0.4572
Iteration 7:  rho = 0.4572

Cochrane-Orcutt AR(1) regression -- iterated estimates
```

Source	SS	df	MS
Model	.009277287	3	.003092429
Residual	.022776194	925	.000024623
Total	.03205348	928	.00003454

Number of obs	=	929
F(3, 925)	=	125.59
Prob > F	=	0.0000
R-squared	=	0.2894
Adj R-squared	=	0.2871
Root MSE	=	.00496

csadcw	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
absrm	.2072673	.033117	6.26	0.000	.1422743	.2722604
rm2	-3.277718	.5684356	-5.77	0.000	-4.393291	-2.162145
vm	2.56e-14	1.53e-15	16.76	0.000	2.26e-14	2.86e-14
_cons	.0094114	.000494	19.05	0.000	.0084419	.0103809
rho	.4572409					

```
Durbin-Watson statistic (original)    0.804169
Durbin-Watson statistic (transformed) 1.625554
```

Source: Authors' calculation

From the Fig 4.4 we can find out that the Durbin-Watson statistic of new transformed model is 1.6 which is not perfect 2, but it is still a good result compared to the original 0.8.

The Fig. 4.4 shows that the iteration from 4 to 7 does not change and considering the stock market work five days out of seven, we choose iteration 5 (0.4572409) as ρ in our new model. According to the equation (3.26) we get our new variables, ${Y_t}^* = Y_t - 0.4572409Y_{t-1}$ and ${X_{i_t}}^* = X_{i_t} - 0.4572409X_{i_{t-1}}$, the new variables are named as ny, nx1, nx2, and nx3. The regression and autocorrelation test are shown below.

Figure 4. 5 New model regression and DW test

```
. regress ny nx1 nx2 nx3

      Source |       SS           df       MS      Number of obs   =       929
-------------+----------------------------------   F(3, 925)       =    125.59
       Model |  .009277287         3  .003092429   Prob > F        =    0.0000
    Residual |  .022776194       925  .000024623   R-squared       =    0.2894
-------------+----------------------------------   Adj R-squared   =    0.2871
       Total |  .032053481       928   .00003454   Root MSE        =    .00496

------------------------------------------------------------------------------
          ny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         nx1 |   .2072673   .0331169     6.26   0.000     .1422743    .2722604
         nx2 |  -3.277717   .5684356    -5.77   0.000     -4.39329   -2.162144
         nx3 |   2.56e-14   1.53e-15    16.76   0.000     2.26e-14    2.86e-14
       _cons |   .0051081   .0002681    19.05   0.000     .0045819    .0056344
------------------------------------------------------------------------------

. dwstat

Number of gaps in sample:  246

Durbin-Watson d-statistic(  4,   929) =  1.625554
```

Source: Authors' calculation

From the Fig. 4.5 we can find that after remove the autocorrelation, the R-square also decreased from 0.43 to 0.28.

Then we use White test to see is there any heteroscedasticity.

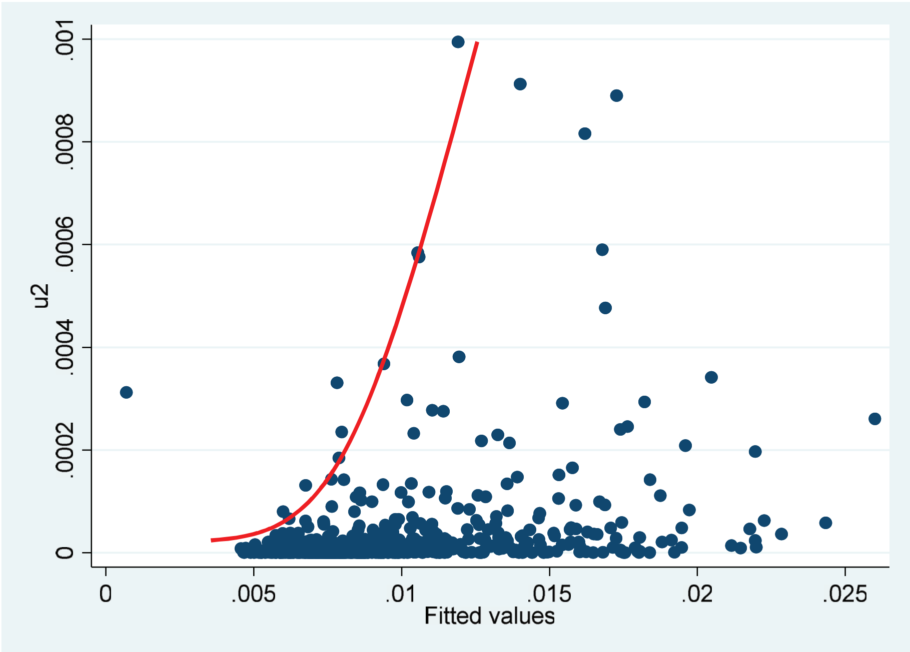
Figure 4. 6 White test

```
. estat imtest,white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(9)      =    174.09
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |     174.09      9    0.0000
            Skewness |      44.58      3    0.0000
            Kurtosis |      11.67      1    0.0006
---------------------+-----------------------------
               Total |     230.35     13    0.0000
---------------------------------------------------
```

Source: Authors' calculation

From the Fig. 4.6, we can find that P-value=0<0.05, which means we can reject the H_0, in other words, the error term is heteroscedasticity. To reduce the heteroscedasticity, we need to know the type of it. The graphical test is used as a tool, the results are below.

Figure 4. 7 Graph test



Source: Authors' calculation

From the Fig. 4.7, we can interpret that there is quadratic dependence on estimated Y, so we use estimated Y as weight to regress. The results of weighted least squares method are shown in Fig. 4.8.

Figure 4. 8 Weighted regression results

```
. regress ny nx1 nx2 nx3 [aweight = Y_pre]
(sum of wgt is 8.582318671804387)
```

Source	SS	df	MS			
				Number of obs	=	929
				F(3, 925)	=	134.24
Model	.013203422	3	.004401141	Prob > F	=	0.0000
Residual	.030327569	925	.000032787	R-squared	=	0.3033
				Adj R-squared	=	0.3011
Total	.043530991	928	.000046908	Root MSE	=	.00573

ny	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
nx1	.2583002	.035771	7.22	0.000	.1880984	.3285019
nx2	-4.635834	.594371	-7.80	0.000	-5.802306	-3.469362
nx3	2.41e-14	1.39e-15	17.38	0.000	2.14e-14	2.68e-14
_cons	.0052542	.000316	16.62	0.000	.0046339	.0058744

Source: Authors' calculation

The Fig. 4.8 shows that our new model fits the T-test and F-test, which means the variables and model are significant. And R-adjusted is 0.3.

After these we test for the multicollinearity.

Figure 4. 9 Correlations between variables

```
. pwcorr ny nx1 nx2 nx3
```

	ny	nx1	nx2	nx3
ny	1.0000			
nx1	0.1925	1.0000		
nx2	0.0861	0.8936	1.0000	
nx3	0.5091	0.2478	0.1923	1.0000

Source: Authors' calculation

We can see from the Fig. 4.9 that most correlation between variables are lower than 0.8, except the correlation between nx1 and nx2, which is decided by the model, because the nx2 is the squared nx1.

Figure 4. 10 VIF test

```
. vif
```

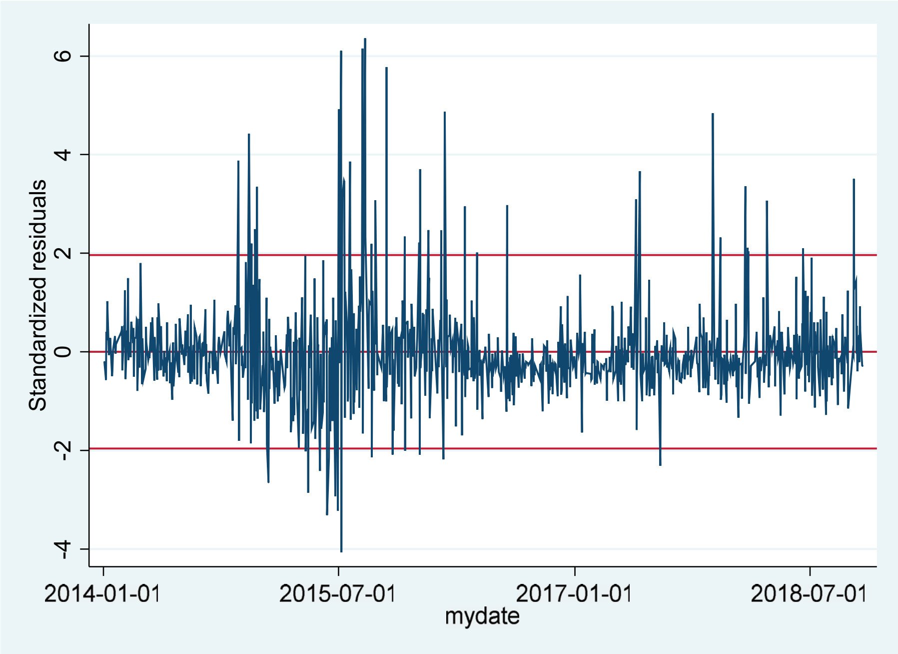
Variable	VIF	1/VIF
nx1	5.55	0.180023
nx2	5.44	0.183809
nx3	1.06	0.941051
Mean VIF	4.02	

Source: Authors' calculation

And the VIF is 4.02, which is very close to 1. Both tests show that there is no multicollinearity in our model.

So, we can continue with the normality. We first start from the graphical test.

Figure 4. 11 Graghical test for model specification



Source: Authors' calculation

Fig. 4.11 shows the development of standardized residuals. The best situation is the development should be random and be in confidence interval, between the two red lines [-1.96, 1.96]. In picture above, the majority of the blue line located between the two red lines, but there are some outliners, so we believe that even though our model specification is not perfect, it is good.

Figure 4. 12 Linktest

```
. linktest
(sum of wgt is 8.582318671804387)
```

Source	SS	df	MS		
				Number of obs =	929
				F(2, 926) =	205.83
Model	.013396361	2	.006698181	Prob > F =	0.0000
Residual	.03013463	926	.000032543	R-squared =	0.3077
				Adj R-squared =	0.3062
Total	.043530991	928	.000046908	Root MSE =	.0057

ny	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_hat	1.630591	.2636948	6.18	0.000	1.113082	2.1481
_hatsq	-24.39536	10.019	-2.43	0.015	-44.05794	-4.732777
_cons	-.0035627	.0015615	-2.28	0.023	-.0066273	-.0004982

Source: Authors' calculation

From Fig. 4.12, the test of both estimated Y (_hat) and squares of estimated Y(_hatsq) are significant, because the P-value is lower than 0.05. This is to say that linktest rejects the assumption that the model is specified correctly. We believe it is because the big fluctuate in the market.

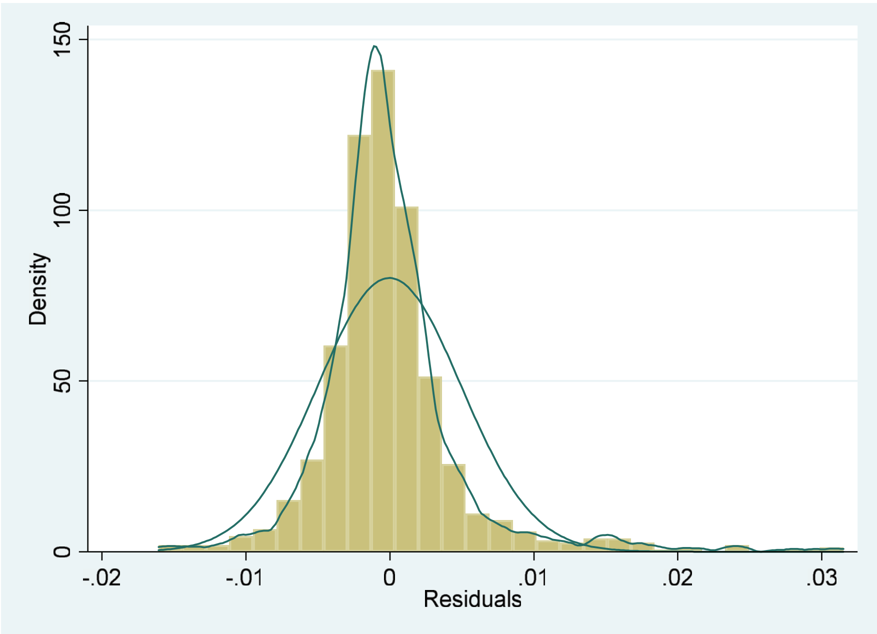
Figure 4. 13 Ovtest

```
. ovtest

Ramsey RESET test using powers of the fitted values of ny
       Ho:  model has no omitted variables
                 F(3, 922) =      7.70
                  Prob > F =      0.0000
```

Source: Authors' calculation

The ovtest performs a regression specification error test (RESET) for omitted variables. The results indicate that there are omitted variables, because Prob > F =0.0000 < 0.05.

Figure 4. 14 Residuals distribution compared with normal distribution



Source: Authors' calculation

Figure 4. 15 Sktest

```
. sktest u_t

                   Skewness/Kurtosis tests for Normality
                                                          ------ joint ------
    Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
-------------+---------------------------------------------------------------
         u_t |        929     0.0000        0.0000            .        0.0000
```

Source: Authors' calculation

From the Fig. 4.14 and Fig. 4.15, we can find that the residuals distribution in our model has positive skewness and the kurtosis is almost zero.

From statistical testing normality testing, we can find that Prob > chi2 = 0.0000 < 0.05, which means that we can reject the H_0, in other words, residuals are not normal distributed. In order to get the best model, we look back to see each assumption of different model. The results are shown in Tab. 4.2

Table 4. 2 Comparison of different models in Shanghai A-share Market

	R-adjusted	Durbin-Watson	White test Prob>chi2	Vif	Linktest _hatsq	Ovtest Prob > F	Sktest Prob>chi2
Model1	0.4375	0.8042	0.0000	4.11	0.001	0.0000	0.0000
Model2	0.2871	1.6255	0.0000	3.72	0.041	0.0000	0.0000
Model3	0.3011	1.6150		4.02	0.015	0.0000	0.0000

Source: Authors' calculation

In the Tab. 4.2, model 1 is the original model, model 2 is the model reduced autocorrelation, model 3 is the model reduced autocorrelation and heteroscedasticity. Some numbers in the Tab. 4.2 are underlined, which means the test results are not ideal and the model does not fit the assumptions.

From Tab. 4.2 we can get the final conclusion that the model 2 is best for now. All three models have problem with model specification and normality. However, when we consider the linktest, the second model, the P-value is closer to 0.05. Model 2 does not remove the heteroscedasticity, and compared with the original model, the R-adjusted are lower, because we remove some autocorrelation. The detailed model regression is shown in Fig. 4.16.

Figure 4. 16 Shanghai A-share regression (removed autocorrelation)

```
. regress ny nx1 nx2 nx3
```

Source	SS	df	MS			
Model	.009277287	3	.003092429	Number of obs	=	929
Residual	.022776194	925	.000024623	F(3, 925)	=	125.59
				Prob > F	=	0.0000
				R-squared	=	0.2894
				Adj R-squared	=	0.2871
Total	.032053481	928	.00003454	Root MSE	=	.00496

ny	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
nx1	.2072673	.0331169	6.26	0.000	.1422743	.2722604
nx2	-3.277717	.5684356	-5.77	0.000	-4.39329	-2.162144
nx3	2.56e-14	1.53e-15	16.76	0.000	2.26e-14	2.86e-14
_cons	.0051081	.0002681	19.05	0.000	.0045819	.0056344

Source: Authors' calculation

As Fig. 4.16 shown. Both our model and variables are significant. And the final regression result is equation (4.1).

$$Y = 0.2073X_1 - 3.2778X_2 + 2.56e^{-14}X_3 + 0.0051$$
$$(0.0000) \quad (0.0000) \quad (0.0000) \quad (0.0000) \qquad (4.1.)$$

Because the coefficient for X_2 is negative, there is herding behavior in the Shanghai A-share market. But there are still some shortcomings in our model, the similar problems also shown in following market. Some of them can be solved. The calculation procedures in following five markets are similar to Shanghai A-share market. All market regressions have autocorrelation, we solve it by using Cochrane-Orcutt method. We reduce the heteroscedasticity by using weighted least squares method in five market. But there are still some problems we cannot fixed including the model specification and normality. So, we just provide the comparison tables and final regression results.

4.4.2 Shanghai B-Share Market

Now, we start from the Shanghai B-share Market. All assumption test results for three different model are shown in Tab. 4.3.

Table 4. 3 Comparison of different models in Shanghai B-share Market

	R-adjusted	Durbin-Watson	White test Prob>chi2	Vif	Linktest _hatsq	Ovtest Prob > F	Sktest Prob>chi2
Model1	0.6693	1.0769	0.0000	5.15	0.027	0.1016	0.0000
Model2	0.5442	1.5449	0.0000	3.83	0.010	0.0185	0.0000
Model3	0.6441	1.4934		3.97	0.001	0.0005	0.0000

Source: Authors' calculation

From Tab. 4.3, we can find that the situation in Shanghai B-share Market are similar to Shanghai A-share Market. The only difference is the ovtest indicated that in the original model there are no omitted variables, which does not make too much difference in final model choice. Considering other test results and model comparations, we still choose the second model which remove the autocorrelation as our final model in Shanghai B-share market. The detailed result is shown in Fig. 4.17.

Figure 4. 17 Shanghai B-share regression (removed autocorrelation)

```
. regress ny nx1 nx2 nx3

      Source |       SS           df       MS      Number of obs   =       929
-------------+----------------------------------   F(3, 925)       =    370.31
       Model |  .006928462         3  .002309487   Prob > F        =    0.0000
    Residual |  .005768861       925  6.2366e-06   R-squared       =    0.5457
-------------+----------------------------------   Adj R-squared   =    0.5442
       Total |  .012697323       928  .000013682   Root MSE        =     .0025

------------------------------------------------------------------------------
          ny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         nx1 |   .2301153   .0159481    14.43   0.000     .1988168    .2614139
         nx2 |  -3.111175   .2024568   -15.37   0.000    -3.508503   -2.713847
         nx3 |   4.08e-11   1.75e-12    23.28   0.000     3.73e-11    4.42e-11
       _cons |   .0031955   .0001097    29.14   0.000     .0029802    .0034107
------------------------------------------------------------------------------
```

Source: Authors' calculation

From Fig. 4.17 we can know the final model is shown below as equation (4.2).

$$Y = 0.2301X_1 - 3.1112X_2 + 4.08e^{-1}\ X_3 + 0.0032$$

$$(0.0000) \quad (0.0000) \quad (0.0000) \quad (0.0000) \qquad (4.2.)$$

Both our model and variables are significant. Similar as Shanghai A-share market, because the coefficient for X_2 is negative, there is herding behavior in the Shanghai B-share market. But there are still some shortcomings in our model. The linktest and ovtest indicate that the model is not correctly specified. And the sktest show the residuals are not normal distributed. Because we choose Model 2, there are also heteroscedasticity in the model.

4.4.3 Shenzhen A-Share Market

The test results of assumptions for different model in Shenzhen A -share market are shown in Tab. 4.4.

Table 4. 4 Comparison of different models in Shenzhen A-share Market

	R-adjusted	Durbin-Watson	White test Prob>chi2	Vif	Linktest _hatsq	Ovtest Prob > F	Sktest Prob>chi2
Model1	0.4552	1.0153	0.0000	4.84	0.117	0.0000	0.0000
Model2	0.3122	1.6243	0.0000	3.78	0.001	0.0005	0.0000
Model3	0.3723	1.4408		3.64	0.000	0.0000	0.0000

Source: Authors' calculation

For the same reason of previous markets, we choose the second model as final results. More detailed results are shown in Fig. 4.18.

Figure 4. 18 Shenzhen A-share regression (removed autocorrelation)

```
. regress ny nx1 nx2 nx3

      Source |       SS           df       MS      Number of obs   =       929
-------------+----------------------------------   F(3, 925)       =    141.42
       Model |  .0063326         3  .002110867   Prob > F        =    0.0000
    Residual |  .0138066       925  .000014926   R-squared       =    0.3144
-------------+----------------------------------   Adj R-squared   =    0.3122
       Total |  .020139199       928  .000021702   Root MSE        =    .00386

------------------------------------------------------------------------------
          ny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         nx1 |   .1472482   .0209908     7.01   0.000     .1060531    .1884434
         nx2 |  -2.963351   .2941696   -10.07   0.000    -3.540668   -2.386034
         nx3 |   3.33e-14   2.01e-15    16.61   0.000     2.94e-14    3.73e-14
       _cons |   .0041585   .0002342    17.76   0.000     .0036989    .0046181
------------------------------------------------------------------------------
```

Source: Authors' calculation

From Fig. 4.18 we can know the final model is shown below. As Fig. 4.18 shown. Both our model and variables are significant.

$$Y = 0.1472X_1 - 2.9633X_2 + 3.33e^{-14}X_3 + 0.0042$$
$$(0.0000) \quad (0.0000) \quad (0.0000) \quad (0.0000) \qquad (4.3.)$$

Because the coefficient for X_2 is negative, there is herding behavior in the Shenzhen A-share market. But there are still same shortcomings as our previous market model.

For now, we have results in the three markets: Shanghai A-share market, Shanghai B-share market and Shenzhen A-share Market. In all three market we all get the negative coefficient for X_2, which means even though our models are not perfect, the herding behavior exists in all three markets. Limited by the meaning of our variables and economic model, we need to tradeoff between economic meaning and statically significant. Thus, all models have common problem of heteroscedasticity, model specification and residuals normal distribution, which cannot be reduced.

4.4.4 Shenzhen B-Share Market

Then we analyze the Shenzhen B-share Market. Shenzhen B-share market has better match with the model assumption. We can see the results first and then show more in procedures. The test comparation for models is shown in Tab. 4.5.

Table 4. 5 Comparison of different models in Shenzhen B-share Market

	R-adjusted	Durbin-Watson	White test Prob>chi2	Vif	Linktest _hatsq	Ovtest Prob > F	Sktest Prob>chi2
Model1	0.5931	1.0753	0.0000	3.82	0.657	0.3563	0.0000
Model2	0.4440	1.4556	0.0000	2.82	0.128	0.2184	0.0000
Model3	0.5734	1.4568		3.01	0.002	0.0000	0.0030

Source: Authors' calculation

From Tab. 4.5, we can get the final conclusion that the model 2 is best for now. The detailed regression is shown in Fig. 4.19

Figure 4. 19 Shenzhen B-share regression (removed autocorrelation)

```
. regress ny nx1 nx2 nx3

      Source |       SS           df       MS      Number of obs   =       929
-------------+----------------------------------   F(3, 925)       =    247.97
       Model |  .006489729         3  .002163243   Prob > F        =    0.0000
    Residual |  .008069413       925  8.7237e-06   R-squared       =    0.4457
-------------+----------------------------------   Adj R-squared   =    0.4440
       Total |  .014559142       928  .000015689   Root MSE        =    .00295

------------------------------------------------------------------------------
          ny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         nx1 |   .2943821   .0198405    14.84   0.000     .2554446    .3333196
         nx2 |  -3.962942   .2917323   -13.58   0.000    -4.535476   -3.390408
         nx3 |   9.62e-12   5.15e-13    18.69   0.000     8.61e-12    1.06e-11
       _cons |   .0034992   .0001382    25.32   0.000      .003228    .0037704
------------------------------------------------------------------------------
```

Source: Authors' calculation

From Fig. 4.19 we can know the final model are shown in equation (4.4). Both our model and variables are significant.

$$Y = 0.2944X_1 - 3.9629X_2 + 9.62e^{-12}X_3 + 0.0035$$

$$(0.0000) \quad (0.0000) \quad (0.0000) \quad (0.0000) \qquad (4.4.)$$

Because the coefficient for X_2 is negative, there is herding behavior in the Shenzhen B-share market. The detailed procedures are shown from Fig. 4.20 to Fig. 4.27.

Figure 4. 20 Cochrane-Orcutt method

```
. prais csadcw absrm rm2 vm

Number of gaps in sample:  248
(note: computations for rho restarted at each gap)

Iteration 0:  rho = 0.0000
Iteration 1:  rho = 0.3125
Iteration 2:  rho = 0.3627
Iteration 3:  rho = 0.3718
Iteration 4:  rho = 0.3735
Iteration 5:  rho = 0.3738
Iteration 6:  rho = 0.3739
Iteration 7:  rho = 0.3739
Iteration 8:  rho = 0.3739
Iteration 9:  rho = 0.3739

Prais-Winsten AR(1) regression -- iterated estimates
```

Source	SS	df	MS			
				Number of obs	=	1,178
				F(3, 1174)	=	563.99
Model	.015339933	3	.005113311	Prob > F	=	0.0000
Residual	.010643843	1,174	9.0663e-06	R-squared	=	0.5904
				Adj R-squared	=	0.5893
Total	.025983777	1,177	.000022076	Root MSE	=	.00301

csadcw	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
absrm	.3205702	.0176001	18.21	0.000	.2860391	.3551014
rm2	-4.035968	.2462937	-16.39	0.000	-4.519193	-3.552743
vm	9.97e-12	4.18e-13	23.88	0.000	9.16e-12	1.08e-11
_cons	.0058642	.0001902	30.84	0.000	.0054911	.0062373
rho	.3738846					

```
Durbin-Watson statistic (original)    1.075318
Durbin-Watson statistic (transformed) 1.612121
```

Source: Authors' calculation

From Fig. 4.20, we can see clearly that the Durbin-Watson statistic for original model is 1.0753(ideal Durbin-Watson statistic should be 2), which means we need to remove the autocorrelation. If we get new variable by using fitted rho, following equation mentioned in theory part, we get $Y_t^* = Y_t - 0.3738846Y_{t-1}$, $X_{i_t}^* = X_{i_t} - 0.3738846X_{i_{t-1}}$. After the transformation the new Durbin-Watson statistic is 1.6121, higher than original one. So, we do reduced autocorrelation. The regression result is in Fig. 4.19 above. Considering we get new variables, we need to test the stationarity of new variables, the results are shown in Fig. 4.21.

Figure 4. 21 stationarity test

```
. dfuller ny

Dickey-Fuller test for unit root                    Number of obs   =       682

                               ---------- Interpolated Dickey-Fuller ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -26.970            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller nx1

Dickey-Fuller test for unit root                    Number of obs   =       682

                               ---------- Interpolated Dickey-Fuller ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -22.753            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller nx2

Dickey-Fuller test for unit root                    Number of obs   =       682

                               ---------- Interpolated Dickey-Fuller ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -13.175            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller nx3

Dickey-Fuller test for unit root                    Number of obs   =       682

                               ---------- Interpolated Dickey-Fuller ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -14.013            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

Source: Authors' calculation

From the Fig. 4.21, we can find that MacKinnon approximate p-value for Z(t) is equal to zero, which is lower than 0.05, which means that for 95% confidence level that all variables are stationary.

Then we use white test to see is there any heteroscedasticity.

Figure 4. 22 White test

```
. estat imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(9)      =    132.20
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	132.20	9	0.0000
Skewness	21.00	3	0.0001
Kurtosis	2.71	1	0.0996
Total	155.91	13	0.0000

Source: Authors' calculation

From the Fig. 4.22, we can find that P-value=0<0.05, which means we can reject the H_0, in other words, the error term is heteroscedasticity. To reduce the heteroscedasticity, we should to know the type and use weighted least squares method to remove the heteroscedasticity. But if we remove the heteroscedasticity, the model will have problem of specification. So, we choose the model with heteroscedasticity and continues to test other assumptions.

Figure 4. 23 VIF test

```
. vif
```

Variable	VIF	1/VIF
nx1	3.77	0.265341
nx2	3.59	0.278440
nx3	1.11	0.899159
Mean VIF	2.82	

Source: Authors' calculation

We can see from the Fig. 4.23 that the VIF is 2.82, which is between 1 and 10, which means there is no multicollinearity in the model. Then we continue with the linktest, which is used to test whether the model is correctly specified.

Figure 4. 24 Linktest

```
. linktest

      Source |       SS           df       MS      Number of obs   =       929
-------------+----------------------------------   F(2, 926)       =    374.46
       Model |  .006509899         2  .00325495   Prob > F        =    0.0000
    Residual |  .008049243       926  8.6925e-06   R-squared       =    0.4471
-------------+----------------------------------   Adj R-squared   =    0.4459
       Total |  .014559142       928  .000015689   Root MSE        =    .00295

------------------------------------------------------------------------------
          ny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   .8855384    .083581    10.59   0.000     .7215083    1.049568
      _hatsq |    5.61216   3.684295     1.52   0.128    -1.618376     12.8427
       _cons |   .0004533   .0003859     1.17   0.240     -.000304    .0012107
------------------------------------------------------------------------------
```

Source: Authors' calculation

From Fig. 4.24, the test of estimated Y is significant, because P-value for _hat is lower than 0.05. And the test of squares of estimated Y is not significant, because P-value for _hatsq is higher than 0.05. This is to say that linkest cannot reject the assumption that the model is specified correctly. Thus, we can conclude that the model is correctly specified.

Next is ovtest. The result is shown below.

Figure 4. 25 Ovtest

```
. ovtest

Ramsey RESET test using powers of the fitted values of ny
       Ho:  model has no omitted variables
                 F(3, 922) =      1.48
                  Prob > F =      0.2184
```

Source: Authors' calculation

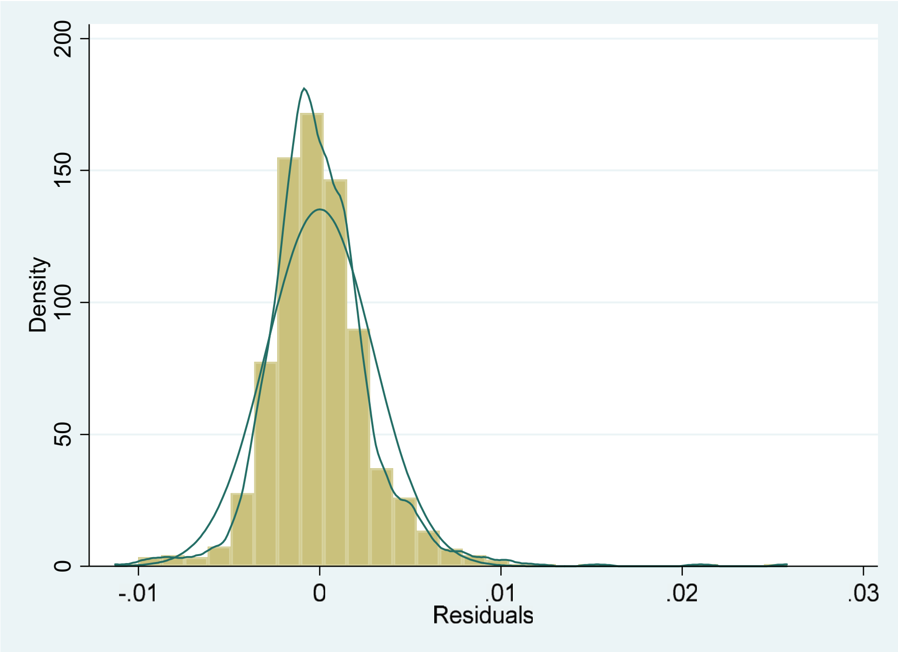
The ovtest performs a regression specification error test (RESET) for omitted variables. The ovtest command indicates that there are no omitted variables, because Prob > F =0.2184 > 0.05. Finally, we come to the sktest for normal distribution of residuals.

Figure 4. 26 Sktest

```
. sktest ut

                    Skewness/Kurtosis tests for Normality
                                                          ------ joint ------
    Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
-------------+---------------------------------------------------------------
          ut |        929     0.0000        0.0000             .        0.0000
```

Source: Authors' calculation

Figure 4. 27 Residuals distribution compared with normal distribution



Source: Authors' calculation

From statistical testing normality testing, we can find that Prob > chi2 = 0.0000 < 0.05, which means that we can reject the H_0, in other words, residuals are not normally distributed. And from graphical test, we can find that the residuals distribution in our model is leptokurtic.

4.4.5 Shenzhen Second Board Market

The results for the Shenzhen second board market are shown in Tab. 4.6.

Table 4. 6 Comparation of different model in Shenzhen Second Board Market

	R-adjusted	Durbin-Watson	White test Prob>chi2	Vif	Linktest _hatsq	Ovtest Prob > F	Sktest Prob>chi2
Model1	0.4103	0.8451	0.0000	5.33	0.000	0.0000	0.0000
Model2	0.4000	1.7149	0.0000	4.18	0.000	0.0000	0.0000
Model3	0.5117	1.5453		4.09	0.000	0.0000	0.0030

Source: Authors' calculation

From Tab. 4.6, we can may think model 3 is best for now. All three models have problem with model specification and normality. And model 3 have better sktest results. But if we look in to linktest details, Model 2 is better than Model 3. Because even though the squares of estimated Y is significant in both models, in model 2 the estimated Y is significant, which means the model worked well. However, in Model 3, linktest shows the estimated Y is not significant while the squares of estimated Y is significant, which means both new created variables are not fit for the ideal situation. Also, considering the comparation between different market, we choose Model 2 as final results. And the detailed model regression is shown in Fig. 4.28.

Figure 4. 28 Second Board regression (removed autocorrelation)

```
. regress ny nx1 nx2 nx3

      Source |       SS           df       MS      Number of obs   =       929
-------------+----------------------------------   F(3, 925)       =    207.20
       Model |  .011187365         3  .003729122   Prob > F        =    0.0000
    Residual |  .016647824       925  .000017998   R-squared       =    0.4019
-------------+----------------------------------   Adj R-squared   =    0.4000
       Total |   .02783519       928  .000029995   Root MSE        =    .00424

------------------------------------------------------------------------------
          ny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         nx1 |   .2257706   .0190605    11.84   0.000     .1883638    .2631773
         nx2 |  -4.211259   .2539431   -16.58   0.000     -4.70963   -3.712887
         nx3 |   1.15e-13   8.03e-15    14.31   0.000     9.91e-14    1.31e-13
       _cons |    .003251   .0002498    13.01   0.000     .0027608    .0037413
------------------------------------------------------------------------------
```

Source: Authors' calculation

As Fig. 4.28 shown. Both our model and variables are significant. And the final regression result is equation (4.5).

$$\mathrm{Y} = 0.2258X_1 - 4.2113X_2 + 1.15e^{-13}X_3 + 0.0033$$

$$(0.0000) \quad (0.0000) \quad (0.0000) \quad (0.0000) \qquad (4.5.)$$

Because the coefficient for X_2 is negative, there is herding behavior in the Shenzhen second board market. However, there are still some shortcomings in our model, the similar problems also shown in previous markets including the heteroscedasticity, model specification and normality.

4.4.6 Results Overview and Comparation

In this part we will summarize five final regression models and make comparation among them. Then we will give some suggestions of improving the model in the future empirical analysis.

Table 4. 7 Final model comparation among five markets

	β_1	β_2	β_3	R-adjusted	Model
Shanghai A-share	0.2072	−3.2777	$2.56e^{-14}$	0.2871	2
Shanghai B-share	0.2301	−3.1112	$4.08e^{-11}$	0.5442	2
Shenzhen A-share	0.1472	−2.9633	$3.33e^{-14}$	0.3122	2
Shenzhen B-share	0.2944	−3.9629	$9.62e^{-12}$	0.4440	2
Second-board	0.2258	−4.2113	$1.15e^{-13}$	0.4000	2

Source: Authors' calculation

From Tab. 4.7 we can find that the coefficient for $|R_{mt}|$ (β_1) are positive, which indicates that there is a linear increasing relationship between the cross-sectional absolute deviation ($CSAD_t$)and the market return rate ($|R_{mt}|$) at the 5% significance level. In other words, when the market return rate increases, the cross-sectional absolute deviation will increase as well. If herd behavior exists in the market, there are two form of representation in the regression equation. One is that the stock return deviation will decrease with the increase of the market yield (β2 is negative); or it can be expressed as the decrease of the deviation degree (β1 is negative). In our model, the coefficients for squares of the market return rate (β_2) are negative for five markets. It can be seen that as the market yield increases, the deviation between the stock return rate and the market yield rate shows a trend of shrinking, and this also indicates to some extent that when the market yield stocks rise or fall, the market investors, there is a trend of behavior that pursues market returns, that is, there is a certain herd effect behavior. β_3 is coefficient of trading volume, the size of β_3 reflects the size and activity of the market. In our model the trading volume variable is significant, indicating that investors are also considering the market's historical information such as the activity level of the transaction while chasing the market returns.

If we only look in to Shanghai market, we can find that the herd behavior is more significant in A-share instead of B-share. But in Shenzhen market, B-share have more serious herd behavior. Second board market have most obviously herd behavior in all five markets. The main reason is the Chinese second-board are not mature. The ChiNext market was inaugurated in October 2009 and only developed for ten years. Chinese economy are still in the transformation period and government are still trying to improve market mechanism. Another reason is the high weight of speculators in second-board. Due to the high risk and high returns of the second board market, many short-term investors have been attracted. The high proportion of individual investors is also one of the reasons. These people are more likely to be influenced by the environment to change their own decisions without a professional investment philosophy.

Our model also including some shortcomings, one of is low R-adjusted, which means there is only a small part of deviation can be explained by our model. And the final model we use does not fit the assumption of heteroscedasticity, model specification and normality. In future work we can take some measures to modify our model in order to overcome these existing disadvantages. First, we can include more variables that people are able to observe before they make investment decision in stock market, such as amplitude. The more variable includes may also increase the degree of fit. Then, we can improve the model by controlling variables unit and magnitude. And we can change the trading volume in to turnover rate, thus the coefficient will be more readable.

5 Conclusion

In traditional economy, rational asset pricing models arguing that investors are rational. The hypothesis of rational man require people acts rationally and with complete knowledge, who seeks to maximize personal utility or satisfaction. Behavioral finance believes that human reason is limited. Because of bounded rationality, investors disregard their own beliefs in order to conform to market consensus - also known as herd behavior. The focus of this paper is on the observation of investment behavior in China's stock market including Shanghai and Shenzhen markets. Significant herd behavior among investors has the potential to push prices away from fundamentals and may cause price volatility, leading to excessive stock market volatility and a bubble-like pattern.

The objective of the work is to test the presence of herding behavior in the Chinese stock market. There are five chapters in this thesis. The first chapter is the general structure and organization of this thesis. To achieve the aim, we first two introduced the literature related to herd behavior, including the efficient market hypothesis, behavioral finance, the definition, the causes and classification of herd behavior. Then we introduced the five economical methods (LSV, PCM, CSSD, CSAD and adjusted CSAD method) and econometric method (linear regression model) to measure the herd behavior. In chapter four, we applied adjusted CSAD model in the empirical market analysis. We show the results calculated by STATA software and make comparison of five submarkets.

In this thesis, herd behavior in stock market is measured by adjusted CSAD model. Based on the CSAD model, this paper tests the correlation between the deviation index and the market return rate by constructing the deviation index between stock return rate and weighted market return rate, and verifies the phenomenon of herd phenomenon in China's stock market as a whole. The five regression shows that as the market yield increases, the deviation between the stock return rate and the market yield rate shows a trend of shrinking, because of negative coefficient ($\beta 2$). This also indicates that when the market yield stocks fluctuate more seriously, the difference between market investors' investment decision become smaller, they all trend to pursues the same action as market shows, that is, there is a certain herd behavior. And if we look into the serious of herding behavior, the second-board market have the most significant herd behavior. The reason is the short development time for second-board and high proportion of speculators and individual investors.

Therefore, in terms of rational investment, investors need to improve their ability to acquire and analyze information, accurately determine the timing of investment, and take the time to break through the market effect of the herd effect, while diversifying investment to reduce systemic risks. In terms of system improvement, it is recommended that relevant management can strengthen the disclosure of stock investment information, achieve open, complete, transparent and effective transmission; further strengthen guidance and education investors to make rational investment, while learning from and learning from the development experience of mature capital markets, Strengthen the integration with international mature markets.

Bibliography

[1.]ANDRIKOPOULOS, P., G. N. GREGORIOU and V. KALLINTERAKIS. *Handbook of frontier markets*. Boston: Academic Press, 2016. ISBN 978-012-8037-768.

[2.]BARBERIS, Nicholas and Richard, THALER. Handbook of the Economics of Finance. Amsterdam: North Holland, 2003. ISBN 978-0444513632

[3.]BURTON, T. Edwin and Sunit, N. SHAH. *Behavioral finance: understanding the social, cognitive, and economic debates*. Hoboken: Wiley, 2013. ISBN 978-1-118-30019-0.

[4.]CHIANG, C. T.，J. LI. and T. LIN. Empirical Investigation of Herding Behavior in Chinese Stock Markets： Evidence from Quantile Regression Analysis. *Global Finance Journal*. 2010，NO.21, p. 111-24. Available on https://ideas.repec.org/a/eee/glofin/v21y2010i1p111-124.html

[5.]CHANG, C. E.，J. CHENG and A. KHORANA. An Examination of Herd Behavior in Equity Markets： An International Perspective. *Journal of Banking & Finance*，2000，NO.24, p.1651-79. Available on https://www.sciencedirect.com/science/article/pii/S0378426699000965

[6.]DEMIRER, Riza and Ali M. KUTAN. Does Herding Behavior Exist in Chinese Stock Markets? *International Financial Markets， Institutions and Money*. 2006，NO.16, p.123-42. Available on https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2148613

[7.]HWANG, Soosung and Mark，SALMON. A New Measure of Herding and Empirical Evidence. *Warwick Business School，Financial Econometrics Research Centre*. 2001，Coventry Working Paper No. 01. Available on http://wrap.warwick.ac.uk/1816/

[8.]LAKONISHOK, J., A. SHLEIFER, and R.VISHY. The Impact of Institutional Trading on Stock Prices. *Journal of Financial Economics*. 1992, NO.32, p. 23-43. Available on https://scholar.harvard.edu/files/shleifer/files/impact_instit_trading.pdf

[9.]LUX Tomas. Herd Behaviour, Bubbles and Crashes. *The Economic Journal*，1995，NO.105, p.881-96. Available on https://econpapers.repec.org/article/ecjeconjl/v_3a105_3ay_3a1995_3ai_3a431_3ap_3a881-96.htm

[10.] CIPRIANI, Carco and Antonio, GUARIANO. Estimating a Structural Model of Herd Behavior in Financial Markets. *American Economic Review.* 2014, NO.104, p. 224-251. Available on https://www.jstor.org/stable/42920693?seq=1#page_scan_tab_contents

[11.] ROLL, Richard. A Possible Explanation of the Small Firm Effect. *Journal of Finance*. 1921, NO. 36, p. 879-888. Available on https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1981.tb04890.x

[12.] RICHARDS Tim. *Investing psychology: the effects of behavioral finance on investment choice and bias.* Hoboken: Wiley, 2014. ISBN 978-1-118-72219-0.

[13.] KOLLER, T., R.DOBBS and B. HUYETT. *Value: The Four Cornerstones of Corporate Finance.* John Wiley & Sons, 2010. ISBN 9780470424605

[14.] BANZ, W. Rolf. The Relationship between Return and Market Value of Common Stocks. *Journal of Financial Economics.* 1981, NO. 9, p.3-18. Available on http://www.business.unr.edu/faculty/liuc/files/BADM742/Banz_sizeeffect_1980.pdf

[15.] WERMERS Russ. Mutual Fund Herding and the Impact on Stock Prices. *The Journal of Finance*. 1999, NO. 54, P.581- 622.
Available on https://onlinelibrary.wiley.com/doi/full/10.1111/0022-1082.00118

[16.] CHRISTIE G. William and Rodger, D. HUANG. Following the Pied Paper: do Individual Returns Herd around The Market? *Financial Analysts Journal.* 1995, NO. 51, p.31-37. Available on https://www.jstor.org/stable/4479855?seq=1#page_scan_tab_contents

List of Abbreviations

CSAD	Cross-sectional Absolute Dispersion of Returns
CSSD	Cross-sectional Standard Deviation of Returns
LSV	Lakonishok, Shleifer and Vishny
PCM	Portfolio-change Measurement
VIF	Variance Inflation Factor
OLS	Ordinary Least Squares

Declaration of Utilisation of Results from the Diploma Thesis

Herewith I declare that

- I am informed that Act No. 121/2000 Coll. – the Copyright Act, in particular, Section 35 – Utilisation of the Work as a Part of Civil and Religious Ceremonies, as a Part of School Performances and the Utilisation of a School Work – and Section 60 – School Work, fully applies to my diploma (bachelor) thesis;
- I take account of the VSB – Technical University of Ostrava (hereinafter as VSB-TUO) having the right to utilize the diploma (bachelor) thesis (under Section 35(3)) unprofitably and for own use ;
- I agree that the diploma (bachelor) thesis shall be archived in the electronic form in VSB-TUO's Central Library. I agree that the bibliographic information about the diploma (bachelor) thesis shall be published in VSB-TUO's information system;
- It was agreed that, in case of VSB-TUO's interest, I shall enter into a license agreement with VSB-TUO, granting the authorization to utilize the work in the scope of Section 12(4) of the Copyright Act;
- It was agreed that I may utilize my work, the diploma (bachelor) thesis or provide a license to utilize it only with the consent of VSB-TUO, which is entitled, in such a case, to claim an adequate contribution from me to cover the cost expended by VSB-TUO for producing the work (up to its real amount).

Ostrava dated 2019. 4. 20

Xinran Chen 陈欣冉

Student's name and surname

List of Annexes

Annex 1: Original Data Sample

Annex 2: Input Data Sample

Annex 1: Original Data Sample

Security code	Trading date	Open price	Close price	Daily return	Market code
		RMB	RMB		
600000	2015-12-31	18.52	18.27	-0.016155	1
600000	2015-12-30	18.71	18.57	-0.009072	1
600000	2015-12-29	18.63	18.74	0.002139	1
600000	2015-12-28	19.37	18.7	-0.032091	1
600000	2015-12-25	19.02	19.32	0.014706	1
600000	2015-12-24	18.85	19.04	0.008475	1
600000	2015-12-23	18.91	18.88	-0.001058	1
600000	2015-12-22	19.01	18.9	-0.012539	1
600000	2015-12-21	18.79	19.14	0.017544	1
600000	2015-12-18	18.49	18.81	0.017307	1
600000	2015-12-17	18.26	18.49	0.017052	1
600000	2015-12-16	18.36	18.18	-0.0071	1
600000	2015-12-15	18.44	18.31	-0.007588	1
600000	2015-12-14	18.31	18.45	-0.008065	1
600000	2015-12-11	19.01	18.6	-0.034769	1
600000	2015-12-10	19.34	19.27	0.00208	1
600000	2015-12-09	19.58	19.23	-0.043284	1
600000	2015-12-08	19.28	20.1	0.033419	1
600000	2015-12-07	18.71	19.45	0.039551	1
600000	2015-12-04	19.27	18.71	-0.028052	1
600000	2015-12-03	19.7	19.25	-0.030227	1
600000	2015-12-02	18.51	19.85	0.063773	1
600000	2015-12-01	18.35	18.66	0.000536	1
600000	2015-11-30	18.44	18.65	0.006476	1
600000	2015-11-27	19.16	18.53	-0.043366	1
600000	2015-11-26	19.5	19.37	-0.015252	1
600000	2015-11-25	18.9	19.67	0.032546	1
600000	2015-11-24	18.9	19.05	-0.000525	1
600000	2015-11-23	18.51	19.06	0.031944	1
600000	2015-11-20	17.81	18.47	0.031844	1
600000	2015-11-19	17.3	17.9	0.029327	1
600000	2015-11-18	17.47	17.39	-0.002867	1
600000	2015-11-17	17.46	17.44	0.001148	1
600000	2015-11-16	17.34	17.42	-0.009101	1
600000	2015-11-13	17.2	17.58	0.012673	1
600000	2015-11-12	17.8	17.36	-0.020869	1
600000	2015-11-11	18.41	17.73	-0.034839	1
600000	2015-11-10	18	18.37	0.012121	1
600000	2015-11-09	17.92	18.15	0.015101	1
600000	2015-11-06	17.52	17.88	0.009599	1
600000	2015-11-05	16.85	17.71	0.051663	1
600000	2015-11-30	18.44	18.65	0.006476	1
600000	2015-11-27	19.16	18.53	-0.043366	1

Annex 2: Input Data Sample

SH A	CSAD(cw)	ABS(Rm)	(Rm)^2	Vm
2014-01-02	0.01326991	0.003652	1.3337E-05	6.1838E+10
2014-01-03	0.01307186	0.012794	0.00016369	7.2227E+10
2014-01-06	0.02035788	0.017324	0.00030012	7.2707E+10
2014-01-07	0.01106342	0.000536	2.873E-07	5.4503E+10
2014-01-08	0.01347672	0.00189	3.5721E-06	6.2832E+10
2014-01-09	0.01476169	0.008318	6.9189E-05	6.7474E+10
2014-01-10	0.0182966	0.006796	4.6186E-05	6.0926E+10
2014-01-13	0.01489292	0.001248	1.5575E-06	5.5497E+10
2014-01-14	0.01378797	0.008328	6.9356E-05	5.6606E+10
2014-01-15	0.01251899	0.001814	3.2906E-06	5.7667E+10
2014-01-16	0.01144617	0.000039	1.521E-09	6.2743E+10
2014-01-17	0.01400842	0.008646	7.4753E-05	5.6969E+10
2014-01-20	0.01310797	0.006745	4.5495E-05	4.8212E+10
2014-01-21	0.01030902	0.008089	6.5432E-05	5.193E+10
2014-01-22	0.01166823	0.021348	0.00045574	8.3907E+10
2014-01-23	0.0112426	0.005338	2.8494E-05	7.5923E+10
2014-01-24	0.01250123	0.005813	3.3791E-05	8.2178E+10
2014-01-27	0.01487976	0.010532	0.00011092	8.1782E+10
2014-01-28	0.01389668	0.002507	6.285E-06	6.5615E+10
2014-01-29	0.01305774	0.005573	3.1058E-05	6.7569E+10
2014-01-30	0.01352915	0.008136	6.6194E-05	5.8158E+10
2014-02-07	0.01515776	0.004902	2.403E-05	6.7666E+10
2014-02-10	0.01447746	0.019667	0.00038679	1.125E+11
2014-02-11	0.01586463	0.009039	8.1704E-05	1.2502E+11
2014-02-12	0.01518673	0.002671	7.1342E-06	1.1057E+11
2014-02-13	0.01714441	0.004571	2.0894E-05	1.2416E+11
2014-02-14	0.01313899	0.007804	6.0902E-05	9.7201E+10
2014-02-17	0.01382789	0.008805	7.7528E-05	1.2601E+11
2014-02-18	0.01670185	0.00826	6.8228E-05	1.2953E+11
2014-02-19	0.01786869	0.012298	0.00015124	1.2766E+11
2014-02-20	0.02059815	0.001042	1.0858E-06	1.3074E+11
2014-02-21	0.0146948	0.011723	0.00013743	9.9378E+10
2014-02-24	0.02287635	0.017975	0.0003231	1.0695E+11
2014-02-25	0.02126542	0.019327	0.00037353	1.2751E+11
2014-02-26	0.01627002	0.0036	0.00001296	9.586E+10
2014-02-27	0.02223394	0.004245	1.802E-05	1.1265E+11
2014-02-28	0.01585871	0.003422	1.171E-05	9.5483E+10
2014-03-03	0.01767769	0.008436	7.1166E-05	1.0475E+11
2014-03-04	0.01515549	0.001663	2.7656E-06	1.0092E+11
2014-03-05	0.01416526	0.009324	8.6937E-05	9.1449E+10
2014-03-06	0.01490618	0.003356	1.1263E-05	9.238E+10
2014-03-07	0.01586759	0.000075	5.625E-09	8.9224E+10
2014-03-10	0.02052224	0.0282	0.00079524	9.4821E+10
2014-03-11	0.01492419	0.00084	7.056E-07	7.6748E+10
2014-03-12	0.01573402	0.001808	3.2689E-06	7.9177E+10