Title
Essays on Stock and Options Market

Author
He, Han

Publication Date
2021

Peer reviewed|Thesis/dissertation

Essays on Stock and Options Market

By

HAN HE

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Aaron Smith, Chair

Wing Thye Woo

Shu Shen

Oscar Jorda

Committee in charge

2021

Han He
June 2021
Economics

Essays on Stock and Options Market

Abstract

As the second-largest economy globally, China's economy has developed rapidly in recent years, and China's stock market has developed particularly fast as one of the most representative and important markets. Its total market capitalization had grown from 4.03 trillion US Dollars in 2010 to 8.52 trillion US Dollars in 2019. It is an important investment channel for individual investors on the supply side. It also provides financing opportunities for enterprises on the demand side. The fast growth has brought many systemic problems that need to be solved urgently, like imperfect regulation and informed trading. Considering these characteristics, I choose China's stock market as the theme for deep research and study.

The first chapter studies the informed trading problem in China's A-Share market. It investigates the abnormal drop of stock returns on announcement day for the listed firms on A board in China's stock market. It finds that the firms who reported a huge irrational goodwill impairment experienced a more significant decline before the announcement day, and their rebounds in abnormal return are, on average, more powerful after the announcement day. There is a significant negative jump in stock return on announcement day for these abnormal firms, while the other firms' jump is not significant. The difference in difference model over various cut-off days suggests that external reasons like informed trading are not the dominating factor determining the decline in abnormal returns of the abnormal firms before announcement day. Rather, the internal difference between firms plays a more important role in the stock disaster.

The second chapter focuses on the profit potential of China's stock indexes by applying neural network models. The chapter uses a neural network approach to predict China's stock indexes' returns from the Shanghai Stock Exchange(SSE) and the Shenzhen Stock Exchange(SZSE). It compares the ARIMA model, Multiple Layer Perception(MLP) model, and Recurrent Neural Network(RNN) model in terms of mean squared error and forecast accuracy to study the added value of improvement from model architecture. It turns out that MLP and RNN models failed to provide a significantly better prediction than ARIMA models if using historical information of stock prices alone. The paper also uses a standard quantitative trading strategy to backtest the value of predictions from these three models. The paper finds that the ARIMA model predicts the SSE Composite Index well, while the RNN model is the best in predicting Industrial Index and Composite Index. Most of the strategy's annualized returns surpassed the return of the benchmark stock index in the bear market from 2017 to 2020, which offers investors a better choice in their stock trading.

While my first two chapters focus on China's stock market, my third part of the dissertation studies the US's options market. China has not yet established a developed options trading system like the US, and the options market is an indispensable supplement of a complete stock market. Actually, China's options trading for stock indexes started in 2019. As the first stock index options product in China's domestic market, CSI 300 stock index options were listed and traded on the China Financial Futures Exchange on Dec 23rd, 2019. Considering the fact, Chapter 3, joint work with my colleague Yitian Xiao, study the US's options market. It is well known that the options market enables traders to hedge positions in asset markets, thereby reducing risk exposure. Traders can choose from many different strike prices – analogous to the coverage level in an insurance contract – when hedging. In Chapter 3, we argue that the hedging motive leads a typical hedger to take positions slightly out of the money, i.e., just below the asset's current value. We develop a simple theoretical model and validate its predictions using data on S&P 500 options. In particular, we show that trading volume follows the natural position of hedgers. Our results imply that hedging is fundamental to the value of options markets because the trading follows the hedgers. Moreover, we show empirically that gambling motivation could be a good supplement to explain our stylized facts.

To My Families

Contents

List of Figures

List of Tables

Chapter 1

Goodwill Impairment and Informed Trading

1.1 Introduction

In January 2019, big news broke out in China's stock market. Around 100 Chinese listed firms disclosed announcements about their performance in 2018, they all claimed a huge profit loss in their announcements, and the main reason these firms provided is goodwill impairment. Announcing profit loss is normal for managers to update shareholders' expectations and potential investors on a firm. Still, three features make this event become big news. First, the magnitude of the profit losses given by these firms is huge and irrational. Second, the main reason for the profit loss is all goodwill impairment. Last, these announcements all occurred together within one week. These characteristics make the event an interesting question worth studying.

Therefore, this paper aims to study the event and investigate the reason for its appearance. The crucial questions the paper focuses on are: are these firms who claimed crazy goodwill impairment different from the firms who did not claim such an announcement? Further, are these firms behave differently only around announcement day or behave differently all the time? These questions could potentially be answered from different perspectives. On the one hand, based on the observation of clustering negative announcements, it is natural for investors to suspect the existence of informed trading. Key information about the huge goodwill impairment may have already been leaked before announcement day so that its stock price behaves differently from other firms. On the other hand, it is also possible that these firms who made irrational goodwill impairment are naturally different from the other firms who did not. They have attributes and characteristics essentially different from other firms, so that they chose the timing of announcement and reason for profit loss differently. Actually, if these firms only behave differently on announcement day, this would be considered evidence for informed trading because it is abnormal for firms to change their behavior only around the announcement date. Otherwise, we could conclude that firms are heterogeneous so that they behave differently due to their own attributes rather than informed trading.

Since informed trading is an indispensable topic in the paper, defining the concept before the further discussion is beneficial. The informed trading here refers to trading based on the information not reflected in stock prices yet. The information includes more accurate signals about the stock market in the future. It could be information leak from the manager of a firm or information about other firms' stock prices, but it does not include the tacit understanding across different firms. For example, if one firm announced goodwill impairment, other firms may follow the forerunner and use the same excuse for free automatically. Such tacit understanding may not need any concrete collusion behaviors. Actually, they could follow the forerunner because they are naturally very similar firms, and they face a similar situation like profit loss. This could not be considered informed trading formally, but it is harmful to China's stock market's sustainable development.

To precisely differentiate the firms who made a crazy announcement from other firms who did not, I define "target firm" and "non-target firm" to name them in the paper. Specifically, a target firm is a firm whose announced profit loss is greater than 1 billion Chinese Yuan(RMB), while a non-target firm is a firm whose announced profit loss is less than the bar. The "target" and "non-target" do not mean acquisition in accounting. Rather, they are used to differentiate two groups of firms, one abnormal group whose firms all reported announced profit loss is greater than 1 billion RMB and one normal group whose firms did not claim irrational announcement. The reason for the profit loss given by the target firms is goodwill impairment. Goodwill impairment is an accounting terminology referring to the goodwill carrying value of an asset exceeding its fair value. It is created to inform investors that one firm is not worth as much as they thought. Usually, there is a drop in stock price after one firm reported a goodwill impairment in its announcement. It is normal and legal. However, it is very abnormal to observe many listed firms claimed large amounts of goodwill impairment together within one week of Jan 2019. The clustering goodwill impairments caused heavy losses for investors in China's stock market. For example, Tianshen Entertainment claimed that its profit loss ranges from 7.3 to 7.8 billion RMB on Jan 30. As a result, its close price plummeted from 4.73 to 4.26 within one day. Still, some firms did not experience a sharp drop in stock price. Ningbo Donly claimed a 1.7 billion goodwill impairment with 700 million extra impairment from one of its subsidiary companies, which leads to an overall 2.5 billion loss, but its stock price surprisingly rose after the announcement came out. This is because its stock price has already reached the bottom of the historical price in the second half-year of 2018, and some investors think it is good timing to conduct a bottom-fishing strategy. Since different firms have different performances on stock prices after an announcement comes out, the paper's central question is to test the difference between the target firms and the non-target firms. The empirical analysis and hypothesis test are discussed in later sections.

The question is attractive from many perspectives. Firstly, China's stock market is an important investment channel for Chinese investors due to China's capital constraint. Huge irrational goodwill impairments hurt investors not only through declining stock prices directly but also through negative impact on future expectations indirectly. The wealth of investors shrank dramatically over the announcement period. So the question proposed in the paper is worth investigating from an investor's point of view. Secondly, this is not the first time such a phenomenon happens in China's stock market. Listed Firms on A board cyclically reported a huge loss in intangible assets at the end of each accounting year. They did not report negative announcements together historically. The magnitude of loss was also not as large as this time in the year 2019. So the question itself is representative and worth studying.

Finally, China's stock market is a large growing and developing market, and it does not have complete and detailed regulations about the provision of the intangible asset. Studying this question is also important to improve the market's legal system and provide a fair environment for all investors. Therefore I studied the question in the paper. I used panel regression and a difference in difference model to empirically test the difference between target and non-target firms in pre-announcement and post-announcement periods. I found the firms who reported a huge irrational goodwill impairment experienced a more significant decline before announcement day. Their rebound in abnormal return is, on average, more powerful after announcement day. There is a significant negative jump on announcement day for these target firms, while other firms' jump is not significant. All the findings are presented and discussed in empirical sections. The rest of the paper is organized as follows. I list the existing relevant literature in section 1.2 and introduce the dataset in section 3.3. Section 1.4 shows the empirical finding. I discuss and conduct the hypothesis test in Section 1.5 and 1.6. Corresponding robustness check is shown in section 2.6. Finally, I draw conclusions in section 3.5 and list all the references in section 1.9.

1.2 Literature Review

The paper's central question is to study the effect of goodwill write-off, and there is no existing literature studied the event of clustering goodwill impairment in China yet. Therefore I review all the relevant literature about goodwill write-offs. I find that there is a lot of literature that studied the effect of goodwill impairment on the whole financial market and a firm's own performance. Most of them conclude that goodwill write-offs have a negative influence on the stock price. I list the most relevant literature as follows.

First, many research studied goodwill account from an agency theory perspective. They investigated the implementation of goodwill write-offs under certain political regulations. Ramanna & Watts (2012) studied the usage of unverifiable estimates in goodwill impairment and found that empirically goodwill impairment could not be explained or predicted by agency theory. They used a sample of collected firms to test the methodology and the motivation of using Statement of Financial Accounting Standards No.142(SFAS 142), which requires managers to use the current fair value of goodwill to determine an appropriate level and frequency of goodwill write-offs. Ideally, the manager would apply goodwill write-offs to convey their private information and expectation of a firm's performance in the future. The flexibility of managers would end up with manipulation of reports according to agency theory. However, the hypothesis is not verified in their paper, so they concluded that managers prefer to avoid timely goodwill write-offs due to their own benefit or the SFAS 142. However, this paper failed to explain the implementation of the SFAS 142 and did not take the 2008-2009 financial crisis into consideration. The paper's finding is in line with Henning & Stock(2004), which found that US managers prefer to delay their goodwill write-offs while UK managers prefer to conduct timely goodwill write-offs.

Second, some other literature focused on the motivation and functions of goodwill write-offs. Elliott & Shaw (1988) emphasized that goodwill write-off functions more like a tool for the manipulation of balance account. The authors used 240 firms with huge write-offs. They studied their balance account to test the relation between write-offs and accounting indicators like earning-to-assets, share returns, change in analyst's forecast, etc. It turns out that all these indicators are affected by write-offs in the sample period. It is especially true when the economy experienced a bear market. However, the paper failed to show the plausibility of the timing and size of write-offs selected in their sample. Francis, Hanna & Vincent(1996) studied the decisive factors of write-offs and the corresponding wealth effect. The level of impaired assets and incentives for accounting management are the two most important factors determining the level and timing of write-offs. Empirical tests showed that both factors exist in samples, but their importance may vary across different accounts. Incentives factor affect the goodwill write-off significantly while it almost does not influence inventory write-offs. The paper also tested the market response to different types of write-offs. On average, write-offs lead to a negative market response in terms of return, and the direction of reactions is different across various types of write-offs.

Last, a branch of literature studied the relationship between goodwill write-offs and their effect on the market's performance. Li, Shroff, Venkataraman & Zhang(2011)studied the market response to goodwill impairment write-offs and examined the losses' key information. It turned out that goodwill impairment is carrying new information of a firm, and goodwill impairment announcement does affect investor's expectation. Empirically, investors would adjust their expectations downward after impairment loss. Also, the negative impact is robust, and goodwill impairment affects the firm's future performance. The effects are different across various policy regimes: pre-SFAS-142 period, transition period, and post-SFAS-142 period. The paper also showed that the historical overpayment factor could predict goodwill impairment. Actually, Bens, Heltzer & Segal(2007) and Liberatore & Mazzi(2010) found similar results under this topic. Besides, Hirschey & Richardson(2003) quantified the loss in stock prices. They focused on market response to goodwill write-offs announcements. They found that the effect of write-offs in goodwill account on stock price ranges from -2.94% to -3.52%, but the effect could be around -11.02% one year after write-off announcements, which implies that investors need a long time to respond to write-off signals fully. They under-react in the short post-announcement period.

All the literature provided helpful insight about goodwill impairment, but they only focused on developed financial markets. Studying the goodwill impairment in the s-tock market in a developing country like China would be a good extension for the research. Besides, there is no literature particularly studied the event of clustering goodwill impairment in China yet. Therefore, the paper provides a different perspective on China's stock market development, which offers a deeper interpretation of the market. I introduce the data sources in section 3.3 and all the empirical findings are discussed in later sections.

1.3 Data

1.3.1 Data Source

In China's stock market, stocks of listed firms on A board are traded in two stock exchanges: Shanghai Stock Exchanges(SSE) and Shenzhen Stock Exchange(SZSE). There are overall around 3000 stocks listed on A board. To study the difference among all the firms, a desirable dataset should include all the daily stock transactions' historical data. Therefore I obtain the historical transaction data from the website of Wang Yi Finance. The website is available at "https://money.163.com/". The biggest advantage of this database is that it includes historical transaction data of stock trading for all the existing stocks and includes the content of historical announcements from these firms. The key variables are "stock

ID", "date", "last price", "highest price", "lowest price", "change in price", "open price", "trading volume", "market value" etc. Moreover, the firms' historical announcements' information is obtained and tested; the variable "date of announcement" is summarized accordingly.

Actually, the dataset helps analyze the motion of stock prices in the pre-announcement period and post-announcement period. As introduced in the section 2.1, I define "target firm" as a firm whose announced profit loss is greater than 1 billion RMB, and "non-target firm" as a firm whose announced profit loss is less than the bar. A total of 95 firms meet the requirement, so the number of target firms is supposed to be 95, but not all of them made announcements in Jan 2019. Considering the paper's goal is to investigate the event of clustering announcements in the month, I focus on the firms who made announcements about goodwill impairment and loss in Jan 2019. After filtering the dataset, there are 76 target firms kept at last. Besides, since the news of clustering goodwill impairment was in January 2019, I choose the sample period from October 8th, 2018 to April 4th, 2019, to have enough observations to investigate the motion of stock prices before and after announcement days. All the empirical analysis and discussion are based on the stock prices and announcement dates from the database.

1.3.2 Stock Returns

The goal of the paper is to study the different performances of the target and the non-target firms. Daily stock return is a direct and normative metric of it. The stock return is defined in equation 2.4, where P_t is the stock price on day t and P_{t-1} is the stock price on day $t-1$. These metrics play a role of a signal telling us the market value and the firm's potential profitability, and it is crucial to test whether stock returns of target firms really changed after the announcement day of goodwill impairment. Specifically, to compare the stock returns before and after announcement day T, a good proxy variable is needed to show the average daily stock return level before announcement day. Actually, many variables are good candidates for it. The most straight forward one is the stock return right before announcement day. Namely, the stock return on $T-1$. This is a direct measurement of the strength of the announcement effect on stock return. I show the densities of return on T

and $T-1$ together in the figure 1.1.

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{1.1}$$

The figure shows that the range of stock returns is approximately from -0.15 to 0.05. The red line represents the density of stock returns of the target firms on announcement day T, and the blue line represents the density of the returns one day ago. Actually, more statistics of them could provide a better insight of the densities, the mean of the returns on the day T is -0.067 and the corresponding standard deviation is 0.039, while the mean of the returns on the day $T-1$ is -0.026 and the corresponding standard deviation is 0.034. A clear leftward shift is observed in the figure, which indicates that the target firms experienced a negative shock in terms of a stock return due to the announcement of their goodwill impairment. The peak of the density curve shifts leftward from around -0.02 to -0.1. The shift distance is even greater than the standard deviations of the returns. To test the statistical significance of the difference between these two densities, I conduct Kolmogorov-Smirnov Test and Stochastic Dominance Test to compare them. The results are shown in the table 1.1, all the P-values are recorded in parenthesis. The D-statistic in the K-S test is 0.547, and the corresponding P-value is 0.000, which supports the rejection of the null hypothesis. Namely, the distributions of the stock returns on T and $T-1$ are different statistically. However, the Somers' D statistic in the Stochastic Dominance Test is 1.12 with a P-value of 0.263. It is not statistically significant at a 5% significance level. This finding suggests that the stock returns on the day $T-1$ do not have first-order stochastic dominance over the stock returns on the day T.

According to the finding in the table 1.1, the distributions in the figure 1.1 are different statistically, but they do not have a relationship of first-order stochastic dominance. It means that the difference between the two distributions is not as large as first-order stochastic dominance. Actually, the result is in line with prior expectations because most of the target firms experienced declining stock prices. Therefore, it is not surprising to see a big shift in stock returns on announcement day.

Nevertheless, the conjecture above has not excluded the effect of the benchmark

Table 1.1: Stochastic Dominance Test and Kolmogorov-Smirnov Test

Variable	Kolmogorov-Smirnov Test D-stat	Stochastic Dominance Test Somers' D stat
Stock Return	0.547	1.12
	(0.000)	(0.263)
Abnormal Return	0.579	2.07
	(0.000)	(0.038)

The table shows the Stochastic Dominance Test and Kolmogorov-Smirnov Test on the stock returns and abnormal returns. K-S statistics are all significant at a 5% significance level. This suggests that the distributions of both returns on the day T and $T-1$ are different. Also, the Stochastic Dominance Test shows that stock returns on the day $T-1$ do not have the first-order stochastic dominance over the stock returns on the day T, but the abnormal returns on the day T have the first-order stochastic dominance over the abnormal returns on the day $T-1$.

market yet. In fact, the huge shift in the target firms' return may be due to the bad market performance. That is, the leftward shift of daily return on announcement day T might come from the effect of the bear market, and the benchmark drives all the individual stock returns. Therefore, it is important to rule out the market effect and obtain each firm's pure stock return. Specifically, I compute residual daily stock return by using equation 1.2 and equation 1.3 from Danielle & Ryan(2013). I estimate the residual daily stock return y_it by running a standard CAPM regression in the equation 1.2. Here R_{it} is the stock return of firm i on day t, M_t is the market benchmark return on day t, α_i and β_i are the corresponding α and β factor of firm i. Based on the CAPM model in the equation, empirical estimate for the residual daily stock return is shown in the equation 1.3 where y_{it} is the residual stock return of firm i on day t, $\hat{\alpha_i}$ and $\hat{\beta_i}$ are the estimates for the α and β factor of firm i. Therefore for each firm i on the day t, we could get a residual return, and I name it as "abnormal stock return."

After replacing the raw stock returns from equation 2.4 by the abnormal stock returns from equation 1.3, I obtain a new density plot in figure 1.2. The figure indicates that the pattern we found in figure 1.1 still holds in abnormal return. I highlight the density of abnormal stock return on announcement day T in red. The distribution of abnormal return shifts leftward significantly and has a "wider" spread from $T-1$ to T. It ranges from around -0.15 to 0.1, and most of the target firms have negative abnormal returns on announcement day T. The mean of the abnormal returns on the day T is -0.064, and the corresponding standard deviation is 0.044, the mean of the abnormal returns on the day $T-1$ is -0.021, and the standard deviation is 0.034. All these statistics are very close to the ones in figure 1.1. The results of the K-S test and stochastic dominance test for abnormal returns are also shown in the table 1.1. The K-S test supports the conclusion that the distributions of abnormal returns on the day T and $T-1$ are statistically different. Moreover, the Somers' D statistic is also significant at a 5% significance level, which means that the abnormal returns on the day T has first-order stochastic dominance over the abnormal returns on the day $T-1$. Therefore it tells us that the finding in Figure 1.1 is robust after ruling out the market effect on stock returns. It also indirectly suggests that the market environment does not drive the abnormal returns of these target firms, rather, their bad performance on daily stock returns are mainly driven by the announcements they claimed.

$$R_{it} = \alpha_i + \beta_i M_t + \epsilon_{it} \tag{1.2}$$

$$y_{it} = R_{it} - E(R_{it}|M_t) = R_{it} - \hat{\alpha}_i - \hat{\beta}_i M_t \tag{1.3}$$

1.3.3 Summary of Returns

The last section's finding clearly shows that the abnormal return of the target firms declined dramatically on announcement day. However, static comparison in abnormal stock returns is not sufficient. It is crucial to test the pattern dynamically. Actually, exploring the time-series features of these target firms' abnormal returns would provide more insights into the decline in abnormal returns. Therefore I plot the time series graph for all the target

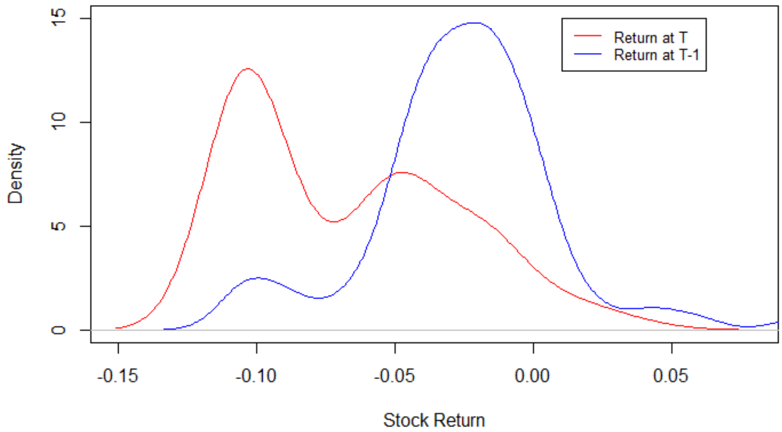


Figure 1.1: Stock Returns of Target Firms on the Day T and $T-1$

This figure shows the difference in the distributions of daily stock returns on announcement day T and on the last day before announcement day $T-1$. It corresponds to the realized daily stock return computed from the target firms. The mean of stock returns on T is -0.067, and the corresponding standard deviation is 0.039. The mean of stock returns on $T-1$ is -0.026, and the corresponding standard deviation is 0.034. The significant leftward shift of the density from $T-1$ to T indicates that the target firms experienced sharp drops in terms of stock return after their announcements were disclosed.

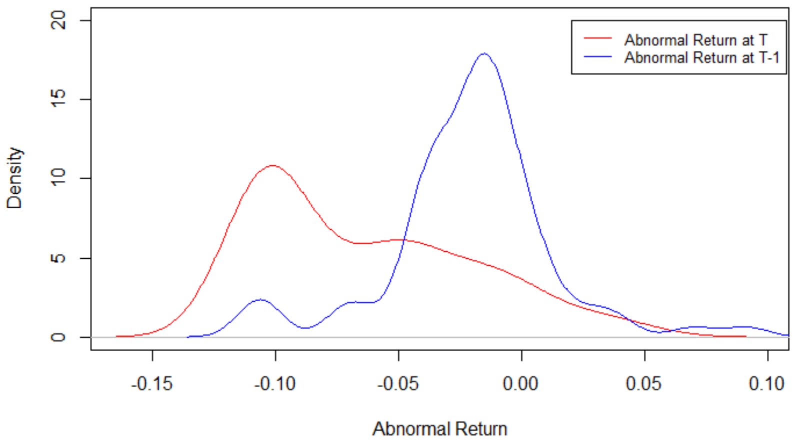


Figure 1.2: Abnormal Returns of Target Firms on the Day T and $T-1$

This figure shows the difference in the distributions of abnormal stock return on announcement day T and on the last day before announcement day $T-1$. It corresponds to the realized abnormal stock returns computed from equation 1.3. The mean of the abnormal returns on the day T is -0.064, and the standard deviation is 0.044, the mean of the abnormal returns on the day $T-1$ is -0.021, and the standard deviation is 0.034. All these statistics are very close to the ones in figure 1.1, and the pattern is robust after ruling out the market effect on stock returns.

firms, and the plot of the first 3 representative firms are shown as an example in figure 1.3. It turns out that all these target firms experienced a similar pattern in the last week before their own announcement days. For example, The abnormal return of firm 2 in the figure 1.3 was around positive 0.04 one week before its announcement day, the day is labeled as -7 in the figure, but it began to decline to a negative return and ended up with around -0.08 on announcement day T, the day is labeled as 0 in the figure. In fact, all the target firms experienced a similar pattern even though they have various announcement dates. This finding is shown in figure 1.4 where the control group and treatment group's abnormal returns are plotted separately. It clearly shows the difference in abnormal returns between the two groups. The target firm group experienced a deeper decline than the non-target firm group before announcement day. The abnormal return of the non-target firm group ranges from -0.01 to 0, while the abnormal return of the target firm group ranges from -0.06 to 0.01. This finding suggests that the target firms are different from non-target firms in abnormal return before announcement day, but they share some similarities within the target firm group.

The similarity in figure 1.4 could be explained in different ways. On the one hand, these target firms might be very similar essentially and naturally, so that they automatically behave similarly before announcement day. On the other hand, these target firms are different naturally, but some exogenous reasons like informed trading make them behave similarly before announcement day. No matter which hypothesis is true, it is necessary to test the heterogeneity of the target firms. Therefore I plot the distribution of abnormal returns of the target firms in figure 1.5. This figure is designed to show the heterogeneity of the target firms from a cross-sectional perspective. The abnormal returns of each firm are shown with its 95% confidence interval. The confidence interval is computed using the abnormal stock returns of a firm in the whole sample period, namely, from October 8th, 2018 to April 4th, 2019. The figure shows that most of the firms have negative abnormal returns, and there is a certain level of difference in the lengths of these confidence intervals. Specifically, some firms like No.15, No.53, and No.62 distribute widely in abnormal returns distribution while other firms do not. Even though they have a difference in distribution, these target firms share two common features: First, most of them have negative abnormal

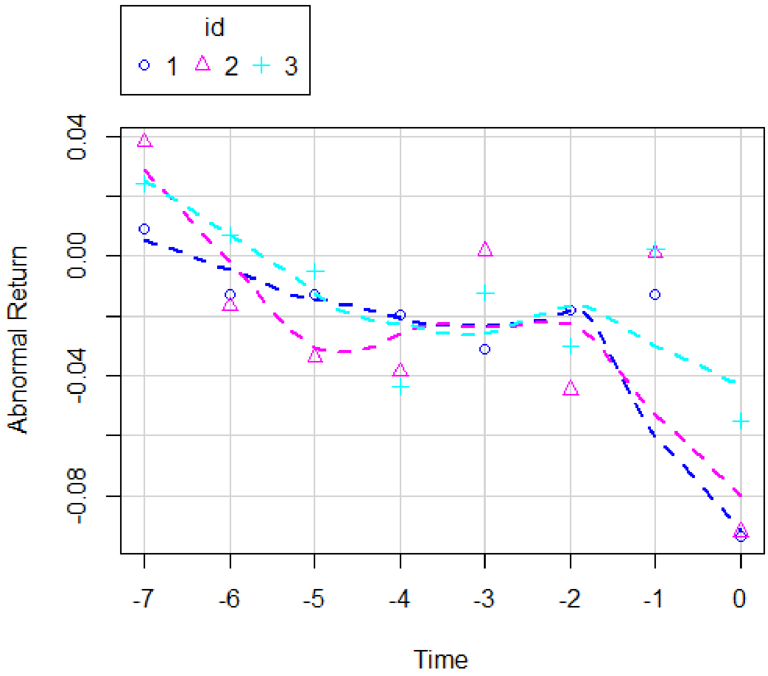


Figure 1.3: Abnormal Returns of the Firm No.1, No.2 and No.3

The figure shows the dynamic motion of abnormal return for the first 3 target firms in their last week before announcement day. The dashed lines are the fitted smoother line from local polynomial regression fitting. The corresponding mean and variance of the lines are estimated nonparametrically. The figure shows that these 3 firms' abnormal returns declined from positive values to around -0.05 on announcement day. This pattern is robust over different target firms.

returns. Second, most of the target firms have their abnormal returns lie in the interval (-0.05,0.02). According to the finding so far, there is not enough evidence to draw any conclusion on the firms' heterogeneity from a cross-sectional point of view. Thus further empirical regression and hypothesis tests are discussed in the following sections.

1.4 Empirical Finding

According to the last section's finding, a target firm experienced a gradual decline in abnormal stock return one week before announcement day and an extremely large drop on announcement day. This pattern holds for almost all the target firms. To further study the dynamic procedure of abnormal stock return for the target firms and quantify each trading day's time effect before announcement day, the most direct method is to run

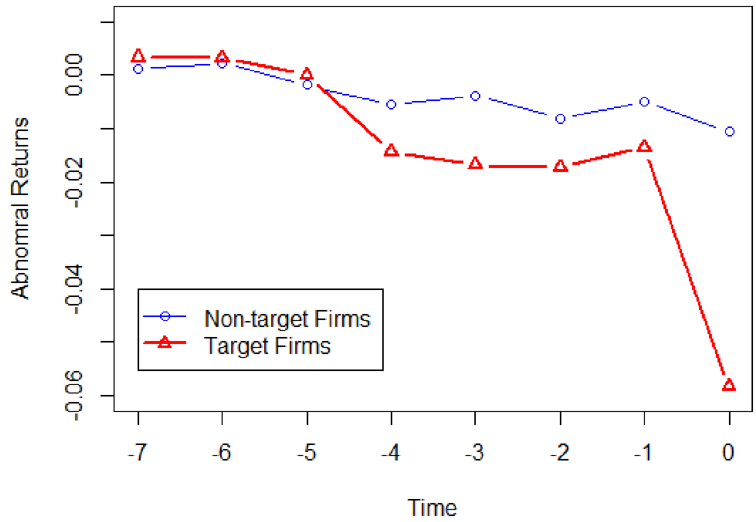


Figure 1.4: Heterogeneity Across Time for the Target and the Non-target Firms

The figure shows the dynamic motion of abnormal returns of all the target firms in the last week before their announcement days. The pattern found in the figure is the same as that in figure 1.3. The figure depicts the general pattern of abnormal returns of the target firms: Started with positive or zero abnormal return from one week before announcement day, and then gradually declined to around -0.06 on announcement day. Further, the non-target firm group's abnormal return ranges from -0.01 to 0, while the target firm group's abnormal return ranges from -0.06 to 0.01.

a panel regression. Considering the question of interest of the paper is to investigate the target firms who claimed irrational announcement in Jan 2019 rather than a potential group of firms from the population, a fixed-effect model is more suitable for the topic, and the specification of the fixed effect model is shown in equation 1.4.

There are overall 7 periods in the panel dataset. I define a list of time dummy variables. They are D_{-1}, D_{-2},..., D_{-7}. These variables are defined to identify the time effect of each day before announcement day. Specifically, D_{-1} is defined as $D_{-1} = 1$ if $t = T - 1$ otherwise $D_{-1} = 0$. The rest 6 variables D_{-2},..., D_{-7} are defined similarly. Moreover, firm's individual fixed effect is captured by γ_i and dependent variable y_{it} is the abnormal return the firm i on day t. The coefficients of the time dummies jointly show

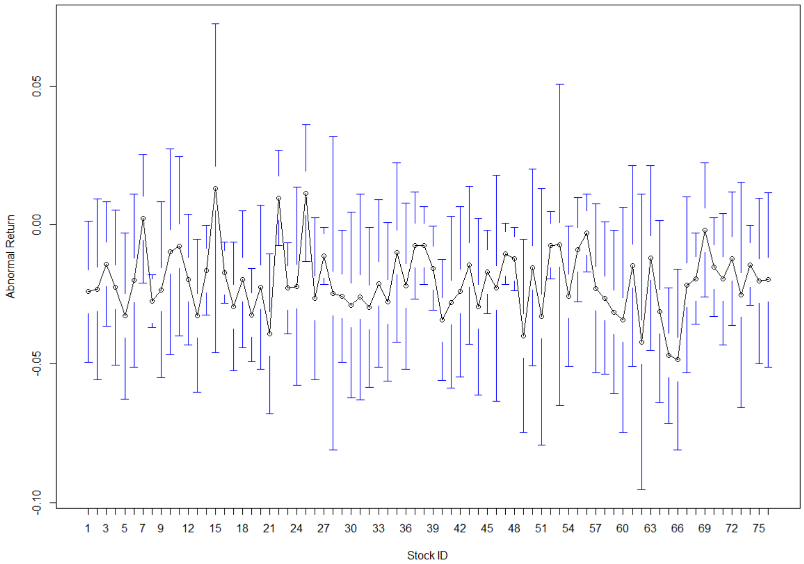


Figure 1.5: Heterogeneity Across the Target Firms

The figure shows the heterogeneity in abnormal returns across the target firms, 95% confidence intervals are plotted for each firm. Most of the firms have negative abnormal returns. Some firms, like No.15, No. 53 and No. 62, have a wider distribution of abnormal return. The majority of the target firms have their abnormal returns lie in the interval (-0.05,0.02).

the dynamic time effect on the target firms' abnormal return. In this model, I choose the abnormal return on announcement day T as the base. Therefore a positive β_j of variable D_{-j} suggests that the abnormal return on the day $T-j$ is greater than the abnormal return on announcement day T. Similarly, a negative β_j suggests the abnormal return on the day T is greater than that on the day $T-j$. In addition, the relative levels of these βs could depict the motion of abnormal return of a target firm before its announcement day. Actually, the fixed-effect model's advantage is its potential to describe a nonlinear trend's features in a certain period. Quantified estimates from the model would provide better insights into the trend in abnormal returns.

The regression result is shown in the table 1.2 below. The estimates for different firms are recorded in two different columns. The P-value of each estimate in the t-test is listed in the corresponding parenthesis. The table shows that the β coefficients range from 0.04 to 0.06, and all the dummy variables are significant at a 5% significance level. Since the abnormal return on announcement day, T is the base. These positive βs tell us that all the abnormal returns before announcement day are greater than that on announcement day.

For example, The estimate of slope coefficient β_1 is 0.0482 for target firms, which implies that the daily abnormal return on the day $T - 1$ is 4.82% greater than the daily abnormal return on announcement day T. Another interesting finding is that estimates of β_7 and β_6 are greater than other estimates of βs. This reflects that the earlier the date is, the higher the abnormal return it is. This downward sloping curve of abnormal return supports the pattern summarized in the last section. The first column in the table 1.2 verifies that the target firms' abnormal returns have already been declining gradually one week before announcement day.

$$y_{it} = \beta_0 + \sum_{j=1}^{7} \beta_j D_{-j} + \gamma_i + \epsilon_{it} \tag{1.4}$$

As a comparison, the empirical regression result of the same fixed effect model by using all the 712 firms is shown in the second column of the table 1.2. Here the 712 firms include both target and non-target firms who made announcements in Jan 2019. Actually, a very similar pattern of coefficient estimates is observed there. All the estimates are positive, the estimates of β_7 and the β_6 are greater than other estimates, and the estimates are all significant at a 5% significance level.

Nevertheless, there is one big difference between the results in these two columns of the table, namely, the coefficients in the second column are much smaller than that in the first column. This difference means that most of the firms which announced in Jan 2019 experienced a very slight or even no drop in abnormal return in the last week before announcement day, and the trend of the abnormal returns of all the firms is not as clear as that of target firms. Actually, there is no downward sloping trend for all the firms according to the second column of the table 1.2. Specifically, the estimates of β_7, β_6 and β_1 are all greater than that of β_2 and β_4. This finding exactly reflects that the motion of the abnormal return of the target firms is different from that of non-target firms. However, running the fixed effect model alone for the pooled dataset is not sufficient to draw a final conclusion. Separate empirical regressions and hypothesis tests on both the target and the non-target firms are discussed in section 1.5 and 1.6.

Table 1.2: Fixed Effect Model

Variable	Target Firms	All Firms
	b/p	b/p
D_{-1}	0.0482	0.0090
	(0.0000)	(0.0000)
D_{-2}	0.0407	0.0050
	(0.0000)	(0.0000)
D_{-3}	0.0422	0.0099
	(0.0000)	(0.0000)
D_{-4}	0.0403	0.0073
	(0.0000)	(0.0000)
D_{-5}	0.0563	0.0122
	(0.0000)	(0.0000)
D_{-6}	0.0600	0.0164
	(0.0000)	(0.0000)
D_{-7}	0.0605	0.0147
	(0.0000)	(0.0000)

The table shows the regression result of the fixed-effect model in equation 1.4 over different samples. The positive β estimates verify that the abnormal returns before announcement day are larger than that on announcement day. The target firms experienced a declining abnormal return in the last week before announcement day, but the pattern is not observed in the sample's regression with all 712 firms.

1.5 Discontinuity Tests

According to the finding in section 1.4, the target firms experienced a larger decline in abnormal return than the decline estimated using the whole sample, but pooling the target firms with non-target firms together is not sufficient. It is necessary to test the abnormal returns of the target and non-target firms separately. In fact, the magnitude and direction of the discontinuity on announcement day are good metrics to compare the difference between these two groups of firms. Following this idea, three empirical tests on the discontinuity of the abnormal returns around announcement day are conducted in this section. From this section, I define the target firms as the treatment group and define the non-target firms as the control group. Therefore there are 76 firms in the treatment group and 636 firms in the control group. In addition, the benchmark regressions introduced in the section use the sample from $T-30$ to $T+10$, where T represents the announcement day. Specifically, I use each firm's abnormal returns from 30 days before its announcement day to 10 days after its announcement day. This is because the dataset only includes trading days in China's stock market, so 10 days after announcement day actually means two trading weeks, and 30 days before announcement day guarantees that I have the historical abnormal returns of each firm for the most recent one month before its announcement day. More details about the results of choosing different sample sizes are discussed and presented in the robustness section 1.7.

The empirical tests are organized as follows. I first test the abnormal return trend in the control group before announcement day and then test whether there is any significant jump in abnormal return. After that, the same tests are conducted on the treatment group in section 1.5.2. Finally, the Difference in Difference (DID) model is applied to the two groups of firms to compare the overall difference between the treatment and control groups. The corresponding results are discussed in section 1.5.3.

1.5.1 Discontinuity in the Control Group

I first investigate the trend of abnormal returns in the control group before announcement day and test the discontinuity on that day in the group. The discontinuity tells us the true effect of the announcement of goodwill impairment on abnormal returns. To study the discontinuity empirically, I use the fixed-effect model in equation 1.5. Variable t is a time index variable: $t = 0$ represents the announcement day, and t is negative in the pre-announcement period while it is positive in the post-announcement period. For example, $t = -3$ represent the day three days before announcement day, $t = 5$ represent the day five days after announcement day. The time index variable centralizes the abnormal returns of different firms along the time axis so that they could be compared with each other even though they have different announcement dates. Moreover, variable D is the dummy variable for announcement day: $D = 1$ if $t \geq 0$ and $D = 0$ if $t < 0$. Therefore β_1 measures the difference in abnormal return before and after the announcement day, β_3 represents the adjustment in the time trend after the announcement day, β_2 shows the direction of the time trend in abnormal return before the announcement day. The fixed effect of firm i is captured by δ_i. The model's empirical finding is summarized in the first column of the table 1.3.

The estimate of β_1 is -0.0013. It is not significant at the 5% significance level. This means that there is no significant jump on announcement day in the control group, which is consistent with the expectation that the discontinuity in the abnormal return of the non-target firms is not large compared with that of target firms. Moreover, the estimate of the coefficient of t is -0.00036; it is significantly negative. So there is a slightly downward sloping time trend in the pre-announcement period after controlling each firm's individual fix effect.

Also, the time trend adjustment β_3 is estimated to be 0.00093, which is greater than -0.00036. So the time trend is positive after announcement day, which implies that the time trend in abnormal return recovered slightly from negative to positive after the shock on announcement day. The table's finding supports the conjecture that the non-target firms did not experience a huge decline when their announcement was disclosed, but they still have a V-shaped time trend around announcement day. However, the magnitude of the jump on announcement day is tiny, and it could be considered as a continuous flat line approximately. More discussions about this are shown in the following sections.

$$y_{it} = \beta_0 + \beta_1 D + \beta_2 t + \beta_3 Dt + \delta_i + \epsilon_{it} \tag{1.5}$$

1.5.2 Discontinuity in the Treatment Group

Similarly, I applied the model in equation 1.5 to the treatment group. The results are shown in the second column of the table 1.3. It indicates that the estimate for the coefficient of the dummy variable D is -0.02483, and it is significant at a 5% significance level. I still observe a significant negative time trend -0.00109 before announcement day and a positive time trend 0.01028 - 0.00109 = 0.00919 after announcement day. The magnitude of the estimates is greater than that in the first column of the table, which proves that the target firms experienced a more significant decline in abnormal return before announcement day, but their abnormal returns rebounded more powerful than that of the non-target firms after the announcement day.

This powerful rebound could be explained and interpreted in different ways. On the one hand, it could be interpreted as a recovery of a correction in the stock market, a correction in the market is damaging in the short run, but it corrects the overvalued stock prices, so it is a good adjustment in the long run. Therefore some investors believe that the quick and powerful rebound in abnormal return of the target firms shows the efficiency and the potentials of these target firms. On the other hand, the quick recovery of the treatment group results from the bottom fishing strategy of investors. Because the sharp decline in the abnormal return is very unusual, investors believe that it is good timing to buy these stocks at an extremely low price, and their prices would rebound in the future. As a result, many investors join the game and push the stock prices up soon. In fact, the abnormal return could rebound because investors are buying the target firms' stocks so that the prices of these firms go up after their announcement was disclosed. Even though the interpretations on the powerful recovery could be very different, I personally prefer the second conjecture. Namely, some investors used the bottom fishing strategy in the market so that the stock prices of the target firms recovered faster than other firms. These target firms experienced a much larger decline in abnormal return, so it is more profitable for the investors to buy the stocks of these target firms if they want to use the bottom fishing strategy. In addition, I have mentioned in the introduction section of the paper that this is not the first time many listed firms on A board in China's stock market claimed a huge profit loss in their announcements. Chinese investors witnessed similar situations already in past years, so they are confident that the stock prices would rise soon in the future according to their historical experience, making the investors more aggressive in the stock trading after announcement day. This conjecture is a possible candidate for the reason for the powerful rebound of the target firms. This could be a good extension for further research, but it is not in the scope of this paper.

In addition, the coefficient estimate -0.02483 of D in the treatment group is much greater than its counterpart -0.0013 in the control group. A clearer description of the trends and discontinuities in these two groups is shown in figure 1.6 below. The red line represents the time trend of abnormal returns in the treatment group, while the black line represents the time trend in the control group. Most interestingly, the downward jump of the abnormal return in the control group is not statistically significant, while the jump in the treatment group is. This finding is strong evidence supporting the difference in the motions of abnormal return between the control and treatment groups.

1.5.3 Difference in Difference Test

Following the test methodology in the past two subsections, I further investigate the discontinuity of abnormal return around announcement day using the Difference in Difference(DID) Model. The model enables me to introduce both group dummy and time dummy variables and analyze all the firms within a single framework. Actually, running a DID model makes it easier to compare the difference between these two groups directly. The specification of the model is shown in equation 1.6. In addition to the time index variable t and announcement day dummy variable D introduced in the section 1.5.1, a group dummy variable S is generated in the model. $S = 1$ if a firm is in the treatment group and $S = 0$ if a firm is in the control group.

$$y_{it} = \beta_0 + \beta_1 D + \beta_2 S + \beta_3 t + \beta_4 DS + \beta_5 Dt + \beta_6 St + \beta_7 DSt + \alpha_i + \epsilon_{it} \tag{1.6}$$

The empirical result of the DID model is shown in the third column of the table 1.3. It turns out that the estimate of β_4 is -0.02353, which proves that the abnormal return in the treatment group after announcement day is lower than that in the control group before announcement day. This finding is consistent with our hypothesis that the target firms have smaller daily abnormal returns after disclosing negative announcements. In addition, the estimate of β_7 is 0.00935, which indicates that the target firms have a more significant rebound after the bad news of the announcement came out. Once again, the finding in

Table 1.3: Fixed Effect Model of Control and Treatment group

Variable	Control	Treatment	DID Test
	b/p	b/p	b/p
Intercept	-0.00462	-0.01280	-0.00054
	(0.000)	(0.000)	(0.000)
D	-0.00130	-0.02483	-0.00130
	(0.060)	(0.000)	(0.069)
t	-0.00036	-0.00109	-0.00037
	(0.000)	(0.000)	(0.000)
Dt	0.00093	0.01028	0.00093
	(0.000)	(0.000)	(0.000)
DS			-0.02353
			(0.000)
St			-0.00073
			(0.000)
DSt			0.00935
			(0.000)

The table shows the trends of abnormal returns across different groups and the jumps of the abnormal returns on announcement day. P-values are shown in parenthesis. The control group and the treatment group both have a V-shaped time trend curve in abnormal return around announcement day, but the magnitude of the decline and the treatment group's recovery are all greater than those in the control group. The downward jump of the abnormal return on announcement day in the treatment group is statistically significant, while its counterpart in the control group is not. The pattern is supported by the difference in difference test.

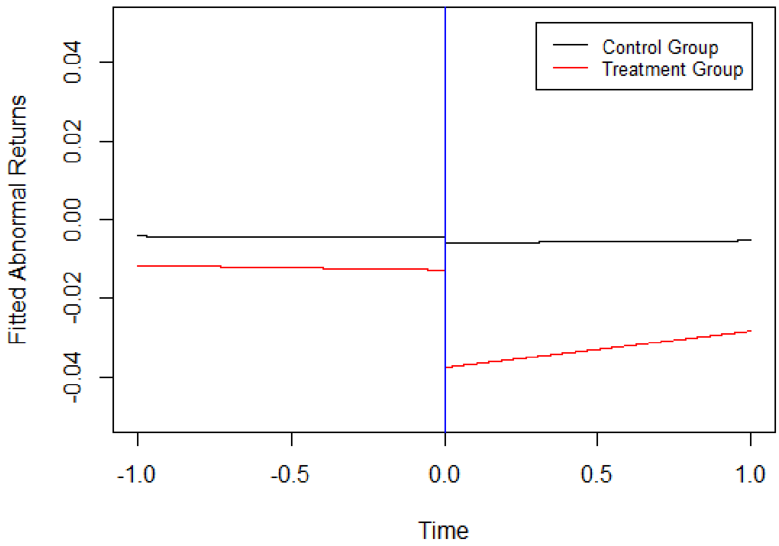


Figure 1.6: Comparison of Control and Treatment Groups

The figure shows the fitted time trend lines of the target and the non-target firms separately in the pre-announcement and the post-announcement period. The slope of the time trend in the control group before announcement day was -0.00036, and it rose to 0.00057 after the announcement was disclosed. As a comparison, The slope of the time trend in the treatment group before announcement day was -0.00109, and it rose to 0.00919 after announcement day. The jump on announcement day in the treatment group is statistically significant at a 5% significant level while the jump in the control group is not. This finding verifies that the target firms experienced a downward plummet due to their announcements of goodwill impairment.

DID regression tells us the same story as that in the fixed effect regression. Even though the target and non-target firms have different performances in abnormal return over time, the results in section 1.5 are not informative to detect informed trading. Therefore a real hypothesis test on the informed trading is conducted and discussed in section 1.6.

1.6 Hypothesis Test

The difference in abnormal returns between the target and non-target firms could be attributed to different reasons. On the one hand, a possible reason is informed trading. People suspect that there are informed trading behaviors before announcement day. Specifically, the information of profit loss was already known by informed traders before announcement day, so the abnormal return dropped gradually before the announcement news came out. The informed traders sold their stock shares in advance and avoided a capital loss. As a result, the motion of abnormal return looks like that in figure 1.4. On the other hand, the difference in abnormal return between the treatment and the control groups simply comes from the firms' heterogeneity. It is possible that the target firms are different from other firms naturally due to their own characteristics like industrial features, number of listed years, etc. Therefore they have a different motion of abnormal returns before and after announcement day.

To study the hypothesis empirically, I test the hypothesis in this section by running the difference in difference regression in the equation 1.6 with various cut-off days. Specifically, I choose a sequence of days before the announcement day, like 5 days before the announcement day, 10 days before the announcement day, 15 days before the announcement day, etc. I test the statistical significance of the coefficient β_6 in the equation 1.6 for all the cut-off days. If the coefficients are significant for all these cut-off days, it shows that the heterogeneity across the firms rather than informed trading plays an important role in figure 1.6. Otherwise, informed trading may exist in the sample of the target firms. Here I assume that the information leak, if it exists, does not occur one month earlier than the announcement day because the informed traders may not be able to predict precisely the motion of stock price in the future if they sell their shares too early, they certainly could change their positions in the stock market right before announcement day so that they could reduce unnecessary uncertainty. So I choose one month as the maximum bar and conduct the test accordingly.

I list the F-statistics and the corresponding P-value of each cut-off day in the table 1.4. In this table, T represents the announcement day of a firm. Therefore $T-5$ represents the day 5 days before the announcement day; other cut-off days are defined similarly. The table indicates that all the P-values are less than 0.05, which shows that all the fitted lines' slopes are different between the treatment group and the control group before these cut-off days. It supports the second conjecture that the heterogeneity across the target and the non-target firms is the dominating factor in explaining the jump on announcement day in the dataset. In addition, I have to concede that the finding of these significant estimates for β_6 does not mean there is no informed trading at all in the market. Actually, it only tells us that there is no strong evidence to detect informed trading behaviors according to the sample.

A more careful looking at the target firms' abnormal returns is shown in figure 1.7. There are 6 different announcement dates in Jan 2019; Jan 24, Jan 25, Jan 28, Jan 29, Jan 30, Jan 31. On each announcement day, some target firms claimed irrational goodwill impairment. I plot the average abnormal returns of these firms for the week before their announcement day. For example, there are 27 target firms announced on Jan 30, so I plot the average abnormal returns of these firms from Jan 23 to Jan 30 in red with the number of firms labeled in the figure. Similarly, the firms' average abnormal returns on other announcement days are plotted separately in the figure.

It turns out that the average abnormal returns are closer to each other as the announcement day is closer to Jan 31. The red line and black line represent the average abnormal returns of the firms announced on Jan 30 and Jan 31 separately. The numbers of the firms announced on Jan 30 and 31 are the most. Actually, 27 firms made an announcement on Jan 30, and 47 firms made an announcement on Jan 31, while the numbers of the firms that made announcements on other days in the figure are all less than 4. The reason for the clustering announcements could be interpreted in different ways. On the one hand, Jan 31 is the deadline for the announcement for a listed firm to summarize its performance in 2018 according to the regulation in China's stock market. These firms are similar so that they all want to report their profit loss on the last day. On the other hand, these firms made announcements at the end of Jan 2019 is due to informed trading. It is possible that the target firms who made an announcement earlier showed a good example for other firms and inspired them to make a similar announcement about goodwill impairment. From the follower's angle, it is attractive to follow the forerunner because they already have profit loss in their accounting reports. The real difficulty is how to convince shareholders and the public naturally. Therefore taking goodwill impairment and following other similar firms would be a good way to go. They will not face big punishment from the government since there are too many such firms.

According to figure 1.7, there is no unique pattern for all these 6 announcement

Table 1.4: Hypothesis Test for Different Cut-off Days

Cut-off Time	T	T-5	T-10	T-15	T-20	T-25	T-30
$F - stat$	102.99	11.23	3.9	7.89	19.87	7.17	10.11
$P - value$	0.0000	0.0008	0.0482	0.0050	0.0000	0.0074	0.0015

The table shows the hypothesis test for different cut-off days, T represents the announcement day, and $T - i$ represents the day that i days before the announcement day. All the P-values are less than 0.05, and they jointly show that all the slopes of fitted lines are different between the treatment group and the control group before these cut-off days. The heterogeneity across the target and non-target firms rather than informed trading is the dominating factor in explaining the jump on announcement day in the sample.

days. Actually, the motions of average abnormal returns are clearly different. This could be considered as support to the first conjecture. Namely, the firms share similarities naturally, so they all have profit loss in 2018, and they decided to make an announcement about their loss in the last week before the deadline. Besides, there is no line in the figure that closely follows another line, which suggests that the effect of the situation described in the second conjecture above, if any, is limited. According to the figure, the firms who made an announcement earlier did not have a strong effect on the firms who made an announcement later. In summary, the key point of the conclusion is that these target firms are somehow different from the rest of the list firms naturally, so they share more similarities within the treatment group rather than with the non-target firms in the control group. The pattern found in the figure 1.6 should not be attributed to informed trading alone. Actually, the tacit understanding across these target firms is a more realistic explanation for the clustering announcement of goodwill impairment in Jan 2019.

To conclude robustly, I test the difference in the trends of abnormal return between the treatment group and the control group by running a fixed effect regression of the abnormal return on the time index variable t, the group dummy variable S, and their interaction term St. The model's specification is shown in the equation 1.7, and the corresponding

Table 1.5: Test on Time Trend and Chow Test

Tests	Sample Size	30	40	50	60	70	80
Difference in Trend	t-stat	-0.0011	-0.0005	-0.0002	-0.0002	-0.0001	-0.0001
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Chow Test	F-stat	139.54	65.83	38.04	49.13	33.46	39.83
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The table shows the Time Trend Difference Test and the Chow Test. The t-stat is the t-statistics of β_3 in the equation 1.7. All the estimates of β_3 are significant at a 5% significance level so that all the firms in the treatment group and the firms in the control group have different slopes of time trend before the announcement day. The finding is robust over different sample sizes, and it is supported by the Chow Test.

fitting result is recorded in the table 1.5. Here I use the samples with different sample sizes from 30 to 80. Each sample includes a certain number of observations of abnormal returns right before the announcement day. For example, sample size 30 means that the sample has the observations from day $T - 30$ to day T, where T is the announcement day; other sample sizes could be interpreted in the same way.

$$y_{it} = \beta_0 + \beta_1 S + \beta_2 t + \beta_3 St + \alpha_i + \epsilon_{it} \tag{1.7}$$

It turns out that all the estimates of β_3 are significantly different from zero in standard t-test, which suggests that the firms in the treatment group and the firms in the control group have different slopes before the announcement day. The negative β_3 tells us that the target firms experienced a larger decline in abnormal return before the announcement day than the non-target firms, so this finding is in line with the results in previous tables, and it holds over different sample sizes. In addition, The Chow test of the difference in the slopes between the two groups shows the same conclusion, and the corresponding results are also recorded in the table. The P-value of all the F-statistics in the Chow Test is 0 over different sample sizes, verifying the conclusion in the section.

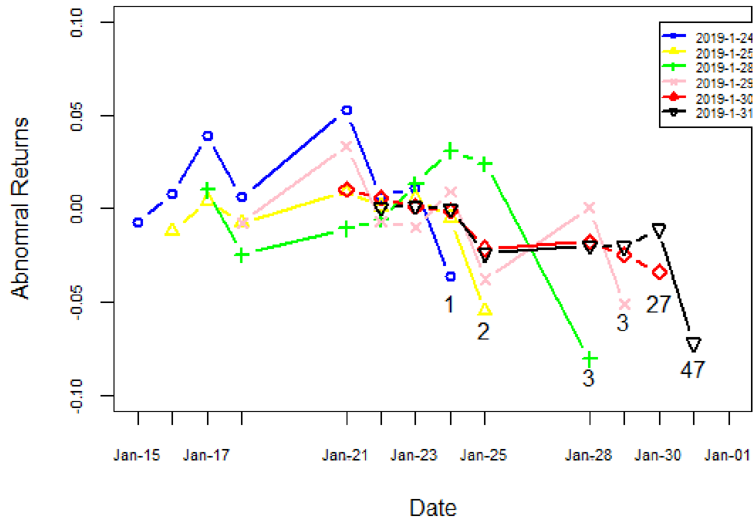


Figure 1.7: Average Returns of firms clustered on Different Announcement Dates

The average abnormal returns are closer to each other as the announcement day is closer to Jan 31. There is no unique pattern for all these 6 announcement days; the firms share similarities naturally, so they all have profit loss in 2018, and they decided to announce their loss in the last week before the deadline.

1.7 Robustness

I conduct a robustness test from three different perspectives. Firstly, I change the sample size in the benchmark regression and test whether the pattern in benchmark regression still holds across different sample sizes in section 1.7.1. Secondly, I test whether there is any difference in the slopes of time trend between the two groups before announcement day in section 1.7.2. Finally, I investigate the target firms' industrial distribution to explore their natural difference in the last section 1.7.3.

1.7.1 Sample Size

According to the methodology of the analysis in section 1.5, I choose the research period from $T - 30$ to T, where T is the day a firm made its announcement in Jan 2019. To make the finding in the benchmark analysis robust, it is necessary to test the finding's sensitivity to the changes in sample size. Therefore I examine the robustness of the results in section 1.5 by changing the ex-event window's length and the post-event window's length. The summary of the test is shown in the table 1.6.

The table indicates that the pattern found in the benchmark analysis is quite robust, and the estimate of β_6 in the DID model is always statistically significant negative. Specifically, the estimates of β_6 over different sample sizes are shown in the table's first row. Here sample size is shown in the form of interval. For example, (-10,+10) represents the sample of abnormal returns from 10 trading days before the announcement day T to 10 trading days after the announcement day T. The estimate of β_6 in the equation 1.6 rises from -0.0016 to -0.0007 as the length of the sample increases. It suggests that once more observations are included in the fitting, the difference in the slopes of the time trend in abnormal return varies across the treatment group and the control group before the announcement day. The target firms and the non-target firms behave more and more differently in abnormal return as the announcement day comes; this is a new finding in the section. Therefore, it is interesting to test whether the difference in the time trends between the treatment and control groups is statistically significant before the announcement day. I could even further ask whether the difference itself is changing over time. If the answer to this question is yes, it would be indirect evidence supporting the conjecture that the target firms and the non-target firms behave differently even before their announcement days. Actually, the answer to this question would provide more dynamic insights into the difference between the two groups.

1.7.2 Difference in Slopes

As discussed at the end of the last section, it is interesting and necessary to test whether the difference in the slopes of time trend between the treatment group and the control group before announcement day is constant or not; therefore, I run the model in the equation 1.6 by using the sample of abnormal returns of each firm before announcement day. Specifically, I achieve this goal by defining variable D differently, actually I define a list of days(t_1, t_2, t_3) before the announcement day T, defining $D = 0$ if t is in the time interval (t_1, t_2) and $D = 1$ if t is in interval(t_2, t_3). For example, given a time interval $(T-80, T-50, T-20)$, I define $D = 0$ if the observations are in the interval $(T-80, T-50)$

Table 1.6: Robustness Test for Different Window Sizes

Time Interval	(-10,+10)	(-20,+10)	(-20,+20)	(-30,+10)	(-30,+20)	(-30,+30)
β_6	-0.0016	-0.0011	-0.0011	-0.0007	-0.0007	-0.0007
$F-stat$	22.17	76.41	76.13	102.97	103.97	99.12
$P-value$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The table shows the estimates of β_6 in the DID model across different samples. The F-statistics and the corresponding P-values are listed. The estimate of β_6 rises from -0.0016 to -0.0007 as the sample size increases. Once more observations are included in the fitting, the difference in the slopes of the time trend in abnormal return varies across the treatment group and the control group before the announcement day.

and $D = 1$ if the observations are in the interval$(T-50, T-20)$. Based on the definition of D, I could test if there is any change in time trend on the day $T-50$ rather than on the announcement day T. If there is any non-regular shock in the stock market, like either the abnormal return of the target firms or that of the non-target firms changed, I expect to see the slope difference between the two groups in the time interval $(T-80, T-50)$ should be different from the slope difference in the time interval $(T-50, T-20)$.

Following the idea introduced above, I list the empirical result of the slope difference test in the table 1.7. It turns out that as t_3 goes closer to the announcement day T, the slope difference between the treatment group and the control group decreases from the interval (t_1, t_2) to (t_2, t_3), which means that the target firms in the treatment group experienced an increasingly sharp decline in abnormal return relative to the non-target firms in the control group as the announcement day comes. The slope difference decreases from 0.00034 to -0.00055 as t_3 goes from $T-20$ to $T-1$. Here I used two 30 days time intervals. A similar pattern is observed if we use a short time interval of 20 days. Therefore, I conclude that the target firms behaved similarly to the non-target firms initially before announcement day. Later, they experienced a more significant slump in abnormal return than the non-target firms as the announcement day comes. The significance of the estimate for β_7 in the equation 1.6 varies over time, but it is interesting to see that the estimate becomes more and more significant as it is closer to T. Moreover, the abnormal returns of both types of firms responded in advance to their announcements. This finding is not surprising because

Table 1.7: Difference in Slopes Test

	Length	30	Length	20
	Time Interval	β_7	Time Interval	β_7
		b/p		b/p
	(T-80,T-50,T-20)	0.00034	(T-60,T-40,T-20)	0.00062
		(0.000)		(0.000)
	(T-75,T-45,T-15)	0.00006	(T-55,T-35,T-15)	0.00010
		(0.478)		(0.518)
Sample Length	(T-70,T-40,T-10)	0.00008	(T-50,T-30,T-10)	-0.00045
		(0.331)		(0.007)
	(T-65,T-35,T-5)	-0.00014	(T-45,T-25,T-5)	-0.00085
		(0.092)		(0.000)
	(T-63,T-33,T-3)	-0.00051	(T-43,T-23,T-3)	-0.00108
		(0.000)		(0.000)
	(T-61,T-31,T-1)	-0.00055	(T-41,T-21,T-1)	-0.00134
		(0.000)		(0.000)

The table shows the difference in the time trends between the treatment and control groups before the announcement day. As t_3 goes closer to the announcement day T, the slope difference between the treatment group and the control group decreases from interval (t_1, t_2) to (t_2, t_3). It reflects that the target firms experienced an increasingly sharp decline in abnormal return relative to the non-target firms as the announcement day comes.

people do not expect the difference among various firms is always constant. The finding in the table gives us a bigger picture of the trends of abnormal returns before announcement day.

1.7.3 Industrial Distribution

If the heterogeneity across the firms in the treatment and the firms in the control groups plays an important role in determining their difference in abnormal return around announcement day, summarizing the target firms' features is a good perspective to understand these firms better. Actually, I summarized the industrial distribution of the target firms in the table 1.8. The table presents the number of target firms in each industry category and the total number of listed firms within that category. It depicts the target firms' distribution across different industries. I only list the industrial category whose target firms number is greater than 1. Some interesting patterns could be found in the table. On the one hand, the machinery industry and electronic information industry have the highest absolute number of target firms, but this result is probably due to the large number of listed firms in these two industries, 211 and 247 respectively, so that their abnormal ratios are not the highest in the sample actually. On the other hand, few other industries have both the large absolute number of target firms and the large abnormal ratio, like media entertainment, non-ferrous metal, integrated industry, shipbuilding, and textile industry. The clustering distribution of the target firms is a necessary condition of the hypothesis that the natural difference between the target firms and the non-target firms is the dominating factor, and it could be considered as indirect support for the hypothesis. Moreover, the abnormal ratio in the table ranges from 2.29% to 25%, the large scope of the abnormal ratio also supports the conclusion in section 1.6.

1.8 Conclusion

In this paper, I studied the abnormal drop of the stock returns on announcement days for China's listed firms on A board. Firstly, I found that the target firms, the firms who reported a huge irrational goodwill impairment, experienced a sharper decline before announcement day. Their rebound in abnormal return is, on average, more powerful after announcement day. The powerful rebound of the target firms may come from the bottom fishing strategy of investors. Actually, the target firms experienced a much larger decline in abnormal return, so it is more profitable for the investors to buy these target firms' stocks if they want to use the strategy. Besides, the empirical finding suggests that there is a statistically significant negative jump in abnormal return on announcement day for the target firms. In contrast, the jump for the non-target firms is not significant.

Secondly, I conducted the hypothesis test in section 1.6 to investigate the possible reasons for the phenomenon. It turns out that the heterogeneity across the target firms and the non-target firms plays a more important role than the exogenous reason like informed

Table 1.8: Distribution of the Target Firms Across Different Industries

Industrial Category	Target firms	Total listed firms	Abnormal Ratio
Transportation	2	87	2.29%
Media Entertainment	5	40	12.50%
Agriculture	2	64	3.13%
Power Equipment	2	65	3.08%
Real Estate	3	123	2.44%
Non-ferrous Metal	5	72	6.94%
Machinery Industry	10	211	4.74%
Vechicle Manufacturer	5	103	4.85%
Biological Pharmacy	4	155	2.58%
Electrical Appliance	4	58	6.89%
Electronic Information	14	247	5.67%
Electronic Equipment	8	152	5.26%
Textile Industry	3	42	7.14%
Integrated Industry	3	33	9.09%
Shipbuilding Industry	2	8	25.00%
Finance Industry	2	51	3.92 %

The table shows the distribution of target firms across different industries. Count of the target firms, the total count of the listed firms on A board, and the abnormal ratio are listed in the table. Some industries have both the larger number of target firms and the larger abnormal ratio, like shipbuilding, media entertainment, integrated industries, etc.

trading. The difference between the target firms and the non-target firms exists not only on announcement day but also on other cut-off days before it. So There is not enough evidence found in the aggregate level data to identify the existence of informed trading in the market.

Finally, I tested the distribution of the target firms over different industrial categories. Some industries have a larger number of target firms and a larger abnormal ratio, like the shipbuilding industry, media entertainment, integrated industries, etc., which supports the finding in section 1.5 and 1.6. The key conclusion of the paper is that these target firms are somehow different from the rest of the list firms naturally, so they share more similarities within the treatment group rather than with the non-target firms in the control group, and the empirical difference in these two groups of firms should not be attributed to informed trading alone. More empirical research on the existence of informed trading in China's stock market correction would be a great extension of the paper.

1.9 References

- Bens, D., Heltzer, W., & Segal, B. (2007). The information content of goodwill impairments and the adoption of SFAS 142. Working paper, University of Arizona.
- Elliott, J., & Shaw, W. (1988). Write-offs as accounting procedures to manage perceptions. Journal of Accounting Research, 40, 91–119.
- Francis, J., Hanna, J., & Vincent, L. (1996). Causes and effects of discretionary asset write-offs. Journal of Accounting Research, 34, 117 – 134.
- Henning, S., & Stock, T. (2004). The Amount and Timing of Goodwill Write-Offs and Revaluations: Evidence from U.S. and U.K. Firms. Review of Quantitative Finance and Accounting
- Hirschey, M., & Richardson, V. (2003). Investor underreaction to goodwill write-offs. Financial Analyst Journal, 59(6), 75–84.
- Lee, David S.(2008). Randomized Experiments from Non-random Selection in U.S. House Elections, Journal of Econometrics, 142 (2), 675–697.
- Lee, David S. & Lemieux,Thomas (2010). Regression Discontinuity Designs in Economics. Journal of Economic Literature 48 (2): 281–355.
- Li,Z., Shroff,P., Venkataraman,R. & Zhang,X.(2011). Causes and consequences of goodwill impairment losses. Review of Accounting Studies
- Liberatore, G. & Mazzi,F.(2010). Goodwill write-off and financial market behaviour: An analysis of possible relationships. Advances in Accounting, 26(2), 333-339.
- Romanna, K. & Watts, Ross L.(2012). Evidence on the use of unverifiable estimates in required goodwill impairment. Review of Accounting Studies. 17, 749-780.

Chapter 2

Predicted Return of Recurrent Neural Network: Evidence from China's Stock Market

2.1 Introduction

Machine learning and Artificial Intelligence have been increasingly popular in recent years due to the capacity to solve intelligence tasks. Neural Network, as the key technology of deep learning, is often applied to various industrial applications like facial recognition, automobile, fraud detection, etc. The emergent technology has also been studied and used in financial applications like the development of quantitative trading strategy, prediction of equity returns. This paper focuses on two different neural network models: the Multilayer Perceptron(MLP) model and the Recurrent Neural Network(RNN) model. These models are compared with the standard ARIMA model in the forecast of stock return.

The paper focuses on empirical analysis, and it mainly answers two big questions of interest. First, what are the prediction performances of these three models empirically when they are used to predict China's stock returns, how are the models compared with each other? Second, what are the corresponding trading performances if these models are adopted in realistic quantitative trading? I choose ARIMA, MLP, and RNN models in the paper because they are designed gradually closer to real time-series data. Actually, the order of the ARIMA, MLP, and RNN model exactly shows the improvement in model architecture. Specifically, the ARIMA model only allows a linear relationship between the historical stock price and today's stock price. MLP model introduced a nonlinear relationship, neural network structure, into the analysis. It allows a more complex structure. It is more realistic because stock price does not always necessarily follow a linear relationship across different periods. Still, the MLP model does not have a memory characteristic; that is, the state of network architecture at period t only depends on the input in this period. It is not associated with the state of the neural network in past periods. The RNN model overcomes this shortage. RNN model assumes that the state of the neural network model at period t not only depends on the input in the period but also depends on the state of the network in past periods. It is a combination of input at period t and all the past history; the memory feature makes it closer to real time-series data.

The empirical analysis of the two big questions of interest could indirectly show the improvement from assuming realistic model architecture. Namely, what is the net gain of switching a linear predictive model like ARIMA to a standard neural network model like the MLP model if we only use historical data as input? Further, it is also interesting to know the net gain of switching a standard neural network model like the MLP model to a RNN model focusing on sequential data.

The above questions are not the central questions to answer in the paper, but empirical analysis of the two questions could provide us with insights. On the one hand, if the MLP and RNN models' predictive performances are similar to that of the ARIMA model, it is a signal telling us that both models fail to provide a better prediction. It implies that the improvement in model architecture is not sufficient to have a significant improvement in prediction. On the other hand, if there is a large difference between the predictions from neural network models and that from the ARIMA model, it is strong evidence for the success of the improvement in model architecture. This is because I only use historical information of stock price in the predictions of the models, and the difference in inputs has already been controlled in the analysis. Actually, the input variables' restrictions are associated with the weak form of the efficient market hypothesis, namely, whether investors could obtain excess profit in realistic stock trading if they only use publicly available historical information. The answer should be "no," according to the week form of the efficient market hypothesis; actually, it is the case in China's stock market.

In summary, two key points distinguish the paper from existing literature. Firstly, I use autoregressive terms of different stock indexes as inputs to predict stock returns. Secondly, I use a standard quantitative trading strategy to backtest the predictions from these models. It turns out that the RNN model provides the best trading performance for the Industrial Index and the Composite Index. MLP model works extremely well for the Business Index and the Property Index. ARIMA model predicts the SSE Composite Index and the Shanghai A-share Index better than other models. I also find the trading strategy using predictions from MLP and RNN models provide a higher return than benchmark stock indexes in the long bear market from 2017 to 2020, which offers investors more options for their portfolio management. I also compare MLP and RNN models with the ARIMA model in terms of their mean squared error and forecast accuracy across various stock indexes in China. There is no unique dominating model in prediction.

The rest of the paper is organized as follows, I list the existing literature in section 2.2 and introduce the dataset in section 2.3. Section 2.4 presents the neural network frames and section 2.5 shows the empirical finding of ARIMA, MLP and RNN models. I also discuss and conduct a robustness test in section 2.6. Finally I draw conclusions in section 2.7.

2.2 Literature Review

Neural Network, as one of the most powerful nonlinear dynamic systems, has been frequently studied in various research. A lot of literature investigated the topic from different aspects of the finance application. Lisboa et al.(2000) reviewed and summarized the business applications of the neural network, their performance, and potential in E-commerce, retail finance, bankrupt prediction, payment card fraud detection, money laundering detection, etc. Dunis & Williams(2002) further discussed whether neural network models' performance is significantly better than traditional forecast models in predicting the EUR/USD exchange rate. They found that the neural network model outperforms the ARMA model in terms of different trading metrics like annualized return, Sharpe ratio, etc., but the neural network model has its own shortage. On the one hand, it is hard to interpret the result given that the neural network is a nonlinear layer structure model. On the other hand, its predictive potential heavily depends on the control variables fed into the model. This is also related to the efficient capital market hypothesis studied in Fama(1970), historical information test, publicly available information test, and all the information test would lead us to the conclusion of weak form, semi-strong form, and the strong form of the efficient capital market respectively.

In addition, a lot of literature focused on the recurrent neural network(RNN) model because of its natural advantage in forecasting time series data. Kamijo & Tanigawa(1990), Tenti(1996), Tino(2001) all studied the performance of recurrent neural network model in forecasting returns of the foreign exchange rate of Deutsche Mark and volatility of DAX. RNN model has been verified to have remarkable forecast potential in the literature. But the findings are not claiming the RNN model always to be the first choice among all neural network models. Actually, Dunis et al.(2006) found that the Multilayer Perceptron(MLP) model performed better than the RNN model in the trading of futures spreads. Dunis et al.(2010) further investigated the difference between MLP, RNN, HONN, and Psi Sigma Models in predicting EUR/USD exchange rate from 1999 to 2006 by only using autoregressive and moving average terms of the raw return sequence, and found that MLP model outperformed other models in terms of Sharpe ratio and annualized return while RNN model performed a little better than traditional ARIMA model. I followed the methodology in the literature and used it to test MLP and RNN models' performance in China's stock market.

Since the paper's motivation is to forecast returns of China's stock indexes by using different models, I searched the literature focusing on China's stock market and neural network model specifically. Actually, many papers studied this topic. Cao et al.(2005) compared the forecast power of the Fama French factor model with that of the artificial neural network model in the study of China's stock market. They found that neural network models outperformed the traditional linear model. Dai et al.(2012) combined nonlinear independent component analysis with a neural network model to predict Asian stock market indexes. They found that their proposed methodology outperformed other methods in the analysis of the Shanghai B-share closing index. All these findings are informative and suggesting that neural network models like RNN and MLP models have the potential to offer better predictions in financial markets. Still, the conclusion needs to be tested and studied systematically for a developing country like China. Therefore, I compared ARIMA, MLP, and RNN models for the 10 most important stock indexes in the Shanghai Stock Exchange and the other 7 most representative stock indexes in the Shenzhen Stock Exchange by only using historical stock price as input. I systematically compare the mean squared error, predictive accuracy, annualized return of strategy using predictions from these models across the stock indexes. Empirical findings and discussions are shown and discussed in the following sections.

2.3 Data Source

To investigate the prediction capacity of neural network models on Chinese Stock indexes, I used stock indexes data from Shanghai Stock Exchange(SSE) and Shenzhen Stock Exchange(SZSE) through Tushare finance database(Tushare.org). "tushare" is a Python package provides API to all the historical trading data of stock indexes in China. This paper chose the 10 most representative stock indexes from the Shanghai Stock Exchange in benchmark analysis. They are "SSE Composite Index", "Shanghai A-Share Index", "Shanghai B-share Index", "Industrial Index", "Business Index", "Property Index","Public Index", "Composite Index", "Shanghai Stock Exchange 180" and "Shanghai Fund Index".

There are a few reasons to choose these indexes in benchmark analysis. On the one hand, these stock indexes represent the most important aspects of China's economy, they are important indicators carrying referenceable signals for investors, and they show the sentiment of investors in different industries. On the other hand, they are the stock indexes created after "China's Stock Share Reform" in 2005, one of China's deepest financial market structural reforms. The feature makes these stock indexes be standardized and more consistent with financial market standards in developed countries. In addition, compared

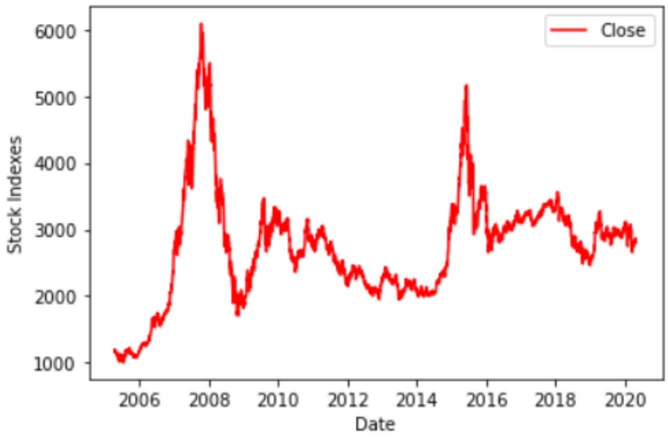


Figure 2.1: SSE Composite Index

The figure shows the SSE Composite Index's close price from April 18, 2005, to April 30, 2020. High volatility is observed. The largest declines are in 2008 and 2015, respectively. Even though the testing dataset from 2017 to 2020 is not that volatile as the training dataset, it is still a bear market in the period.

with many Chinese Stock indexes created after 2010, these stock indexes have longer trading history, which means that their data could be traced to 2005. This is another advantage of the dataset. Specifically, the dataset is from April 18, 2005, to April 30, 2020. There are overall 3658 observations in it. Each observation represents a trading day. This dataset only includes trading days, so all holidays and weekends are not included here. I chose the first 12 years as the training dataset and the last 3 years as the testing dataset. Precisely, the training dataset is from April 18, 2005, to March 22, 2017, the testing dataset is from March 23, 2017, to April 30, 2020. Therefore there are 2900 observations in the training dataset and 758 observations in the testing dataset.

In addition, the dataset includes four different metrics of the stock price; they are "High," "Low," "Open," and "Close," respectively. I used the close price of stock indexes as a benchmark because it includes the information about the strength of both long and short sides within a trading day. Therefore it is more informative for us to learn and predict market returns than other stock prices. The historical close price of the SSE Composite Index is shown in Figure 2.1 below.

In Figure 2.1, I observe the high volatility of the SSE Composite Index. The largest declines are in 2008 and 2015, respectively. Even though the testing dataset from

2017 to 2020 is not that volatile as the training dataset, it is clear that the market is a bear market in the period. Also, as introduced in section 2.1, one contribution of the paper is to investigate the forecast capacity of different models by using historical price information only. That is, I want to test whether the precise prediction of stock returns could be provided by using public available data if we apply neural network models in the analysis. Intuitively, the answer should be "No" due to the random walk feature of stock prices, but this may not hold in China's Stock market. On the one hand, literature like Dai et al.(2012) has already shown that neural network models, combined with other nonlinear independent component analyses, could be applied to the Asian stock market. These models have the potential to outperform traditional time series models in prediction, but they did not provide a general answer for most of China's stock indexes. On the other hand, Dunis et al.(2010) showed that autoregressive and moving average terms of the historical stock price, once applied with neural network models, are informative enough to make a profit in real trading. In this paper, I combine the ideas of the literature. I do not only test the real forecast capacity of both MLP and RNN models in the stock market but also investigate the profitability of the predictions from the models in backtests for the most important and representative stock indexes in China. Forecast comparison and trading performance are discussed in section 2.5.

2.4 Forecast Model

2.4.1 ARIMA Model

As one of the most traditional time series forecast model, the ARIMA model is used as the benchmark model in the analysis in this section. Specifically, the specification of model ARIMA(p,d,q) I used is shown in equation 2.1 below, where p is the maximum number of autoregressive terms, d is the difference order, and q is the maximum number of moving average terms. $y_t = \frac{S_t - S_{t-1}}{S_{t-1}}$ is the return of stock index on day t and it is proofed to be stationary after the ADF test. y_{t-p} and ϵ_{t-q} are autoregressive and moving average terms, ϕ s and θ s are the coefficients of them separately.

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - ... - \theta_q \epsilon_{t-q} \tag{2.1}$$

2.4.2 Multilayer Perceptron Model

Multilayer Perceptron (MLP) model is one of the most popular architectures in the neural network literature. It generates the neural network architecture to forecast the target variable by using a list of explanatory variables. Usually, it has three different types of layers: an input layer, multiple hidden layers, and an output layer. The structure of MLP is shown in Figure 2.2 below. The input layer includes all the explanatory variables; each explanatory variable is represented by a neuron on the input layer. So its dimension is equal to the dimension of the list of explanatory variables k; further, the hidden layer also includes a different number of neurons on it, and one MLP model could have multiple hidden layers. In the figure, there are n neurons on the hidden layer. The output layer is the final forecast output of the MLP model. Since my target is to forecast the daily stock return, my output layer has one neuron, representing the final forecast $\tilde{y}_t$.

All the neurons are connected by the lines in the figure either directly or indirectly with other neurons, each line represents a weight parameter w and bias parameter b, the value of neuron on hidden layer could be computed as nonlinear activation function of a weighted average of neuron values on its last layer. Usually, the nonlinear activation function is the sigmoid function, and the final $\tilde{y}_t$ is a linear function of neurons in hidden layers. The architecture is summarized by the equation 2.2.

$$\begin{aligned} \tilde{y}_t &= \sum_{i=1}^{n} \gamma_i S_{i,t} + \theta_i \\ S_{i,t} &= f(\sum_{j=1}^{k} w_{ji} X_{j,t} + b_{ji}) \\ f(u) &= \frac{1}{1+e^{-u}} \end{aligned} \tag{2.2}$$

Here w_{ji} and b_{ji} are the weight and bias parameter between neuron $X_{j,t}$ and neuron $S_{i,t}$, γ_i and θ_i are the weight and bias parameter between neuron $S_{i,t}$ and final output $\tilde{y}_t$, $i = 1, 2, ..., n$. The target function to minimize is shown in the equation 2.3 where T is

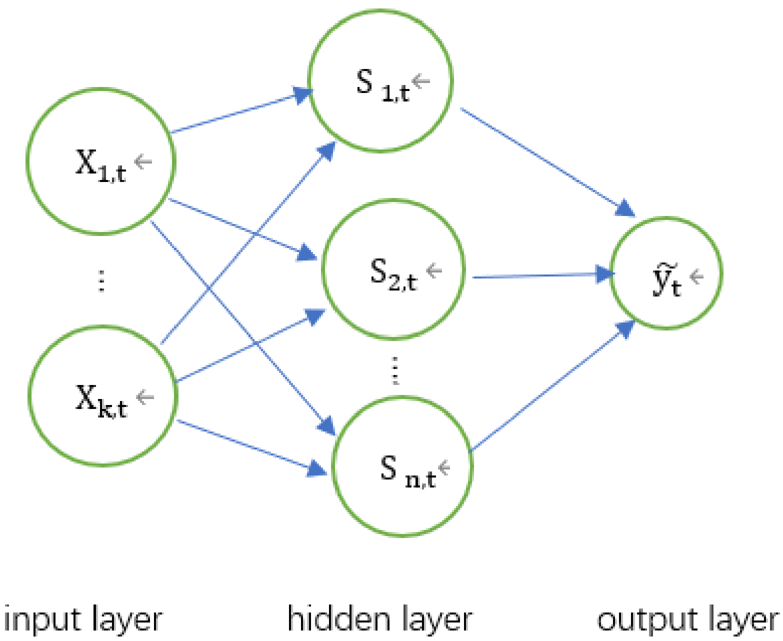


Figure 2.2: Multilayer Perceptron

The figure shows the architecture of the Multilayer Perceptron(MLP) model. It includes an input layer, a hidden layer, and an output layer. The input layer has k input neurons, the hidden layer has n neurons, and the output layer includes the model's output.

the sample size. Optimal weight and bias parameters could be estimated through different methods; the one I use here is the most commonly used one, the backward propagation algorithm following Dunis et al.(2010). The advantage of the MLP model compared with the ARIMA model is that it is a nonlinear forecast model able to capture a more complicated relationship between input and output variables. Its accuracy increases once more neurons and hidden layers are added into the model, and it naturally expanded the linear model's forecast boundary. It is a good supplement to the ARIMA model.

$$min\frac{1}{T}\sum_{t=1}^{T}(y_t - \tilde{y}_t)^2 \tag{2.3}$$

2.4.3 Recurrent Neural Network

The recurrent Neural Network (RNN) model is a special MLP model designed to deal with the sequential dataset because the MLP model does not treat neurons on hidden layers sequentially. Specifically, the MLP model assumes that all the neurons on the hidden layer are independent of the neurons' past states. That is, the MLP model does not take the autocorrelation feature of sequential data into consideration, and all the weights and biases estimated in the MLP model are based on this assumption. This feature of the MLP model naturally does not hold in time series data. Therefore the RNN model was designed to overcome this problem. I showed the RNN model's architecture in figure 2.3 below. The past states of neurons $S_{1,t-1}, ..., S_{n,t-1}$ are all introduced into the fitting in the optimization process. The target of RNN is still minimizing the sum of squared error function shown in the equation 2.3. Theoretically, RNN has an advantage in dealing with sequential data compared with the MLP model due to its memory feature. Particularly, I use Long Short Term Memory(LSTM) method in the fitting of RNN to overcome the gradient vanishing problem, following Dunis et al.(2010). I compare ARIMA, MLP, and RNN models' performances on forecasting the return of various stock indexes; the result and corresponding discussion are shown in the following sections.

2.5 Empirical Results

2.5.1 Data

As introduced in previous sections, I used stock index data from the Shanghai Stock Exchange(SSE) and the Shenzhen Stock Exchange(SZSE) through Tushare finance database(Tushare.org). I chose the 10 most representative stock indexes from SSE as a benchmark. They are "Shanghai Composite Index," "Shanghai A-Share Index," "Shanghai B-share Index," "Industrial Index," "Business Index," "Property Index," "Public Index," "Composite Index," "Shanghai Stock Exchange 180" and "Shanghai Fund Index." Since the paper's motivation is to investigate the forecast capacity of different models by only using historical close price, the input variables are all generated by close price. They are the daily stock returns in the most recent trading days and their volatility, which are defined as in equation 2.4 and 2.5

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{2.4}$$

$$v_t = \sqrt{\sum_{i=1}^{5} \frac{(r_{t-i} - \bar{r})^2}{5}} \tag{2.5}$$

where $\bar{r} = \frac{\sum_{i=1}^{5} r_{t-i}}{5}$

Here we use the SSE Composite Index as an example. The Close price and return sequences of SSE Composite Index are shown in Figure 2.1 and Figure 2.4. The SSE Composite Index's raw close price is neither stationary nor normally distributed, but the return sequence r_t is stationary and normally distributed. Therefore the target variable is the stock return r_t. All the input variables in benchmark models are r_{t-1}, r_{t-2}, r_{t-3}, r_{t-4}, r_{t-5} and v_t. I use maximum lag 5 for daily stock returns and 5-day volatility as input because one week usually has 5 trading days. I treat 5 trading days as a standard period to forecast returns. Here the underlying assumption is that the stock return on the day $t+1$ is only affected by the past 5 stock returns and their volatility within one week. This is assumed in the benchmark model. Robustness checks for longer lag terms are discussed in later sections. With these input variables, I predict daily stock returns of the stock indexes from SSE by using ARIMA, MLP, and RNN models and compare their performance in the next section.

2.5.2 Forecast

To make the MLP and RNN models comparable in terms of forecast capacity, I set the same neural network architecture for them. Namely, MLP and RNN models both have 1 input layer with 6 input variables, 1 hidden layer with 20 neurons, and 1 output layer with 1 neuron. The neuron on the output layer is the predicted return of a stock index. I set this architecture as a benchmark, following the framework of Dunis et al.(2010), where all the neural networks have only 1 hidden layer and the same amount of neurons on it to make models comparable to each other. Also, I chose 20 in the benchmark analysis. Actually, the 20 is not the unique possible choice, it could be any positive number greater than the numbers of input variables in general, and usually, the more neurons there are on the hidden layer, the better the fitting would be. To avoid overfitting, the number of neurons on this hidden layer could not be too large, and the empirical result is robust for different neuron numbers. Therefore 20 is a good benchmark as the beginning.

I used Adam optimizer from Kingma & Ba(2014) in optimization due to its efficiency in computation and a high degree of feasibility. According to Kingma & Ba(2014), the learning rate is set to be the default optimal value of 0.001. As discussed in section 3.3, I focus on the forecast capacity of the models and profitability in realistic trading in the most recent 3 years, from 2017 to 2020. This sample has 758 trading days. I use a training dataset with 2900 observations. It is a rolling window in forecast methodology. Specifically, the sample from April 18, 2005, to March 22, 2017, is used to fit and predict the return of a stock index on March 23, 2017. This is the first training sample, and I could get a predicted stock return on March 23, 2017. Then the sample from April 19, 2005, to March 23, 2017, is used to predict the stock return on March 24, 2017. The training window keeps rolling until the last day of the dataset, April 30, 2020. Therefore, I would get 758 predicted stock returns for each stock index. These predicted stock returns could be referenced as signals for realistic stock trading. Figure 2.6 shows the comparison of the predictions from MLP, RNN, and ARIMA models.

It turns out that the realized return of the SSE Composite Index from 2017 to 2020 ranges from -5% to 5% in most of the times except for some outliers. MLP model provides predictions ranges from -0.5% to 0.5%, RNN model gives predictions fluctuates between -1% and 1%, and ARIMA model provides predictions ranges from -0.4% to 0.4%. All the predictions have less volatility than realized returns. This makes perfect sense because of the characteristics of forecasting. Realized returns usually have larger volatility than predicted returns. In addition, all these models captured the clustering feature in realized stock return, but all these 3 models failed to give good predictions in outliers. For example, there was a sharp drop in daily stock return in Feb 2020. The smallest daily stock return was around -8%. This is a totally unexpected shock due to the COVID-19 pandemic. It is impossible to predict such outliers by only using historical price data.

Nevertheless, we observe something interesting in the predictions from these models, MLP model failed to predict the sharp negative return, but according to it past learning experience, the model knows that it is very abnormal to have such a return. Therefore it predicts that there would be a sharp rebound the next day. This is the intuition why we see a big jump in predictions from the MLP model the next day. As for the RNN model, we could see that it predicts a big positive return on the day right after the "bad news" day, but it quickly predicts a large negative return around -3% on the day after that day. This exactly reflects the difference between MLP and RNN models. As we introduced in section 2.4, MLP does not take the autocorrelation feature of sequential data into consideration while the RNN model used information of past states in fitting optimization, so past deep negative return would have a stronger effect on the prediction than MLP model. Compared with MLP and RNN models, the ARIMA model has smoother predictions. Here I used ARIMA(5,0,5) model to make the input variables comparable with that in MLP and RNN models. To quantify the predictions' performance. Mean Squared Error(MSE) and accuracy of predictions are shown in the table 2.1.

It indicates that the ARIMA model provides the minimum MSE in forecasting daily stock return compared with MLP and RNN models. Let's take the SSE Composite Index as an example, MSE of the MLP model is 1.29×10^{-4}, which is smaller than that of the RNN model 1.4×10^{-4}, but they are all greater than that of the ARIMA model 1.27×10^{-4}. This pattern in MSE from MLP to ARIMA model is robust for almost all the 10 stock indexes. The finding in the table shows that even though the MLP and RNN models have very small MSE and they are all very close to the MSE from the ARIMA model, they still failed to provide a better forecast for stock return as compared with the ARIMA model for the listed stock indexes. To make our evaluation robust, I also compute the accuracy of the predictions from these models. Here the accuracy is defined as the number of correct sign predictions divided by the total number of predictions. This magnitude free metric depicts the performance of these models from another angle. The variance of realized stock returns is listed too. The results are shown in the right panel of Table 2.1.

According to the results shown in Table 2.1, the predictive accuracy of the MLP model is 48.42% for The SSE Composite Index, the predictive accuracy of the RNN model is 49.34%, and the predictive accuracy of the ARIMA model is the highest 49.47%. Actually, the RNN model provides the highest forecast accuracy for most of the stock indexes(4 out of

Table 2.1: Mean Squared Error and Accuracy Comparison

	Mean Squared Error(10^{-4})			Accuracy(%)			Variance(10^{-4})
Stock Index	MLP	RNN	ARIMA	MLP	RNN	ARIMA	
SSE Composite	1.29	1.4	1.27	48.42	49.34	49.47	1.25
Shanghai A-share	1.31	1.33	1.28	48.88	47.29	48.28	1.25
Shanghai B-share	1.23	1.25	1.14	51.66	51.92	46.49	1.13
Industrial	1.51	1.51	1.43	49.08	48.81	49.86	1.41
Business	2.21	2.32	2.13	52.37	48.02	48.54	2.11
Property	2.36	2.37	2.35	51.98	50.92	46.70	2.31
Public	1.49	1.51	1.41	48.74	48.88	46.56	1.4
Composite	1.33	1.41	1.32	47.89	52.24	50.52	1.3
Shanghai 180	1.44	1.47	1.41	50.26	49.21	50.65	1.39
Fund	0.52	0.52	0.51	50.53	51.32	48.94	0.5

The table shows the Mean Squared Error(MSE), the accuracy of the predictions from different models. ARIMA model provides the minimum MSE in forecasting daily stock return as compared with MLP and RNN models. This pattern in MSE from MLP to ARIMA model is robust for almost all the 10 stock indexes. Actually, the RNN model provides the highest forecast accuracy for most of the stock indexes(4 out of 10), and the MLP model and ARIMA model both provide the highest predictive accuracy for 3 stock indexes. The empirical finding suggests that MLP and RNN models are not significantly better than the ARIMA model in forecasting the returns of China's stock indexes if we only use historical information of stock prices.

10), and the MLP model and ARIMA model both provide the highest predictive accuracy for 3 stock indexes. The empirical finding suggests that MLP and RNN models are not significantly better than traditional time series models like ARIMA in forecasting the stock returns in China's stock indexes if we only use historical information of stock prices.

Actually, I also compute the 95% confidence interval of the accuracy to further test whether the accuracy from these models are significantly different from 0.5, I use

$SE = \sqrt{\frac{0.5(1-0.5)}{n}}$ to approximate the standard error of a binomial distribution, where $n = 758$, the corresponding confidence interval is $(0.5 - 2SE, 0.5 + 2SE)$. The accuracies are not statistically different from 0.5 if they are within the confidence interval. It turns out that the confidence interval is approximately (0.4637, 0.5363), which means that all these accuracies are not statistically different from the probability of tossing a fair coin 0.5.

This finding is interesting because it implies only adding the autocorrelation structure in the estimation in the neural network, from the MLP model to the RNN model, is not enough to provide a significant improvement in prediction. Even though, RNN model still shows its potential in the predictive accuracy of stock motion. Besides, it is interesting to see that almost all the predictive accuracies are around 50% in the sample. This is not surprising because all the input variables are from historical stock prices. Stock indexes are usually affected by different factors like fundamentals, unexpected announcements, public sentiment, etc. It is intuitive to see all the accuracies in table 2.1 are around 50%. I also show the "False Positive Rate(FPR)," "True Positive Rate(TPR)," and "Precision" of the models in table 2.2. Almost all the metrics are around 0.5 for these three models, which supports the finding that these models' predictive accuracies are not statistically different from 0.5.

So far, I have shown that we could not predict future stock return precisely by only using historical data of stock indexes; it is actually close to the probability of tossing a fair coin. This is the paper's first conclusion, but it is still possible for us to use the predictions from MLP and RNN models constructively. I would use the predictions in the section to proceed with real trade in the stock market and compare the trading performances of the predictions from these models. Further discussion is shown in the next section.

2.5.3 Trading Performance

In the last section, ARIMA, MLP, and RNN models provide similar performance in predictive accuracy and mean squared error. In order to verify the value of the predictions, I backtest the performance of these predictions in realistic trading by using the most straightforward strategy. Specifically, if the prediction of tomorrow's stock return is

Table 2.2: FPR, TPR, Precision of the models

Stock Index	MLP FPR	MLP TPR	MLP Precision
SSE Composite	0.518	0.486	0.508
Shanghai A-share	0.568	0.540	0.511
Shanghai B-share	0.453	0.485	0.511
Industrial	0.501	0.484	0.517
Business	0.405	0.449	0.514
Property	0.572	0.620	0.499
Public	0.469	0.441	0.470
Composite	0.564	0.521	0.484
Shanghai 180	0.539	0.543	0.510
Fund	0.495	0.505	0.528
	RNN FPR	**RNN TPR**	**RNN Precision**
SSE Composite	0.416	0.411	0.521
Shanghai A-share	0.504	0.452	0.496
Shanghai B-share	0.369	0.405	0.517
Industrial	0.429	0.414	0.517
Business	0.307	0.257	0.444
Property	0.489	0.507	0.488
Public	0.449	0.422	0.470
Composite	0.516	0.560	0.525
Shanghai 180	0.515	0.499	0.500
Fund	0.558	0.578	0.531
	ARIMA FPR	**ARIMA TPR**	**ARIMA Precision**
SSE Composite	0.585	0.567	0.516
Shanghai A-share	0.584	0.549	0.509
Shanghai B-share	0.613	0.603	0.506
Industrial	0.564	0.556	0.524
Business	0.562	0.541	0.478
Property	0.660	0.606	0.458
Public	0.538	0.574	0.500
Composite	0.583	0.596	0.510
Shanghai 180	0.601	0.615	0.514
Fund	0.782	0.768	0.516

The table shows the FPR, TPR, and Precision of the three models. Almost all the metrics are around 0.5 for these three models, which supports the finding that these models' predictive accuracies are not statistically different from 0.5.

positive, I use my capital to buy the stock index when the market opens tomorrow(9:30 am), then I sell all my stock shares when the market closes tomorrow(3:00 pm) and get cash back. If the prediction of tomorrow's stock return is negative, I will not trade. This is a trading strategy that directly measures the performance of predictions of stock return. Here a few points worth mentioning. (1) China's stock market opens at 9:30 am and closes at 3:00 pm on every single trading day, and there is one and half hour lunch break at noon from 11:30 am to 1:00 pm. Since the strategy only focuses on the open and close prices of stock indexes, this break at noon could be ignored in the analysis. (2) China's stock market is a t+1, rather than a t+0 market, which means that if I buy a stock today, I could not sell it within the same trading day. Thus, the strategy I introduced above is not feasible in a realistic market. Still, the simulated return, after taking realistic transaction cost into consideration, could be calculated by using JoinQuant API, which makes the backtest of the strategy feasible in the paper. (3) The standard transaction cost in Shanghai Stock Exchange has been set in the strategy's simulation. The open and close commission per trade is 0.03%, and close tax is 0.1%, while the minimum commission per trade is 5 RMB. All these transaction costs are considered and included in the simulation for all stock indexes.

Based on the methodology, table 2.3 shows the trading performance of the standard strategy on each Chinese stock index by using predicted stock returns from different models. Cumulative return, annualized return, alpha, and premium values are used as metrics to compare the strategy's performance. Cumulative return is the cumulative return of the strategy in realistic trading in the testing period from March 23, 2017, to April 30, 2020. It is a three years sample with 758 observations. Annualized returns are also computed by using the same sample, alpha is the alpha factor of the strategy from CAPM regression, and premium means the difference between the cumulative return and the return of benchmark stock index adjusted by stock weights in a stock index. For example, In the MLP panel of table 2.3, the SSE Composite Index has a cumulative return of 4.32%, but the premium is 18.37%, this is because the Shanghai Composite Index, the benchmark index, declined by 11.87% from 2017 to 2020. The premium here has adjusted the stock weights, so it is not simply the difference between 4.32% and benchmark return -11.87%. As a result, the premium variable directly tells us how far the strategy surpasses the benchmark index.

For example, the strategy's premium by using RNN predictions is 23.80%, which is greater than that from MLP model 18.37% but lower than that from ARIMA 30.83%. It implies that in the 3 years bear market, even though the market index has declined by 11.87%, applying the strategy with predictions from the RNN model provides a cumulative return of 9.11%, and its annualized return is 2.92%. Actually, all the three models supply positive cumulative returns, meaning that they all surpass the performance of the benchmark, the SSE Composite index.

In addition, table 2.3 also includes other stock indexes on the Shanghai Stock Exchange. In fact, it indicates that the RNN model provides the best trading performance for Industrial Index and Composite Index. MLP model works extremely well for Business Index and Property Index. As for the ARIMA model, it predicts the SSE Composite Index and the Shanghai A-share Index successfully. The empirical finding is informative, and its insight could be interpreted in different ways.

Firstly, it tells us which model should be used in predicting returns for different stock indexes. For example, if we trade the SSE Composite Index, the ARIMA model is the best in its annualized return and premium. If we trade Industrial Index, the RNN model would better then. So we could fully use the information in the table to optimize returns. Secondly, there is no unique model always superior to others. This finding indirectly supports the conclusions in past literature, that is, MLP and RNN models perform better than the ARIMA model in predicting stock returns under certain situations. This is not a general conclusion holds for all the cases. The results in table 2.1 and table 2.3 exactly showed this. Actually, all the above finding is an empirical analysis focusing on the 3 years-long bear market from 2017 to 2020, the differences in trading performance between models might change over different periods, and it also could be insignificant, so the key point here is that there is no single correct answer found in this analysis rather than the specific differences in trading performance from different models.

Finally, we see that most of the premiums and annualized returns are positive in table 2.3, which means that the application of RNN, MLP, and ARIMA models all provide us better returns than benchmark stock indexes. This finding is counterintuitive to the result in the table 2.1, because most of the accuracies are around 0.5 or less, but the strategy returns overpassed the return of benchmark index at last. This could be interpreted from different perspectives. First, the strategy in the analysis is an aggressive one, use up all the available capital to buy one stock index when the market opens if I predict it would rise today, and clear my position to get the cash back when the market closes, so it plays a role of maximum possible return I could make from realistic trading. Therefore it is natural to be larger than the return from the benchmark index. Second, the strategy used in the analysis only buys stock index if predicting a rise and does not trade if predicting a fall. Therefore it could avoid loss even though it mispredicts future stock index. For example, if tomorrow's stock will rise, but I predict it would fall, I would not trade in this case, so this wrong prediction does not lead to a real loss to me. Third, the final loss or profit heavily depends on the magnitude of price changes. Specifically, if there is a big jump in stock returns, as long as I correctly predict its direction this time, it could potentially cover the loss from many wrong predictions before. Such randomness makes it possible for strategy return to be higher than the benchmark return.

In addition, I present the trading statistics of the standard strategy in table 2.4. The table shows the trading statistics of the standard strategy, including "Sharpe Ratio," "Profit Trade %," and "Profit/Loss Ratio." Here "Profit Trade %" is defined as the total number of profit trades divided by the total number of trades. "Profit/Loss Ratio" is defined as the total profit divided by the total loss. It turns out that most of the Sharpe ratios are negative, most of the profit trade percentages are greater than 0.5, and almost all the profit/loss ratios are greater than 1. This finding indicates that the predictions from these models could predict the right direction of the stock movement in around 50% of trades. The profit is usually greater than the loss. This implies that the predictions, on average, provide large profits in large trades and small losses in small trades. Specifically, the predictions from the MLP model provide good trading statistics when they are used for trading the business index and the property index, while the predictions from the RNN model work well in the trading of the industrial index and the composite index. The finding is in line with that in table 2.3.

I also conduct a one-way ANOVA test to check the trading performance of these models formally. The results are shown in the table 2.5. It turns out that variable "Annualized return" and "Alpha" from table 2.3 passed Bartlett's equal variance test, the P-values of their Chi2 statistics are all greater than 0.05. In addition, the F-statistics in the ANOVA test is not significant at 5% significance level, which means that there is no significant difference in annualized return or alpha between these models. This finding supports our previous observation that MLP or RNN models failed to provide a better forecast accuracy or lower MSE than the ARIMA model.

Besides, I extend the analysis to strategy evaluation. Actually, the standard strategy I introduced in the section is very aggressive. It would use 100% the available cash if it believes that stock price would rise in the future, but this percentage could be changed to different values. Actually, I generate three different strategies by adjusting the percentage parameter. Strategy 1 is the standard trading strategy using 100% available cash in daily trading. Strategy 2 is a standard strategy using 80% available cash in daily trading. Strategy 3 is a standard strategy using 50% available cash in daily trading. So strategy 3 is the most conservative one out of these three strategies. I evaluate and compare these strategies based on predictions from the MLP, RNN, and ARIMA models. Here I use the "SSE composite Index" as a benchmark index to show the strategies' performance. The result is shown in table 2.6.

The table indicates that a more aggressive strategy leads to a higher return, no matter which model is used in the prediction, but this also leads to a higher maximum drawdown. Conservative strategy usually provides a higher profit/loss ratio, which is also in line with our expectations. Also, the ARIMA model performs better than the other two models in terms of the trading statistics. Given the predictions from the models, the above analysis could be extended to compare different trading strategies.

In fact, the key finding of the empirical analysis is not a comparison across models. The most interesting finding is that most of the premiums are higher than 0. Recall that the sample period from 2017 to 2020 is a very long bear market. Almost all the stock indexes in table 2.3 declined sharply in this period. Under this circumstance, it isn't easy to get positive returns if we trade these stock indexes. The empirical finding in this section provides an insight into the potential of these neural network models. This experience could be generalized to ETF fund stocks in real trading, which would make the simulated returns come true in a realistic market. This finding is one of the key findings in the paper.

2.6 Robustness

As discussed in the last section, I conduct a robustness test from two different aspects. On the one hand, I increase the number of control variables in the benchmark model. Specifically, I introduce more lags of the stock price in the prediction of future stock returns. This is because I assumed that the stock return is associated with its past returns in the most recent 5 days, which may not always hold in a realistic stock market. Therefore adding longer historical information could provide us a deeper understanding of the benchmark result. On the other hand, all the stock indexes in benchmark analysis come from SSE, which may bias the finding in section 2.5. To make the finding robust, I also proceed with the analysis for 7 stock indexes from SZSE. More details are discussed below.

2.6.1 Forecast Evaluation For More Inputs

Table 2.7 shows the mean squared error and accuracy of MLP and RNN models across different lag terms. There are four lag terms included in the table, and the maximum lag is 30 trading days, which is already longer than 1 month since every month generally has 22 trading days. The underlying consideration is that stock return usually is affected by the most recent history. It does not have a memory longer than 1 month. The table offers a better insight into MLP and RNN models. It turns out that the MSE and predictive accuracy from the MLP model is not improved if we add more lag stock returns as input, but the result of the RNN model is different. Both MSE and accuracy are improved with more inputs. Specifically, MSE drops from 1.4×10^{-4} to 1.25×10^{-4}, and its predictive accuracy rises from 49.34% to 53.03%. This is not surprising because the RNN model introduced the autocorrelation of states across different layers; therefore, more historical information is informative and helpful for predicting future stock return. Nevertheless, all the accuracies of the RNN model are still around 50%, and the pattern we found in the benchmark model has been verified again in the table. Only using historical information of stock price is not sufficient for return prediction. Namely, according to the evidence from

Table 2.3: Trading Performance of the ARIMA, MLP, and RNN models

Stock Index	MLP			
	Cum. Return	Ann. Return	Alpha	Premium
SSE Composite	4.32%	1.41%	-0.003	18.37%
Shanghai A-share	-7.17%	-2.43%	-0.038	5.23%
Shanghai B-share	-21.21%	-7.56%	-0.051	27.92%
Industrial	1.57%	0.51%	-0.012	14.29%
Business	41.78%	12.20%	0.126	119.19%
Property	30.81%	9.26%	0.068	35.99%
Public	-6.02%	-2.03%	-0.022	42.14%
Composite	-0.07%	-0.02%	-0.024	2.99%
Shanghai 180	12.00%	3.81%	-0.001	0.42%
Fund	3.90%	1.27%	-0.025	-3.50%
	RNN			
	Cum. Return	Ann. Return	Alpha	Premium
SSE Composite	9.11%	2.92%	0.005	23.80%
Shanghai A-share	-1.20%	-0.40%	-0.02	12.01%
Shanghai B-share	-10.26%	-3.51%	-0.022	45.68%
Industrial	18.84%	5.86%	0.038	33.72%
Business	-9.79%	-3.34%	-0.044	39.46%
Property	12.35%	3.91%	0.012	16.80%
Public	0.87%	0.29%	-0.001	52.56%
Composite	24.97%	7.63%	0.053	28.81%
Shanghai 180	-2.72%	-0.91%	-0.048	-12.78%
Fund	-6.32%	-2.13%	-0.057	-12.99%
	ARIMA			
	Cum. Return	Ann. Return	Alpha	Premium
SSE Composite	15.31%	4.81%	0.033	30.83%
Shanghai A-share	12.32%	3.91%	0.023	27.33%
Shanghai B-share	-17.28%	-6.06%	-0.049	34.30%
Industrial	11.22%	3.57%	0.021	25.15%
Business	13.97%	4.41%	0.062	76.19%
Property	-13.23%	-4.57%	-0.07	-9.78%
Public	4.11%	1.34%	0.015	57.47%
Composite	15.29%	4.80%	0.027	18.83%
Shanghai 180	19.50%	6.05%	0.021	7.15%
Fund	-14.66%	-5.10%	-0.085	-20.74%

The table shows a standard strategy's trading performance on each stock index using the predicted stock returns from MLP, RNN, and ARIMA models. These models all provide better returns than benchmark stock indexes. This finding offers investors an insight into the potential of these neural network models.

Table 2.4: Trading Statistics of the MLP, RNN and ARIMA models

Stock Index	MLP		
	Sharpe Ratio	Profit Trade %	Profit/Loss Ratio
SSE Composite	-0.30	0.57	1.25
Shanghai A-share	-0.69	0.56	1.14
Shanghai B-share	-1.23	0.53	0.97
Industrial	-0.37	0.55	1.20
Business	0.76	0.57	1.56
Property	0.45	0.52	1.35
Public	-0.78	0.50	1.10
Composite	-0.44	0.52	1.20
Shanghai 180	-0.03	0.54	1.33
Fund	-0.76	0.54	1.51
	RNN		
	Sharpe Ratio	Profit Trade %	Profit/Loss Ratio
SSE Composite	-0.15	0.57	1.37
Shanghai A-share	-0.49	0.55	1.20
Shanghai B-share	-0.83	0.54	1.07
Industrial	0.21	0.56	1.41
Business	-0.89	0.51	1.01
Property	-0.01	0.52	1.26
Public	-0.48	0.49	1.19
Composite	0.37	0.56	1.42
Shanghai 180	-0.62	0.55	1.17
Fund	-1.36	0.54	1.20
	ARIMA		
	Sharpe Ratio	Profit Trade %	Profit/Loss Ratio
SSE Composite	0.09	0.56	1.36
Shanghai A-share	-0.01	0.56	1.34
Shanghai B-share	-1.18	0.52	1.01
Industrial	-0.04	0.55	1.28
Business	0.03	0.52	1.25
Property	-0.79	0.48	1.04
Public	-0.31	0.53	1.23
Composite	0.08	0.53	1.31
Shanghai 180	0.26	0.56	1.38
Fund	-1.69	0.52	1.09

The table shows a standard strategy's trading statistics on each stock index using the predicted stock returns from MLP, RNN, and ARIMA models. Most of the Sharpe Ratios are negative, and almost all the profit/loss ratios are greater than 1.

Table 2.5: One-way Anova Test

variable	One-way Anova $F - stat/p$	Bartlett's Test $Chi2 - stat/p$
Ann. Return	0.04	1.28
	(0.96)	(0.53)
Alpha	0.13	1.32
	(0.88)	(0.52)

The table shows the One-way ANOVA test of the annualized return and alpha from ARIMA, MLP, and RNN models. It turns out that Bartlett's equal variance test is passed, and there is no significant difference in annualized return or alpha across these models.

China's stock market, there is no space for excess profit for investors focusing on technical analysis.

2.6.2 Forecast Evaluation For SZSE Data

In this section, I used 7 different representative stock indexes from the Shenzhen Stock Exchange(SZSE) to test the sensitivity of the finding in section 2.5 to stock indexes. The dataset comes from the Tushare finance database too. The seven stock indexes from SZSE are: "CSI 300 Stock Index", "SZSE Component Index", "SZSE Component Total Return Index", "SZSC Index", "Shenzhen A-share Index", "Shenzhen B-share Index" and "GZ 300 Index". To make the analysis comparable, the sample period is still from April 18, 2005, to April 30, 2020. All the models have the same architecture as the ones in benchmark analysis.

The result is shown in table 2.8, it indicates that the pattern found in section 3.4 is not sensitive to the sample collected from SZSE. Specifically, the mean squared error of the ARIMA model is better but close to that of MLP and RNN models. Also, the RNN model performs better in terms of predictive accuracy for the CSI 300 Stock Index and SZSE Component Index. However, its accuracy is still not significantly different from 50%, the probability of tossing a fair coin. Further, there is no unique dominating model in predicting

Table 2.6: Comparison Between Different Strategies

Strategy	Sharpe Ratio	Ann. Ret.	Alpha	Premium	P/L Ratio	Max Drawdown
	MLP					
Strategy 1	-0.297	1.41%	-0.003	18.37%	1.253	8.08%
Strategy 2	-0.401	1.44%	-0.009	18.51%	1.273	5.56%
Strategy 3	-1.054	0.79%	-0.024	16.21%	1.287	2.87%
	RNN					
	Sharpe Ratio	Ann. Ret.	Alpha	Premium	P/L Ratio	Max Drawdown
Strategy 1	-0.150	2.92%	0.005	23.80%	1.373	10.12%
Strategy 2	-0.446	1.63%	-0.012	19.18%	1.337	7.97%
Strategy 3	-1.221	1.05%	-0.024	17.13%	1.395	3.18%
	ARIMA					
	Sharpe Ratio	Ann. Ret.	Alpha	Premium	P/L Ratio	Max Drawdown
Strategy 1	0.087	4.81%	0.033	30.83%	1.361	12.99%
Strategy 2	-0.009	3.93%	0.018	27.54%	1.372	9.56%
Strategy 3	-0.565	2.15%	-0.010	21.01%	1.415	4.67%

The table shows a comparison between three standard trading strategies. The trading results are computed by using the "SSE Composite Index." Strategy 1: using 100% available cash in daily trading, strategy 2: using 80% available cash in daily trading, strategy 3: using 50% available cash in daily trading. The table indicates that a more aggressive strategy will lead to a higher return, no matter which model is used in the prediction. Also, the ARIMA model provides the highest return out of the three models.

returns for the stock indexes listed in the table. In summary, the pattern in benchmark analysis still holds for the stock indexes from SZSE. This verified that improvement in the model's architecture is insufficient to offer a better prediction for future stock returns if we only use historical price information.

2.7 Conclusion

In this paper, I studied the MLP and RNN models and compared them with the traditional ARIMA model according to their performances in predicting daily stock returns. I used autoregressive terms of stock return as inputs in the empirical analysis

Table 2.7: Forecast of MLP and RNN Model With More Inputs

Model	Metrics	Input Variable Lags 5	10	20	30
MLP	MSE(10^{-4})	1.29	1.36	1.49	1.46
	Accuracy(%)	48.42	49.34	47.49	49.87
RNN	MSE(10^{-4})	1.4	1.34	1.31	1.25
	Accuracy(%)	49.34	50.66	51.45	53.03

The table shows the mean squared error and accuracy of MLP and RNN models across different lag terms. The MSE and predictive accuracy from the MLP model are not improved if we add more lag stock returns as input, but the RNN model's result is different. Both MSE and accuracy are improved with more inputs.

to study the value-added of switching linear model(ARIMA) to nonlinear neural network model(MLP, RNN). The empirical finding indicates that MLP and RNN models are not as good as the ARIMA model in terms of mean squared error, and all three models have predictive accuracy of around 50%. This means that only switching from linear model to neural network model(ARIMA to MLP) or further adding realistic autocorrelated structure within neural network model(MLP to RNN) are not sufficient for us to predict future stock return precisely. This is also verified in robustness tests. In fact, this finding is consistent with the weak form efficiency of the efficient market hypothesis. Using historical information alone could not provide a better stock trading return, but the empirical result is informative from different perspectives.

Firstly, it tells us which model should be used in predicting returns for different stock indexes. ARIMA model works well in predicting the SSE Composite Index, while the RNN model is better than MLP and ARIMA if we trade Industrial Index. This information is helpful from the investor's point of view. Secondly, I verify there is no unique winner

Table 2.8: Forecast Evaluation For SZSE Data

	Mean Squared Error(10-4)			Accuracy(%)		
Stock Index	MLP	RNN	ARIMA	MLP	RNN	ARIMA
CSI 300 Stock	1.64	1.66	1.55	47.89	51.32	51.05
SZSE Component	2.23	2.2	1.24	48.15	50.53	50.13
SZSE Comp. Tot. Ret.	2.23	2.24	2.1	49.07	48.15	49.60
SZSC	2.26	2.17	2.05	50.66	50.66	50.65
Shenzhen A-share	2.21	2.17	2.06	52.37	51.06	50.26
Shenzhen B-share	1.14	1.14	1.09	49.60	50.92	51.32
GZ 300	1.63	1.67	1.57	45.25	51.32	48.28

The table shows the MSE and accuracy of the predictions from these models by using the SZSE data. The pattern found in SSE data still holds in the SZSE data. RNN model performs better in terms of predictive accuracy for the CSI 300 Stock Index and SZSE Component Index. However, its accuracy is still not significantly different from 50%. There is no unique dominating model in the prediction for the stock indexes listed.

among these three models according to the paper's empirical finding, which is in line with literature like Dunis& Williams(2002). Finally, almost all the strategy returns are greater than the return of benchmark stock indexes, which provides investors a better choice of trading stock indexes and make positive returns even in the long bear market period.

2.8 References

- Cao, Q., Leggio, K.B. & Schniederjans, M.J. (2005), A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market", Computers & Operations Research, 32(10), 2499-2512.

- Dai, W., Wu, JY. & Lu, CJ. (2012), Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes, Expert

Systems with Applications, 39(4), 4444-4452.

- Dunis, C.L., Laws, J. & Evans, B. (2006), Trading Futures Spreads: An application of Correlation and Threshold Filters, Applied Financial Economics, 16(12), 903-914.
- Dunis, C.L., Laws, J. & Sermpinis, G. (2010), Modelling and trading the EUR/USD exchange rate at the ECB fixing, European Journal of Finance, 16(6), 541-560.
- Dunis, C.L. & Williams, M. (2002), Modelling and Trading the EUR/USD Exchange Rate: Do Neural Network Models Perform Better? Derivatives Use, Trading and Regulation, 8(3), 211-239.
- Fama, E.F.(1970), Efficient capital markets: a review of theory and empirical work, Journal of Finance, 25(2), 383-417.
- Hochreiter, S. & Schmidhuber, J.(1997), Long Short-Term Memory, Neural Computation, 9(8), 1735-1780
- Kamijo, K. & Tanigawa, T. (1990), Stock Price Pattern Recognition: A Recurrent Neural Network Approach. 1990 IJCNN International Joint Conference on Neural Networks
- Kingma, D.P. & Ba, Jimmy (2014), Adam: A Method for Stochastic Optimization.
- Lisboa, P. J. G. & Vellido, A. (2000), Business Applications of Neural Networks: The State-of-the-Art of Real-World Applications, World Scientific,Singapore.
- Tenti, P. (1996), Forecasting Foreign Exchange Rates Using Recurrent Neural Networks, Applied Artificial Intelligence, 10, 567-581.
- Tino, P., Schittenkopf, C. & Doffner, G. (2001), Financial Volatility Trading Using Recurrent Networks, IEEE Transactions in Neural Networks, 12(4), 865-874.

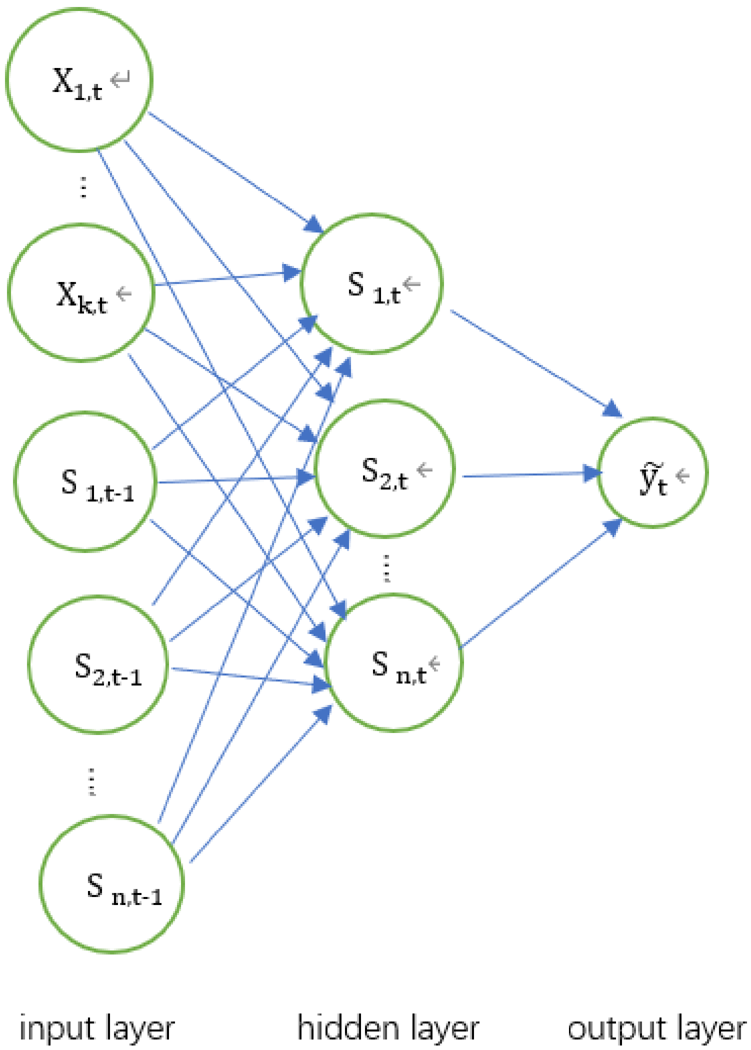


Figure 2.3: Recurrent Neural Network

The figure shows the architecture of the Recurrent Neural Network(RNN) model. It includes an input layer, a hidden layer, and an output layer. The model includes the past states of neurons as input in the fitting so that it has a memory feature, which is designed to deal with time-series data.

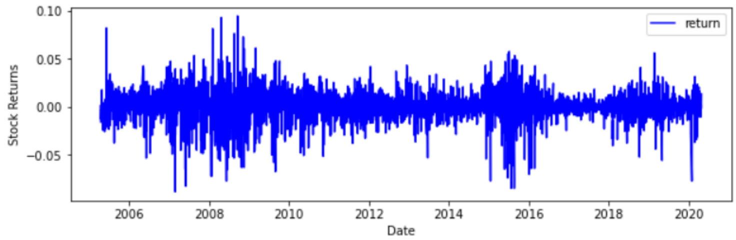


Figure 2.4: Daily return of the SSE Composite Index

The figure shows the sequence of daily returns of the SSE Composite Index. The sequence is stationary in the sample period and has a feature of clustering volatility.

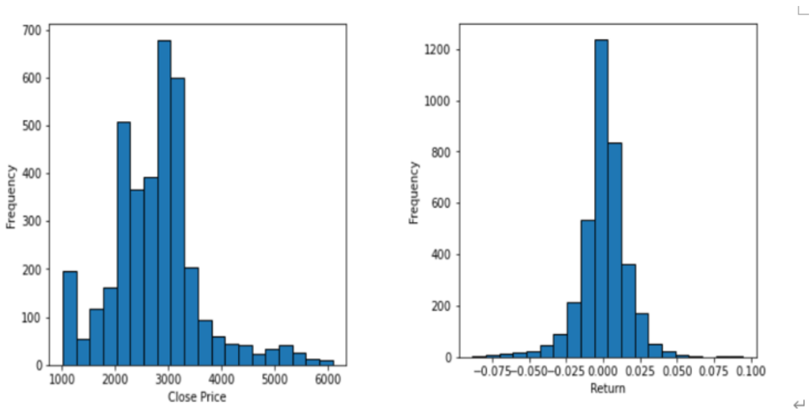


Figure 2.5: Histogram of Close Price & Return

The figure shows the histogram of close price and returns of the SSE Composite Index. The return variable has a normal distribution compared with the close price. Therefore we use return rather than close price as an input variable in the analysis.

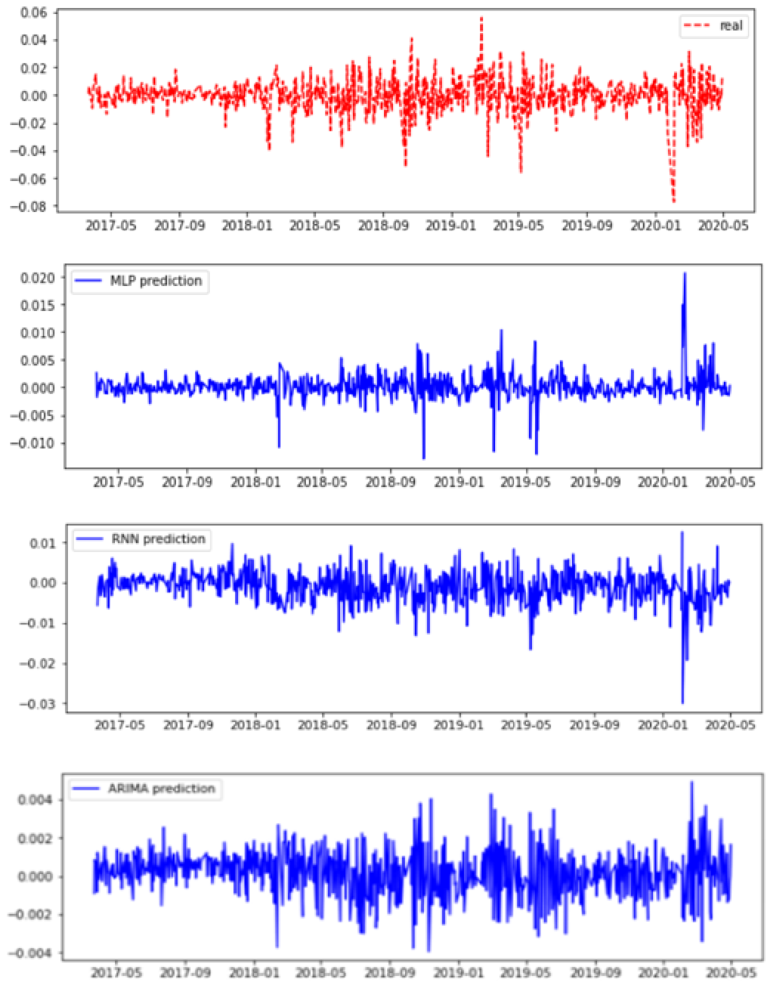


Figure 2.6: The Predictions for SSE Composite Index

The figure shows the difference between the predictions from ARIMA, MLP, and RNN models and compares them with the realized stock return in the test period. All the predictions have less volatility than realized returns. These models captured the clustering feature in realized stock return, but all these 3 models failed to give good predictions in outliers.

Chapter 3

Distribution of Contracts over Strikes in Options Market

3.1 Introduction

Options are attractive to investors due to their high leverage, special risk exposure, and low transaction costs. They have the highest leverage among all financial assets because the money invested is a premium of potential price movements instead of prices of the underlying assets. They enable investors to limit losses and bet on volatility changes instead of price fluctuations. The richness in the variety of underlying assets and the availability of a good range of strike prices around the spot price sharply reduces the transaction costs.

Both investors and researchers have noticed these unique and desirable characteristics in options, so we see high trading volumes and active participants in the options market, and plenty of research from academia. Most of the literature on options focused on their pricing, like Merton (1973) and the relationship between the options market and the underlying stock market like Jones (2003). Few of them looked at trading volume, open interest, or the aggregate trading pattern of investors as questions of interest. At the same time, almost no literature empirically studied the choice of options from the investors' point of view. Previous literature generally investigated how to allocate capital across different assets in one's portfolio. They could not answer what specific options to purchase and hold when there are hundreds of them at tens of different strike prices in the market. We thus attempt to fill in the gap in the literature. This paper is trying to depict the behavioral patterns of the investors in real options trading. Specifically, our goal is to capture the features of the contracts that are traded and held most frequently.

We observe that slightly-out-of-the-money options are the most popular in terms of having the highest daily trading volume and open interest shares in our dataset of S&P 500 options trading. We use trading volume and open interest as weights and calculate the weighted average strike prices of put and call options so that we know the position of the distribution of trading volume and open interest over strike. The weighted average strike prices closely follow the underlying index. Options are also slightly out of the money on the aggregate level. The weighted average call strike is almost always above the underlying index. In contrast, the weighted average put strike is almost always below the underlying index.

We start our empirical study by first looking at how the weighted average strike prices co-move with the underlying S&P 500 index. We use a structural vector auto-regression (SVAR) model to study the short-run association between option strikes and the underlying index. We assume the same-day impact between the two is unidirectional, where the underlying index affects the option strikes but not vice versa. By looking at the coefficient estimates and the impulse response functions, we find that option strikes move slowly in the same direction as the underlying index. The initial reactions are small and become larger as time passes. Option trading behavior is more sensitive to broad-market shocks than option writing and settling.

We then estimate a vector error-correction model (VECM) to test the relationship between the underlying index and the weighted average strike in the long run. We find evidence for cointegration and an almost one-to-one mapping between the changes in both variables.

We continue to ask whether the position of the weighted average strike relative to the underlying index is related to the implied volatility of the latter. We regress the relative position of the weighted average strike on VIX. We calculate the standard deviation of the daily distribution of open interest and trading volume and regress them on VIX. We find that when the implied volatility is higher, option strikes move further from the underlying index both on the aggregate level and on the contract level.

After that, we try to rationalize the preference for slightly-out-of-the-money options. We do that by simulating the return series of three trading strategies, naked option, full hedge, and delta hedge, with options at different strikes. We find that the return series of the two hedging strategies is closer to what we observe in the data. The naked option strategy provides insight into how gambling motivation can also be the reason for choosing out-of-the-money options.

Lastly, we investigate gambling motivation and look for options that are lottery-like assets. We find that out-of-the-money options have low premiums, high return volatility, and skewness, so they make perfect speculation tools.

The remainder of the paper is organized as follows. We first review the existing literature on option trading in section 3.2. We then introduce the dataset of S&P 500 option transactions and the stylized facts in section 3.3. Our empirical work are in section 3.4, with section 3.4.1 talking about the SVAR model, section 3.4.2 on the VECM, section 3.4.3 on the relative position and implied volatility, section 3.4.4 on the simulation of the three strategies, and 3.4.5 on the gambling motivation. Section 3.5 concludes the paper.

3.2 Literature Review

As one of the most active financial derivatives markets, the options market has always been an important subject for researchers to study. Trading volume as an important indicator in the market is the focus of many scholars. Actually, numerous papers studied trading volumes from different angles.

One branch of literature studied the effect of the trading volume of options on the underlying stock market. Anthony (1988) studied the relation between stock trading volume and call option trading volume empirically and found that options volume Granger causes trading in underlying stock with a one-day lag and options volume is a good proxy for the rate of information arrival, but the paper failed to rule out the possibility of technical hedging, investors could trade options and underlying stock sequentially for hedging motivation. The flaw inspired us that importance of the hedgers could never be neglected in studying trading volume in the options market. Jennings and Starks (1986) studied the options market's effect from a price adjustment point of view. They investigated the effect of options trading on the underlying stock price adjustment to the release of earnings announcements for both firms with or without listed options. They found that non-option firms need a much longer time to adjust the stock price after claiming announcements. The paper provides us more insights into channels offered by the options market and highlighted the relation between stock price adjustment and option trading volumes, which strengthens the importance of dynamic motion of stock price and options trading in both the short run and the long run, actually, the application of SVAR and the VECM methods in our paper is exactly following this idea.

In addition, some other literature focused on the prediction power of information in the options market. Bernile et al. (2017) constructed an index called volume-weighted strike-spot price ratio (VWKS) to predict equity returns. They found the VWKS index has a better performance than existing return predictors. The good performance of VWKS in backtesting inspired us, and we referenced their method of constructing VWKS in our empirical analysis of option trading volumes. Nevertheless, there is hardly any literature using trading volume in the options market directly as a question of interest until Roll et al. (2010). They first studied the relative trading volume, the ratio between option volume and stock volume, and explored the determinants of the relative trading volume empirically. Their finding indicates that their panel regression model could explain almost half of the variation in options trading volume, which offers us many insights into the determinants of options trading volume.

Moreover, another branch of literature explored the essence and motivation of options trading. They are interested in investigating the speculation in options trading, particularly, a lot of literature showed that gambling trading is one of the most popular motivations for option trading behaviors. Bauer et al. (2009) found that most investors experienced larger losses in their option investment than the losses they experienced in equity investment. The authors verified that it is due to the bad timing of options trading for investors. The paper also ruled out hedging as a key motivation for the trading behaviors in the options market. Rather, it claims that gambling motivation is the dominant one in explaining the market. This paper broadens our horizons in options trading and introduced gambling motivation into our vision. We empirically test the gambling motivation in the last section of the paper following this literature. Meanwhile, some other literature further investigated the relationship between gambling preference and option trading behaviors. Byun and Kim (2016) studied the topic and found that options with higher lottery potential are overpriced and have lower returns. Blau et al.(2016) found that the ratio of call option volume over total options volume strongly associates with investors' preference for stocks with lottery potentials. They also claimed that such preference has a significant effect on the volatility of the underlying asset. Kumar (2009) even listed 3 key features to select lottery-like assets: low asset price, high volatility, and high skewness of return. Our paper referenced this literature and used their selection methods in our empirical work.

Even though all the above papers provide us many thorough insights about trading behaviors in the options market, none of them answered the question we mentioned above that what kind of options contracts are preferred and traded frequently by investors, and how investors would choose their most profitable strike prices in options market when there are hundreds of contracts available to them. Therefore we use this paper to fill in the gap and answer the question empirically.

3.3 Data and Stylized Facts

3.3.1 Data Description

The goal is to study the distribution of trading volumes in options markets. The desired dataset should contain contract-level data, where we can see daily fluctuations in prices and trading volumes of options with different strike prices and maturities. To see such fluctuations, we want the market to be very active. That is to say; we want sufficient participants in the market and high daily trading volumes. At the same time, we want to study the behavior of a representative investor in a broad market rather than in a market of some specific security or commodity. Thus this paper uses S&P 500 options data from the Commodity Research Bureau (CRB), which satisfies all our needs mentioned above. The

CRB collects the data from the Chicago Board Options Exchange (CBOE). In this dataset, we have more than 20 years of daily data of S&P 500 options, from January 2, 1990, to June 6, 2012. The dataset includes the open price, close price, last bid, last ask of all S&P 500 options with different strike prices and maturities. Daily trading volume and open interest are also available for both put and call contracts in the dataset. The entire CRB dataset contains a wide variety of classes of options, such as SPX, SPXW, LEAPS, XSP, etc. [1][2] To make our analysis more representative and make our models easier to build, we follow Metaxoglou and Smith (2011), only use standardized contracts and drop classes SXZ, SPB, LSW, LSX, LSY, LSZ, XSC, XSB, XSK, XSL, XSO, and XSP.

The *strike price* of an option is the price fixed by the contract at which the option can be exercised. A holder of a call option can buy the underlying asset from the opponent at the strike price. A holder of a put option can sell the underlying asset to the opponent at the strike price, both before the contract expires on the maturity date.

Options are categorized into three groups using the strike price, in the money, at the money, and out of the money. In-the-money options are those that have positive payoffs if exercised immediately. Call options with strikes lower than the current underlying asset price and put options with strikes higher than the current underlying asset price are in the money. The immediate payoff from exercising an in-the-money option is called the intrinsic value of the option. Out-of-the-money options, on the other hand, are those that have negative payoffs if exercised immediately, and as a result, they will not be exercised immediately. Call options with strikes higher than the current underlying asset price and put options with strikes lower than the current underlying asset price are out of the money. An out-of-the-money option has 0 intrinsic value. The remaining options are at the money. They are options with strikes equal or very close to the current underlying asset price.

When an investor buys an option, he is said to be in a long position of that option, and the seller of this option, called the writer, is said to be in a short position. An

[1]SPX is the flagship contract of CBOE's suite of S&P 500 index products. SPXW are the contracts with weekly expiration dates. LEAPS is short for Long-term Equity AnticiPation Securities, and LEAPS options have expiration dates up to three years into the future. XSP contracts are 1/10 the size of the standard SPX contract.

[2]For more details, see http://www.cboe.com/products/stock-index-options-spx-rut-msci-ftse/s-p-500-index-options.

investor's long positions and short positions of the same option cancel out each other, and those positions are said to be closed. Positions are also closed when options are exercised. American options can be exercised before the maturity date, and European options can only be exercised on the maturity date. All options in our dataset are European. The positions that are not closed in one of the two ways described above are said to be outstanding or open. *Open interest* is the total number of outstanding options at the end of a trading day in our dataset. *Volume* is the total number of contracts traded on a trading day. *Last price* is, as the name suggests, the price at which the last option is traded on a trading day. The bid price is the highest price in the order book at which someone is willing to buy an option, and the ask price is the lowest price in the order book at which someone is willing to sell an option. *Last bid* and *last ask* are, again as their names suggest, the last bid price and the last ask price in the order book on a trading day that cannot find a counterpart to lead to a trade. Therefore typically on any trading day, the last bid is lower than or equal to the last price, and the last ask is higher than or equal to the last price. *Return* is the realized return of a contract, calculated by the payoff divided by the last price minus 1. If an option is exercised, its payoff is the absolute difference between its strike price and the underlying index on the maturity day. If an option is not exercised, its payoff is 0, and thus its return is -1.

In our dataset, each observation identifies a call or put option on a trading day, which matures on a specific future date with a specific strike. More than half of the observations are considered inactive options in that they have 0 trading volume or a 0 last price on that day. After dropping those inactive options, the total number of observations is 413390 for call options and 614402 for put options. The summary statistics of the dataset are given in table 3.1. It should be noticed that our dataset contains only European options, so we are narrowing down our question of interest to the distribution of trading volumes in European options markets, and later in this paper, our theoretical model is based on European option assumptions.

The mean and median underlying S&P 500 index are 1041.37 and 1116.21 for call options and 1051.76 and 1125.07 for put options. The fact that the S&P 500 index has different means and standard deviations for the two types is due to the nature of this

Table 3.1: Summary statistics of S&P 500 options

		Count	Mean	Median	S.D.	Min	Max
Call Option	S&P 500	431679	1041.374	1116.21	304.6558	295.46	1565.15
	Strike Price	431679	1068.598	1145	333.963	100	3000
	Open Interest	431679	13781.6	5893	23323.6	0	295814
	Volume	431679	1098.768	110	3269.651	1	158635
	Last Price	431679	41.23456	17	74.72939	.01	1294.4
	Last Bid	431679	40.62989	16.38	74.67551	0	1286.4
	Last Ask	431679	42.13959	17.9	75.2224	0	1290.2
	Return	431679	-.0461376	-1	18.79643	-1	7021.8
Put Option	S&P 500	635336	1051.763	1125.07	309.9013	295.46	1565.15
	Strike Price	635336	957.177	1005	313.3035	50	3000
	Open Interest	635336	17047.02	6989	27729.94	0	370769
	Volume	635336	1259.152	125	3772.888	1	200777
	Last Price	635336	28.73371	10	64.61985	.01	7617
	Last Bid	635336	28.00411	9.25	63.66305	0	1890.6
	Last Ask	635336	29.34724	10.25	64.28291	0	1894.5
	Return	635336	-.4963176	-1	5.674132	-1	2746.467

The table shows the summary statistics of S&P 500 European options.

contract-level dataset. The mean is not the true historical daily average of the index but the average of the underlying index of all historical contracts for both put and call options. The fact that both mean and median S&P 500 are higher for put options is suggesting that investors might prefer trading call options, the right to buy, when the underlying index is lower, and they might prefer trading put options, the right to sell, when the underlying index is higher. The mean and median open interest of the put options are 17047.02 and 6989, which are much larger than those of the call options of 13781.60 and 5893. The mean and median daily trading volumes of the put options are 1259.15 and 125, which are also larger than those of the call options of 1098.77 and 110. From comparing the mean and median open interest and trading volume, we can see that investors clearly prefer trading put options. This preference is likely because most investors in the options market are using options as risk management and hedging tools for their long positions in the equity market.

Table 3.2 exhibits the structure and a part of the trading data on September 18, 2006, which contains put and call options of 6 strikes (from 1300 to 1335) that mature in 33

Table 3.2: Snapshot of the dataset on Sept. 18, 2006

	Expiration Date	Strike Price	Open Interest	Volume	Last Price	S&P 500
Call	21-Oct-06	1300	85408	7647	34.6	1321.18
	21-Oct-06	1305	4940	9	32	1321.18
	21-Oct-06	1310	10693	51	27	1321.18
	21-Oct-06	1315	12607	116	23	1321.18
	21-Oct-06	1320	9639	2562	19.5	1321.18
	21-Oct-06	1325	60017	15327	17	1321.18
	21-Oct-06	1330	12425	5473	13.5	1321.18
	21-Oct-06	1335	20526	1887	11.2	1321.18
Put	21-Oct-06	1300	84742	15301	8	1321.18
	21-Oct-06	1305	15169	1276	8.7	1321.18
	21-Oct-06	1310	20897	8939	10	1321.18
	21-Oct-06	1315	7277	701	11.6	1321.18
	21-Oct-06	1320	9461	1015	13.2	1321.18
	21-Oct-06	1325	10028	13744	15.1	1321.18
	21-Oct-06	1330	2704	70	16.4	1321.18
	21-Oct-06	1335	1533	19	20.4	1321.18

The table shows the snapshot of the dataset on Sept. 18, 2006.

days (on October 21, 2006). The underlying S&P 500 index was 1321.18. The put option prices increased with the strike, from 11.2 to 34.6 dollars per contract. Put options are rights to sell the underlying index at the strikes. The higher the strike is, the more value this right has to the holder. Similarly, the call option prices decreased with the strike, from 8 to 20.4 dollars per contract. The highest trading volume of put options was 15301, which appeared at the strike of 1300, and that of the call options was 15327, which appeared at the strike of 1325. Both types of options saw their highest trading volumes at strikes that were slightly out of the money. At the same time, the second-highest trading volume of put options was 13744 at the strike of 1325, where we saw the highest trading volume of call options. The second-highest trading volume of call options was 7647 at the strike of 1300, where we saw the highest trading volume of put options. We observe investors preferring strikes of one type of option (call or put) where the opposite type is the most active. This is probably because option market participants use strategies such as a straddle to place volatility bets of the underlying asset.[3]

We use trading volume rather than open interest in our analysis because trading volume is a more sensitive and direct measure of investors' trading behavior in different contracts. A lot of literature, such as Roll et al. (2010) and Chan et al. (2002), have shown that trading volume does contain useful information in explaining behaviors in financial markets. Trading volume and open interest are highly correlated. We would check our results using open interest data in future analysis.

3.3.2 Stylized Facts

We observe the following stylized facts in the S&P 500 index option dataset.

1. We find that the average strike price weighted by open interest is always closely following the underlying S&P 500 index. This pattern holds for both call and put options. The average strike prices for call and put options are calculated respectively using equation 3.1.

$$\mu_{K,t} = \frac{\sum_{i=1}^{n_t} K_{it} oit_{it}}{\sum_{i=1}^{n_t} oit_{it}} \tag{3.1}$$

   $\mu_{K,t}$ is the average strike price of all call or put options traded on the day t. K_{it} is the strike price of contract i traded on day t. A contract is uniquely identified by its strike price and maturity. The corresponding open interest for that contract is oit_{it}. n_t is the total number of different call or put options traded on the day t. In addition, we also computed the average strike price weighted by trading volume and plotted their trends for both put and call options in figure 3.1.

2. The average put strike price is almost always lower than the underlying S&P 500 index, while the average call strike price is almost always higher. Investors in the

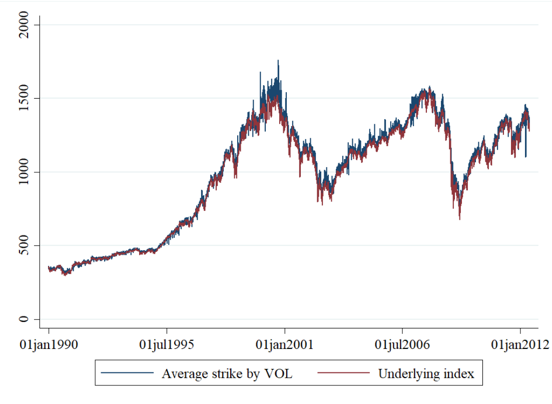
[3] A straddle is simultaneously buying a put option and a call option with the same strike and maturing on the same date. A straddle is profitable when there are big swings in the underlying asset price, regardless of the direction of the swings, so it is a volatility bet.

options market on the aggregate level prefer contracts out of the money for both put and call options. The implication is that investors probably use options as hedging tools. When investors hold short positions in the underlying index, they want to insure themselves from unexpected price increases and have the right to buy the underlying index at a lower strike price. Similarly, when investors hold long positions in the underlying index, they want to hedge against unexpected price drops and have the right to sell the underlying index at a higher strike price. This hedging incentive is consistent with our observation in table 3.1 that investors prefer put options over call options because they use options to hedge their long positions in the equity market. In addition, this finding not only holds for average strike price weighted by open interest, but also holds for average strike price weighted by trading volume. It turns out that using trading volume as weight provides us a more volatile time series plot because trading volume is more volatile than open interest. This is because open interest represents the subjective willingness of opening new positions, while trading volume represents both opening and closing positions. Investors could trade passively if they want to close their existing positions. Therefore, open interest would be a smoother and preciser metric for our target. We used open interest rather than trading volume as our benchmark through the analysis in the paper.

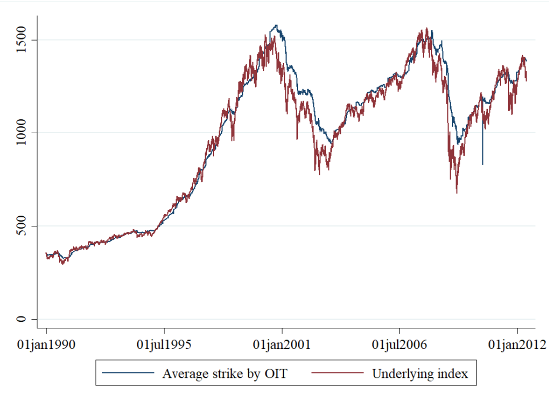
3. Slightly-out-of-the-money options are traded and held the most. We plot the time series of trading volume shares and open interest according to their moneyness on each trading day. Moneyness describes the relative position of the underlying index (S_0) against the option strike (K), and here we measure moneyness with the ratio S_0/K. For call options, we define that one option is in the near-out-of-the-money (NOTM) group if it satisfies $0.8 \leq S_0/K_{\text{call}} < 1$, and it is in the middle-out-of-the-money (MOTM) group if it satisfies $0.6 \leq S_0/K_{\text{call}} < 0.8$. We define one option is in the near-in-the-money (NITM) group if it satisfies $1 < S_0/K_{\text{call}} \leq 1.2$, and it is in middle-in-the-money (MITM) group if it satisfies $1.2 < S_0/K_{\text{call}} \leq 1.4$. Similarly, we define those groups for put options. We can easily see from figure 3.2 that on any trading day, the shares of these 4 moneyness groups sum up to almost 1, meaning that

Figure 3.1: Option average strikes and the underlying S&P 500 index

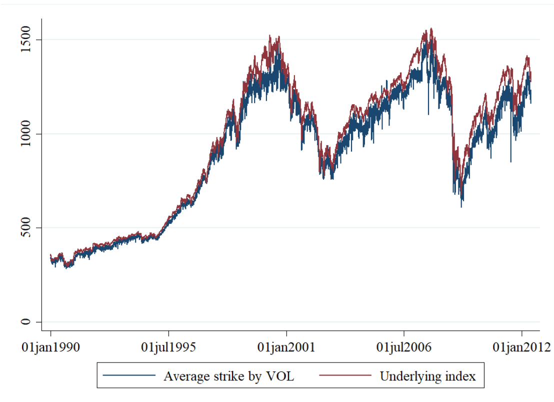
(a) Call, weighted by trading volume



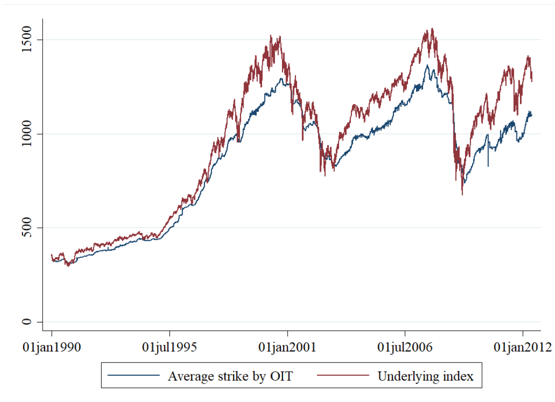
(b) Call, weighted by open interest



(c) Put, weighted by trading volume



(d) Put, weighted by open interest

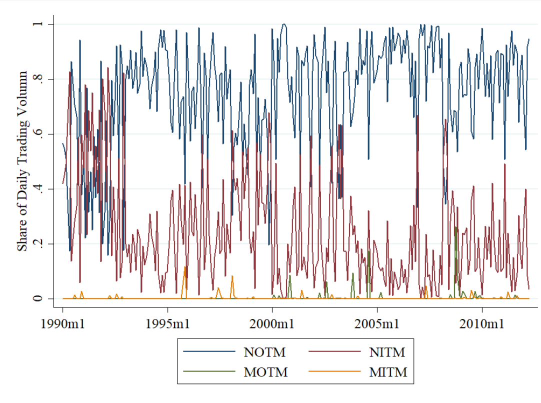


The figure shows the relative positions of the option average strikes and the underlying S&P 500 index. The average put strike price is almost always lower than the underlying S&P 500 index, while the average call strike price is almost always higher. Investors in the options market on the aggregate level prefer contracts out of the money for both put and call options.

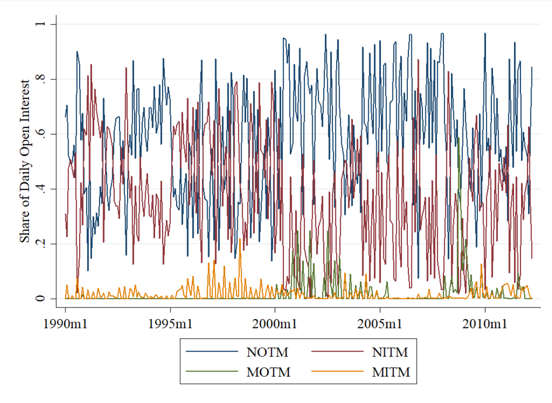
there are few active contracts that are deep in or out of the money. Deep-in-the-money and deep-out-of-the-money groups are defined by S_0/K being less than 0.6 or greater than 1.4, and they are not shown in the plot. The NOTM group almost always has the highest share by a significant margin no matter what metrics we use. Very few investors choose to trade options that are far out of or in the money, making those shares almost zero. The shares of daily open interest among four moneyness groups are closer to each other than the shares of daily trading volumes, and the pattern is robust in both call and put options. That is, investors prefer to hold and trade slightly-out-of-the-money options, both puts and calls.

Figure 3.2: Shares of options trading volume by moneyness group

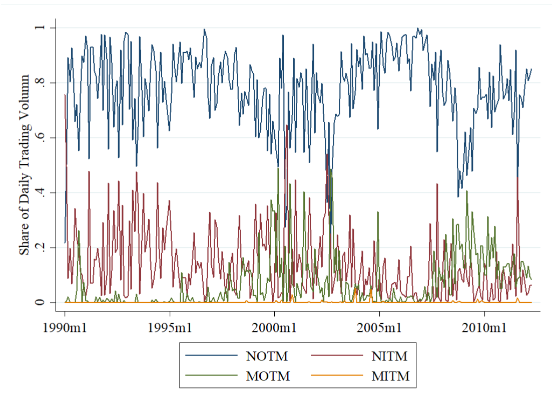
(a) Call options trading volume shares



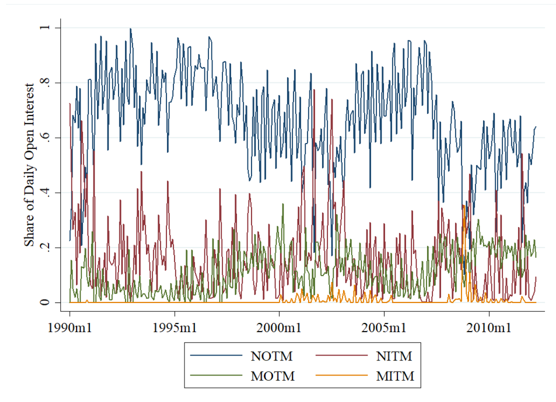
(b) Call options open interest shares



(c) Put options trading volume shares



(d) Put options open interest shares



The figure shows the shares of options trading volume by moneyness group from 1990 to 2012. The NOTM group almost always has the highest share by a significant margin, no matter what metrics are used. The shares of daily open interest among four moneyness groups are closer to each other than the shares of daily trading volumes, and the pattern is robust in both call and put options.

We thus want to study the investors' preference for slightly-out-of-the-money options empirically, and we hope to provide some realistic guidance for investors in choosing options. In section 3.4, we first look into how the distribution of option open interest moves with the underlying index using a structural vector auto-regression (SVAR) model and a vector error-correction model (VECM). Then we study the relationship between the implied volatility of the underlying index and the relative position of the weighted average strike. After that, we simulate the return series of three strategies to show that slightly-out-of-money options are the most favorable ones to investors. Lastly, we provide evidence for using gambling motivation as a candidate explanation for the preference for slightly-out-of-the-money options.

3.4 Empirical Results

3.4.1 Structural Vector Auto-regression (SVAR) Model

We start with further investigating the pattern we see in figure 3.1, and want to study how the distribution of option open interest, particularly the options' weighted average strike price is associated with the underlying index. We estimate a VAR model based on daily data for $y_t = (\ln S_t, \mu_{\ln K,t})'$. $\ln S_t$ is the natural log of the underlying S&P 500 index at time t. $\mu_{\ln K,t}$ is the open-interest-weighted average log strike price defined as

$$\mu_{\ln K,t} = \frac{\sum_{i=1}^{n_t} \ln K_{it} oit_{it}}{\sum_{i=1}^{n_t} oit_{it}}. \tag{3.2}$$

Both variables are in logs so that numbers in the entire time horizon have the same scale. We estimate the model

$$A_0 y_t = \alpha + \sum_{i=1}^{L} A_i y_{t-i} + \epsilon_t, \tag{3.3}$$

to study the short-run relationship between the two variables of interest. We restrict ourselves only to use put options that mature in one month or less because firstly, we are looking at the short-run impacts of exogenous shocks. Secondly, this restriction enables us to have data on every trading day without overlapping, which means that contracts on the same trading day must mature on the same day too. L is the maximum lag in the VAR model. The structure or the short-run constraint imposed on the model is

$$A_0 = \begin{bmatrix} 1 & 0 \\ -a_{21} & 1 \end{bmatrix}, \Omega_\epsilon = \begin{bmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{bmatrix}.$$

Essentially we are assuming that the same-day impact between the underlying index and the average log strike price is unidirectional. Movements in the underlying index affect investors' choices of put option strike prices but not vice versa. Therefore shocks can be orthogonally decomposed into two parts, one broad-market shock, and one option-specific shock. The broad-market shock may include fluctuations in the general economy and the introduction of new policies. On the other hand, option-specific shocks can be changes in investors' risk appetites and expectations on market volatility. Broad-market shocks have instantaneous effects on both the S&P 500 index and investors' choices of put options strikes, while there is at least a one-day lag in the response of the underlying index to option-specific shocks. Algebraically, the reduced-form errors e_t can be written as

$$e_t = \begin{pmatrix} e_t^{\ln S_t} \\ e_t^{\mu_{\ln K,t}} \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ -a_{21} & 1 \end{bmatrix}^{-1} \begin{pmatrix} \epsilon_t^{broad-market} \\ \epsilon_t^{option-specific} \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ a_{21} & 1 \end{bmatrix} \begin{pmatrix} \epsilon_t^{broad-market} \\ \epsilon_t^{option-specific} \end{pmatrix}. \quad (3.4)$$

Our primary interest is in a_{21}, how broad-market shocks can impact the choices of strike prices. We first set L to 30 and estimate equation 3.3, and then we pick the optimal lag as the smallest optimal lag from AIC and SBIC and re-estimate the model. Table 3.3 shows the result in the first row, where the average log strike price $\mu_{\ln K,t}$ is weighted by open interest and the endogenous variables are $\ln S_t$ and $\mu_{\ln K,t}$. A broad-market shock causing a 1% increase in the underlying index causes a 0.040% increase in the average log strike. The two endogenous variables move in the same direction, but the aggregate position in put options held by the market is quite insensitive to broad-market shocks. Contracts are traded but not easily written or settled. Positions are opened and closed on the individual level but not much on the market level.

This finding is confirmed by the rest of the table 3.3. The levels are non-stationary so we take the first difference of them, replace the right-hand side with $y = (\Delta \ln S_t, \Delta\mu_{\ln K,t})'$, and re-estimate equation 3.3. The second row in table 3.3 shows very similar results to the first row. The third and fourth rows show results when we use trading volume instead of open interest as weight in equation 3.2 to calculate the average log strike. If, instead of looking at what contracts the market holds, we look at what contracts the market trades, a_{21} becomes 10 times as large. A broad-market shock causing a 1% increase in the underlying index causes a 0.408% increase in the average log strike. Trading behavior is much more reactive to broad-market shocks than writing and settling contracts. Table 3.3 echos with figure 3.1 where we see the trading-volume-weighted average strike in the left panel follows the S&P 500 index closer than the open-interest-weighted average strike in the right panel.

Table 3.3: Impact of broad-market shocks on put option average log strike price

Weight	Endogenous variables	L	a_{21}	s.e.	95% CI	
Open interest	Level	2	0.040	0.015	(0.011,	0.070)
Open interest	First difference	2	0.044	0.015	(0.014,	0.074)
Trading volume	Level	23	0.408	0.024	(0.362,	0.454)
Trading volume	First difference	21	0.418	0.024	(0.371,	0.464)

The table shows the impact of broad-market shocks on put option average log strike price, average log strike price $\mu_{\ln K,t}$ is weighted by open interest and trading volume separately. A broad-market shock causing a 1% increase in the underlying index causes a 0.040% increase in the average log strike for the strike price weighted by open interest. The two endogenous variables move in the same direction, but the aggregate position in put options held by the market is quite insensitive to broad-market shocks. Contracts are traded but not easily written or settled. Positions are opened and closed on the individual level but not much on the market level.

Figure 3.3 exhibits the impulse response functions corresponding to the models in table 3.3. The top two plots use open-interest-weighted average log strike, while the bottom two use trading-volume-weighted average log strike. The left two plots are the level models, while the right two are the first-difference models. Each of the plots is in the same order as A_0^{-1}. We confirm our findings above with the plots and further notice that the impulse responses in the left panel have long terms. Moreover, we see in the left two plots that the response of the average log strike on the impulse of the underlying index (in the bottom left) increases with time. Investors in the options market adjust their positions slowly and gradually with the underlying index, and the market adjusts its aggregate position slower.

We provide similar but weaker evidence using call options data in table 3.4 and figure 3.4 below.

Table 3.4: Impact of broad-market shocks on call option average log strike price

Weight	Endogenous variables	L	a_{21}	s.e.	95% CI	
Open interest	Level	3	0.010	0.016	(-0.021,	0.040)
Open interest	First difference	2	0.009	0.016	(-0.023,	0.040)
Trading volume	Level	5	0.349	0.022	(0.306,	0.392)
Trading volume	First difference	18	0.354	0.022	(0.310,	0.398)

The table shows the impact of broad-market shocks on call option average log strike price. Similar but weaker evidence has been found in the case of call options.

3.4.2 Cointegration and Vector Error-Correction Model (VECM)

We have seen the short-run association between the S&P 500 index and the average strike price in section 3.4.1. It is necessary for us to continue to test the relationship between those two variables in the long run. Given the non-stationarity shown in the plots in section 3.3.2, the cointegration test and the corresponding VECM are the best and most straight forward approaches to serve the purpose. Therefore we organize the section into three parts. First, we use the Johansen Test to test the existence and order of the cointegration relationship between $\mu_{\ln K,t}$ and $\ln S_t$. Then we use the VECM to test the strength of this relationship empirically. Last, we use the ARIMA model to interpret and further discuss the results.

In this section, we still use the S&P 500 index put options that mature within one month. The $\mu_{\ln K,t}$ and $\ln S_t$ are both I(1) processes. The results of the Augmented Dickey-Fuller (ADF) test are shown in table 3.5. The statistics and 5% critical values of the ADF test and corresponding p-values are recorded for each variable. We include a time trend in all the ADF tests. The p-values of $\mu_{\ln K,t}$ and $\ln S_t$ are 0.80 and 0.98 respectively, so these two variables are both non-stationary, but their first differentiation terms $\Delta \ln S_t, \Delta\mu_{\ln K,t}$ are stationary. Then we test the existence of the cointegration relationship between these two variables. We use the Johansen test to check the hypothesis, and the results are shown in table 3.5. The maximum lag is 2, according to the BIC lag selection method. It turns out that when the rank is 0, both trace statistic and max statistic are greater than the 5%

critical values, which indicates that we should reject the null hypothesis that there is no cointegration relationship. When the rank is 1, the corresponding statistics are both 2.63 and less than 5% critical value of 3.76. We could not reject the null hypothesis that there is one cointegration relationship. Therefore the Johansen test suggests the existence of one cointegration equation.

Table 3.5: Augmented Dickey-Fuller Test and Johansen Test

	ADF test			
Variable	Z-stat	5% c.v.		P-value
$\mu_{\ln K,t}$	-1.56	-3.41		0.80
$\ln S_t$	-0.63	-3.41		0.98
$\Delta\mu_{\ln K,t}$	-67.79	-2.86		0.00
$\Delta \ln S_t$	-58.85	-2.86		0.00
	Johansen Test			
Rank	trace-stat	5% c.v.	max-stat	5% c.v.
0	638.38	15.41	635.75	14.07
1	2.63	3.76	2.63	3.76

The table shows the results of the ADF test and the Johansen Test for the S&P 500 index put options that mature within one month. The $\mu_{\ln K,t}$ and $\ln S_t$ are both I(1) processes, their first differentiation terms $\Delta \ln S_t, \Delta\mu_{\ln K,t}$ are stationary. In addition, the Johansen test suggests the existence of one cointegration equation.

Next, we use the VECM method to find the strength of the relationship. Given the findings in our stylized facts and ADF tests, we use the VECM specified in the equation 3.5

$$\Delta y_t = \alpha(\beta' y_{t-1} + \mu) + \Sigma_{i=1}^{p-1}\Gamma_i \Delta y_{t-i} + \gamma + \epsilon_t, \tag{3.5}$$

where $y_t = (\mu_{\ln K,t}, \ln S_t)'$. α is a $2 \times r$ adjustment coefficient matrix, and β is a $2 \times r$ vector of the parameters in the cointegrating equation. Here r is the number of independent cointegrating vectors, and $p = 2$ is the lag number for the underlying VAR model selected from the BIC method. Γ_i's are 2×2 matrices of coefficients. ϵ_t is a 2×1 vector of errors. We cannot estimate (α, β) without further assumptions. Therefore we assume that there is a linear trend in the original time series of $\mu_{\ln K,t}$ and $\ln S_t$ and there is not a trend but a constant intercept in the cointegration equation. $\alpha\mu$ is a $2 \times r$ matrix that captures the linear trend in y_t, and γ is a 2×1 vector that captures the constant intercept in the cointegration equation. γ is orthogonal to $\alpha\mu$ so that the parameters are all identifiable in the specification of VECM. The cointegration equation estimated from the VECM is $\mu_{\ln K,t} - 0.9732 \ln S_t - 0.1355 = 0$. A 1% increase in the underlying index is associated with an around 1% increase in the weighted average strike. Such mapping strengthens our empirical test results and successfully captures the feature of the plots in section 3.3.2. So far, we have shown that $\mu_{\ln K,t}$ moves closely with $\ln S_t$ in both the short run and the long run.

3.4.3 Relative Position of Average Strike

There is one question left unanswered in sections 3.4.1 and 3.4.2. Now that we know how the overall distribution of option contracts held and traded by the market moves with the underlying index, we are interested in further depicting the position of the average strike relative to the underlying index and asking whether this relative position is connected with the volatility of the underlying index. Does the highest open interest contract tend to be some percent away from the index, or does it tend to be some standard deviations away from the index? In other words, if volatility in the underlying index doubles, does the weighted average strike also move further from the index?

We continue to use options that mature in one month or less. The daily open-interest-weighted average strike is defined by equation 3.1 and the relative position can be defined as

$$P_t^1 = \frac{\mu_{K,t}}{S_t} - 1. \tag{3.6}$$

Alternatively, the daily open-interest-weighted average log strike is defined by equation 3.2 and the relative position can be defined as

$$P_t^2 = \mu_{\ln K,t} - \ln S_t. \tag{3.7}$$

As we are going to show below, these two definitions give very similar results. Table 3.6 summarizes the two relative positions for put and call options and figure 3.5 shows the distributions. Depending on the definition, the put option weighted average strike is 8 to 9% below the underlying index, with a standard deviation of 5 to 6% and a median of 7 to 8% below the underlying index. The call option weighted average strike is around 1% above the underlying index, with a standard deviation of around 5% and a median of around 1% above the underlying index. We confirm the stylized fact that the aggregate positions are slightly out of the money. Put options are deeper out of the money on the aggregate level than call options. In other words, put option weighted average strike is further from the underlying index. This is consistent with the right panel in figure 3.1.

Table 3.6: Summary Statistics of Relative Positions

		Count	Mean	S.D.	Skewness	Kurtosis	Median
Put	P^1	5326	-7.84%	5.18%	0.064	5.165	-7.50%
	P^2	5326	-9.08%	6.12%	-0.740	4.960	-8.32%
Call	P^1	5324	1.58%	5.25%	1.856	12.501	0.96%
	P^2	5324	1.04%	4.90%	1.036	9.889	0.67%

The table shows the summary statistics of relative positions for both put and call options. The put option weighted average strike is 8 to 9% below the underlying index, with a standard deviation of 5 to 6% and a median of 7 to 8% below the underlying index. The call option weighted average strike is around 1% above the underlying index, with a standard deviation of around 5% and a median of around 1% above the underlying index. Put options are deeper out of the money on the aggregate level than call options.

All relative positions are very concentrated around the mean. For put options, P^1 falls within one standard deviation from the mean 73.75% of the time, and P^2 falls within one standard deviation from the mean 74.54% of the time. For call options, the numbers are 79.32% and 76.99%, respectively.

Then we try to answer whether the weighted average strike moves further away from the index as volatility increases by running a simple linear regression of the relative position on VIX, the implied volatility measure of the S&P 500 index. The prior is that the relative position of the weighted average strike will be deeper on days when the implied volatility is higher, whether in or out of the money. Both relative positions and the VIX are stationary according to the Dicky-Fuller test. Table 3.7 presents the results and agrees with our prior. The coefficient on VIX always has the same sign as the relative positions, whether for puts or calls, in or out of the money. Table 3.7 also shows that besides the mean and the median, another important difference between the distributions of relative positions of puts and calls is that there are more days on which the weighted average strike is in the money for calls than for puts.

Table 3.7: Regressions of Relative Position on VIX by Option Type and Moneyness

	Put				Call			
	OTM		ITM		ITM		OTM	
	P^1	P^2	P^1	P^2	P^1	P^2	P^1	P^2
	b/t	b/t	b/t	b/t	b/t	b/t	b/t	b/t
VIX	-0.087***	-0.126***	0.215***	0.158***	-0.053***	-0.040***	0.364***	0.305***
	(-9.828)	(-11.743)	(12.675)	(9.519)	(-5.095)	(-3.786)	(56.968)	(54.692)
R^2	0.019	0.026	0.419	0.354	0.012	0.006	0.501	0.496
N	5101	5159	225	167	2087	2285	3237	3039

The table presents the regression results of running relative position on VIX by option type and moneyness group. The coefficient on VIX always has the same sign as the relative positions, whether for puts or calls, in or out of the money. Besides the mean and the median, another important difference between the distributions of relative positions of puts and calls is that there are more days on which the weighted average strike is in the money for calls than for puts.

We can imagine ourselves looking at all the daily distributions of open interest over strike during the sample period. Table 3.7 tells us that when the distribution is centered around some point on the left of the underlying index, higher implied volatility in the index tends to move the distribution further to the left, and vice versa. We then continue to ask, apart from the effect of implied volatility on the mean, whether implied volatility can also affect the standard deviation of the distribution. In other words, we now know that volatility makes the relative position of the weighted average strike more spread out on the aggregate level. We wonder if volatility does the same for the daily distributions on the contract level.

We define the standard deviations of the daily distributions of open interest, $\sigma_{K,t}$ and $\sigma_{\ln K,t}$, in equations 3.8 and 3.9, corresponding to our definitions of $\mu_{K,t}$ and $\mu_{\ln K,t}$ in equations 3.1 and 3.2,

$$\sigma_{K,t} = \sqrt{\frac{\sum_{i=1}^{n_t}(K_{it} - \mu_{K,t})^2 oit_{it}}{\sum_{i=1}^{n_t} oit_{it}}}, \tag{3.8}$$

and

$$\sigma_{\ln K,t} = \sqrt{\frac{\sum_{i=1}^{n_t}(\ln(K_{it}) - \mu_{\ln K,t})^2 oit_{it}}{\sum_{i=1}^{n_t} oit_{it}}}. \tag{3.9}$$

Then we run a simple linear regression of the standard deviation on VIX. We are expecting that on days when the implied volatility is higher, the distribution of open interest has a higher standard deviation and is thus more spread out, which means options in deeper positions are more likely to be held by the market. Both standard deviations are stationary according to the Dicky-Fuller test. Table 3.8 presents the results. We see that, indeed, VIX has positive and significant effects on the standard deviations of the daily open interest distributions. Implied volatility not only shifts the distribution's position but also changes its shape.

3.4.4 Preference for Slightly-out-of-the-money Options with Simulations

In section 3.4.3, we have seen how the relative position of the weighted average strike moves with the implied volatility. We now proceed to show why investors prefer slightly-out-of-the-money options using simulations. We simulate the daily return series of three trading strategies, naked option, full hedge, and delta hedge, with historical data. The

Table 3.8: Regressions of Standard Deviation of Daily Open Interest Distribution on VIX by Option Type

	Put		Call	
	$\sigma_{K,t}$ b/t	$\sigma_{\ln K,t}$ b/t	$\sigma_{K,t}$ b/t	$\sigma_{\ln K,t}$ b/t
VIX	378.320***	0.429***	314.827***	0.253***
	(40.684)	(46.587)	(36.859)	(34.645)
R^2	0.237	0.290	0.203	0.184
N	5326	5326	5324	5324

The table presents the regression result of running standard deviation of daily open interest distribution on VIX by option types. VIX has positive and significant effects on the standard deviations of the daily open interest distributions. Implied volatility not only shifts the distribution's position but also changes its shape.

naked option strategy means buying one option each day and selling it the next day. The full hedge strategy means buying/selling one share of the underlying index and buying one put/call option at the same time each day, and selling the portfolio the next day. The delta hedge strategy means buying one share of the underlying index and buying $-\frac{1}{\Delta}$ put/call options at the same time each day and selling the portfolio the next day. Δ is the option price sensitivity to the underlying index, calculated from the Black-Scholes model. Δ is negative for put options and positive for call options, so buying $-\frac{1}{\Delta}$ put options means a long position in put options and buying $-\frac{1}{\Delta}$ call options means a short position in call options.

We choose an option at the strike closest to $X\%$ moneyness in each of the three strategies. Here $X\%$ is the relative position of the strike defined as a percentage of the underlying index,

$$X = \frac{K}{S} * 100. \tag{3.10}$$

Put options are out of the money when $X < 100$, and call options are out of the money when $X > 100$. We compute the daily return series of the three strategies for different X values ranging from 92 to 108. We summarize the simulation results in table 3.9.

Table 3.9: Simulation Results of Three Trading Strategies

Naked Option								
	Put				Call			
X	Mean	S.D.	VaR	Median	Mean	S.D.	VaR	Median
92	-0.693	48.963	-50.000	-8.000	-0.427	15.894	-21.371	-0.376
94	-1.055	44.025	-50.000	-7.708	-0.374	16.738	-23.084	-0.447
96	-1.531	39.650	-47.761	-7.063	-0.314	19.029	-26.586	-0.532
98	-1.135	35.763	-44.270	-5.438	-0.541	20.684	-32.326	-0.826
100	-0.973	29.130	-40.406	-4.138	-0.003	26.813	-38.920	-1.442
102	-1.312	24.183	-36.252	-3.499	1.396	43.152	-49.223	-2.632
104	-0.948	20.960	-31.964	-2.396	2.726	52.947	-53.000	-4.954
106	-0.705	18.522	-29.177	-1.620	2.924	56.562	-53.846	-4.153
108	-0.602	17.240	-26.136	-1.298	3.792	59.947	-53.846	0.000
Full Hedge								
92	0.005	1.010	-1.556	0.018	0.036	0.872	-1.229	0.035
94	0.006	0.922	-1.430	0.014	0.055	0.810	-1.107	0.041
96	0.010	0.841	-1.281	0.005	0.050	0.772	-1.017	0.045
98	0.003	0.754	-1.104	-0.007	0.035	0.732	-1.013	0.058
100	0.003	0.636	-0.899	-0.014	0.035	0.733	-1.083	0.082
102	-0.017	0.622	-0.871	-0.008	0.030	0.843	-1.302	0.089
104	-0.014	0.616	-0.921	-0.015	0.032	0.968	-1.514	0.084
106	-0.002	0.651	-0.962	-0.009	0.036	1.048	-1.625	0.080
108	0.010	0.737	-1.041	-0.004	0.041	1.117	-1.722	0.081
Delta Hedge								
92	-0.249	2.708	-3.245	-0.292	0.053	0.915	-1.210	0.036
94	-0.101	1.161	-1.597	-0.168	0.072	0.836	-1.072	0.038
96	-0.052	0.761	-0.963	-0.106	0.074	0.804	-0.953	0.041
98	-0.031	0.592	-0.727	-0.063	0.058	0.697	-0.864	0.043
100	-0.024	0.502	-0.593	-0.041	0.047	0.532	-0.608	0.055
102	-0.050	0.633	-0.844	-0.045	0.036	0.603	-0.816	0.080
104	-0.041	0.681	-1.002	-0.046	0.039	0.721	-1.049	0.099
106	-0.024	0.700	-1.053	-0.032	0.050	0.831	-1.289	0.123
108	-0.007	0.804	-1.135	-0.026	0.050	0.908	-1.456	0.113

The table shows the simulation results of the naked option, full hedge and delta hedge trading strategies.

Table 3.9 shows the summary statistics of the daily return series of the three trading strategies. All numbers are percentages. We look at the mean, standard deviation, value at risk (VaR) as the 5th percentile and the median of the return series. Ideally, the investor looks for a mean as high as possible, a standard deviation as low as possible, a VaR as high as possible, and a median as high as possible.

Almost immediately, an investor who has some risk aversion prefers hedging over taking naked option positions. The standard deviations of the naked option strategy are 10 to 50 times higher than those of the hedging strategies, and the VaR's 20 to 50 times lower. However, the "Naked Option - Call" panel is worth further attention. The mean return of buying naked call options increases with X and reaches a stunning 3.792% when we choose the deepest out of the money options from our list, $X = 108$. At the same time, the median return of the same strategy decreases with X and can be as low as -4.153% when $X = 106$. When $X = 108$, the call options are very deeply out of the money so that their premiums are quite insensitive to daily underlying index changes and have a small trading volume on many of the trading days. The mode and the median return of such options are thus 0%. When we exclude such 0's from the return series, the median becomes -13.42%. Therefore we know the positive mean returns of the out-of-the-money call options are driven by large positive outliers. The return distributions of those options are likely to be highly skewed. Because of this unique feature, investors may want to use options as their speculation tools, or lottery. We further discuss this in section 3.4.5.

Now we focus on the lower two panels of table 3.9 and compare among different X's. The "Full Hedge - Put" panel sees the highest mean and median returns in the out-of-the-money options. Although when $X = 108$ the mean return is also 0.010%, the median return is negative, making the in-the-money options less attractive. The standard deviations are the lowest when X is near 100, and so are the VaR's. Thus, if an investor is looking for some positive return and wants to avoid too much risk, it is probably the most reasonable to choose a put option strike near 96-98% of the underlying index. Namely, a slightly-out-of-the-money put option.

Similarly, in the "Full Hedge - Call" panel, we see the highest median return when $X = 102$. The risk measures, standard deviation, and VaR are the lowest when X is near 100. Again an investor doing some risk and return trade-off probably prefers a call option strike near 100-102% of the underlying index, thus a slightly-out-of-the-money call option. We see the same patterns in the "Delta Hedge" panel. It is reasonable to believe that a risk-averse investor will choose a put option strike near 98-100% of the underlying index and a call option strike near 100-102% of the underlying index.

Comparing what we observe in table 3.9 to table 3.6 and figure 3.5, we see that the full hedge strategy and the delta hedge strategy are closer to the behavior of the market, and our simulation results are very consistent with the data. A real investor in the market may choose some strategy between full hedge and delta hedge and multiple options at different strikes simultaneously.

3.4.5 Gambling Motivation for Trading

.

So far, we have studied the investors' preference for slightly-out-of-the-money options and the stylized fact that the average strike price always closely follows the underlying index. In section 3.4.4, we find that the hedging motivation can explain such behavioral patterns, and at the same time, we find a hint on the skewness of the option return and the possibility that investors use options as speculation tools and the gambling motivation can also lead to the choice of slightly-out-of-the-money options. As introduced in the literature review, gambling motivation can be a good candidate motivation for options trading since some investors trade options for leverage and risk exposure provided by options, and they consider options as lotteries and gamble with options. The gambling motivation does not contradict with our hedging motivation. They could coexist in realistic options trading. Actually, we also find empirical evidence about the gambling motivation in our analysis, it provides us a different angle on understanding options trading, and it also indirectly proves the conclusion about gambling trading in the options market from past literature. All these findings could be included in the framework of our analysis and let us see a bigger picture of realistic option trading behaviors.

According to Kumar (2009), there are three key features for an asset to be defined as a lottery-like asset: low asset price, high return volatility, and high return skewness. If we consider options, we know that out-of-the-money options have low premiums because of their zero intrinsic value. If we measure moneyness with the index-strike ratio S/K, call premiums increase monotonically with moneyness and put premiums decrease monotonically with moneyness. Now we want to see what options have high return volatility and high return skewness. We look at monthly put options that mature in exactly one month from January 1990 to June 2012, so on any given day, the only difference between the contracts is their strike prices, thus moneynesses. The put options are divided into five groups using the index-strike ratio of S/K as the criterion. A put option is deeply out of the money (DOTM) if $S/K \geq 1.2$, out of the money (OTM) if $1.01 \leq S/K < 1.2$, at the money (ATM) if $0.99 \leq S/K < 1.01$, in the money (ITM) if $0.8 \leq S/K < 0.99$, and finally deeply in the money (DITM) if $S/K < 0.8$. Trading-volume-weighted average returns are calculated for each of the five groups, followed by 2-year or 24-month historical return variance and skewness, which are plotted in figure 3.6.

Figure 3.6 only shows plots for three of the five groups, excluding the DOTM and DITM groups, because both groups have incomplete data. On most of the days during our sample period, there are neither DOTM nor DITM put options traded, and even if there are any, the trading volumes are very low and account for very low shares of put options traded on that day, as shown in figure 3.2.

In the top panel of figure 3.6, we can observe that when put options generally have low return volatility, all three groups have similar return volatility. During such periods of history, any group can have the highest return volatility. Based on the plot, it seems that the ATM group has the highest return volatility most often. However, during periods when put options have high return volatility, which is right at or after the underlying index's periodical bottom according to figure 3.1, the OTM group, without doubt, has the highest return volatility that can easily be several times of the return volatility of the other two groups. The ITM group almost always has the lowest return volatility.

In the bottom panel of figure 3.6 patterns are even clearer. The three groups' Return skewness move closely, with the OTM group line almost always being the highest and the ITM group line almost always being the lowest.

From the discussion above, we see that OTM options have low premiums, high return volatility due to the low premiums and the resulting high leverages, and high return skewness due to the low probabilities of being executed. Therefore, OTM options best fit the three features of a lottery-like asset and can best serve speculators' gambling purpose.

3.5 Conclusion

In this paper, we start by identifying a very robust stylized fact in the options market that slightly-out-of-the-money options are the most popular in trading volume and open interest. The open-interest- and trading-volume- weighted average strikes closely follow the underlying index, being slightly below the index for calls and slightly above the index for puts. We try to give rationality to this fact and provide instructive insights to investors for their choices of strike prices. This topic has been rarely covered in the options literature.

We investigate the association between the weighted average strike price and the underlying index using SVAR and the VECM with the CBOE dataset. Our empirical work reveals the mechanism through which the weighted average strike price co-moves with the index. The choice of option strike moves slowly in the same direction as the underlying index. The effect of a broad-market shock on the trading-volume-weighted average strike is 10 times as large as that on the open-interest-weighted average price. From the impulse response functions, we see that the responses of weighted average strikes are small initially and grow with time. There is evidence for cointegration in the long run, and a 1% change in the underlying index is associated with an increase close to 1% in the weighted average strike.

We try to answer whether the weighted average strike tends to be some fixed percent away from the underlying index or some standard deviations away. Our results show that when the implied volatility is higher, option strikes move further from the underlying index whether in or out of the money, for put or call option, on the aggregate level or the contract level.

We use simulations with historical data to guide the choice of strikes. We find that hedging strategies are more consistent with the patterns in our data where investors, taking both risk and return into consideration, most likely prefer slightly-out-of-the-money options.

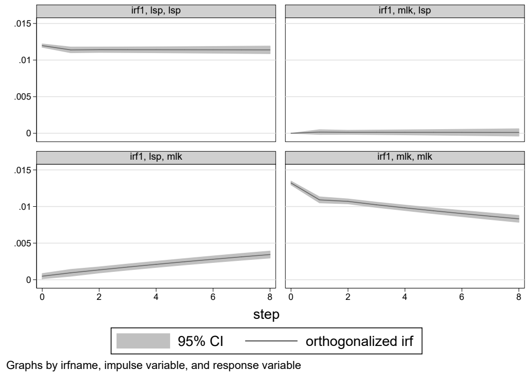
On the other hand, the naked option strategy leads us to think of gambling motivation as a parallel explanation for such preferences. Therefore, we investigate the gambling motivation and find that out-of-the-money options have low premiums, high return volatility, and skewness, so they satisfy the definitions of lottery-like assets and best fit the need for speculation tools. Both the hedging motivation and gambling motivation can justify the choice of options that are slightly out of the money.

3.6 References

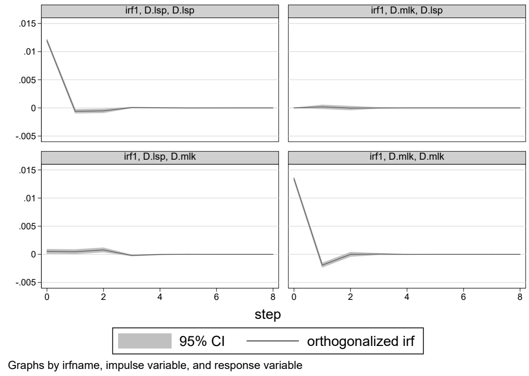
- Anthony, J.H. (1988) The Interrelation of Stock and Options Market Trading-Volume Data. Journal of Finance, 43(4), 949-964.
- Bauer, R., Cosemans, M. & Eichholtz, P.M.A. (2009) Option Trading and Individual Investor Performance. Journal of Banking & Finance, 33(4), 731-746.
- Bernile, G., Gao, F. & Hu, J.F. (2017) Center of Volume Mass: Does Aggregate Options Market Opinion Predict Future Equity Returns?
- Blau,B.M., Bowles, T.B. & Whitby, R.J. (2016) Gambling Preferences, Options Markets, and Volatility. Journal of Financial and Quantitative Analysis, 51(2), 515-540.
- Byun, S.J. & Kim, D.H. (2016) Gambling preference and individual equity option returns. Journal of Financial Economics, 122(1), 155-174.
- Holt, C. A. & Laury, S.K. (2002) Risk Aversion and Incentive Effects. The American Economic Review, 92(5), 1644-1655.
- Hu J.F. (2014) Does option trading convey stock price information? Journal of Financial Economics, 111(3), 625-645.
- Jennings,R. & Starks, L.(1986) Earnings Announcements, Stock Price Adjustment, and the Existence of Option Markets. Journal of Finance, 41(1), 107-125.
- Jones, C.S.(2003) The dynamics of stochastic volatility: evidence from underlying and options markets. Journal of Econometrics, 116(1-2), 181-224.

- Kumar, A. (2009) Who Gambles In The Stock Market? Journal of Financial and Quantitative Analysis, 64(4), 1889-1933.
- Merton, R.C. (1973) Theory of Rational Option Pricing. The Bell Journal of Economics and Management Science, 4(1), 141-183.
- Rabin, M. (2000) Risk Aversion and Expected-Utility Theory: A Calibration Theorem. Econometrica, 68(5), 1281-1292.
- Roll, R., Schwartz, E. & Subrahmanyam, A. (2010) O/S: The relative trading activity in options and stock. Journal of Financial Economics, 96(1), 1-17.

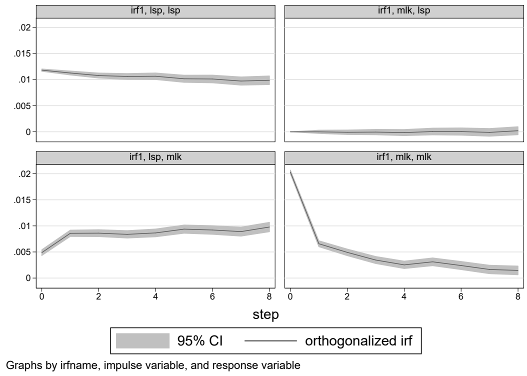
Figure 3.3: Impulse response functions of put option average log strike and log of the underlying index



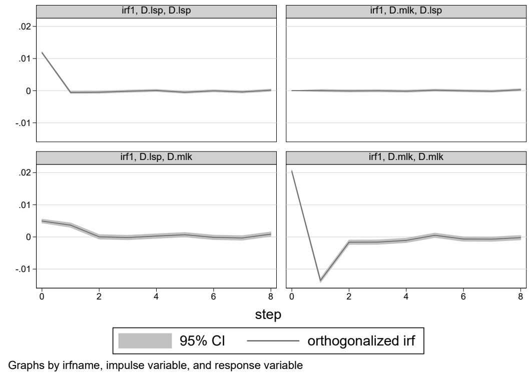
(a) Open-interest-weighted average log strike, level



(b) Open-interest-weighted average log strike, first difference



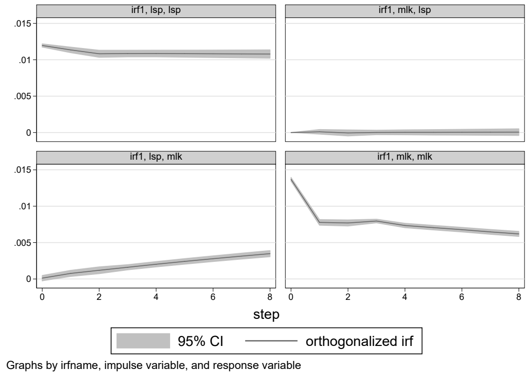
(c) Trading-volume-weighted average log strike, level



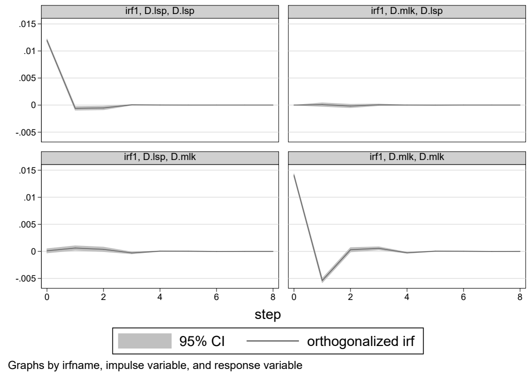
(d) Trading-volume-weighted average log strike, first difference

The figure shows the IRF plots of put option average log strike and log of the underlying index. The impulse responses in the left panel have long terms. The response of the average log strike on the impulse of the underlying index (in the bottom left) increases with time. Investors in the options market adjust their positions slowly and gradually with the underlying index, and the market adjusts its aggregate position slower.

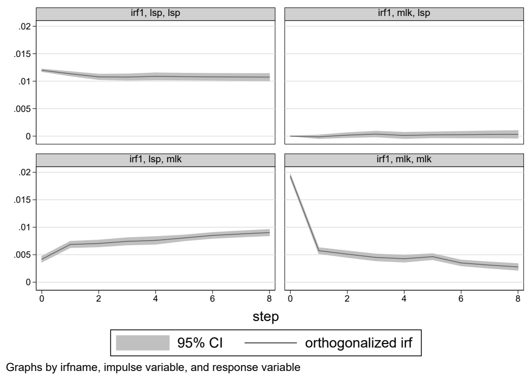
Figure 3.4: Impulse response functions of call option average log strike and log of the underlying index



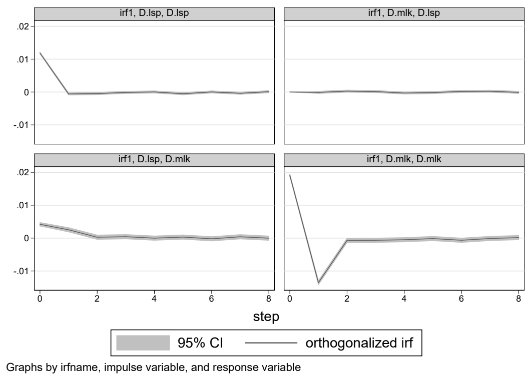
(a) Open-interest-weighted average strike, level



(b) Open-interest-weighted average strike, first difference



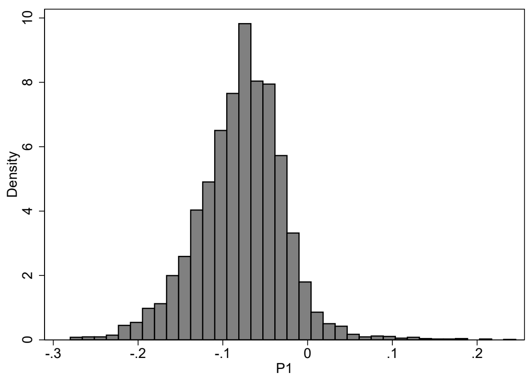
(c) Trading-volume-weighted average strike, level



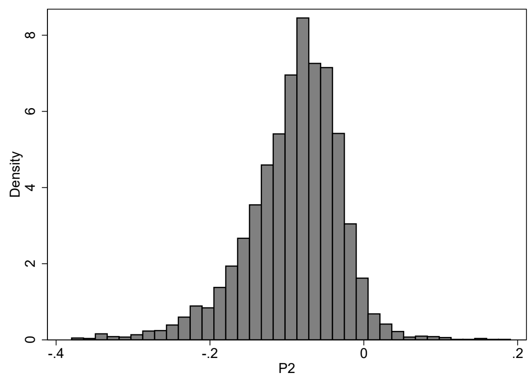
(d) Trading-volume-weighted average strike, first difference

The figure shows the IRF plots of call option average log strike and log of the underlying index. The pattern here is similar to that in put options.

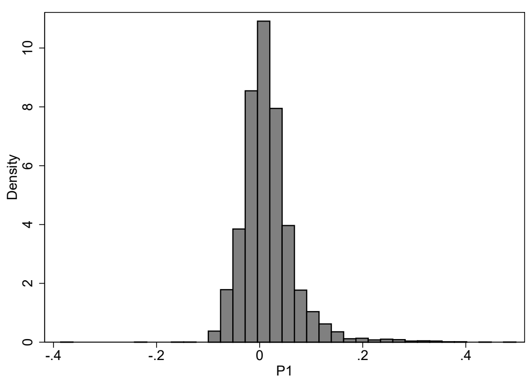
Figure 3.5: Distribution of Relative Positions



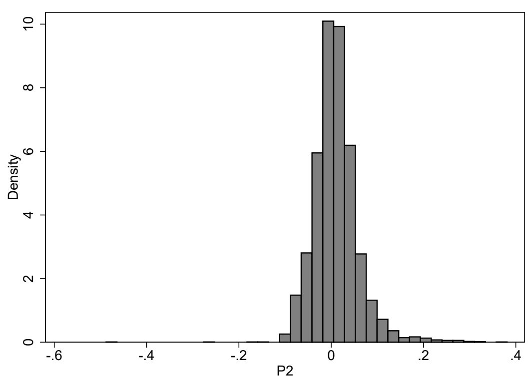
(a) P^1 of Put



(b) P^2 of Put



(c) P^1 of Call

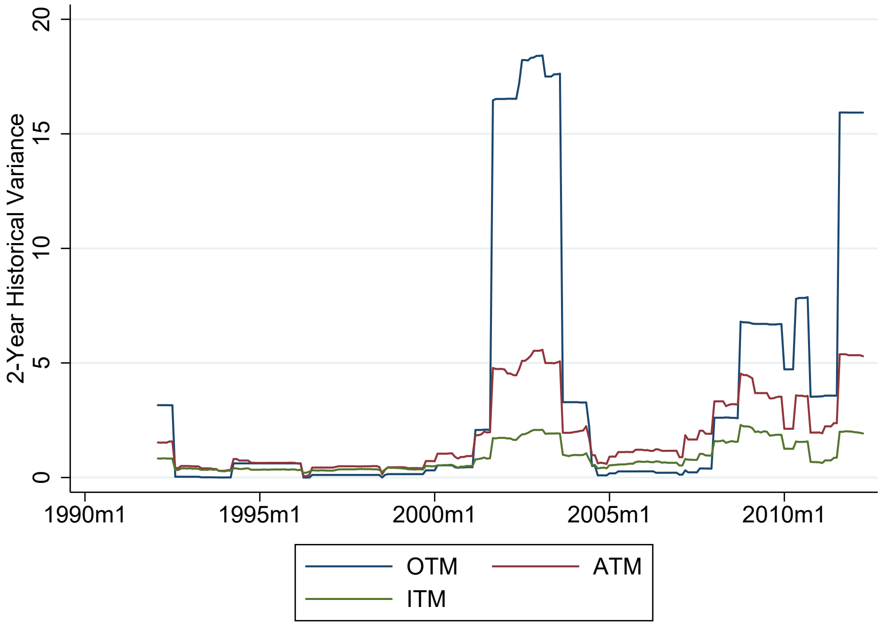


(d) P^2 of Call

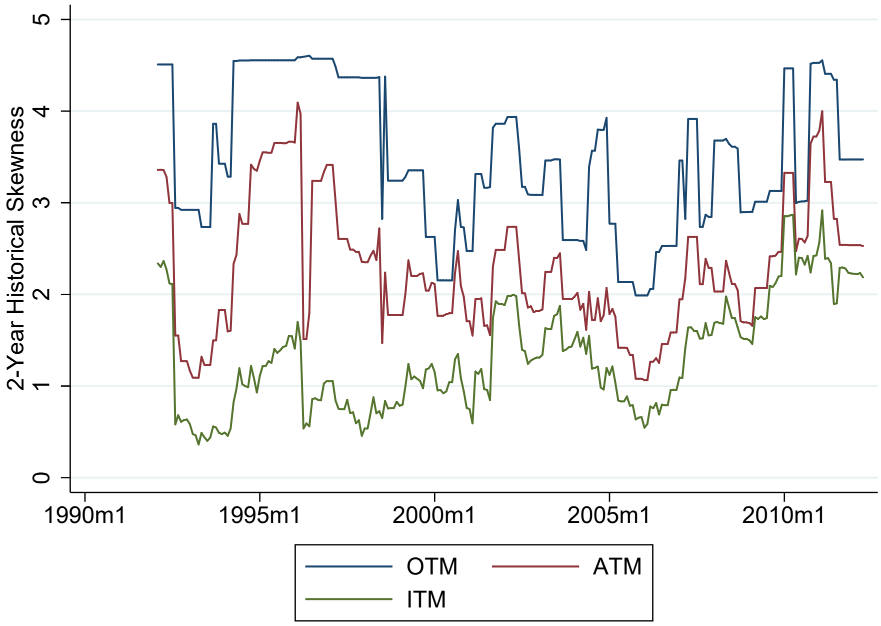
The figure shows the distributions of relative positions of the price for both put and call options.

Figure 3.6: Put options historical return variance and skewness by moneyness group

(a) Variance



(b) Skewness



The figure shows the put option's historical return variance and skewness by moneyness group. OTM options have low premiums, high return volatility due to the low premiums and the resulting high leverages, and high return skewness due to the low probabilities of being executed. Therefore, OTM options best fit the three features of a lottery-like asset and can best serve speculators' gambling purpose.

Appendices