TWO ESSAYS ON ASSET PRICING AND RETAIL INVESTORS

DU QINGJIE

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University

School of Accounting and Finance

Two Essays on Asset Pricing and Retail Investors

Du Qingjie

A thesis submitted in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

April 2022

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

Signature: ____________________

Name of Student: Du Qingjie

ABSTRACT

This thesis contains two essays. The commonality of the two essays is that I investigate the role and impact of the retail trading activities on the asset pricing in both essays. In the first essay, I focus on the option market, and investigate the impact of retail investors trading activities of the underlying stocks on the corresponding cross-sectional option returns. In the second essay, I decompose the aggregate retail order imbalance into pure buying order and pure selling order, and examine the pricing impact on the cross-sectional stock returns separately.

Specifically, in the first essay, I examine the relation between the cross-sectional delta-hedged option returns and the retail investor trading activities of the underlying stocks. I hypothesize that the option price could be affected either by the retail investor's gambling appetite or affected by the noise trader risk associated with retail trading activities. Empirically, I find that both call option returns and put option returns decrease as the retail trading volume of the underlying stocks increases. Further analysis shows that the retail trading order imbalance (i.e. the betting direction) does not predict future option returns, but that the volatility of retail trading activities does. The pricing effect of retail trading volume becomes stronger when retail trading is more volatile or when the stock's arbitrage cost is high. A test using Abel Noser data suggests that institutional trading activities do not affect the delta-hedged option returns. Overall, the results suggest that the trading activities of

retail investors increase options' hedging costs and hedging difficulty, and option writers charge higher prices to compensate for this noise trader risk.

In the second essay, I decompose the retail order imbalance into aggregate selling orders and aggregate buying orders, and document an asymmetric pricing effect between the selling orders and buying orders. Specifically, the long-short hedge portfolio formed based on retail selling orders generates about 10 bps abnormal return each day, i.e., 2% each month. However, the aggregate retail buying orders cannot predict the cross-sectional stock returns. The previous documented positive relationship between cross-sectional stock return and retail order imbalance could be mainly driven by the selling side. Stocks with intensive aggregate retail selling orders continue to underperform in the future, receive excess retail selling pressure, and are associated with drying-up liquidity. The pricing effect of retail selling orders becomes stronger when the VIX is high or when market is bearish, and when the stock is hard to value, but disappears on Fridays when the investor mood is high.

To summarize, the two essays provide new evidence on the role and impact of the retail investors in the financial market. The results help both the market participants and the policy makers to better understand the impact of retail investors on the asset pricing.

ACKNOWLEDGEMENTS

First and foremost, I want to express my sincerest gratitude to my chief supervisor, Prof. K.C. John Wei, for his supervision, advice, and guidance throughout my Ph.D. study. His patient training of my research skills and critical thinking will benefit me the entire life. Without him, this thesis would never have been completed.

I would also like to show my thanks for the entire AF family, including all the excellent faculty members and lovely classmates. The AF department provides me plenty of research support. The discussions with the other classmates and faculty members really help me to improve a lot. I am so lucky to be one of the AF family, and the life experience in the AF department is extraordinary.

Finally, I would like to thank my family. Without their love and support, I can never become who I am today. It is my fortune to always have them in my life.

Table of Contents

ABSTRACT.....I
ACKNOWLEDGEMENTS.....III
LIST OF TABLES.....VI
LIST OF FIGURES.....VIII
Chapter 1: Retail Investor Trading Activities and the Cross-Section of Option Returns.....1
1.1 Introduction.....1
1.2 Data and Variable.....9
1.2.1 Option Data.....9
1.2.2 Delta-hedged Option Return.....10
1.2.3 Retail Trading Activities.....12
1.3 Baseline Empirical Analysis.....16
1.3.1 Delta-hedged Option Return and Retail Trading Volume: Regression Analysis.....16
1.3.2 Portfolio-level Analysis: Profitability of Holding Options.....18
1.3.3 Robustness Tests.....20
1.4 Explore the Underlying Mechanism.....21
1.4.1 Lottery Premium Hypothesis: The Role of Retail Order Imbalance.....21
1.4.2 Noise Trader Risk Hypothesis: The Role of Volatility of the Retail Trading Activities.....24
1.4.3 Noise Trader Risk Hypothesis: The Role of Arbitrage Cost.....27
1.4.4 The Role of Institutional Trading Activities on Option Returns.....28
1.4.5 Evidence From Out-of-The-Money Options.....31
1.5 Further Discussion.....34
1.5.1 The Impact of Retail Option Trader.....34
1.5.2 Other Possible Channels.....35
1.6 Conclusion.....36
Chapter 2: Dissecting Retail Trading Orders.....39
2.1 Introduction.....39
2.2 Data and Sample.....48
2.2.1 Retail Trading Data.....48

2.2.2 Discussion on the retail trading measures 50

2.2.3 Control Variables 53

2.3 Empirical Results 54

2.3.1 Summary statistics of retail trading activities and stock characteristics 54

2.3.2 Dissecting Retail Trading Activities: Portfolio Sorts 57

2.3.2.1 Portfolio sorts using retail buying and selling measures separately 57

2.3.2.2 10×10 Portfolio sorts using retail buying and selling measures 59

2.3.3 Dissecting Retail Trading Activities: Fama-MacBeth Regression .. 61

2.4 Long term analysis 63

2.4.1 Weekly rebalanced portfolio analysis 63

2.4.2 Buy and Hold Return 64

2.5 Mechanism Test 65

2.5.1 Concentrated selling and dispersed buying 66

2.5.2 Persistent selling and transitory buying 67

2.5.3 Investor sentiment and stock characteristics 69

2.5.4 Investor mood: weekday effects 70

2.6 Alternative explanation 72

2.6.1 Asymmetric liquidity provision 72

2.6.2 Asymmetric informed trading 73

2.6.3 Other possible explanations 76

2.7 Conclusion 76

Figures and Tables 78

References 123

LIST OF TABLES

Table 1.1. Summary Statistics of Options

Table 1.2. Summary Statistics of Stock Characteristics

Table 1.3. Baseline Results: Fama-MacBeth Cross-section Regressions

Table 1.4. Returns of Option Portfolios Sorted by Total Retail Trading Volume

Table 1.5. Risk-adjusted Option Portfolio Returns

Table 1.6. Robustness Tests: Alternative Measures of Option Returns and Retail Trading Volume

Table 1.7. Lottery Premium Hypothesis Test: The Role of Retail Order Imbalance

Table 1.8. Noise Trader Risk Hypothesis Test: The Role of Retail Trading Volatility

Table 1.9. Noise Trader Risk Hypothesis Test: The Impact of Trading Volatility on the Pricing Effect of Retail Trading Volume

Table 1.10. Noise Trader Risk Hypothesis Test: The Impact of Arbitrage Cost on the Pricing Effect of Retail Trading Volume

Table 1.11. The Impact of Abel Noser Institutional Trading Activities on Option Returns

Table 1.12. Evidence from out-of-money options

Table 2.1: Summary Statistics

Table 2.2. Daily Portfolio Analysis Based on Aggregate Retail Orders

Table 2.3. 10×10 Double Sorted Portfolios Based on Aggregate Retail Selling and Retail Buying Orders

Table 2.4. Fama-MacBeth Cross-sectional Return Regressions

Table 2.5. Weekly Rebalanced Portfolio Analysis Based on Aggregate Retail Orders

Table 2.6. Buy and Hold Returns in the Following 21 Trading Day

Table 2.7. Portfolio Order Imbalance in Each 10×10 Double Sorted Portfolios

Table 2.8. Portfolio Performance and Retail Trading Activities Around Intensive Individual Trading Activities

Table 2.9. Additional Analysis on Fear or Greed Sentiment

Table 2.10. Investor Mood: Weekday Effects

Table 2.11. Retail Trading Activities Around S&P 500 Index Addition and Deletion

Table 2.12. Portfolio Return and Individual Trading Around the Earnings Announcement Day

Table 2.13. Retail Trading and Earnings Announcements: Fama-MacBeth Regression

LIST OF FIGURES

Figure 1.1. Portfolio Returns of Delta-hedged Options of Stocks with Different Retail Trading Volume

Figure 1.2. Portfolio Returns of Delta-hedged Options of Stocks with Different Retail Trading Order Imbalance

Figure 2.1. Daily Decile Rsel Hedge Portfolio Performance around the Formation Day

Figure 2.2. Daily Decile Rbuy Hedge Portfolio Performance around the Formation Day

Figure 2.3. Daily Decile OIbjzz Hedge Portfolio Performance around the Formation Day

Chapter 1: Retail Investor Trading Activities and the Cross-Section of Option Returns

1.1 Introduction

The extreme stock price surges and dips of the meme stocks in January 2021 heighten the role of retail investors as a non-negligible market participants in the US equity market. Collectively, these retail “army” could force professional investors to take huge losses or even change their risk control policies (Li, 2021; Ronalds-Hannon and Surane, 2021). The flood of speculative retail money has also changed the Wall Street trading behaviors, making option trading more and more popular (Winck, 2020; McCabe, 2021; Scuffham and Ahmed, 2021). In this paper, I comprehensively investigate the impact of retail investor trading activities on the cross-section of option returns.

In the traditional finance studies, researchers always think retail investors act as noise traders in the stock market. Compared with institutions, retail investors are generally uninformed and less sophisticated. They lack professional financial knowledge and trading skills, suffer from the human behavioral biases, and lose money in the stock market (Black, 1986; Barber and Odean, 2000; Barber et al., 2009). In contrast, some studies show that the aggregate retail order flow could positively predict future stock returns, suggesting that on aggregate the retail investors are informed (Kaniel, Saar, Liu, and Titman, 2012; Kelley and Tetlock, 2013; Boehmer et al., 2022). However, recent studies inspired by the surge in zero-commission retail investors rise new doubt about the role of retail investors. Their results seem to support the noise trader argument (Barber, Huang, Odean, and Schwarz, 2022; Welch, 2020;

Eaton, Green, Roseman, and Wu, 2021). Unlike previous studies focusing on the stock market, I extend the testing venue and investigate the impact of retail trading activities on the pricing of stock options. As an equity derivative instrument, the pricing of options involves considerable and complex financial knowledge, which is generally beyond the scope of most retail investors. In this paper, I hypothesize that the retail trading activities of the underlying stocks could increase the noise trader risk of the stocks, and further affect the cross-sectional option pricing, and I label this as the *noise trader risk hypothesis*.

Specifically, the noise trader risk hypothesis argues that stocks with high retail investor trading activities are associated with significant noise trader risk for option writers and other arbitragers. In the imperfect market, option writers face intense arbitrage costs, such as capital constrains, transaction costs, and the impossibility of continuous trading/hedging (Figlewski, 1989; Garleanu, Pedersen, and Poteshman, 2009). Cao and Han (2013) find that higher idiosyncratic volatility of the underlying stocks lead to more significant negative future delta-hedged option returns. Boyer and Vorkink (2014) document that there exists a negative relation between ex ante skewness and option returns, suggesting that financial intermediaries are compensated for bearing unhedgeable risks for writing lottery-like options. Huang et al. (2019) show that the volatility of volatility negatively predicts the delta-hedged option payoffs. In line with this argument, I argue that the existence of noise traders amplifies the hedging cost and hedging difficulty for option writers. The collective action of those noise traders could possibly move the stock price far away from the intrinsic value. For example, the Wall Street Journal reports that Melvin Capital Management, a hedge fund, lost 53% on its investments in

January 2021 due to the soaring stock price of GameStop driven by retail investors (Chung, 2021). Thus, the noise trader risk hypothesis states that option writers charge higher prices to compensate for the noise trader risk associated with retail investors, which leads to lower expected option return in future.

Alternatively, the retail trading activities could also affect the cross-sectional option pricing through another channel, and I label it as the *lottery premium hypothesis*. It argues that stocks with high retail investor trading activities are lottery-like stocks. Kumar (2009) documents that individual investors prefer to trade stocks with lottery features. Grinblatt and Keloharju (2001, 2009) suggest that retail investors trade stocks for the purpose of sensation-seeking, a psychological trait linked to gambling and risk-taking. The implicit leverage effect of stock options satisfies retail investors' speculative appetite, which enables them to use a relatively small investment to bet on potentially extreme future returns. In addition to the leverage effect, Coval and Shumway (2001) argue that the curvature of option payoffs related to return volatility makes options non-redundant assets. Thus, even retail investors can home-make the leverage effect, to speculate on the expected volatility can only be achieved via option trading. In line with the lottery literature, I hypothesize that retail investors gambling on the stocks are also willing to pay a lottery premium for the corresponding stock options, leading to lower expected option returns in future.

To test and distinguish the above hypotheses, I examine the cross-sectional delta-hedged option returns of individual stocks. In each month, for each optionable stock, I pick one call option and one put option that are closest to being at-the-money and are near maturity (about 50 calendar days in the sample). I then calculate the delta-hedged option returns until maturity (Cao and Han, 2013; Cao et al., 2021).[1] Delta-hedging is the most commonly used trading strategy in the option market, which removes the impact of stock price changes on the corresponding option pricing and allows to focus on the volatility effect. To capture the retail investor trading activities, I follow Boehmer et al. (2022) and extract the retail trading orders from the TAQ dataset. After applying some commonly used filters to select data, the final sample contains about 150,000 monthly delta-hedged option returns for about 4,000 stocks from 2010 to 2017.

Empirically, I find that higher retail trading volumes of the underlying stocks predict significant lower future option returns. The monthly Fama-MacBeth cross-sectional regression reports significant and negative coefficients for the retail trading volume measure. The coefficients from the univariate regressions are -0.66 (t-stat = -11.81) and -0.52 (t-stat = -11.82) for call options and put options, respectively. This negative relation continues to hold after I control the different option return predictors documented in previous literature. Referring to the economic magnitude, I test the zero-cost long-short hedge portfolio returns constructed based on the retail trading activities. In details, I first calculate the average decile portfolio returns for holding the delta-hedged call (put) option. At the end of each month, I sort all the delta-hedged call (put) options by the retail trading volume of the stock in previous month, and form

1 All results are similar if calculating the delta-hedged option returns until the month end. Following previous literature, I only select the near-maturity options, but require the time-to-maturity to be longer than one month. Thus, the option maturity is generally within 46 to 52 calendar days. As the option becomes near-maturity, the trading cost to close the position becomes relatively large compared to the option price, thus holding the option until maturity could partially avoid the trading cost to close position.

decile portfolios. To make the zero-cost long-short hedge portfolio, I will buy the high retail trading decile, and sell the low retail trading decile, and focus on the average return spreads between this top and bottom decile portfolios. The empirical results suggest that all the long-short hedge returns are all negative and statistically significant at the 1% level, ranging from -0.83% to -2.17% for different weighting schemes. Further, the long-short hedge return spreads are not explained by the common risk factors, including the Fama-French three factors, the momentum factor, two volatility-based risk factors, or a combination of these factors. The results are also robust to alternative option return measures and retail trading volume measures.

To further explore the mechanism underlying the return predictability of retail trading volume, I conduct additional analyses to distinguish the lottery premium hypothesis and the noise trader risk hypothesis. Although both hypotheses support a negative relationship, they can have different further predictions. In the lottery premium hypothesis, retail investors bet on future extreme payoffs. At the aggregate level, the total retail order imbalance should reveal the betting directions, i.e., on average, whether the retail investors are betting on positive returns or negative returns. The speculation direction should generate asymmetric return predictions on the call options and put options. When the retail investors buy the stocks to bet on the positive returns, the corresponding call options should enjoy a lottery premium, but the put options should not. Thus, the aggregate retail order imbalance should play a more important role than the total retail trading volume under the lottery premium hypothesis. However, the noise trader risk hypothesis does not lead to such a *directional prediction*. As long as the noise traders actively trade the stock, the potential noise trader risk will always affect the option writers' hedging cost and hedging difficulty. Furthermore, when the noise traders' trading activities become more volatile, the hedging cost increases. Thus, the noise trader risk hypothesis predicts that the volatility of the retail trading activities should play a role, i.e., the options (both call and put) should be more expensive when the retail trading activities of the underlying stocks are more volatile.

To test the above predictions, I further calculate the aggregate retail order imbalance and the volatility of the retail trading activities, including the volatility of the trading volume and the volatility of the retail order imbalance. I then conduct the portfolio analysis using these measures. The empirical results show that when the stocks are associated with large net retail buying orders, the call options become more expensive; however, the put options also become more expensive. When the stocks are associated with large net retail selling orders, both the call options and put options also become more expensive. In short, the options are always more expensive regardless of whether the underlying stocks are associated with extreme retail buying orders or retail selling orders. This U-shape option price distribution helps reject the lottery premium hypothesis.

I further test the noise trader risk hypothesis and form decile option portfolios sorted by the volatility of the retail order imbalance and the volatility of the retail trading volume. In contrast to the results obtained from using directional order imbalance measures, the volatility of the retail order imbalance shows a strong negative relation with future option returns. When the stocks are associated with high volatility of the retail trading order imbalance, both the call options and put options become more expensive. The relation also holds when

I use the volatility of the retail trading volume or different weighting schemes. To summarize, these results together suggest that the direction of the retail trading activities does not affect the option pricing; but the volatility of the retail trading does. Thus, the noise trader risk explanation is more plausible to explain the role of retail investors in the option pricing. Further supporting the noise trader risk hypothesis, I find that the negative relation between retail trading volume and cross-sectional option returns becomes more significant when the retail trading activities are more volatile and when the stocks are associated with high arbitrage costs (i.e., high idiosyncratic volatility or high short interest). All these results consistently provide support for the noise trader risk hypothesis.

To provide more confidence of the findings on the role of retail investor trading activities, I conduct two additional tests using different testing samples. First, I explore whether the institutional trading activities affect the cross-sectional delta-hedged option returns by extracting the daily institutional trading orders from the Abel Noser data. Not surprisingly, I do not find any significant relation between the institutional trading activities and the cross-sectional option returns. Consistent with the notion that institutions are more rational, the test results support the argument that retail trading activities increase the noise trader risk in the option market. Second, I also examine the impact of different retail trading activities on the pricing of out-of-money options. Similar as the results obtained from the at-the-money option sample, I find that the retail trading volume and the volatility of the retail trading activities negatively predict future delta-hedged option returns. But the retail order imbalance does not affect the cross-sectional option returns. These two out-of-sample tests give us more confidence on the conclusion that retail investors act as noise trader risk in the option market.

This paper contributes to several strands of the literature. First, I conduct a comprehensive study of the impact of retail trading activities on the cross-sectional equity option pricing. A previous study by Choy (2015) examines the retail clientele in the option market. However, he only constructs the retail trading volume measure, and concludes that the retail investors' gambling demand increases the option prices. I further investigate the impacts of retail order imbalance and the volatility of the retail trading activities. More importantly, the results reject the gambling demand explanation (i.e., the lottery premium hypothesis), and highlight the noise trader risk explanation. Thus, this study provides new understanding of the underlying mechanism on how the retail investors' trading activities in the stock market could affect the option pricing.

Second, this study contributes to the literature on the cross-sectional option return predictability. Previous literature mainly focuses on the impacts of different kinds of volatilities on the option returns, including the volatility deviation (Goyal and Sarreto, 2009), idiosyncratic volatility (Cao and Han, 2013), the market volatility of volatility (Huang et al., 2019), the individual stock's volatility of volatility (Ruan, 2019; Cao, Vasquez, Xiao, and Zhan, 2021), and the volatility term structure (Vasquez, 2017). By investigating the trading activities of retail investors, I show that the composition of the market participants can also affect the option pricing. Furthermore, the trading activities of retail investors may provide a potential source of volatility or amplify the different kinds of volatilities.

Finally, this paper contributes to the general literature on retail investors. I show that the directional aggregate retail order imbalance of the underlying stocks does not predict future delta-hedged option returns. In contrast, the level and volatility of the retail trading activities matter. The more retail investors trade the stocks, the more expensive the corresponding options, both call options and put options. The results suggest that the retail investors act as noise traders in the option market, or at least, the option writers treat the retail investors as noise traders.

Section 1.2 in this paper describes the sample construction and variable definitions. Section 1.3 presents the main empirical results. Section 1.4 explores the underlying mechanisms. Section 1.5 provides more discussion, and Section 1.6 concludes the paper.

1.2 Data and Variable

1.2.1 Option Data

The option data is obtained from the OptionMetrics Ivy database, which covers all equity options from 1996. The database provides detailed option related information at the daily level, including the daily closing bid and ask quotes, the option open interest, daily trading volume, implied volatility, and option Greeks, such as delta, vega, and gamma.

I apply several commonly used filters to screen the options used in the sample (Cao and Han, 2013; Cao, Vasquez, Xiao, and Zhan, 2021). First, to avoid the illiquidity issue, I exclude the options if the bid quote equals to 0, the average bid and ask quotes is less than \$0.125, or the ask quote is equal to or less than the bid quote. Second, I only keep the options with moneyness (defined as the ratio of strike price and stock's close price) between the range of 0.8 to 1.2 and options with remaining maturity longer than one month but shorter than two months. Third, I exclude options that violate the no-arbitrage restrictions. Fourth, I exclude options whose underlying stocks will pay dividend during the option's remaining life, which ensures that the options are assemble to European style options. Finally, I select one call option and one put option that are closest to being at-the-money and are with the shortest maturity beyond one month. To make the option returns comparable among different stocks, in each month, I further exclude options with maturity date different with that of the majority of stock options.[2] The options in the final sample are generally associated with a maturity between 46 and 52 calendar days, and with the moneyness between 0.95 and 1.05.

1.2.2 Delta-hedged Option Return

Different with the stock price changes, the risk-return nature of option consists two separate components: the leverage component and the volatility component (Coval and Shumway, 2001). To remove the leverage effect, I calculate the delta-hedged option gains so that the option returns are not

2 In each month, the options of most stocks will mature at the end of the third week. But some stocks will have options expiring in each week. In this case, I only choose the options with the most common maturity, i.e. mature at the end of the third week.

sensitive to the stock price changes (Bakshi and Kapadia, 2003; Cao and Han, 2013; Choi, 2015).

To construct a delta-hedged call option portfolio, it will consist of a long position of one call option contract and a short position of delta shares of the underlying stock, where delta is the call option's Black-Scholes' delta. Under the continuous setting, the delta-hedged option gain for the call option from time t to time t+τ could be expressed as:

$$\prod(t, t+\tau) = C_{t+\tau} - C_t - \int_t^{t+\tau} \Delta_u dS_u - \int_t^{t+\tau} r_u(C_u - \Delta_u S_u) d_u \quad (1.1)$$

where C is the call option price, S is the stock price, Δ is the call option delta, and r is the risk-free rate. As the continuous hedging is impossible in reality, I transform the above expression into discrete format. Suppose the option is hedged N times over the period [t,t+τ] and the stock position is rebalanced at each time t_n. The discrete version of equation (1) will become:

$$\begin{aligned}\prod(t, t+\tau) = C_{t+\tau} - C_t - \sum_{n=0}^{n=N-1} \Delta_{c,t_n}[S(t_{n+1}) - S(t_n)] \\ - \sum_{n=0}^{n=N-1} r_{t_n}\left[C(t_n) - \Delta_{c,t_n} S(t_n)\right] \frac{\alpha_n}{365}\end{aligned} \quad (1.2)$$

where C is the call option price at the end of each date, S is the stock close price, Δ is the call option delta, r is the risk-free rate, and α is the number of calendar days between each rebalance. After obtaining the delta-hedged option gain, the delta-hedged call option return is defined as:

$$Oret_{t,t+\tau} = \frac{\prod(t, t+\tau)}{|C_t - |\Delta_t S_t||} \quad (1.3)$$

Where the denominator is the net initial investment to establish the long position of call option and the short position of the stocks. When calculating the delta-hedged put option return, I just replace the call option price and call option delta by the put option price and put option delta in equations (1.2) and (1.3).

[Insert Table 1.1 here]

Table 1.1 presents brief summary of the option sample. Besides the option return data, I also report the distribution of other option characteristics, including the days to maturity, moneyness, option vega (defined as the Black-Scholes' vega scaled by stock price), and the bid-ask spread (defined as the difference between bid and ask quotes scaled by the midpoint of the bid and ask quotes). The detailed variable construction is listed in Table 1.1. Panel A and Panel B show the sample distribution for delta-hedged call options and put options respectively. Consistent with Bakshi and Kapadia (2003) and Cao and Han (2013), the average return of both the call options and put options are significantly negative. The median returns for the call option and put option held until the maturity are -1.01% and -0.88% respectively. Both the mean and median of the time to maturity are 50 calendar days. The average option bid-ask spread is about 30%, with the median at about 20%. To be short, all the statistics are comparable with those reported in Cao and Han (2013).

1.2.3 Retail Trading Activities

Earlier studies on the retail trading activities either use the account-level data with small coverage or identify the retail orders based on the size of trading volume (Lee and Radhakrishna, 2000; Barber and Odean, 2000). In recent days, Campbell et al. (2009) find that the small trading orders are more likely to be submitted by institutions due to the adoption of computer algorithms. To identify the retail trading orders more precisely, I follow Boehmer et al. (2022) and extract the retail trading orders based on the concept of "price improvement".

In details, most of the trading orders submitted by retail investors are executed either by their brokerage house (i.e., internalized with broker's inventory) or by other wholesalers. These orders will be reported to a FINRA Trade Reporting Facility (TRF) with an exchange code "D". Due to the regulatory restrictions (e.g., Reg 606T and Reg NMS), trading orders from the retail investors will receive a price improvement relative to the National Best Bid or Offer (NBBO) price as a reward for providing order flows. As a result, the recorded trading price of retail orders are more likely to be in a format of XX.XXxx, which enables researchers to identify the retail trading orders. The price improvement is usually a very small fraction of a cent, such as 0.01 cent, 0.1 cent, or 0.2 cent. For example, TAQ data records a trading order of Agilent Technologies with a volume of 800 shares at a price of $41.0799 at 9:37:12.766288 on Jan 04, 2016. Although the dollar size of the order is very large, it can be inferred that this order comes from a retail investor based on the recorded transaction price.

To further distinguish whether the retail order is initiated by a retail buyer or by a retail seller, I extract the subpenny part of the recorded trading price. I define $\mathrm{Z} \equiv 100 \times \mathrm{mod}(\mathrm{P}, 0.01)$. Following Boehmer et al. (2022), if Z is larger than 0 and smaller than 0.4, the order is a retail seller initiated transaction. If Z is larger than 0.6 and smaller than 1, the transaction is initiated by retail buyers. Otherwise, I define the transaction as un-identified order flow. The criterion is relatively conservative, but makes the identified retail order more accurate. After identifying all the retail trading orders, I further construct different retail trading activity measures, including total retail trading volume (Rtrd), net retail buying volume (Rnby), and the retail trading order imbalance measure (OIbjzz) used in Boehmer et al. (2022).[3] I define the above measures for each stock *i* on each trading day *t* as followings:

$$R_{trd(i,t)} = \frac{Total\ retail\ buying\ dollar\ volume_{i,t} + Total\ retail\ selling\ dollar\ volume_{i,t}}{Total\ dollar\ trading\ volume_{i,t}} \quad (1.4)$$

$$R_{nby(i,t)} = \frac{Total\ retail\ buying\ dollar\ volume_{i,t} - Total\ retail\ selling\ dollar\ volume_{i,t}}{Total\ dollar\ trading\ volume_{i,t}} \quad (1.5)$$

$$OI_{bjzz(i,t)} = \frac{Total\ retail\ buying\ dollar\ volume_{i,t} - Total\ retail\ selling\ dollar\ volume_{i,t}}{Total\ retail\ buying\ dollar\ volume_{i,t} + Total\ retail\ selling\ dollar\ volume_{i,t}} \quad (1.6)$$

To match with the monthly option return data, I further calculate the average retail trading volume and average retail order imbalance in each month. The standard deviation of the above measures in each month are also calculated to reflect the volatility of the corresponding retail trading activities.

[Insert Table 1.2 here]

3 Both Rnby and OIbjzz reflect the degree of retail order imbalance. But the two measures are constructed using different denominators. The OIbjzz measure is equivalent to the ratio of retail buying orders and retail selling orders, thus ignores the absolute retail trading magnitude. See the transformation: $\frac{x-y}{x+y} = \frac{x/y-1}{x/y+1} \rightarrow \frac{x}{y}$

Table 1.2 reports the summary of the retail trading variables and other control variables. Panel A and Panel B show the distribution and correlation matrix respectively. The sample period is from 2010 to 2017, which is constrained by data availability of the option data and TAQ data. On average, the retail trading volume (Rtrd) accounts for about 6.63% of the total trading volume, which is similar as that (i.e., 6.91%) reported in Boehmer et al. (2022). Not surprisingly, the proportion of retail trading volume during the sample period is smaller than the recent estimation.[4] The net retail buying volume (Rnby) is a bit negative, but very close to zero, suggesting that the proportion of retail buying orders is similar to that of the retail selling orders.

For the control variables, I report the stock size in previous month end (Mep), the momentum effect (Mom), idiosyncratic volatility (IVol), expected idiosyncratic skewness (EIS), the Stambaugh and Yuan (2017) mispricing score (MIS), the stock return of previous month (STR), the maximum daily stock return in previous month (MAX), and the Amihud illiquidity (ILLIQ). The detailed variable construction is listed in Table 2. Panel B reports the Pearson correlation matrix between different variables. For both call options and put options, the correlation coefficients between the option returns and retail trading volume are all negative, suggesting that the option of stocks with higher retail trading volume will have lower future return. For other variables, IVol and EIS

4 For example, Joe Mecane, the head of execution services at Citadel Securities, estimates that retail investors make up about 10% of the market in 2019, and this proportion increase to about 20% or 25% in middle 2020. See more details at: https://www.bloomberg.com/news/videos/2020-07-09/citadel-s-mecane-says-volatility-behind-rise-in-retail-investing-video

are negatively correlated with the option returns, consistent with previous literature (Cao and Han, 2013; Boyer and Vorkink, 2014).

1.3 Baseline Empirical Analysis

I report the main result in this section. First, I present the Fama-Macbeth cross-sectional regression results. I then develop an option trading strategy based on the retail trading volume to explore the economic magnitude, and also examine the risk-adjusted excess returns for the zero-cost long-short portfolio. Lastly, some robustness tests are reported.

1.3.1 Delta-hedged Option Return and Retail Trading Volume: Regression Analysis

I first examine the relation between delta-hedged option return and the retail trading volume using the monthly cross-sectional Fama-Macbeth regression. In each month, I run the cross-sectional regression using the following model:

$$Oret_{t,t+\tau} = \beta_0 + \beta_1 Rtrd_t + \sum_{n=1}^{n=N} \beta_{n+1} X_t + \varepsilon_t \tag{1.7}$$

Where Oret is the option return of the delta-hedged call option or put option as defined in Equation (1.3); Rtrd is the average daily retail trading volume in previous month; β_1 is the coefficient of interest. X represents a set of control variables, including idiosyncratic volatility (IVol), expected idiosyncratic skewness (EIS), mispricing score (MIS), maximum daily return

(MAX), Amihud illiquidity (ILLIQ), option's bid-ask spread (OPBaspread), and stock's market cap at the end of previous month (Ln(Mep)). All the control variables are winsorized at the top 0.5% and bottom 0.5%. To make the regression coefficients comparable, each month, I also normalize all variables to be N(0,1). The average of the time-series coefficients estimated from the monthly cross-sectional regression are reported in Table 1.3. The corresponding t-statistics are adjusted following Newey and West (1987).

[Insert Table 1.3 here]

Table 1.3 Column (1) reports the univariate regression result of delta-hedged call option return on the average retail trading volume in previous month.[5] The significant negative coefficient (coeff = −0.66; t-stat = −11.81) suggests a strong negative relation between the future cross-sectional option returns and previous retail trading volume of the underlying stocks. Column (2) shows the multivariate regression result. The coefficient of retail trading volume (Rtrd) is still negative and significant at the 1% level. For the control variables, consistent with previous literature (Cao and Han, 2013; Boyer and Vorkink, 2014), the coefficients of idiosyncratic volatility (IVol), expected idiosyncratic skewness (EIS), and option bid-ask spread (OPBaspread) are all negative. More importantly, the coefficient of Rtrd is at least comparable or even larger than those of the other control variables, suggesting that the impact of retail trading volume on cross-sectional option returns is economically meaningful. Column

[5] The main analysis is based on the Fama-Macbeth cross-sectional regression. In each month, I run a cross-sectional regression, and the reported results are the time-series average of the regression coefficients. In this method, the (adjusted) R-square and the number of observations will also be the time-series average across different months. Econometrically, these statistics do not provide much information on the model explanation power. Thus, in the reported results, I omit these statistics.

(3) and (4) report the results for delta-hedged put options. All the results are similar as in Column (1) and (2).

Overall, the regression results provide compelling evidence that the retail trading volume of the underlying stocks is priced in the cross-sectional option returns. Both the call options and put options are associated with lower future returns if the underlying stocks have higher retail trading activities. The negative relation cannot be fully explained by the existing factors, such as idiosyncratic volatility, return skewness, stock mispricing, and illiquidity.

1.3.2 Portfolio-level Analysis: Profitability of Holding Options

The above regression results show that the option prices are higher if the underlying stocks have higher retail trading volume. I further explore the economic magnitude of the pricing impact of retail trading activities by developing an option trading strategy and illustrate the profitability using portfolio analysis in this section. For each option, I construct the delta-neutral option combination, which consists of a long position in the call (put) option and delta shares of short (long) position of the corresponding stock. The delta-neutral option-stock combination will be rebalanced daily, and be held until the option expires.

At the month end, I establish the delta-neutral option-stock combination (buy option and hedged with stocks) for all optionable stocks. Then all the stocks are sorted into decile portfolios based on the corresponding retail trading volume in previous month. The trading strategy is a zero-cost long-short hedge portfolio that long the high retail trading decile and short the low retail trading decile.

[Insert Table 1.4 here]

[Insert Figure 1.1 here]

Table 1.4 reports the average portfolio returns of the decile portfolios and the excess return of the zero-cost long-short hedge portfolio. Panel A and B show the results for call option and put option respectively. I also visualize the results in Figure 1.1. The results show that all the raw returns of the decile portfolios are negative, indicating that holding delta-hedged options have negative return on average. More importantly, the portfolio returns are more negative as the retail trading volume increases for both call option and put option. Table 1.4 Panel A shows that high retail trading decile portfolio is associated with lowest portfolio return and the return spread between the high retail decile and low retail decile ranges from -0.84% (t-stat=-5.75) to -2.17% (t-stat=-12.79) under different weighting schemes. Panel B shows that the portfolio returns of holding put options exhibits similar pattern as those of holding call options in Panel A. All these results together suggest that holding options are associated with negative return, especially holding options of stocks with high retail trading volume.

[Insert Table 1.5 here]

I further examine whether the return spread between high retail decile and low retail decile portfolio could be explained by existing common risk factors, such as the Fama-French three factors, the momentum factor, the zero-beta straddle option return of the S&P 500 index option, and the change in the VIX index. I regress the long-short hedge portfolio return in Table 1.4 on different risk factors or a combination of the factors, and report the regression alphas and factor loadings in Table 1.5. To save space, I only report the results using equal-weighted weighting scheme. Not surprisingly, the stock market risk factors do not provide much explanation power for the option returns. The factor loadings (i.e., coefficients) are generally very small and insignificant. The risk-adjusted alphas are almost the same as the raw long-short excess returns as in Table 1.4. In sum, the portfolio sorting results are consistent with the regression results, and the existing common risk factors do not provide much explanation in the option market.

1.3.3 Robustness Tests

I conduct a set of robustness checks in this section. I use different holding periods to calculate the option returns and different methods to construct the retail trading volume measures, and then re-run the monthly regressions using these alternative measures. In the baseline analysis, once the option position is established, it is held until maturity. Here, I also calculate alternative option returns assuming the option is only held until next month end (rather than until maturity). For the retail trading volume measures, I use three different construction methods. Instead of using the dollar volume of the trade, I also use the number of trades, and define the retail trading volume as the number of retail investor initiated trades scaled by the total number of trades, i.e., variable Rtrd_trades. Secondly, instead of using the total trading volume on that day as the denominator, I also scale the retail trading volume by the average total trading volume in previous one year, and define the variable as Rtrd_vol1y. Lastly, I define the retail trading volume as the total retail trading volume in previous month scaled by the total trading volume in previous month, and define the variable as Rtrd_total.

[Insert Table 1.6 here]

Table 1.6 reports the robustness test results. Column (1) to (4) show the call option results, and Column (5) to (8) show the put option results. In all regressions, the coefficients on the retail trading measures are always negative and significant at the 1% level. Thus, the negative relation between retail trading volume and future cross-sectional option returns is robust using alternative variable constructions.

1.4 Explore the Underlying Mechanism

Section 3 documents a robust and significant negative relation between the option returns and the retail trading volume. The negative relation could be driven by the lottery demand from the retail investors or could be driven by the increased noise trade risk exposure of the option writers. I conduct additional tests to distinguish the underlying mechanism that drives the negative relation.

1.4.1 Lottery Premium Hypothesis: The Role of Retail Order Imbalance

The lottery premium hypothesis suggests that retail investors speculating on the stocks are also willing to pay a lottery premium on the corresponding stock options, which will bid up the option prices and lead to lower future option returns. Generally, the trading direction of retail investors in the stock market and option market will be positively correlated. When the retail investors expect the stock price to go up, they will long the stocks and long the call options. These investors with the "similar belief" collectively will generate excess demand on the call options. Based on the demand-supply relationship, I expect the call options will be sold at a premium.[6] In addition to this main prediction, the lottery premium hypothesis suggests that the direction of the retail speculation should also affect the lottery premium. I conjecture that if the retail investors collectively bet on the price increasing, the call options should enjoy a lottery premium; but the put options should not. To be short, if the lottery premium hypothesis holds, the direction of retail trade (i.e., retail trading order imbalance) should also affect the option pricing.

Empirically, I construct two retail order imbalance measures. Rnby is defined as the daily net retail buying volume scaled by the total trading volume on that day (see Equation 1.5), and I take the average value in the previous month as the first retail order imbalance measure. OIbjzz is defined as the daily net retail buying volume scaled by the total retail trading volume on that day (see Equation 1.6), and I also take the average value during the previous month as the second retail order imbalance measure. Note that, the second measure is equivalent to the ratio of the retail buying volume and retail selling volume. After I construct the two measures, I repeat the portfolio analysis as in Table 1.4 to test the pricing power of the retail order imbalance.

[6] I acknowledge that this assumption needs to be validated, especially for the high retail trading stocks. It is also possible that retail investors long the stocks, and short the call options to hedge. However, due to the data limitation, I am not able to conduct such tests at this stage.

[Insert Table 1.7 here]

[Insert Figure 1.2 here]

Table 1.7 reports the results. I also visualize the results in Figure 1.2. Interestingly, I find a reverse U-shape distribution of the raw returns along the decile portfolios. Take Panel A Column (1) for example, the delta-hedged call option generates -1.56% monthly return for stocks associated with extreme retail selling orders. This return increases to -0.68% and -0.75% when the stocks are associated with moderate or neutral retail order flows; but further decreases to -1.59% when the stocks are associated with extreme retail buying orders. This reverse U-sharp pattern exists for both call options and put options and for different weighting schemes. More importantly, most of the long-short hedge returns are not significant, suggesting that the retail order imbalance is not priced in the cross-sectional option prices.

I further compare the results in Panel A and Panel B. The stocks in the same decile portfolios in the two panels are roughly the same set of stocks. For example, the stocks in the first decile in Panel A should roughly be the same as the stocks in the first decile in Panel B, and these two deciles contain those stocks associated with extreme retail selling volume.[7] In this case, the retail investors sell the stocks on aggregate, suggesting they bet on the future price drops. Thus, call options should not enjoy a lottery premium as the speculation

7 Theoretically, the stocks in the two deciles should be exactly the same. However, not all stocks are associated with a pair of one call option and one put option. Due to the data screening, some stocks in some months will only have one option chosen. Thus, the stocks used in the call option test and put option test are not exactly the same. The different numbers of observation between call option sample and put option sample in Table 1 also indicate this issue.

on call options is weak; in contrast, put options should charge high prices. An asymmetric return pattern should be observed. However, the empirical results suggest both the call options and put options for the same stocks become more expensive, which is inconsistent with the lottery premium hypothesis. Overall, the results suggest that the retail trading order imbalance is not priced in the cross-sectional option returns. When the stocks are associated with either extreme retail buying volume or extreme retail selling volume, both the call options and put options become more expensive. Thus, the lottery premium hypothesis is not able to explain the negative relation between retail trading volume and cross-sectional option returns.

1.4.2 Noise Trader Risk Hypothesis: The Role of Volatility of the Retail Trading Activities

I further investigate whether the noise trader risk hypothesis drives the findings. Different with the lottery premium hypothesis, the noise trader risk hypothesis does not lead to a directional predication. However, it predicts that the more volatile the retail trading activities, the stronger the negative relation between the retail trading activities and cross-sectional option returns. When the retail trading activities become more active and more volatile, the corresponding noise trader risk becomes intensified. Generally, when the option writers sell options, they will hedge in the stock market, making the options covered options. The hedging amount will depend on certain pricing models. The noise risk associated with the retail investor trading activities could affect the volatility of the underlying stocks, making the option pricing model less accurate. The option writers could suffer from such a model risk. In addition, the noise risk will also increase the volatility of the stock volatilities. The option writers may need to hedge more frequently, which incurs high cost. Lastly, the noise traders could also affect the option writers' ability to conduct hedging. For example, it is possible that there's not enough stocks to borrow or to buy from the equity market. Thus, the option writers should charge higher option prices to compensate this risk.

To test this prediction, I construct the retail investor trading volatility measures for the retail trading volume and the two retail trading order imbalance variables. As the retail trading order is at the daily level, I further calculate the standard deviations of the corresponding retail trading volume and the two retail order imbalance measures in the previous month, and define Std(Rtrd), Std(Rnby), and Std(OIbjzz) as the retail trading volatility measures. After that, I first examine whether the trading volatility measures could predict the cross-sectional option returns by conducting the portfolio analysis sorted by these three measures.

[Insert Table 1.8 here]

Table 1.8 reports the results. Panel A and B show the results for call option and put option respectively. For the trading volume volatility measure Std(Rtrd) and the first order imbalance volatility measure Std(Rnby), all the return spreads between high volatility portfolio and low volatility portfolio are statistically significant, suggesting that the volatility of the retail trading activities are priced in the cross-sectional option returns.[8]

To further investigate the impact of retail trading volatility on the relation between retail trading volume and future option returns, I conduct monthly Fama-MacBeth regression with the interaction terms between retail trading volume and the retail trading volatility using the following model.

$$Oret_{t,t+\tau} = \beta_0 + \beta_1 Rtrd_t + \beta_2 Rtrd_t \times VR_t + \beta_3 VR_t + \sum_{n=1}^{n=N} \beta_{n+3} X_t + \varepsilon_t \quad (1.8)$$

Where VR is the volatility of the retail trading activities, and β_2 is the coefficient of interest, which captures the impact of retail trading volatility on the relation between retail trading volume and future option returns. I use four measures to reflect the volatility of retail trading activities, including two high volatility indicators and two continuous volatility variables. The two continuous variables are Std(Rnby) and Std(Rtrd). The two high volatility indicators are defined as 1 if Std(Rtrd) or Std(Rnby) is above the 80% percentile among all stocks in each month, and 0 otherwise.

[Insert Table 1.9 here]

8 The results based on the Std(OIbjzz) measure are a bit weak. Possible explanation is that the OIbjzz measure only considers the ratio between retail buying volume and selling volume, and ignores the absolute retail trading level. For example, the retail investors buy 2 shares and sell 1 share, then OIbjzz will have a value of 2. But this small trading activity may not affect the option writers. In the following Fama-MacBeth regression, I will only use the first two retail trading volatility measures.

Table 1.9 reports the results. Column (1) to (4) show the results for call options and Column (5) to (8) show the results for put options. The coefficients of the interaction terms between the retail trading volume and the two high retail trading volatility indicators are almost two times larger than the coefficients of the retail trading volume itself (Rtrd), suggesting that the impact of retail trading volume becomes much stronger when the retail trading activity is also more volatile. The results are similar when using the continuous volatility measures. Taken together, the results suggest the volatility of the retail trading activities are priced in the cross-sectional option returns, and the pricing impact of retail trading volume becomes much stronger when the retail trading activity is also more volatile. All these results support the noise trader risk hypothesis: the existence of retail investors increases the option writers' hedging cost and hedging difficulty, and the option writers charge higher option price to compensate this noise trader risk.

1.4.3 Noise Trader Risk Hypothesis: The Role of Arbitrage Cost

To provide more evidence on the noise trader risk hypothesis, I further explore the role of arbitrage cost on the pricing impact of retail trading volume. Similar as the impact of retail trading volatility on the relation between retail trading volume and option returns, I conjecture that the pricing impact of retail trading volume should be stronger when the stocks are associated with higher arbitrage cost. When the stocks have high arbitrage cost, option writers will be more sensitive to additional hedging risks due to the convexity of the option returns, making the impact of noise trader risk more significant. To empirically test the prediction, I construct two measures of the arbitrage cost: the idiosyncratic volatility (IVol) and the short interest (SIR). I then run the monthly Fama-MacBeth regression with the interaction terms between retail trading volume and the arbitrage cost measures. Both indicator variables and continuous variables are used in the regression.

[Insert Table 1.10 here]

Table 1.10 reports the results. Column (1) to (4) show the results for call options and Column (5) to (8) show the results for put options. All the interaction terms are associated with negative coefficients which are significant at the 1% level, suggesting that the negative relation between retail trading volume and future option returns becomes much stronger when the stocks are associated with high arbitrage costs. Take Column (1) as an example, the coefficient of the interaction term is -0.35 (t-stat=-7.76), and the coefficient of the retail trading volume (Rtrd) is only -0.20 (t-stat=-4.14). The pricing effect of the retail trading volume almost becomes tripled among the high IVol stocks. Results using put options also show similar patterns. In sum, these results support the noise trader risk hypothesis.

1.4.4 The Role of Institutional Trading Activities on Option Returns

All of the above tests examine the impact of retail investor trading activities on the cross-sectional option returns. The results suggest the retail investors are treated as noise traders, who amplify the noise trader risks faced by the option writers. However, there may still exist some unobservable factors driving the empirical findings. To provide more confidence on the interpretation, I further investigate the role of institutional investors. In details, I examine whether the institutional trading activities in the stock market will affect the cross-sectional option returns in the derivative market. Compared with retail investors, the trading activities of institutional investors should be more rational. Thus, I predict that the institutional trading activities should not affect the delta-hedged option returns.[9] Note that, the delta-hedging strategy largely rules out the impact of stock price changes on the option returns.

In the empirical test, I extract institutional trading orders from the Abel Noser dataset. The Abel Noser is a brokerage firm that provides transaction cost optimization analysis for institutional clients. Their clients mainly covers mutual fund investment managers and plan sponsors. Although the data does not cover all the institutional trading orders, it is still a most widely used dataset in the literature (Puckett and Yan, 2011; Cready, Kumas, and Subasi, 2014). For each stock on each day, I calculate the total institutional buying volume and selling volume across different institutions (i.e., clients). The total institutional trading volume (Itrd) and the net institutional buying volume (Iimb) are calculated as the total institutional trading volume or the net institutional buying volume scaled by the total trading volume on that day. Lastly, I calculate the average trading volume and order imbalance in that month as the variables used

9 A natural question might be that the counterpart of the retail trading is the institutional trading. While this may not be the case in the empirical tests. First, the retail trading orders used in this study only contains the market orders, and the limit orders are excluded. Thus, the retail orders mainly reflect the trading behavior of the active trading retail investors. I also exclude some un-identified orders. Second, a considerable institutional trading is executed in the format of cross-trading - transactions within the same fund family – that are not exposed to an external marketplace (Chan, Conrad, Hu, and Wahal, 2018; Eisele, Nefedova, Parise, and Peijnenburg, 2020).

in the test. I also calculate the standard deviations of the two variables in that month as the volatility of the institutional trading activities.

[Insert Table 1.11 here]

Table 1.11 reports the empirical results. The sample period is from 1999 to 2011, which is constrained by the Aber Noser dataset. Panel A shows the distribution of the institutional trading measures. On average, the Aber Noser institutional trading accounts for 6.30% of the trading volume in the sample. This number is a bit smaller than those (e.g., 8%) reported in previous literature, but still comparable.[10] Panel B shows the monthly regression results of the option returns on different institutional trading activities. Column (1) to (4) show the results for call options and Column (5) to (8) show the results for put options. Four institutional trading measures are used in the tests, including the total trading volume (Itrd), the volatility of the institutional trading volume (Std(Itrd)), the net institutional buying volume (Inby), and the volatility of the net buying volume (Std(Inby)). Not surprisingly, none of the four measures show significant relation with the cross-sectional option returns, suggesting that the trading activities of institutional investors are not priced in the cross-sectional option pricing.

This sharp difference between the results using retail investor trading activities and institutional trading activities provides us more confidence to conclude that the retail investors act as noise traders in the derivative market.

10 For example, Puckett and Yan (2011) report that the Abel Noser trading activities accounts for about 8% of the universe CRSP trading volume. As the sample in this paper only covers optionable stocks, which should be larger than the average CRSP universe, thus the 6.30% proportion estimated in this paper is reasonable and comparable with these reported.

The retail trading activities increase the option writers' hedging cost and hedging difficulty, thus the option writers charge higher option prices. However, it is also a bit surprising that there does not exist any significant relationship. Theoretically, more institutional trading could also have a higher adverse selection risk for option writers. While, compared with the retail investor trading, I expect the impact should be much weaker. The option writers should be easier to hedge against the option positions after they write the options. In addition, there could also exist some data limitations in the current test. For example, the sample period is different with that used for testing the retail investor trading activities; the Abel Noser data only cover a small proportion of the institutions, and the covered institutions are mostly mutual funds. Given these limitations, the insignificant results still provide some suggestive interpretation and support the main argument.

1.4.5 Evidence From Out-of-The-Money Options

In this section, I re-examine the impact of different retail trading activities on the cross-sectional option returns by using the out-of-money options as the testing sample. Previous literature suggests that retail investors prefer stocks with positive skewness in the payoffs (Han and Kumar 2013). Compared with the at-the-money options, out-of-money options are more likely to fit the retail appetite. At the same time, the out-of-money options are less liquid compared with the at-the-money options, making the option writers' hedging activities more costly and difficult. To be short, using the out-of-money options as the testing asset should amplify the impact (if any) of both lottery premium hypothesis and noise trader risk hypothesis, which in turn enables us to further distinguish the two hypotheses.

To construct the out-of-money option sample, I filter the option sample following the same standards as used in the main test, but restrict the moneyness of the call options to be within 1.1 and 1.2, and restrict the moneyness of the put options to be within 0.8 and 0.9. Within this range, I further select call options whose moneyness is closest to 1.15 and select put options with moneyness closest to 0.85.[11] Then, I re-run the monthly Fama-Macbeth regression to test the pricing effects of different retail trading activities, including the total retail trading volume (Rtrd), the net retail buying volume (Rnby), and the two volatility measures of the retail trading activities.

[Insert Table 1.12 here]

Table 1.12 reports the empirical results. Panels A and B show the sample distribution of the out-of-money options. The average return of delta-hedged out-of-money call (put) options held until maturity is -4.82% (-3.92%), which is much larger than that of the at-the-money options as reported in Table 1. The moneyness of call (put) option is about 1.1323 (0.8689), consistent with the pre-set screening standard (1.15 for call options and 0.85 for put options). Not surprisingly, the option vega becomes much smaller compared with the at-the-money options, and the bid-ask spread becomes much larger.

11 The results are similar if I select call options with moneyness closest to 1.1 or 1.2 or select put options with moneyness closest to 0.9 or 0.8.

Panel C shows the monthly regression of the out-of-money option returns on different retail trading activities. I also include all the control variables as in Table 3, but do not report the coefficients (except for IVol). Column (1) to (4) show the results for call options and Column (5) to (8) show the results for put options. All the variables are standardized to N(0,1), thus the regression coefficients are directly comparable. Specifically, the coefficients of the total retail trading volume (-0.79 for call, and -0.70 for put) are much larger than those in Table 3 (-0.36 for call, and -0.37 for put), suggesting that the retail trading volume is associated with much stronger impact on the out-of-money option returns. It is worth to note that the coefficients of the retail trading volume (Rtrd) are also much larger than those for IVol (-0.28 for call, and -0.20 for put). In Column (2), the coefficient of the net retail buying volume is not significant, suggesting that the retail investors' trading direction does not affect the call option prices. Further, the coefficients of the two trading volatility measures are significantly negative, consistent with the prediction of the noise trader risk hypothesis. I do acknowledge that the coefficient of the net retail buying volume is positive in Column (6), indicating that when the retail investors buy the stocks, the corresponding put option price becomes a bit cheaper, which is consistent with the lottery premium hypothesis. However, the magnitude of the coefficient is much smaller compared with those of the other retail trading activities. Overall, Table 1.12 support the noise trader risk hypothesis and reject the lottery premium hypothesis.

1.5 Further Discussion

In this chapter, I document a robust and significant negative relationship between the retail trading volume of the underlying stocks and the expected future options. The empirical analyses collectively support that intensive retail investor trading activities could increase the noise trader risk of the underlying stocks, which further increases the hedging cost and hedging difficulty of the option writers or the market makers. Thus, the options (both call option and put option) will become more expensive, leading to lower expected returns for holding options. In the following section, I will briefly discuss some other potential issues that could affect the documented relationship.

1.5.1 The Impact of Retail Option Trader

The very assumption in this chapter is that the retail investors' trading direction will be the same both in the option market and in the stock market. Generally, the retail investors are thought to be the end users in the option market. Thus, on average, the retail investors will hold long position of options. When the retail investors expect the stock price to go up and long the stocks in the stock market, they are highly possible to buy the corresponding call options in the option market, i.e. retail investors buying the stocks will also buy the call options, and retail investors selling the stocks will buy the put options. However, I acknowledge that this assumption needs to be verified, especially for the high-retail-trading stocks. It is possible that the retail investors long stocks and short options, or vice versa. Due to the lack of option trading data, I can not test this assumption at current stage.

Another related point is that, the retail investors could also trade the options. Thus, the documented relationship could be driven by the retail trading activities of the options, rather than the retail trading of the stocks. While, theoretically, one can think that the retail order flow in the option market could be driven by the retail order flow in the stock market. The retail trading activities in the stock market should be the fundamental driver. In the future, if I could get the retail trading data in the option market, I can add the retail trading of the options as additional control variables and check whether the retail trading in stocks or the retail trading in options exhibits stronger impact.

1.5.2 Other Possible Channels

First, the noise trader risk could also affect the bid-ask spread of the options, which further affect the pricing efficiency of options. At the same time, the retail trading may also affect the stock short selling activities in the equity market, which could further affect the option market makers' or the option writers' hedging difficulty and hedging cost. While, my current empirical tests can not rule out these potential channels. However, to some degree, these channels could be partially related to the noise trade risk associated with the retail investors. The increased noise trader risk could affect the stock volatility or the volatility of volatility, which further affect the option pricing efficiency and the stock short selling activities.

A second concern is that some omitted variables could drive the retail trading activities and the option return together. For example, Cao, Han, Tong, and Zhan (2022) find that the expected returns to writing delta-hedged calls are negatively correlated with the stock price, profit margin, and firm profitability, but positively correlated with cash holding, cash flow variance, new shares issuance, total external financing, distress risk, and dispersion of analysts' forecasts. The retail trading activities could be correlated or be driven by these existing option return predictors. Compared with those firm fundamentals, which are relatively stable among a long period, the retail trading is more likely to be driven by short-term sentiment or short-term irrational reasons. In this sense, the two sets of factors could come from different dimensions, thus the correlation could be low. However, I also agree that certain firms with certain fundamentals could attract more retail investors. Instead of testing the impact of retail trading level, examining whether the change of the retail trading activities of the same firm time to time could provide more evidence. In the current empirical results, Table 1.8 provides some hints. When the retail trading activities become more volatile, regardless of the retail trading levels, the option will become more expensive. This result suggests the time-to-time change of the retail investor trading also affects the option return. It could partially mitigate the concerns on the omitted variables. However, it is better to construct and control for these factors in the empirical analysis.

1.6 Conclusion

The recent pandemic shutdown and the innovation of the zero-trading platform attract a large proportion of retail investors into the US equity market. These retail army have deeply impressed the Wall Street by bidding up the stock price quickly and suddenly, and more importantly, far away from the intrinsic stock value. This trend also makes the option trading or option hedging more popular than before. In this paper, I comprehensively examine the impact of retail trading activities on the cross-sectional option pricing. The findings suggest a negative relation between the stock option returns and the corresponding retail trading activities of the underlying stocks. Both the delta-hedge call and put options of the stocks with high retail trading volume are associated with lower future returns. Further, I find the aggregate retail investors' trading direction (i.e., the order imbalance) is not priced. When the aggregate retail investors bet on the price increases, the corresponding put options do not become cheaper. Thus, the negative relation is not driven by the speculation demand (i.e. gambling demand). In contrast, the negative relation becomes stronger when the retail investors' trading activities are more volatile or the stocks are associated with high arbitrage cost. More interestingly, the institutional trading activities do not affect the cross-sectional option returns. Thus, the findings suggest that the retail trading activities act as potential noise trader risk, which increases the option writers' hedging cost and hedging difficulty. The option writers charge higher option prices.

The empirical results raise new interest on the role of retail investors in the financial market. Although some literature shows the aggregate retail order imbalance could positively predict future stock returns, I show evidence that the option writers still treat the retail investors as noise traders and ignore the potential information released from the aggregate retail order imbalance. Both call options and put options are charged higher prices regardless of whether the underlying stocks are associated with net retail buying orders or net retail selling orders. This study echoes the recent stock market "wars" between retail investors and the Wall Street. But I acknowledge that it is more difficult to draw broader conclusions on whether the retail investors are informed or not. I leave this to future research.

Chapter 2: Dissecting Retail Trading Orders

2.1 Introduction

Retail investors can be classified as informed traders (Kaniel et al., 2012; Kelley and Tetlock, 2013; Boehmer et al., 2022), liquidity providers (Grossman and Miller, 1988; Kaniel, et al., 2008; Barrot et al., 2016), and noise traders (Black, 1986; Barber and Odean, 2000; Barber et al., 2009). Existing research presents conflicting conclusions about the role and performance of retail investors. Severe limitations in the datasets used in earlier studies provide at least some reason for the mixed empirical findings (Battalio and Loughran, 2008; Barber, Odean, and Zhu, 2009; Kelley and Tetlock, 2013).[12] These limitations may lead to biased inferences from retail investors' trading behavior.

In addition, most studies use the aggregate net order flow (i.e., net order imbalance) as a proxy for collective retail trading activity, which assumes the buying orders and selling orders have equal and symmetrical impact on stock prices.[13] However, the assumption of symmetry may be unwarranted. Brennan et al. (2012) estimate the buy- and sell-order illiquidity and find that of the two, sell-order illiquidity shows stronger return predictability. Brennan et al. (2016) document that the probability of informed trading based on bad news (but not good news) significantly affects the cost of equity. The psychology literature

12 The potential data limitations in previous studies include but are not limited to: 1) limited sample periods, 2) small stock coverage, 3) investor heterogeneity across different brokerage firms, 4) distorted order submissions to the NYSE, and 5) use of proprietary data.

13 To construct the order imbalance measure, researchers use aggregate buying orders minus aggregate selling orders and then scale net buying activity using different measures, such as total retail trading volume or total trading volume. See for example, Kaniel et al., (2008), Barber, Odean, and Zhu (2009), Kelley and Tetlock (2013), and Boehmer et al. (2022).

also shows that the emotional effects of fear and greed on investors' decision-making are asymmetric. Motivated by the above arguments, I re-examine the role and performance of retail investors using the high-coverage and publicly available NYSE TAQ data and I investigate aggregate retail buying and selling activities separately. My results suggest that aggregate retail selling orders predict cross-sectional stock returns, whereas aggregate retail buying orders do not. The previously documented positive relation between retail order imbalance and cross-sectional stock returns is mainly driven by the negative effect of retail selling orders on pricing.[14]

Following Boehmer et al. (2022), I use the concept of "price improvement" to identify and extract retail trading orders from the NYSE TAQ data for each individual stock at the daily level. TAQ trading data are publicly available and provide all millisecond transactions for all stocks listed on national exchanges in the U.S. Due to the regulatory restrictions (e.g. Reg 606T and Reg NMS) and institutional arrangements, trading orders from retail investors can receive price improvements, measured in small fractions of a cent per share. As a result, the trading prices of retail orders are more likely to be in the format XX.XXxx.[15] This mechanism enables researchers to identify a large proportion of the retail trading orders (Boehmer et al., 2022). I then aggregate

14 Both the informed trader explanation and liquidity provider explanation predict a positive relation between retail order imbalance and future stock returns.

15 For example, TAQ reports a trading order for Agilent Technologies with a volume of 800 shares at a price of $41.0799 at 9:37:12.766288 on Jan 04, 2016. The market stock price should be $41.08 per share. However, when the retail investor submits a buying order, the dealer gives a small discount, i.e., price improvement, to reward the retail investor for providing liquidity. The final trading price for the retail order is recorded as $41.0799 in this case. The dollar volume is approximately $32,864, which is classified as an institutional order based on the traditional trading size classification algorithm. Thus, the price improvement characteristic enables researchers to identify retail orders more precisely.

the retail trading orders for a specific stock within the same day and construct the aggregate retail order imbalance, aggregate retail buying orders, and aggregate retail selling orders.[16]

My primary results show that the previously documented positive relation between the aggregate retail order imbalance (i.e., net buying orders) and cross-sectional stock returns still holds in my sample. More surprisingly, I find that the pricing ability of the retail order imbalance mainly comes from selling orders instead of from buying orders. The pure aggregate buying orders do not positively predict future stock returns, whereas the pure aggregate selling orders significantly predict negative future stock returns, implying that retail selling activity is more powerful in terms of moving the market. Even after controlling for the aggregate retail order imbalance and other well-known return predictors, aggregate retail selling orders still predict the cross-sectional stock returns. Specifically, a daily rebalanced long–short decile hedge portfolio based on aggregate retail selling orders produces excess returns of -10.42 bps (*t*-stat = -7.10) for an EW portfolio and -9.14 bps (*t*-stat = -5.77) for a VW portfolio, representing approximate -2% per month. When I consider aggregate retail buying orders, the corresponding excess return for the VW portfolio become much smaller and insignificant. When comparing at the weekly horizon, the excess returns for the long-short hedge portfolio based on retail selling orders become -33.07 bps (*t*-stat = -4.19) for an EW portfolio and -44.85 bps (*t*-stat = -4.81) for a VW portfolio, which are still more than -1.3% per month.

16 A natural question is whether the retail trading I identify is merely the counterpart of institutional trading. I further discuss retail trading data in the following sections. Readers can also refer to Boehmer et al. (2022) for a comprehensive discussion.

Meanwhile, the signs on excess returns for a long-short hedged portfolio using retail buying orders flip sign and become -8.11 bps (*t*-stat = -1.01) and -21.72 bps (*t*-stat = -1.99) for EW and VW portfolios, respectively.

This asymmetric return predictability still holds when I use the Fama–MacBeth (1973) regression analysis while controlling for many other return predictors. The coefficients of the aggregate retail selling orders are -2.28 (*t*-stat = -6.22) for an equal-weighted least squares (EWLS) Fama–MacBeth regression and -1.97 (*t*-stat = -3.23) for a value-weighted least squares (VWLS) Fama–MacBeth regression.[17] The coefficients of the retail buying orders are 0.78 (*t*-stat = 2.07) and -1.50 (*t*-stat = -2.47) for the EWLS and VWLS Fama–MacBeth regressions, respectively. The coefficients for the retail buying orders exhibit much smaller magnitudes and even flip signs when the VWLS regression is used. The above results collectively provide strong support for the notion that the pricing power of the aggregate retail order imbalance is derived mainly from aggregate retail selling orders.

This asymmetric pricing effect between retail buying and selling orders raises a great challenge to studies of retail trading. Both the informed retail investor explanation and the liquidity provision explanation have no such asymmetric predictions.[18] To explain my results, I conjecture that buying and

17 To make the coefficients comparable across different characteristics each day, I standardize all characteristics to a N(0,1) normal distribution. Generally, the difference between the top and bottom 10% samples in the distribution is approximately 3.3 times the standard deviation. When converting the regression coefficients into the excess returns of a long-short portfolio, the magnitudes are approximately 7.5 and 6.5 bps for EW and VW hedge portfolios, respectively. These magnitudes are comparable with that in the portfolio analysis.

18 Most previous studies use the net order imbalance, which does not distinguish between buying and selling orders. However, there are notable exceptions. For example, Kelley and Tetlock (2017) find that increased retail short selling predicts lower future stock returns.

selling decisions involve different processes. Brennan et al. (2016) document that "investors who take long positions will be more concerned about informed selling than about informed buying since the former depresses the sale price whereas the latter raises it." Tetlock (2007) finds that fluctuations in negative words in news items from the Wall Street Journal are associated with stronger market reactions than fluctuations in positive words. Brennan et al. (2012) show that sell-order lambdas are generally larger than buy-order lambdas, suggesting that selling orders have more influence on stock prices. Thus, I argue that stocks associated with intensive retail selling orders are more likely to attract the attention of retail investors and experience negative investor sentiment, driving the price to deviate further from the underlying fundamentals (Shleifer and Summers, 1990).[19] Compared with institutional investors, retail investors are less sophisticated, less rational, and more emotional. Thus, retail trading activity is more likely to be affected by investor sentiment (Lee et al., 1991; Barber and Odean, 2000; Barher et al., 2009). Given that retail investors generally only hold long positions, negative sentiment (or fear sentiment) should have more

Boehmer et al. (2008) find that institutional short sellers predict stock returns, whereas other short sellers (such as retail short sellers) do not. However, short selling is totally different from selling; it is considered much riskier and is rare among retail investors, who generally only hold long positions.

[19] To drive the stock price, the retail selling orders could either "move" the price or "predict" the price. When the selling pressure is extremely large, the equity price could be "mechanically" moved down. While, when the trading volume is not so large or when the market is deep enough, the trade direction may reflect some undiscovered information, which will predict the future return. If it is the "moving" argument, we may expect there to be a reverse of the stock price in a short horizon. If it is the "predicting" argument, the price may not rebound. However, observing such a rebound or not is not enough to distinguish the two channels. In my current empirical test, I also examine the long-horizon performance of the portfolios, and the results provide some support for the "predicting" argument. The selling orders could contain some information. While, at the same time, the selling order could also be more powerful to "move" the market as it could exert larger impact on the stock liquidity. It's more likely that the two channels work together.

influence on retail investor trading activities, creating more persistent selling pressure and further driving down stock prices.

In addition to the asymmetric emotional impact, investors (especially retail investors) suffer from a heuristics bias: they generally allocate more effort to buying than to selling decisions due to the psychological signal described as "When I sell, I'm done with it." Consistent with this, using unique institutional account level trading data, Akepanidtaworn et al. (2022) document that portfolio managers are skilled in picking and buying stocks, but perform badly when selling stocks. They argue that portfolio managers spend more time and effort to explore information when buying stocks, but allocate relatively less attention when selling stocks due to the heuristics bias. I argue that retail investors are similarly subject to this human heuristic process (even to a larger degree), which exaggerates the impact of fear sentiment. When a stock experiences intensive retail selling pressure, retail investors are more likely to herd to other retail investors' selling activities without careful thinking, leading to relatively concentrated and persistent retail selling orders.

To further test my conjecture, I first examine the *ex ante* and *ex post* order imbalance in 10×10 double-sorted portfolios based on retail buying and selling orders. I find that long–short hedge portfolios show consistently larger spreads in retail selling orders than retail buying orders, both *ex ante* and *ex post*, indicating that retail selling orders are more concentrated in high-selling stocks. I further examine the portfolio performance and retail trading activities around the portfolio formation day. Long–short hedge portfolios based on retail selling orders continue to receive persistent excess retail selling orders in the month after portfolio formation, indicating that retail investors continue to sell high-selling stocks.[20] The high-selling decile group persistently underperforms by approximately 4 bps per day during the month after portfolio formation. However, long-short hedged portfolios based on retail buying orders only generate excess retail buying orders in the first 2 days after portfolio formation and only achieve positive excess returns on the next day (using equal-weighted portfolios). Excess retail buying orders are more transitory than excess retail selling orders. Also noteworthy is the decrease in average daily dollar trading volumes. For a high-selling decile portfolio, the average daily trading volumes before and after portfolio formation are approximately $4.85M and $4.50M, respectively, a decrease of around 7%, indicating that retail selling activity is associated with decreased liquidity.

If the pricing effect of retail selling orders is indeed due to fear sentiment, I expect the effect to be stronger when the market sentiment is fearful, when stocks are smaller, and when volatility is higher (i.e., hard to value). The empirical results confirm that the predictive power of retail selling orders is strengthened when individual stocks have smaller market capitalization and higher idiosyncratic volatility, and when the market return is below the median or the VIX index is above the median. Furthermore, Birru (2018) and Cao, Chordia, and Zhan (2020) show that investor mood varies according to the weekday: it is low on Mondays and high on Fridays. I find that

[20] It is possible that when retail investors have negative private information, they gradually sell the stocks to minimize the pricing impacts.

the pricing effect of retail selling orders disappears on Fridays, suggesting that high investor mood can mitigate fear sentiment and selling pressure.

Finally, I discuss some alternative explanations. Kaniel, Saar, and Titman (2008) argue that retail investors are compensated for providing liquidity to institutional investors. It is possible that the liquidity provision is asymmetric when buying or selling stocks. I use the constituent turnover of the S&P 500 index as a demand shock for certain stocks. The results show that retail investors are more likely to buy stocks if the stocks are added to the S&P 500 index and to sell stocks that become excluded from the S&P 500 index. Thus, retail investors seem not to provide liquidity to index-tracking institutions around the S&P 500 index turnover. Kaniel et al. (2012) and Boehmer et al. (2022) argue that at least some retail investors are likely to be informed investors. It is possible that retail investors are aware of some bad news and do not receive good news. To test this, I investigate retail trading activity around earnings announcement days (i.e., EAday). I find that retail investors simply buy stocks before EAdays and do not distinguish between positive and negative *ex post* earnings surprises. They are merely attracted by news releases and bet on the earnings surprises. Formal regression tests using a quarterly Fama–MacBeth regression of post-earnings announcement drift on previous retail trading activity show that both high retail selling orders and high retail buying orders during the pre-announcement period predict lower post-announcement returns. Retail order imbalance positively predict the cumulative abnormal return (CAR), but the predicting power mainly comes from the selling orders. Both buying order and selling order negatively predict CAR, but the selling order dominates the buying order. The pre-announcement selling orders are more influential, leading to a positive relation between retail order imbalance and post-announcement return. In summary, both the informed trading and liquidity provision explanations fail to explain the asymmetric return predictability of retail buying and selling orders.

The paper contributes to the retail trading literature by decomposing the order imbalance into aggregate buying and selling orders and documenting an asymmetric pricing effect. Recent research on retail investors shows that the aggregate retail order imbalance can positively predict future stock returns, either because retail investors are informed or because they provide liquidity to institutions. I decompose the order imbalance and find that pricing power mainly comes from the selling side rather than the buying side. While I do not deny that at least some retail investors are informed or they are compensated for providing liquidity, behavioral bias is a more plausible explanation for the documented asymmetric pricing effect. [21]

I conjecture that the asymmetric allocation of attention to buying and selling decisions and the asymmetric emotional effects of fear and greed sentiment together drive the trading behavior of retail investors. Reduced attention to selling decisions and the greater influence of fearful sentiment make retail selling more concentrated and persistent, which in turn drives prices

[21] The retail orders only contain the retail market orders, and do not include the limited orders or other types. In the same time, the current construction is conservative as it ignores the retail orders whose sub-penny price is around 0.5 cent. However, the market order should be most suitable to reflect the opinion of the active retail traders. Thus, to some degree, the measure is valid to reflect the behavior of the retail traders. I do acknowledge that the testing sample does not contain the full retail order sample. The price improvement feature only enables the researchers to identify and extract the retail market order, and is salient on the other retail orders.

lower. By evaluating retail investors' pure buying and pure selling activities, I provide a new angle to explain how retail investors can move the market in the same direction as they trade.

Section 2.2 describes the data and sample construction. Section 2.3 presents the main results. Section 2.4 examines the long-term performance of retail trading activities. Section 2.5 explores the mechanisms underlying my results. Section 2.6 discusses some alternative explanations, and Section 2.7 concludes the paper.

2.2 Data and Sample

All the data comes from standard data sources. The retail investor trading data comes from the NYSE trade and quote data (TAQ). I extract stock return data from the Center of Research in Security Prices (CRSP). The firm accounting data comes from Compustat/NA.

2.2.1 Retail Trading Data

In US, most retail investors' trading orders are not directly fulfilled at the registered exchange. Rather, the retail orders are executed either by their brokerage house (i.e., internalized with broker's inventory) or by other wholesalers. Retail orders fulfilled in either of these two ways are generally reported to a FINRA Trade Reporting Facility (TRF) with an exchange code "D". In addition, these orders will generally receive a small amount of price improvement relative to the National Best Bid or Offer (NBBO) price for providing the order flow. The price improvement is usually a very small fraction of a cent, such as 0.01 cent, 0.1 cent, or 0.2 cent. The merit of this mechanism is that, the orders initiated by institutional investors generally do not receive this price improvement. Instead, the institutional orders are sent to the exchanges' dark pools for matching. The Regulation NMS prohibits these orders in the dark pools from having subpenny limit prices. Thus, the recorded trading prices of institutional orders are more like in round pennies. This price improvement feature enables researchers to distinguish a large proportion of retail orders and institutional orders. The trading price of retail investors is more likely to be in a format like XX.XXxx.[22] Take the previous example, TAQ reports a trading order for Agilent Technologies with a volume of 800 shares at a price of $41.0799 at 9:37:12.766288 on Jan 04, 2016. This order would be more likely to be initiated by an institution in terms of the large trading size. However, the price improvement feature enables us to conclude that it is from a retail investor, and is fulfilled either by the brokerage house or by a wholesaler. To further distinguish whether the retail order is initiated by a buyer or by a seller, I extract the subpenny part of the trading price. I define $\mathrm{Z} \equiv 100 \times \mathrm{mod}(\mathrm{P}, 0.01)$. By construction, Z should fall within [0,1). If Z is larger than 0 and smaller than 0.4, I define the order as a retail seller-initiated transaction. If Z is larger than 0.6, I define that transaction as retail buyer-initiated. If Z is 0 or is between 0.4

22 See Boehmer et al. (2022) for a more detailed discussion. My methodology to identify retail trading orders exactly follows the method they proposed. In their paper, they also validate this method by comparing the retail orders extracted this way with those from some proprietary trading data.

and 0.6, I define this transaction as un-identified order. This criterion is conservative, but makes the identified retail order more accurate.

After I identify all the retail trading orders, I construct three different retail trading measures: retail trading order imbalance, retail buying orders, and retail selling orders. Following Boehmer et al. (2022), I define the retail trading order imbalance for each stock *i* on each trading day *t* as:

$$OI_{bjzz(i,t)} = \frac{Total\ retail\ buying\ dollar\ volume_{i,t} - Total\ retail\ selling\ dollar\ volume_{i,t}}{Total\ retail\ buying\ dollar\ volume_{i,t} + Total\ retail\ selling\ dollar\ volume_{i,t}} \quad (2.1)$$

To construct the aggregate retail buying orders and retail selling orders, I define the corresponding retail trading measures for stock *i* on trading day *t* as:

$$R_{buy(i,t)} = \frac{Total\ retail\ buying\ dollar\ volume_{i,t}}{Total\ dollar\ trading\ volume_{i,t}} \quad (2.2)$$

$$R_{sel(i,t)} = \frac{Total\ retail\ selling\ dollar\ orders_{i,t}}{Total\ dollar\ trading\ volume_{i,t}} \quad (2.3)$$

2.2.2 Discussion on the retail trading measures

The retail trading data extracted from this method has some advantages. First, in earlier studies, trading orders below a certain trade size (e.g. $20,000) will be classified as small retail trading activities (Lee and Radhakrishna, 2000). While this method becomes inaccurate after the adoption of computer algorithms in trading. Campbell et al. (2009) document that trading orders below $2,000 are more likely to come from institutions. Second, different with previous proprietary datasets used in retail investor literature, this TAQ data has the most comprehensive stock coverage with relative longer data period. Researchers can extract the retail trading orders for almost all individual stocks. Second, different with the NYSE data used in Kaniel, Saar, and Titman (2008) and other studies (which do not differentiate market order and limit order), the retail orders identified from my method only contain market order. The distinction between market order and limit order may bias the interpretation (Barber, Odean, and Zhu, 2009; Kelley and Tetlock, 2013). Market orders are more suitable for the study of investor sentiment.

A natural question might be that the retail trading is just the counterpart of the institutional trading. While, different institutions show huge heterogeneity. For example, hedge fund is very different from mutual fund. To some degree, the retail trading may show more commonality. Further, my retail trading measure only contains the market order. Limit order is excluded from the analysis. Thus my measures mainly reflect the active trading behaviors from a fraction of the retail investors. Lastly, a considerable institutional trading is executed in the format of cross-trading – transactions within the same fund family – that are not exposed to an external marketplace (Chan et al., 2018; Eisele et al., 2020).

Besides the data source, the retail trading measure constructions are also different from previous literature. In Boehmer et al. (2022), they use the total retail trading volume as the scaler. Thus, their order imbalance measure is equivalent to the ratio of aggregate retail buying volume and aggregate retail selling volume.[23] This construction method combines the buying activities and selling activities together, and is silent on which component is more important. Besides, the order imbalance measure also ignores the impact of the absolute level of retail trading activities. Large stocks generally are less affected by the retail investors than the small stocks. To construct my retail buying orders and retail selling orders, I cannot use the total retail trading volume as the scaler because it will make both measures be equivalent to the ratio of retail buying orders and retail selling orders (see footnote 18). Thus, I use the total daily trading volume as the scaler to construct the retail buying measure and retail selling measure.

Kaniel, Saar, and Titman (2008) construct the net individual trading by subtracting the individual selling volume from the individual buying volume, and standardize the measure by the average daily dollar trading volume in previous one year. Since individual investors trade more frequently, using the one-year average trading volume may ignore some timely trends that may affect the trading activities. The total market capitalization might be another potential scaler. But for some stocks, a high proportion of the total shares may be held by passive mutual funds, which do not trade frequently. Thus using the total market capitalization is not a best scaler.

In summary, when constructing the retail order imbalance measure, I follow Boehmer et al. (2022), i.e., use the ratio of the buying orders and selling orders. However, when constructing the aggregate retail buying orders or the

23 See the formula for details: $\frac{X-Y}{X+Y} = \frac{\frac{X}{Y}-1}{\frac{X}{Y}+1} \propto \frac{X}{Y}$; and $\frac{X}{X+Y} = \frac{\frac{X}{Y}}{\frac{X}{Y}+1} \propto \frac{X}{Y}$; and $\frac{Y}{X+Y} = \frac{1}{\frac{X}{Y}+1} \propto -\frac{X}{Y}$

aggregate retail selling orders, I use the total daily trading volume as my main scaler. In the unreported analysis, I also show some robustness analysis using different retail measures.

2.2.3 Control Variables

I follow the existing literature to construct stock characteristics that can predict future stock returns, which include market capitalization (Mep), market beta (Beta252), book-to-market ratio (B/M), asset growth (TAG), operating profitability (OP), past one-year stock return from day t-252 to day t−21 (MOM252), past 21 trading day return (STR21), and past one trading day return (Return t). I collect the number of analyst coverage (Analysts) data and the analyst forecasting dispersion (Dispersion) from I/B/E/S and calculate institutional ownership (IO%) from the Thomson Reuters 13F database. I also construct the following measures that proxy for lottery-type features: idiosyncratic volatility (IVOL21) and maximum daily return (MAX21). I estimate the market friction and illiquidity measures using average daily bid-ask spread (Spread) and Amihud illiquidity (ILLIQ). I winsorize all continuous variables at the 1% and 99% levels to remove the influence of outliers. The full details of variable construction are presented in Table 2.1.

After I collect all the data, I merge the retail trading data with the daily stock returns and accounting data from CRSP and Compustat, respectively. I only include common stocks (stock share code 10 or 11) listed on NYSE, AMEX, and NASDAQ. I remove lowprice stocks with stock price less than $1 in previous trading day (Kelley and Tetlock, 2017; Boehmer et al., 2022). To mitigate the measurement error in the retail trading measures, I limit my analysis to stocks with a minimum of two retail trading orders (one buying order and one selling order) on each day.[24] I also exclude stocks with retail buying volume or selling volume greater than the total daily trading volume.[25] My final sample contains about 4.9 million stock-day observations, covering 1762 trading days from 2010 to 2016.

2.3 Empirical Results

I report the main results in this section. I first present the summary statistics of stock characteristics within sub-groups formed by different retail trading measures. Next, I report the daily rebalanced portfolio analysis results. I also show the 10×10 double sorted portfolios' returns based on retail buying orders and retail selling orders, and present the Fama-MacBeth (1973) regression results with controlling for many other stock characteristics in the last.

2.3.1 Summary statistics of retail trading activities and stock characteristics

I present the sample summary statistics in Table 2.1. Panel A shows the distribution of the retail trading activities during my sample period. Panel B, C,

24 Barher, Odean, and Zhu (2009) require the stocks to have at least 10 small trades at the weekly frequency. Kelley and Tetlock (2013) require the stocks to have at least 5 orders each day. My results are similar when using different cut points or using no cut point.

25 My results are materially the same if I (1) exclude stocks with price lower than $5 in the previous day, (2) exclude stocks with market cap below the 20th percentile of NYSE stocks, and (3) include stocks with only one retail trading order identified.

and D show the time-series average of the stocks characteristics within different sub-groups formed by retail selling activities, retail buying activities, and retail order imbalance. Panel E presents the time-series average of the correlation matrix between different variables.

[Insert Table 2.1 here]

The result in Panel A shows that the average daily trading volume for each individual stock is about $38.1 million from 6338 trading orders, indicating that the average dollar trading volume is only a bit higher than $6000 for each order. When I look at the retail trading orders, the average dollar volume is about $2.3 million, or 6% of the total trading volume. The average retail trading amount is about $10,135 per order, which is larger than the average dollar amount of all orders. This result is consistent with the usage of computer algorithms after the early 2000s that enables the institutions to "slice and dice" the large institutional parent orders into a sequence of small child orders. Campbell et al. (2009) also document that small trades are more likely to come from institutional investors in recent period. When comparing the retail buying activities and the retail selling activities, the total dollar volume is very close. The retail selling volume ($1.16M) is about 1% higher than the retail buying volume ($1.15M). While, when looking at the percentage values, the average retail selling (buying) volume becomes 4.28% (4.11%), with a difference of more than 4%. The difference in the dollar volume and the percentage volume provides the first clue that retail selling orders are more concentrated in low trading volume stocks (not necessary small stocks).[26] While overall, the retail trading only account for a small proportion in the US market.

Panel B to D report the time-series averages of stock characteristics in different sub-groups formed on the retail selling orders, retail buying orders, and retail order imbalance, respectively. Comparing the results in Column (10) in Panel B and Panel C, I find some similarity of the stocks associated with intensive retail selling orders or retail buying orders. For example, both stocks have small market capitalization ($0.36B for selling-intense stocks, and $0.5B for buying-intense stocks), higher book to market equity (0.85 VS 0.84), low profitability (0.03 VS 0.03), low institutional holding, fewer analyst following, and higher analyst forecast dispersion. Both of them show lottery-type characteristics, including higher max daily return in previous 21 trading days, higher idiosyncratic volatility, and higher bid-ask spread. They perform well in the previous 21 days, but are the losers in previous 252 days. When looking at Panel D, the results in Column (-10) and Column (10) also confirm the above patterns, but the degree is attenuated. I also note that the extreme sold (or bought) stocks also attract large buying (or selling) orders from retail investors. The main implication is that retail investors' trading focuses on some certain stocks, but the expected stock returns disperse greatly among the retail investors, which triggers both the large buying and large selling. This pattern is

26 On average, the retail buying volume is close to the retail selling volume, but the percentage difference becomes larger, i.e. selling is more concentrated on low trading volume stocks. For example, stock A and B have the same market size, and the trading volumes are $100 and $1000. In stock A, the retail buy is $1, and sell is $9; while in stock B, the retail buy is $9 and retail sell is $1. When using the dollar amount, the average volumes are all $5, while when comparing the percentage, the buying volume is (1/100+9/1000)/2≈1%; the selling volume is (9/100+1/1000)≈4.5%.

also consistent with Kumar and Lee (2006) that documents retail investors trade within their habit.

One interesting pattern worth to note is that, in Panel B and Panel C, when comparing the portfolio return on the formation day (Return *t*) in Column (9) and (10), both the best-performance stocks and worst-performance stocks are traded (either buy or sell) heavily by retail investors. The total retail trading volume is more than 25% in Column (10). The results are somewhat different as in Kaniel, Saar, and Titman (2008) which find that retail investors tend to be contrarian. Rather, my results are more consistent with the argument that retail investors are attracted by exciting news, and they are gambling in the market.

Panel E reports the Spearman correlation and Pearson correlation. The correlation between retail buying and retail selling is higher than 0.5, indicating that some certain type of stocks attract the retail investors' attention. On average, both the retail buying and retail selling show low correlation with the previous day return and previous 21 day return, but high correlation with the max daily return in previous 21 days, suggesting that retail investors are simply gambling in the market.

2.3.2 Dissecting Retail Trading Activities: Portfolio Sorts

2.3.2.1 Portfolio sorts using retail buying and selling measures separately

On each day, I sort stocks into deciles either based on the aggregate retail buying orders or the aggregate retail selling orders, and calculate realized returns of each portfolio in the following trading day. I then calculate the returns of a zero-cost long-short hedge portfolio that is the difference in returns between the top and bottom decile portfolios. I report the average portfolio excess returns (raw return over the risk-free rate) and alphas based on the CAPM, Fama-French three-factor, and Carhart four-factor models using both equal- and value-weighted approaches. I also estimate the excess returns or alphas using the second extreme portfolios (i.e., the 2^{nd} and 9^{th} portfolios).

[Insert Table 2.2 here]

Table 2.2 Panel A presents the portfolio returns sorted by the aggregate retail selling orders. Stocks associated with high retail selling activities show significant lower future returns. The long-short portfolio generates an average daily return of −10.42‱ (t-stat = −7.10) and −9.14‱ (t-stat = −5.77) for EW and VW portfolio returns respectively, which are more than −2% per month. The portfolio alphas using the CAPM model, the three- and four-factor models produce similar patterns. The equal-weighted alphas are −9.52‱ (t-stat = −6.01), −9.58‱ (t-stat = −6.11), and −9.50‱ (t-stat = −6.06), respectively. The value-weighted alphas are a bit smaller, but still higher than 9‱. I also show the long-short portfolio using the second extreme decile portfolios, and all the results are still significantly negative. A robust return pattern emerges across these specifications. Moreover, the high-selling portfolios exhibit significantly negative risk-adjusted average returns, and the low-selling portfolios exhibit significantly positive risk-adjusted average returns. This suggests that the negative pricing effect of retail selling orders is both due to the underperformance of high-selling stocks and due to the outperformance of low-selling stocks.

In Panel B, I repeat the portfolio analysis results using the aggregate retail buying orders. Surprisingly, the return predictability of aggregate retail buying orders are much weaker. Value-weighted portfolios do not generate any significant results. Although the equal-weighted portfolio generate significant positive hedge returns using the extreme decile portfolios, the long-short hedge returns flip signs when using the second extreme decile portfolios. Besides, the daily hedge return and factor-adjusted alphas are much smaller than those based on the aggregate retail selling orders.

In summary, the results show that there exist distinctive differences in the pricing effects of aggregate retail selling orders and retail buying orders. The aggregate retail selling orders are much powerful to move the stock price.

2.3.2.2 10×10 Portfolio sorts using retail buying and selling measures

To further compare the pricing effects of retail buying activities and retail selling activities, I sort stocks into 10×10 portfolios based on both the aggregate retail buying orders and the aggregate retail selling orders. I then calculate realized returns of each portfolio in the following trading day. Specifically, in Table 2.3 Panel A, I first sort all the stocks into decile portfolios by the retail buying orders, and then further sort stocks in each decile portfolio into finer decile portfolios by the retail selling orders. This method ensures that all the individual stocks in each decile portfolio based on retail selling orders have similar retail buying orders. I then calculate the realized long-short portfolio returns in the next trading day. In Panel B, I reverse the sorting process by first sorting stocks by the retail selling orders, and then by the retail buying orders.

[Insert Table 2.3 here]

Table 2.3 presents my analysis results. To save space, I only report the results of 1st, 2nd, 5th, 6th, 9th, and 10th portfolios. Results in Panel A show that when the stocks are associated with similar buying orders, different levels of retail selling orders still predict significant cross-sectional difference in the stocks' realized returns. The pricing effect of retail selling orders are more significant when the stocks have higher retail buying orders (i.e., the stocks are heavily traded by retail investors). Specifically, the long-short hedge portfolio returns for stocks with highest retail buying orders are −16.65‱ (t-stat = −5.30), and −14.79‱ (t-stat = −4.94) for EW and VW, respectively. The magnitudes are almost twice as for the stocks with lowest retail buying orders.

In Panel B, I do similar analysis, but I first sort stocks by the retail selling orders, and then by the retail buying orders. For the equal-weighted 10×10 portfolios, I find similar results as in Panel A. But for the value-weighted portfolios, only the 2nd hedge portfolio generates significant results, indicating that when the stocks are associated with similar retail selling orders, the retail buying orders do not affect the future stock prices.

Combine the results in both Panel A and Panel B, I conclude that the retail selling orders still negatively predict future stock return even after I control for the retail buying orders, but not the other way.

2.3.3 Dissecting Retail Trading Activities: Fama-MacBeth Regression

The portfolio sorting results show strong evidence that there exists distinctive pricing effects of retail selling orders and retail buying orders. However, other characteristics or a combination of characteristics may explain the negative retail selling premium (Fama and French, 2008). To investigate the marginal power of retail selling on expected returns, I estimate cross-sectional Fama-MacBeth regressions. The baseline regression includes controls for total order imbalance, stock size, market beta, book to market equity ratio, operating profitability, asset growth rate, momentum, short-term reversals, maximum daily stock return, idiosyncratic volatility, and stock turnover rates. I estimate both equal-weighted least squares (EWLS) regressions and value-weighted least squares (VWLS) regressions. To compare coefficient estimates across different specifications, I normalize all variables on the same day to have zero mean and standard deviation of one. Table 2.4 reports the time-series averages of the coefficient estimates for the 1762-trading-day period between 2010 and 2016. The t-statistics are adjusted following Newey and West (1987) with up to 12 lags.

[Insert Table 2.4 here]

Panel A and Panel B report the EWLS and VWLS results, respectively. In column (1) to (3), I test the retail trading measures one by one separately, and in column (4) to (6), I test the combination of them. In Panel A, column (1) to

(3), the retail selling orders has the largest coefficients (coeff.=-3.39, t-stat.=-10.45) among the three retail trading measures, suggesting the retail selling orders has the strongest pricing power. Comparing the results in column (5) and (6), the retail buying orders almost lose return prediction power after controlling for the retail order imbalance constructed following Boehmer et al. (2022), i.e., the ratio of retail buying volume to retail selling volume. But the retail selling volume still provides incremental return predicting power (coeff.=-2.28, t-stat.=-6.22).

The results in Panel B are more obvious. When I test the aggregate retail buying orders separately in column (2), the coefficient is not significant (coeff.=-0.15, *t*-stat=-0.29). The coefficient of retail buying orders in column (6) even becomes negative and significant (coeff.=-1.41, t-stat.=-2.23) after I control for the retail order imbalance measure. For the retail selling measures, the coefficients are always significant negative at the 1% levels. After controlling for the retail order imbalance, the coefficient still has large magnitude (coeff.=-1.97, *t*-stat=-3.23), which is equivalent to about 6.5 bps using long-short decile hedge portfolios.

In summary, the portfolio analysis results and the Fama-MacBeth regression results together support that the retail selling activities have stronger pricing effects than the retail buying activities, and the pricing power of retail order imbalance documented in previous literature mainly comes from the selling side.

2.4 Long term analysis

I conduct additional tests to examine the long-term performance of the retail selling orders and retail buying orders. The results in this section make the trading strategy more flexible and help to address the concern on the transaction fee.

2.4.1 Weekly rebalanced portfolio analysis

To construct the weekly retail buying orders and retail selling orders, I aggregate all the retail orders within each week, and construction the weekly measure as the same in equation (1) to equation (3). At the end of each week, I sort stocks into decile portfolio, and estimate the realized portfolio return in the following week.

[Insert Table 2.5 here]

Table 5 reports the results. To save space, I only report the results for the 1^{st}, 2^{nd}, 5^{th}, 6^{th}, 9^{th}, and 10^{th} decile portfolio. Panel A and Panel B present the returns for retail selling orders and retail buying orders, respectively. In Panel A, the long-short hedge returns and alphas for portfolio based on retail selling order range from -29.42‱ (t-stat.=-3.73) to -46.50‱ (t-stat.=-4.88), which are more than -1% each month. While the results in Panel B show that the weekly retail buying orders are negatively associated with future stock returns, and the value-weighted results are marginal significant. Combine the results, I conclude that the transaction costs do not affect the asymmetric pricing power of the aggregate retail selling orders and retail buying orders. What's more, the relation between retail buying orders and future stock returns becomes negative, i.e. stocks associated with intense retail buying orders perform worse in future.

2.4.2 Buy and Hold Return

I also test the persistent of the pricing effect from retail selling orders and retail buying orders. Boehmer et al. (2022) investigate the retail order imbalance using the weekly rebalanced portfolios, and they find the retail order imbalance order can predict future stock returns up to 12 weeks, i.e. once we establish the portfolio, we can hold it without rebalance up to 12 weeks. I do a similar analysis, but use the retail selling orders and retail buying orders separately. On each trading day t, I establish the portfolio based on one of the retail trading measures, and hold the portfolio constant for the following 21 days. I then compare the 21-day buy and hold return for the decile portfolios. In the analysis, the transaction cost should be a minor issue since I do not rebalance the portfolio for the following 21 days.

[Insert Table 2.6 here]

Table 2.6 presents the results. Panel A shows the results based on the retail selling orders, and Panel B shows the results based on the retail buying orders. The long-short returns and alphas of the portfolio based on retail selling orders range from -0.77% (t-stat.=-4.91) to -1.18% (t-stat.=-5.86), which is still comparable to that based on daily portfolios or weekly portfolios.

In Panel B, I find similar pattern as that in the weekly portfolio analysis. All the long-short returns and alphas of the hedge portfolio based on retail buying orders flip signs, i.e. stocks with high retail buying orders underperform in the long run. What's more, the results become significant at the 1% level. I also note that the absolute magnitudes of the hedge returns based on buying orders are much smaller than those based on the selling orders, which helps to reconcile the positive relation between retail order imbalance and future stock returns.

Combine the results in Table 2.5 and Table 2.7, I conclude the retail selling orders negatively predict stock return even in the longer term. But the retail buying orders lose return predicting powers, and even flip signs in longer term. The transaction cost will not affect the asymmetric pricing impact.

2.5 Mechanism Test

In the previous analysis, I document a strong and robust asymmetric pricing effect between aggregate retail selling orders and retail buying orders. I explore some possible channels that could explain the asymmetric relation in this part. I first compare the *ex ante* and *ex post* retail order imbalance in the sub groups based on retail selling orders or retail buying order. I also track the portfolio performance and retail trading activities around the portfolio formation day. Lastly, I examine the heterogeneous pricing power in different subgroups.

2.5.1 Concentrated selling and dispersed buying

The summary statistic in Table 2.1 Panel A provide first evidence that the retail selling orders are more concentrated on low trading volume stocks as the average percentage of retail selling orders are much higher than the retail buying order. This also implies that the spread of retail selling orders should be larger than that of the retail buying order. I formally test this idea by comparing the retail selling order spread and retail buying order spread after I control for the retail buying activities or retail selling activities. On each day, I form the 10×10 double sorted portfolios as I do in Table 2.3, i.e. either first sorting stocks by retail buying order and then by retail selling orders or first sorting stocks by retail selling orders and then by retail buying orders. Instead of reporting the portfolio return, I report the average net retail buying orders in each sub-group on both the portfolio formation day and the next day after portfolio formation.

[Insert Table 2.7 here]

Table 2.7 presents the results. In Panel A, I first sort stocks into decile portfolios, and then further sort stocks in each portfolio into finer decile portfolios. In Panel B, I reverse the sorting sequence. When I control for the aggregate retail buying orders (Panel A), the average retail selling order spread in the high-selling portfolios and low-selling portfolios is about 12.50% *ex ante* (ranked from 7.47% to 22.09%) and 1.01% *ex post* (ranked from 0.75% to 1.40%). When I control for the aggregate retail selling orders (Panel B), the average retail buying order spread is 11.99% *ex ante* (ranked from 6.83% to

22.83%) and 0.56% *ex post* (ranked from 0.40% to 0.70%). The spread of the selling orders is about 0.5% larger than the spread of retail buying orders both *ex ante* and *ex post*. This pattern helps to explain why the retail selling orders are more powerful to move the market.

2.5.2 Persistent selling and transitory buying

To further explore the retail trading activities and stock performance around the portfolio formation day, I first track the portfolio performance in the two-week-before and four-week-after the portfolio formation. I plot the daily portfolio performance in Figure 2.1 to Figure 2.3.

[Insert Figure 2.1 to Figure 2.3 here]

Figure 2.1 shows the daily portfolio returns of the decile portfolios based on the aggregate retail selling measures. On day 0, I rank all the stocks into decile portfolios by the aggregate retail selling orders on that day. I find the high-selling decile portfolio perform worst on day 0, but perform best during the previous 9 trading days except the portfolio formation day (i.e., day 0). When I look at the post-formation performance, I find the high-selling portfolio underperform by about 10 bps in the following trading day, and continues to underperform by about 5 bps each day starting from day 2 until day 21. The pattern is strongly persistent.

The results in Figure 2.2 and Figure 2.3 show different patterns. In Figure 2.2, the decile portfolios based on aggregate retail buying orders, I do not find any persistent patterns. The long-short hedge returns are close to zero in the first several days (except day 1) and then becomes negative in longer window. Figure 2.3 shows the decile portfolios based on the retail order imbalance (the ratio of retail buying orders to retail selling orders). All the hedge returns are positive, but after one week of the portfolio formation, the magnitudes become much smaller and close to zero.

[Insert Table 2.8 here]

I present the formal comparison of the portfolio performance in Table 2.8.To save space, I only report the results for the 1st and 10th decile groups and the long-short hedge portfolio. For each decile group, I report the average daily dollar trading volume, aggregate retail buying orders and selling orders, the difference in the buying and selling orders, and the average daily portfolio return. Panel A shows the results for retail selling orders, and Panel B shows the results for retail buying orders. The long-short portfolio returns in both panels show similar pattern as in Figure 2.1 and Figure 2.2. For the retail selling orders, all the daily returns before the portfolio formation are positive, indicating the high-selling stocks perform well in the pre-formation period. While on the formation day, the pattern reverses. The high-selling stocks experience a dramatic drop in the stock price, with the average return of -19.51 bps (*t*-stat.=-10.82). The trend continues after the portfolio formation, with the first day return of -10.42 bps (*t*-stat.=-7.10), and then drops to about 4 bps each day for the next 20 days. Interestingly, the high-selling stocks have already experienced excess selling pressure before the portfolio formation, ranging from 0.24% to 0.49% each day, and continues to receive excess selling orders in the post-formation period, ranging from 0.38% to 0.65% each day. It seems the extreme selling orders (7.22%) on the formation day trigger the crash of the high-selling stocks. A last point to mention is the daily trading volume. Before the portfolio formation, the 10th decile group (high-selling group) have a daily average dollar trading volume of $4.85M, and the volume drops to $4.50M in the post-formation period, indicating that the high-selling stocks are also associated with drying-up liquidity.

When I look at the decile portfolios formed by the aggregate retail buying orders, I do not find any persistent patterns. The long-short portfolio does not show positive return except the three-day window around the portfolio formation. Although the high-buying decile group receive excess buying order before the portfolio formation, this trend does not hold since the 3rd day in the post-formation period. The trading volume is much larger in the high-selling decile than that in the high-selling decile.

Overall, the results in Table 2.8 suggest that the retail selling activities are more concentrated and more persistent than the retail buying activities. Besides, the retail selling activities are also associated with drying-up liquidity. These evidences help to explain the asymmetric pricing effects of the retail buying activities and retail selling activities.

2.5.3 Investor sentiment and stock characteristics

The above analysis shows that the concentrated and persistent retail selling activities move the market to the same direction as the retail investors trade. To further explain the phenomena, I conjecture that the asymmetric emotional impact of fear and greed sentiment drives this result. To proxy for the potential sentiment effects, I use two firm-level measures and two macro-level measures. Small firms have relatively more asymmetric information environment, thus are more likely to be affect by the fear sentiment. Similarly, stocks with high idiosyncratic volatility are more uncertain, and more likely to be affected by the fear sentiment. For the macro-level environment, I use the VIX index and market return. Specifically, I define trading days with VIX above the period median as high VIX period, and trading days with market return lower than the period median as low MKT period. The retail selling activities should become more powerful on the high VIX period and bearish market period. I then run the Fama-MacBeth regression including the interaction terms.

[Insert Table 2.9 here]

Table 2.9 presents the results. Panel A and Panel B show the EWLS results and VWLS results. I find that the pricing effect of retail selling activities becomes stronger when the stocks are small and with high idiosyncratic volatility. The effects are also stronger when the market return is lower and VIX index is higher. The negative pricing effects are almost doubled. These results support that retail investors' trading are driven more by the fear sentiment.

2.5.4 Investor mood: weekday effects

A prominent phenomena in the psychology literature is that human mood increases from Thursday to Friday, while decreases on Monday. People/investors become more optimistic of future prospects when they are in good mood than when they are in bad mood (Wright and Bower, 1992). Birru (2018) shows that stock market anomalies whose speculative leg is the short leg experience the highest returns on Monday, and becomes very weak or insignificant on Friday. Cao, Chorida, and Zhan (2020) also document that the IVol effect mainly occurs on Monday, and even reverses on Friday. I conduct similar tests to check whether the investor mood affect the retail selling orders' pricing ability by investigating the weekday effects. At the end of each week, I estimate the average retail selling measures for each stock for the past one week, and rank all the stocks into decile portfolios. I then hold these decile portfolios for one week, and report the daily portfolio returns on each day in the following week.

[Insert Table 2.10 here]

Table 2.10 presents the results. Panel A and Panel B show the EW and VW results. I find that the pricing effect of retail selling activities becomes insignificant on Fridays, as the long-short returns or alphas becomes insignificant or flip signs. The negative relation between retail selling orders and future stock return hold from Monday to Thursday, with Tuesday and Thursday show the strongest effects. Taken together, the high investor mood mitigates the fearful sentiment associated with intensive retail selling activities.

2.6 Alternative explanation

I briefly discuss some alternative explanations in this section, including asymmetric liquidity provision, asymmetric informed trading on good or bad information, and some other explanations.

2.6.1 Asymmetric liquidity provision

Kaniel, Saar, and Titman (2008) argue that retail investors are compensated by providing liquidity to the institutional investors, thus the retail order imbalance could positively predict future stock return. It might be that the institutions' buying and selling activities are different, leading to asymmetric liquidity provisions from the retail investors. In this section, I investigate the retail trading activities around the S&P 500 index constituents' turnover. There exist large amount of index-tracking funds that follow the index constituents. Once a stock is added into the S&P 500 index, it will attract a lot of institutional buying activities. Practically, the S&P index change will be announced one month before the real effective day, thus I focus on a period one-week before the accouchement day and four-week after the effective inclusion day for each S&P 500 index addition or deletion, i.e., (-25, 20) window around the effective turnover day.

[Insert Table 2.11 here]

I present the results in Table 2.11. Panel A reports the results around index addition, and Panel B reports the results around index deletion. I report the average dollar trading volume, retail buying orders, retail selling orders, the difference between retail buying and retail selling, and the stock performance.

Firstly, I note that both the retail buying and selling activities drop greatly during the three-day window (-2, 0) before the effective day. Since I use the total trading volume as the scaler, this dramatic drop is more likely to be driven by the increased institutional trading volume. Except this point, I do not find any significant pattern during this 9-week period. The net retail buying orders are positive in Panel A, and are negative in Panel B, suggesting that the retail investors are also buying the stock added into the index, and selling the stocks that are deleted from the index. The market adjusted stock returns for the added or deleted stocks are insignificant from 0 either. These results suggest that retail investors are not liquidity providers, at least not during the S&P 500 index addition or deletion window. While I admit that, due to the low-frequency of S&P 500 index turnover, my tests in this part may not be so informative.

2.6.2 Asymmetric informed trading

Kaniel, Liu, Saar, and Titman (2012) investigate the retail trading activities around the firms' Earnings Announcement days (EA days), and argue the retail investors trade as they are informed. Thus, it might be that the retail investors perform asymmetrically for the good news stocks and bad news stocks, i.e., retail investors know bad news but do not know good news. Although this explanation is not plausible, I do a similar analysis as for the S&P 500 index turnover to investigate the retail trading activities around the EA days.

[Insert Table 2.12 here]

Table 2.12 presents my results. In each quarter, I estimate the standardized earnings surprises (SUEs) for each stocks, and rank the stocks into quintile groups based on their SUE. Panel A to C report the results for the extreme positive SUE quintile (Q5), moderate SUE quintile (Q3), and extreme negative SUE quintile (Q1). For each quintile, I report the average daily trading volume, retail buying orders, retail selling orders, the difference between retail buying and selling, and the average stock returns.

I find that just before the EA days, retail investors buy the stocks for all stocks. The net buying for the three quintile groups are 0.52% (*t*-stat.=8.86), 0.26% (*t*-stat.=8.59), and 0.27% (*t*-stat.=3.83) for Q5, Q3, and Q1 respectively. It seems the retail investors do not distinguish whether the stock can beat the forecast, rather they are attracted by the EA news, and simply bet on the earnings announcement. After the earnings news are released, the retail investors start to sell the stocks, as the net buying for the three groups become -0.41% (*t*-stat.=-10.37), -0.15% (*t*-stat.=-6.68), and -0.42% (*t*-stat.=-6.30). I also show that the Post Earnings Announcement Drift (PEAD) still exist in recent period, but the drift becomes much weaker after the first week.

To conduct a formal regression test on the informed trading explanation, I estimate both the cumulative stock returns (CAR) using different periods after the EA day and retail trading activities in the pre-announcement period. Then I conduct a quarterly Fama-MacBeth regression, i.e. I regress CARs among different horizon on the retail selling, retail buying, and retail order imbalance measures in the pre-announcement period quarter by quarter, and report the average coefficients. I do the regression separately for different retail measures with all the control variables as in Table 2.4, but report the coefficients on retail trading measures into one table, and omit the coefficients of control variables for brevity.

[Insert Table 2.13 here]

Table 2.13 reports the results. Panel A to Panel C reports the results based on retail trading measures among different period before the EA days. In all three panels, I find similar and consistent patterns. Both the retail selling measures and retail buying measures negatively predict future CARs after the earnings announcement, but the negative effect is stronger, leading to a positive relation between retail order imbalance and future CARs. Following the informed trading, my results could suggest that retail investors only know bad news, but are cheated by the good news as the retail buying activities negatively predict future CARs, which is not convincing.

To sum up, my results are similar as in Kaniel, Liu, Saar, and Titman (2012), which also document that retail investors are attracted by the EA events and start to buy the stocks before the EA days. But my results suggest the retail investors do not distinguish the good news stocks and bad news stocks. They just bet on the news. I do not find any evidence to support the argument that there exist asymmetric informed trading patterns among the retail investors.

2.6.3 Other possible explanations

There may still exist some alternative explanations to explain the asymmetric pricing effect between retail buying activities and retail selling activities. For example, the litigation enforcement against insider trading on good news and bad news may be different. It might be harder to detect the inside trading on bad news, thus the retail investors could acquire some inside information and start to sell the stocks. Another possible explanation may be that the institutions buy and sell stocks in asymmetric ways. When the institutions buy stocks, they "slice and dice" the block orders to hide their trading directions, but when the institutions sell the stocks, they may become less patient to do so. Due to the less trading constrain or reputation constrain of retail investors, they can even ride on the institutions and trade much faster.

While, I cannot exclude these possibilities, the asymmetric emotional impact of fear and greed sentiment together with the asymmetric attention allocation are more plausible to explain my results. The drying-up liquidity associated with the retail selling activities also exaggerates the fearful sentiment, which makes the retail selling activities more powerful to move the market.

2.7 Conclusion

In this paper, I use the unique feature that retail orders will receive price improvement in the US equity market to identify retail investor trading activities and investigate the impact of their trading orders. I find consistent results that the aggregate retail trading order imbalance could positively predict future stock returns. I further decompose the retail order imbalance into retail buying orders and retail selling orders, and I find the pricing effect of retail order imbalance mainly comes from the selling side. Using the pure retail selling orders, the daily rebalanced long-short portfolio could generate more than 10 bps abnormal returns each day. Moreover, this negative pricing effect is still strong and persistent in longer terms. After considering the transaction cost, the trading strategy based on the pure retail selling orders still generate more than 1% abnormal return each month. While when I do the same analysis using the retail buying orders, I do not find any meaningful results. The retail buying orders do not positively predict stock return in short term, and the relation even becomes negative in the longer periods.

I explore several possible explanations for the asymmetric pricing effect between retail buying orders and retail selling orders, and conjecture that the asymmetric impact of fear emotion and greed emotion together with the asymmetric attention allocation drive this asymmetric pricing effect. Retail investors are more likely to be affected by the fear emotion, and they also suffer from the cognitive biases (i.e., allocate more attention on the buying decision and less attention on selling decision). These two factors combined makes the retail selling orders more concentrated and more persistent, which move the stock price in the same direction as the retail investor trade.

Figures and Tables

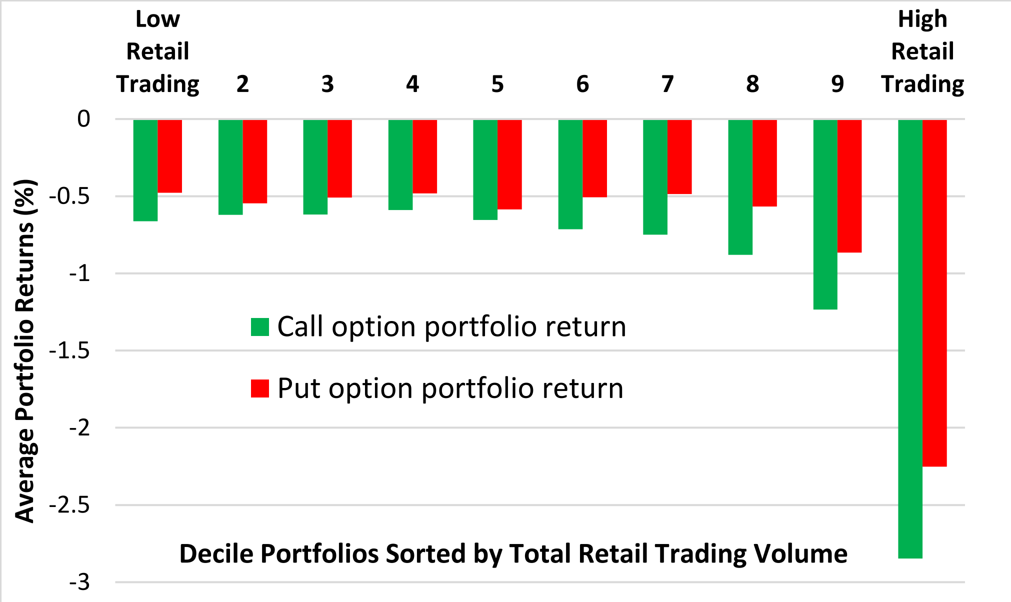

Figure 1.1. Portfolio Returns of Delta-hedged Options of Stocks with Different Retail Trading Volume

This figure shows the average returns for decile option portfolios formed based on the level of total retail trading volume of the underlying stocks in previous month. At the end of each month, I sort all the call options or put options into decile portfolios based on the stocks' average total retail trading volume in previous month, and report the average delta-hedged option return for each decile portfolio, assuming the option is held until maturity. The delta-hedging is conducted daily until the option maturity. The top decile portfolio contains stocks with the highest retail trading volume, and the bottom decile portfolio contains stocks with the lowest retail trading volume.

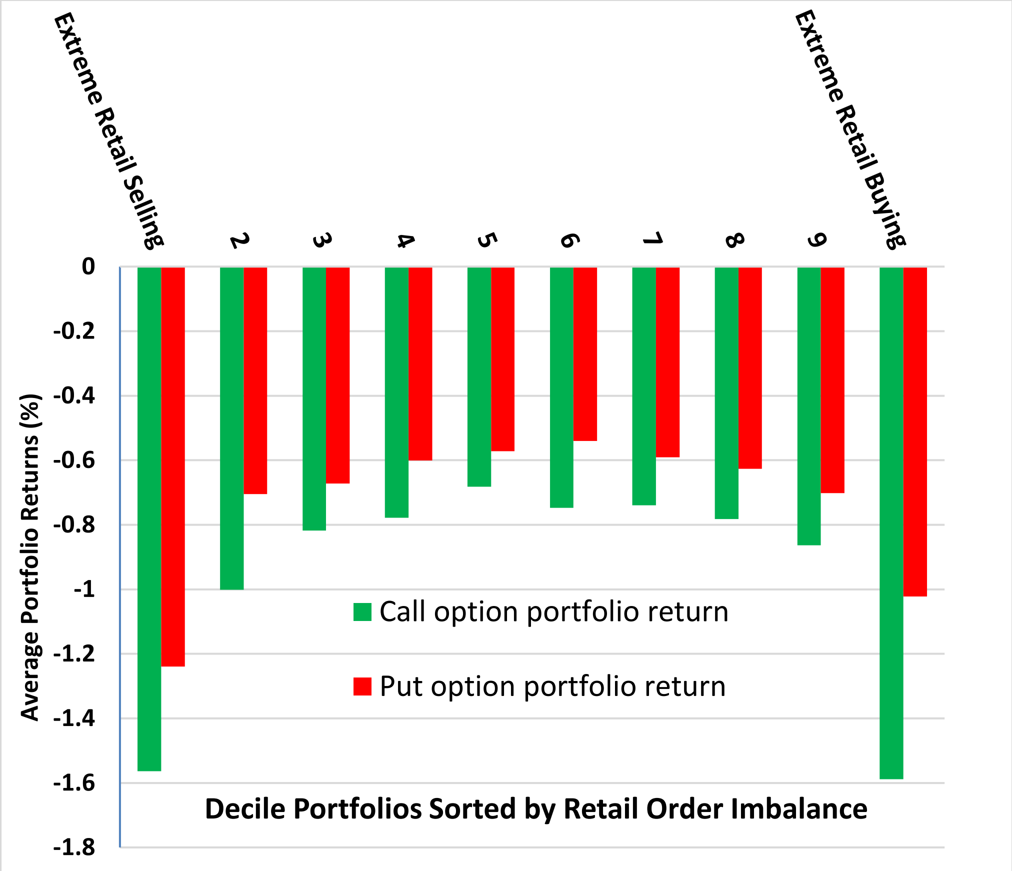

Figure 1.2. Portfolio Returns of Delta-hedged Options of Stocks with Different Retail Trading Order Imbalance

This figure shows the average return for decile option portfolios formed based on the level of retail order imbalance of the underlying stocks in previous month. At the end of each month, I sort all the call options or put options into decile portfolios based on the stocks' retail trading order imbalance in previous month, and report the average delta-hedged option return for each decile portfolio, assuming the option is held until maturity. The delta-hedging is conducted daily until the option maturity. The top decile portfolio contains stocks with extreme retail net buying volume, and the bottom decile portfolio contains stocks with extreme retail net selling volume.

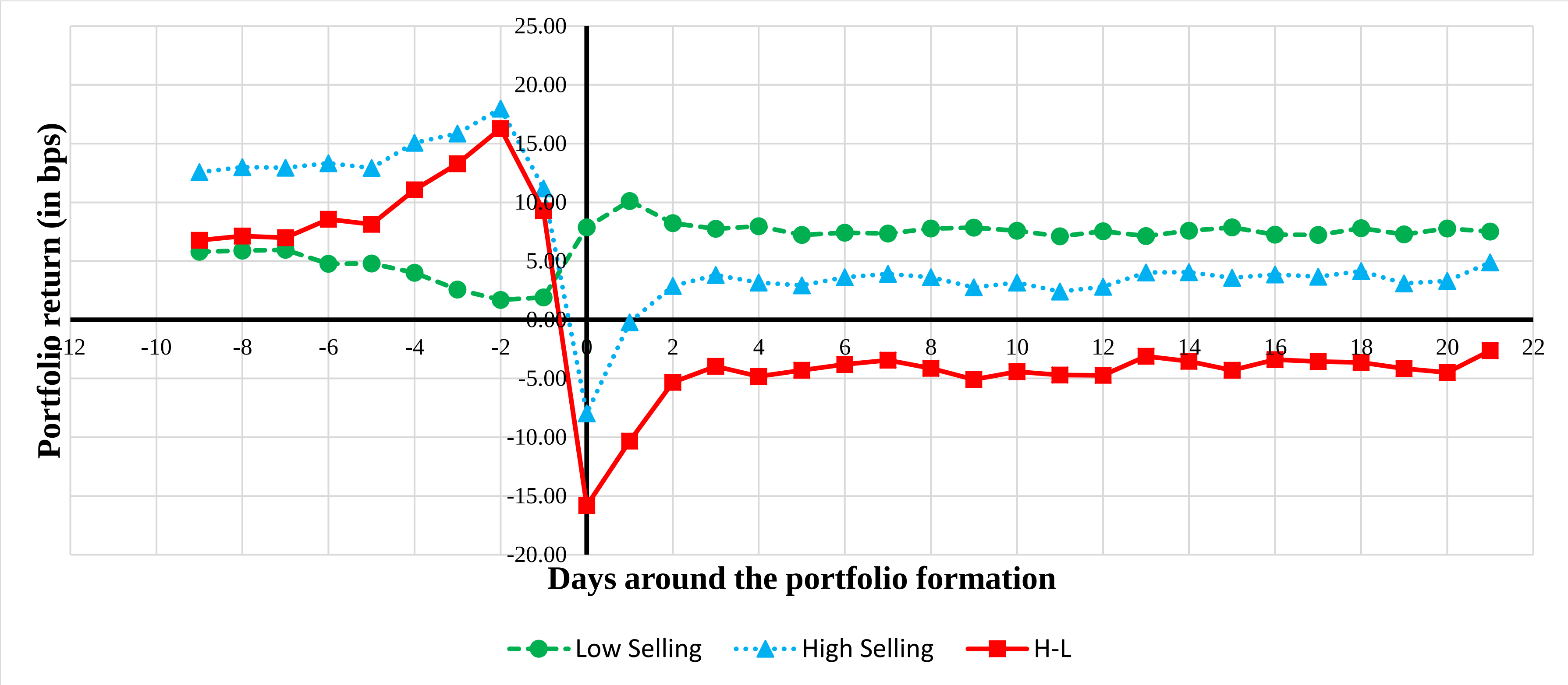

Figure 2.1. Daily Decile Rsel Hedge Portfolio Performance Around the Formation Day

This figure shows the daily portfolio performance for the decile portfolios based on retail selling orders (Rsel) and the long-short Rsel portfolio. On each day, all stocks are ranked and assigned to one of the ten decile portfolios, and then I track the daily returns in the (-10,20) window around the portfolio formation day (i.e. day 0). The long-short portfolio is a zero-cost portfolio that longs the largest Rsel decile stocks and shorts the smallest Rsel decile stocks.

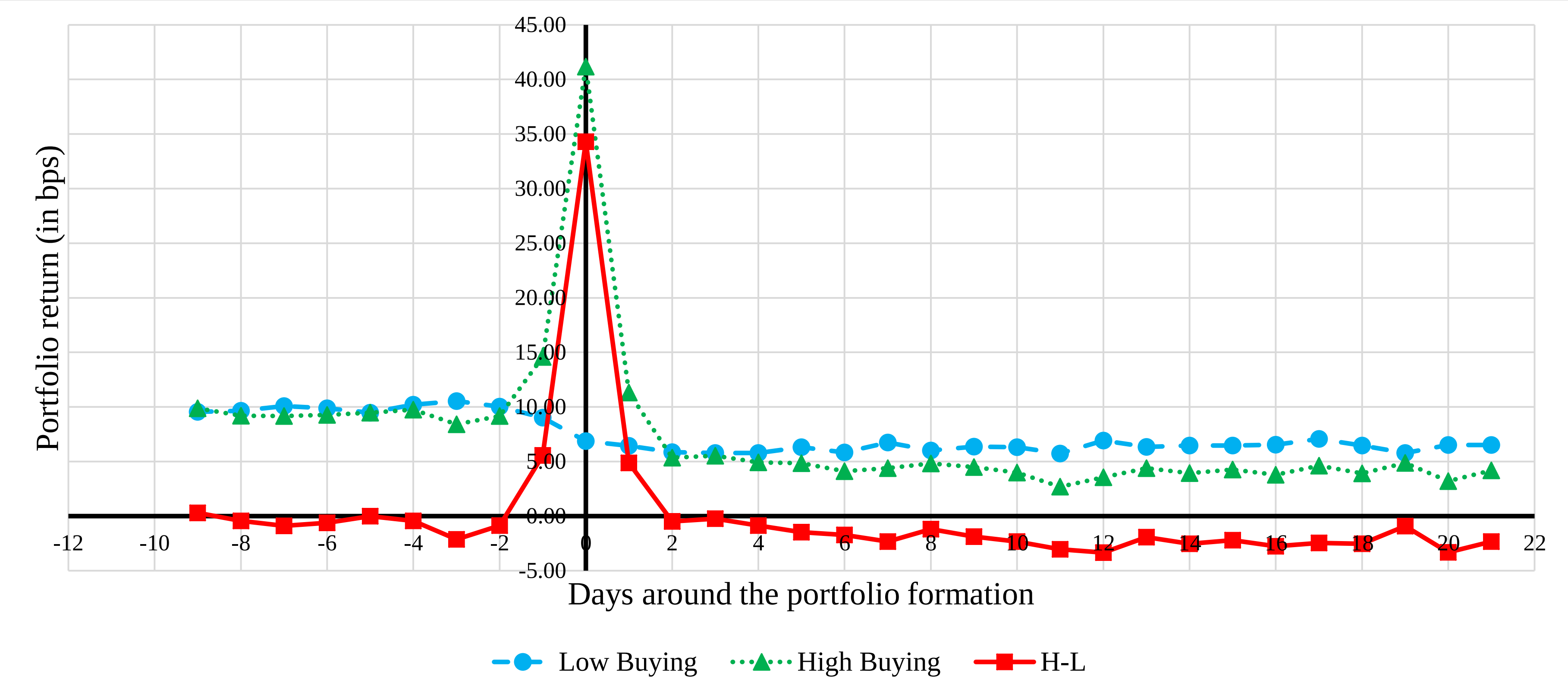

Figure 2.2. Daily Decile Rbuy Hedge Portfolio Performance Around the Formation Day

This figure shows the daily portfolio performance for the decile portfolios based on retail buying orders (Rbuy) and the long-short Rbuy portfolio. On each day, all stocks are ranked and assigned to one of the ten decile portfolios, and then I track the daily returns in the (-10,20) window around the portfolio formation day (i.e. day 0). The long-short portfolio is a zero-cost portfolio that longs the largest Rbuy decile stocks and shorts the smallest Rbuy decile stocks.

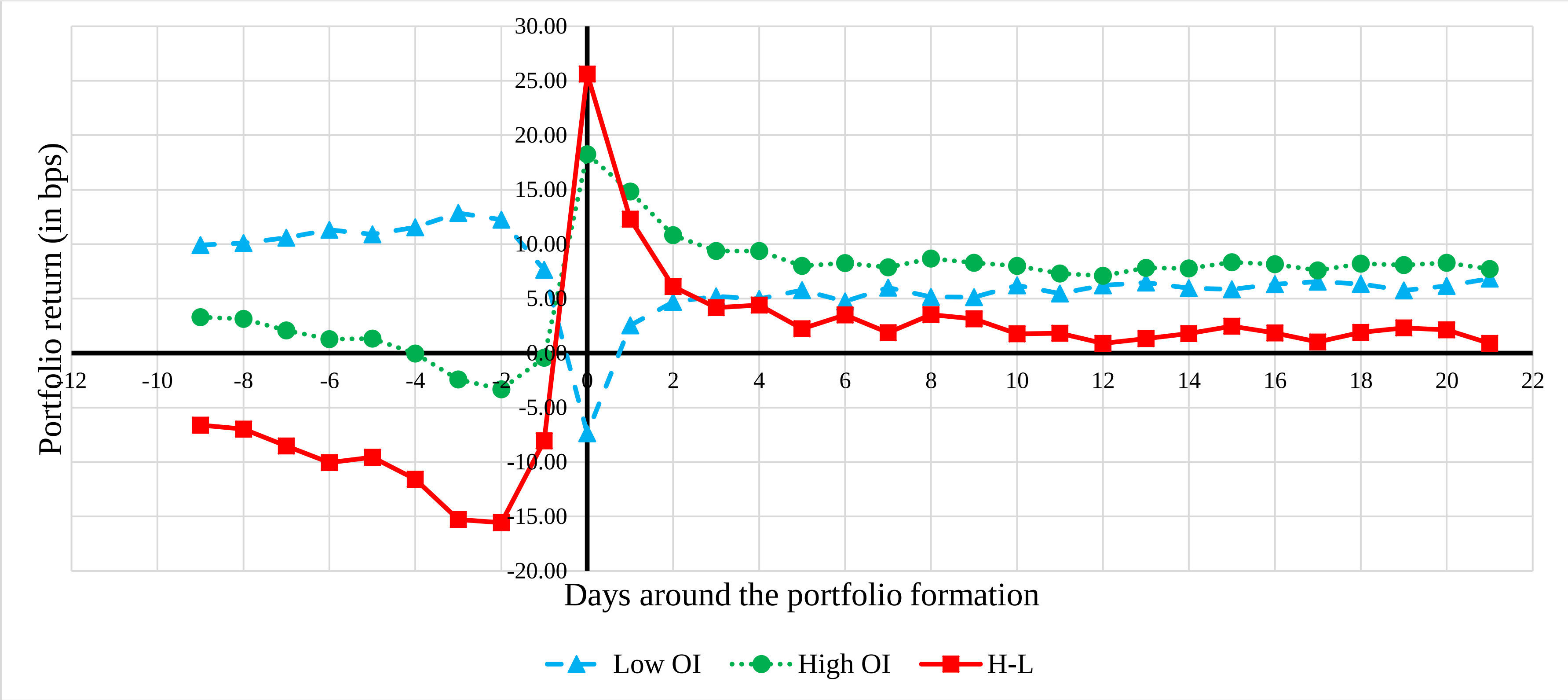

Figure 2.3. Daily Decile OIbjzz Hedge Portfolio Performance Around the Formation Day

This figure shows the daily portfolio performance for the decile portfolios based on retail order imbalance measure (OIbjzz) by Boehmer et al. (2022) and the long-short Rsel portfolio. On each day, all stocks are ranked and assigned to one of the ten decile portfolios, and then I track the daily returns in the (-10,20) window around the portfolio formation day (i.e. day 0). The long-short portfolio is a zero-cost portfolio that longs the largest OIbjzz decile stocks and shorts the smallest OIbjzz decile stocks.

Table 1.1. Summary Statistics of Options

Panels A and B report the summary for call options and put options, respectively. The option returns are calculated as the delta-hedged gains over the initial investment. The delta-hedged gain is the value change over the next month or until option maturity in the value of a portfolio consisting of one contract of long position of call (put) option and delta shares of short position of the underlying stock. The delta-hedging is rebalanced daily. The initial investment to establish the positions is (Δ*S-C) for calls and (P-Δ*S) for puts, where Δ is the Black-Scholes option delta, S is the underlying stock price, and C (P) is the option price for call (put) options. Oret_ex is the option return (in percentage) until maturity. Oret_ed is the option return (in percentage) over the next month. Maturity is the number of calendar days until the option expires. Moneyness is the ratio of the stock price to option strike price (multiplied by 100). Vega is the option vega according to the Black-Scholes model, scaled by the stock price. OPBaspread is the option bid-ask spread, estimated as the ratio of the difference between best ask and best bid quotes of the option to the midpoint of the bid and ask quotes (multiplied by 100) at the end of each month. All these variables are winsorized at the 1% level in each month. The sample period is from January 2009 to December 2017.

Panel A: Call option (156,159 obs)

Variables	Mean	Media	Std	P10	P25	P75	P90
Oret_ex (%)	-0.95	-1.01	6.72	-5.99	-3.10	0.87	3.63
Oret_ed (%)	-0.59	-0.71	5.48	-4.42	-2.27	0.81	3.09
Maturity (days)	50	50	2	46	49	51	52
Moneyness (%)	99.88	100.0	5.05	94.34	97.81	102.1	105.3
Vega	0.14	0.14	0.01	0.12	0.14	0.15	0.15
OPBaspread (%)	30.98	19.05	33.24	4.55	9.52	40.00	74.29

Panel B: Put option (148,508 obs)

Variables	Mean	Media	Std	P10	P25	P75	P90
Oret_ex (%)	-0.72	-0.88	5.52	-5.05	-2.65	0.84	3.49
Oret_ed (%)	-0.45	-0.66	4.40	-3.85	-2.02	0.71	2.84
Maturity (days)	50	50	2	46	49	51	52
Moneyness (%)	100.0	99.93	4.94	94.55	97.79	101.9	105.3
Vega	0.14	0.14	0.01	0.13	0.14	0.15	0.15
OPBaspread (%)	29.57	18.18	32.13	4.33	9.23	36.36	68.57

Table 1.2. Summary Statistics of Stock Characteristics

This table reports the statistics for different stock characteristics and their correlation between the option returns. Panel A reports the distribution of different stock characteristics, and Panel B reports the time-series average of the cross-sectional correlation between different stock characteristic and the delta-hedged option returns. Oret_ex is the option return (in percentage) until maturity. For the stock characteristics, Rtrd is the average ratio of total retail trading volume over the total daily trading volume in previous month. Rnby is the average ratio of net retail buying volume over the total daily trading volume in previous month. Mep is the stock's market capitalization at the end of previous month. MOM is the return momentum, defined as the cumulative stock return in previous 12 months except the previous 1 month. IVOL is the idiosyncratic volatility, estimated as the standard deviation of residuals from the regression of daily excess stock return on the Fama-French three factors in previous month. EIS is the expected idiosyncratic skewness, estimated as in Boyer, Mitton and Vorkink (2010). MIS is the mispricing score, estimated as in Stambaugh and Yuan (2017). STR is the short term reversal, defined as the monthly stock return of previous month. MAX is the maximum daily return in previous month. ILLIQ is the Amihud illiquidity. All these variables are winsorized at the 1% level in each month. The sample period is from January 2009 to December 2017.

Panel A: Stock characteristics

Variables	Mean	Median	Std	P10	P25	P75	P90
Rtrd (%)	6.63	4.97	4.75	2.71	3.52	7.97	13.16
Rnby (%)	-0.11	-0.07	0.77	-0.93	-0.43	0.24	0.65
Mep ($B)	7.47	1.66	18.91	0.26	0.59	5.32	16.82
Mom	0.20	0.13	0.54	-0.31	-0.09	0.37	0.72
IVol	0.28	0.22	0.18	0.11	0.15	0.34	0.51
EIS	0.45	0.35	0.64	-0.25	0.01	0.78	1.30
MIS	49.82	49.36	12.50	33.55	40.64	58.55	66.73
STR	0.01	0.01	0.11	-0.12	-0.05	0.07	0.14
MAX	0.05	0.04	0.04	0.02	0.03	0.06	0.09
ILLIQ($*10^{-9}$)	6.28	0.98	16.30	0.07	0.23	4.17	15.23

(Table 1.2 continued)

Panel B: Time-series average of cross-sectional Pearson correlations

			(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Oret_ex		(1)		-0.106	-0.096	-0.084	-0.033	-0.065	0.066	Put
Rtrd (%)		(2)	-0.102		0.495	0.500	0.191	0.506	-0.401	Option
IVol		(3)	-0.099	0.493		0.483	0.253	0.297	-0.473	
EIS		(4)	-0.106	0.516	0.482		0.318	0.474	-0.575	
MIS		(5)	-0.046	0.185	0.246	0.314		0.109	-0.250	
ILLIQ	Call	(6)	-0.073	0.521	0.294	0.481	0.105		-0.533	
Ln(Mep)	Option	(7)	0.090	-0.416	-0.468	-0.582	-0.247	-0.534		

Table 1.3. Baseline Results: Fama-MacBeth Cross-section Regressions

This table reports the baseline results. Each month, I regress the call option returns (in percentage) or put option returns (in percentage) on the total retail trading volume and other control variables. The time-series average of the cross-sectional regression coefficients are reported. All control variables are standardized (N~(0,1)) to make the results comparable. Column (1) and (2) show the results for call option, and Column (3) and (4) report the results for put option. The explanatory variable of interest is the total retail trading volume (*Rtrd*). The control variables include: the idiosyncratic volatility (*IVOL*), expected idiosyncratic skewness (*EIS*), the Stambaugh and Yuan (2017) mispricing score (MIS), maximum daily return in previous month (*MAX*), Amihud illiquidity (*ILLIQ*), the option bid-ask spread between the best-offer quote and best-ask quote scaled by the bid-ask midpoint (OPBaspred), and the natural logarithm of stock's market capitalization (*Ln*(*SIZE*). The *t*-statistics are shown in parentheses using the Newey and West (1987) corrected standard errors. ***, **, and, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

	Delta-hedged Call option return until expire		Delta-hedged Put option return until expire	
	(1)	(2)	(3)	(4)
Rtrd	-0.66***	-0.36***	-0.52***	-0.37***
	(-11.81)	(-7.74)	(-12.46)	(-11.82)
IVol		-0.39***		-0.25***
		(-6.90)		(-5.33)
EIS		-0.23***		-0.08
		(-3.37)		(-1.11)
MIS		-0.05		-0.01
		(-1.56)		(-0.43)
MAX		0.05		-0.02
		(0.89)		(-0.49)
ILLIQ		0.02		0.02
		(0.51)		(0.64)
OPBaspread		-0.22***		-0.08*
		(-8.27)		(-1.87)
Ln(Mep)		0.06		-0.03
		(1.16)		(-0.73)

Table 1.4. Returns of Option Portfolios Sorted by Total Retail Trading Volume

This table reports the average returns (in percentage) of holding short-maturity at-the-money call options (Panel A) or put options (Panel B) on stocks associated with different level of total retail trading volume. At the end of each month, I rank all optionable stocks into decile groups based on the corresponding retail trading volume of stocks in that month. For each stock, for the call option, I buy one contract of call option against a short position of delta shares of the underlying stock, where delta is the Black-Scholes call option delta. For the put option, I buy one contract of put option against a long position of delta shares of the underlying stock. The delta hedging is rebalanced each day. For each stock in each month, I compound the daily returns of the rebalanced delta-hedged call options or put options over the remaining maturity to estimate the delta-hedged option returns. I use three weighting schemes to compute the average portfolio returns: equal weighted (EW), weighted by the market capitalization of the underlying stock (SVW), and weighted by the option open interest at the end of previous month (OVW). (H – L) represents the long-short hedge return of a zero-cost portfolio that longs the high decile retail trading volume option portfolio and short the low decile portfolio. The *t*-statistics shown in parentheses are computed based on standard errors with Newey-West corrections.

Panel A. Portfolio returns for call options

	LOW	2	3	4	5	6	7	8	9	HIGH	H-L
EW	-0.66	-0.62	-0.62	-0.60	-0.65	-0.71	-0.75	-0.88	-1.25	-2.83	-2.17
	(-5.42)	(-4.42)	(-4.96)	(-4.74)	(-5.21)	(-6.30)	(-5.80)	(-7.03)	(-8.58)	(-14.09)	(-12.79)
SVW	-0.60	-0.48	-0.47	-0.39	-0.37	-0.25	-0.31	-0.27	-0.51	-1.44	-0.84
	(-5.87)	(-3.96)	(-3.93)	(-3.61)	(-3.33)	(-1.83)	(-3.23)	(-2.21)	(-3.94)	(-8.51)	(-5.75)
OVW	-0.95	-0.88	-0.77	-0.91	-0.67	-0.72	-0.62	-0.60	-1.22	-2.57	-1.62
	(-9.14)	(4.13)	(-5.28)	(-6.80)	(-4.92)	(-5.33)	(-5.20)	(-5.35)	(-5.50)	(-7.86)	(-5.57)

Panel B. Portfolio returns for put options

	LOW	2	3	4	5	6	7	8	9	HIGH	H-L
EW	-0.48	-0.54	-0.51	-0.48	-0.59	-0.50	-0.49	-0.55	-0.88	-2.25	-1.78
	(-3.78)	(-3.81)	(-3.69)	(-3.34)	(-4.69)	(-4.35)	(-4.18)	(-5.12)	(-6.04)	(-9.62)	(-11.87)
SVW	-0.46	-0.52	-0.53	-0.52	-0.53	-0.40	-0.40	-0.47	-0.55	-1.29	-0.83
	(-3.95)	(-4.30)	(-4.14)	(-5.21)	(-4.92)	(-4.50)	(-4.61)	(-4.93)	(-5.96)	(-9.71)	(-8.02)
OVW	-0.64	-0.83	-0.65	-0.78	-0.53	-0.62	-0.55	-0.62	-0.83	-1.99	-1.35
	(-4.08)	(-4.86)	(-4.34)	(-4.15)	(-5.28)	(-5.89)	(-4.80)	(-4.50)	(-4.24)	(-8.20)	(-8.35)

Table 1.5. Risk-adjusted Option Portfolio Returns

This table presents risk-adjusted long-short option portfolio returns (i.e., alphas) and the corresponding loadings on the risk factors. I form the zero-cost long-short portfolio as in Table 1.4, and regress the excess portfolio return on different risk factors. Panels A and B present the results for call option portfolio and put option portfolio, respectively. Alpha is the intercept term from the regression of the excess return on the following risk factors: (1) the CAPM market factor, (2) the Fama-French three factors, (3) the Fama-French three factors plus the Carhart momentum factor, (4) the zero-straddle S&P500 index return, (5) the change in the VIX index, and (6) all the factors. The *t*-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ***, **, and, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

Panel A: Call option portfolio return

	(1)	(2)	(3)	(4)	(5)	(6)
Alpha	-2.10***	-2.12***	-2.11***	-2.18***	-2.18***	-2.11***
	(-13.93)	(-14.48)	(-14.21)	(-10.76)	(-12.60)	(-11.96)
MKTRF	-0.07**	-0.05	-0.05			-0.03
	(-2.29)	(-1.20)	(-1.26)			(-0.62)
SMB		-0.12**	-0.11**			-0.12**
		(-2.10)	(-1.98)			(-2.14)
HML		-0.00	-0.02			-0.02
		(-0.02)	(-0.28)			(-0.30)
MOM			-0.04			-0.04
			(1.04)			(-1.16)
ZB_STRAD_SP500				0.00		0.00
				(0.09)		(0.26)
ΔVIX					0.05*	0.03
					(1.77)	(0.65)

Panel B: Put option portfolio return

	(1)	(2)	(3)	(4)	(5)	(6)
Alpha	-1.72***	-1.71***	-1.74***	-1.84***	-1.77***	-1.78***
	(-10.48)	(-10.91)	(-10.96)	(-12.07)	(-11.62)	(-9.17)
MKTRF	-0.05	-0.05**	-0.05**			-0.05
	(-1.59)	(-1.98)	(-1.99)			(-1.15)
SMB		0.03	0.03			0.03
		(0.58)	(0.46)			(0.46)
HML		-0.01	0.01			0.01
		(-0.35)	(0.34)			(0.33)
MOM			0.06			0.06
			(1.01)			(1.04)
ZB_STRAD_SP500				-0.01		-0.00
				(-1.40)		(-0.83)
ΔVIX					0.04	-0.01
					(1.45)	(-0.18)

Table 1.6. Robustness Tests: Alternative Measures of Option Returns and Retail Trading Volume

This table reports the regression results, using alternative construction methods of delta-hedged option returns and retail trading volume. Column (1) and (5) show the results for option returns over the next month (Oret_ed) with the baseline retail trading volume measure (Rtrd), and other columns report the results using the baseline delta-hedged option returns (Oret_ex) but with alternative retail trading volume measures. Rtrd is the baseline retail trading volume measure. *Rtrd_trades* is defined as the monthly average of the ratio of the number of retail trades and total number of trades in the same day. *Rtrd_vol1y* is defined as the monthly average of the ratio of the retail trading volume and the average trading volume in previous one year. *Rtrd_total* is defined as the ratio of the total retail trading volume in previous month and the total trading volume in previous month. The control variables are the same as in the baseline specification. *t*-statistics are shown in parentheses using the Newey and West (1987) corrected standard errors. ***, **, and, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

	Delta-hedged Call option return				Delta-hedged Put option return			
	(1) Oret_ed	(2) Oret_ex	(3) Oret_ex	(4) Oret_ex	(5) Oret_ed	(6) Oret_ex	(7) Oret_ex	(8) Oret_ex
Rtrd	-0.24***				-0.29***			
	(-6.22)				(-12.21)			
Rtrd_trades		-0.20***				-0.33***		
		(-4.68)				(-8.36)		
Rtrd_vol1y			-0.14***				-0.09***	
			(-3.92)				(-5.53)	
Rtrd_total				-0.35***				-0.37***
				(-7.17)				(-13.14)
IVol	-0.32***	-0.46***	-0.45***	-0.38***	-0.24***	-0.28***	-0.34***	-0.24***
	(-7.61)	(-7.45)	(-7.73)	(-6.64)	(-5.52)	(-5.44)	(-6.66)	(-4.96)
EIS	-0.14***	-0.27***	-0.29***	-0.23***	-0.03	-0.11	-0.14*	-0.08
	(-2.71)	(-3.91)	(-4.33)	(-3.41)	(-0.51)	(-1.49)	(-1.91)	(-1.14)
MIS	-0.01	-0.05	-0.05*	-0.05	-0.01	-0.02	-0.02	-0.01
	(-0.54)	(-1.66)	(-1.75)	(-1.58)	(-0.70)	(-0.74)	(-0.78)	(-0.53)
MAX	0.09**	0.05	0.07	0.05	0.00	-0.03	-0.00	-0.02
	(2.14)	(0.95)	(1.40)	(0.89)	(0.06)	(-0.56)	(-0.09)	(-0.49)
ILLIQ	0.01	-0.02	-0.10**	0.01	0.04*	0.03	-0.11***	0.00
	(0.18)	(-0.41)	(-2.26)	(0.18)	(1.73)	(0.96)	(-3.79)	(0.14)
OPBaspread	-0.23***	-0.21***	-0.21***	-0.23***	-0.11**	-0.07*	-0.06	-0.08*
	(-7.42)	(-8.19)	(-7.84)	(-8.50)	(-2.51)	(-1.70)	(-1.37)	(-1.91)
Ln(Mep)	-0.03	0.09*	0.08	0.06	-0.01	0.00	-0.01	-0.03
	(-0.90)	(1.76)	(1.55)	(1.20)	(-0.30)	(0.08)	(-0.34)	(-0.65)

Table 1.7. Lottery Premium Hypothesis Test: The Role of Retail Order Imbalance

This table reports the average returns (in percentage) of holding short-maturity at-the-money call options (Panel A) or put options (Panel B) on stocks associated with different level of retail trading order imbalances. I use two measures to estimate the retail trading order imbalance. Rnby is the net retail buying volume, scaled by the total daily trading volume on that day. OIbjzz is the total retail buying volume minus the total retail selling volume, but scaled by the total retail trading volume on that day (Boehmer et al., 2022). Both measures are estimated at the daily level, and are then averaged to the monthly level to match with the option returns. At the end of each month, I rank all optionable stocks into decile groups based on the corresponding retail order imbalance in that month. For each stock, for the call option, I buy one contract of call option against a short position of delta shares of the underlying stock, where delta is the Black-Scholes call option delta. For the put option, I buy one contract of put option against a long position of delta shares of the underlying stock. The delta hedging is rebalanced each day. For each stock in each month, I compound the daily returns of the rebalanced delta-hedged call options or put options over the remaining maturity to estimate the delta-hedged option returns. I use three weighting schemes to compute the average portfolio returns: equal weighted (EW), weighted by the market capitalization of the underlying stock (SVW), and weighted by the option open interest at the end of previous month (OVW). (H – L) represents the long-short hedge return of a zero-cost portfolio that longs the high decile portfolio and short the low decile portfolio. The *t*-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ***, **, and, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

Panel A: Delta-hedged call option

	Rnby			OIbjzz		
	(1) EW	(2) SVW	(3) OVW	(4) EW	(5) SVW	(6) OVW
1 (Net sell)	-1.56	-0.52	-1.47	-1.07	-0.57	-1.01
2	-1.00	-0.40	-0.91	-0.92	-0.41	-0.85
3	-0.82	-0.37	-0.74	-1.01	-0.40	-1.02
4	-0.78	-0.34	-0.74	-0.93	-0.37	-0.74
5 (moderate)	-0.68	-0.43	-0.86	-0.85	-0.37	-0.74
6 (moderate)	-0.75	-0.43	-0.96	-0.94	-0.38	-1.03
7	-0.74	-0.29	-0.78	-0.89	-0.34	-0.80
8	-0.78	-0.40	-0.80	-1.05	-0.43	-0.98
9	-0.86	-0.37	-0.80	-0.96	-0.38	-1.23
10 (Net buy)	-1.59	-0.59	-2.05	-0.94	-0.44	-1.39
H – L	-0.03	-0.08	-0.58*	0.13	0.13	-0.38
(*t*-stat)	(-0.20)	(-1.56)	(-1.72)	(1.27)	(1.44)	(-1.36)

Panel B: Delta-hedged put option

	Rnby			OIbjzz		
	(1) EW	(2) SVW	(3) OVW	(4) EW	(5) SVW	(6) OVW
1 (Net sell)	-1.24	-0.53	-1.36	-0.79	-0.48	-1.07
2	-0.70	-0.38	-0.78	-0.72	-0.42	-0.66
3	-0.67	-0.44	-0.64	-0.72	-0.39	-0.76
4	-0.60	-0.40	-0.58	-0.69	-0.45	-0.67
5	-0.57	-0.52	-0.83	-0.75	-0.49	-0.82
6	-0.54	-0.48	-0.55	-0.73	-0.51	-0.76
7	-0.59	-0.51	-0.63	-0.72	-0.50	-0.68
8	-0.63	-0.57	-0.78	-0.79	-0.53	-0.75
9	-0.70	-0.46	-0.82	-0.70	-0.60	-1.15
10 (Net	-1.02	-0.68	-1.45	-0.64	-0.51	-0.98
H – L	0.22***	-0.16**	-0.09	0.15*	-0.03	0.09
(*t*-stat)	(3.69)	(-2.04)	(-0.30)	(1.94)	(-0.37)	(0.46)

Table 1.8. Noise Trader Risk Hypothesis Test: The Role of Retail Trading Volatility

This table reports the average returns (in percentage) of selling short-maturity at-the-money call options (Panel A) or put options (Panel B) on stocks associated with different volatility level of the retail trading activities. Three retail trading activities are used, inclusion the total retail trading volume (Rtrd), the net retail buying volume (Rnby), and the Boehmer et al. (2022) retail order imbalance (OIbjzz). Std(*) is the standard deviation of the three measures in the previous month. At the end of each month, I rank all optionable stocks into decile groups based on the corresponding volatilities of the three retail trading activities. I use three weighting schemes to compute the average portfolio returns: equal weighted (EW), weighted by the market capitalization of the underlying stock (SVW), and weighted by the option open interest at the end of previous month (OVW). (H – L) represents the long-short hedge return of a zero-cost portfolio that longs the high volatility decile and short the low volatility decile portfolio. The *t*-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ***, **, and, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

Panel A: Delta-hedged call option

	Std(Rtrd)			Std(Rnby)			Std(OIbjzz)		
	(1) EW	(2) SVW	(3) OVW	(4) EW	(5) SVW	(6) OVW	(7) EW	(8) SVW	(9) OVW
LOW	-0.54	-0.41	-0.74	-0.54	-0.39	-0.74	-0.93	-0.32	-0.82
2	-0.56	-0.30	-0.59	-0.51	-0.28	-0.63	-0.94	-0.33	-0.88
3	-0.56	-0.27	-0.58	-0.59	-0.41	-0.68	-0.89	-0.48	-1.12
4	-0.55	-0.39	-0.65	-0.60	-0.34	-0.76	-0.91	-0.38	-1.02
5	-0.69	-0.46	-0.88	-0.71	-0.34	-0.94	-0.90	-0.47	-1.13
6	-0.74	-0.34	-0.97	-0.88	-0.47	-1.05	-0.88	-0.52	-1.39
7	-0.87	-0.50	-1.12	-1.00	-0.57	-1.80	-0.98	-0.63	-0.82
8	-1.20	-0.70	-1.99	-1.10	-0.57	-1.47	-0.98	-0.59	-1.17
9	-1.61	-0.71	-1.96	-1.51	-0.75	-2.64	-1.06	-0.61	-1.64
High	-2.24	-0.94	-3.00	-2.13	-0.66	-2.47	-1.08	-0.75	-1.57
H – L	-1.71***	-0.53***	-2.26***	-1.59***	-0.26***	-1.73***	-0.16	-0.43***	-0.75***
(*t*-	(-13.01)	(-4.37)	(-6.66)	(-14.14)	(-2.71)	(-4.08)	(-1.36)	(-3.43)	(-4.26)

Panel B: Delta-hedged put option

	Std(Rtrd)			Std(Rnby)			Std(OIbjzz)		
	(1) EW	(2) SVW	(3) OVW	(4) EW	(5) SVW	(6) OVW	(7) EW	(8) SVW	(9) OVW
LOW	-0.46	-0.50	-0.59	-0.49	-0.51	-0.51	-0.90	-0.49	-0.72
2	-0.51	-0.42	-0.46	-0.53	-0.50	-0.73	-0.78	-0.44	-0.64
3	-0.43	-0.42	-0.46	-0.47	-0.45	-0.59	-0.74	-0.58	-0.74
4	-0.49	-0.49	-0.80	-0.48	-0.46	-0.61	-0.69	-0.45	-0.86
5	-0.46	-0.49	-0.59	-0.54	-0.43	-0.75	-0.69	-0.49	-1.03
6	-0.62	-0.55	-0.88	-0.58	-0.48	-0.82	-0.64	-0.46	-1.04
7	-0.63	-0.54	-0.93	-0.67	-0.47	-1.13	-0.69	-0.53	-1.01
8	-0.78	-0.60	-1.26	-0.82	-0.52	-1.16	-0.58	-0.53	-1.02
9	-1.12	-0.59	-1.60	-1.01	-0.60	-1.63	-0.74	-0.53	-1.27
High	-1.77	-0.86	-2.91	-1.66	-0.73	-2.43	-0.84	-0.58	-1.37
H – L	-1.31***	-0.36***	-2.31***	-1.17***	-0.22**	-1.92***	0.06	-0.09	-0.64***
(*t*-	(-12.25)	(-4.98)	(-9.72)	(-11.66)	(-2.52)	(-5.23)	(0.51)	(-0.97)	(-2.93)

Table 1.9. Noise Trader Risk Hypothesis Test: The Impact of Trading Volatility on the Pricing Effect of Retail Trading Volume

This table tests the impact of volatility of retail trading activities on the pricing effect of retail trading volume. Each month, I regress the delta-hedged call option returns (in percentage) or put option returns (in percentage) on the total retail trading volume and the interaction terms with the trading volatility variables. The time-series average of the cross-sectional regression coefficients are reported. All control variables are standardized (N~(0,1)) to make the results comparable. Column (1) to (4) show the results for call option, and Column (5) and (8) report the results for put option. I use two measures to estimate the volatility of retail trading activities: the volatility of the retail trading volume (Std(Rtrd)), and the volatility of the net retail buying volume (*Std(Rnby)*). The variables of interest are the total retail trading volume (*Rtrd*) and the interaction terms. The interaction terms are based on both indicator variables and continuous variables, including high trading volume volatility dummy (Std(Rtrd)>*Q80%*) and high volatility of retail order imbalance dummy (*Std(Rnby)*>*Q80%*). The *t*-statistics are shown in parentheses using the Newey and West (1987) corrected standard errors. ***, **, and, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

(Table 1.9 continued)

	Call option				Put option			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rtrd	-0.14***	-0.18***	-0.33***	-0.36***	-0.12***	-0.20***	-0.25***	-0.27***
	(-4.04)	(-3.28)	(-6.55)	(-6.29)	(-3.35)	(-5.08)	(-6.10)	(-8.14)
Rtrd × Std(Rtrd)>Q80%	-0.43***				-0.46***			
	(-6.41)				(-6.23)			
Rtrd × Std(Rnby)>Q80%		-0.38***				-0.34***		
		(-8.94)				(-4.11)		
Rtrd × Std(Rtrd)			-0.11***				-0.14***	
			(-5.98)				(-4.25)	
Rtrd × Std(Rnby)				-0.09***				-0.11***
				(-5.23)				(-3.33)
Std(Rtrd)	0.15*		0.15***		0.07		0.02	
	(1.93)		(3.43)		(0.87)		(0.48)	
Std(Rnby)		0.15***		0.18***		0.06		0.00
		(2.96)		(4.83)		(0.90)		(0.03)
IVol	-0.41***	-0.41***	-0.40***	-0.39***	-0.27***	-0.27***	-0.28***	-0.28***
	(-7.39)	(-7.28)	(-7.17)	(-6.82)	(-5.75)	(-6.00)	(-5.89)	(-6.00)
EIS	-0.24***	-0.24***	-0.24***	-0.23***	-0.09	-0.08	-0.09	-0.09
	(-3.52)	(-3.51)	(-3.56)	(-3.52)	(-1.24)	(-1.21)	(-1.30)	(-1.26)
MIS	-0.06*	-0.06*	-0.05	-0.05	-0.02	-0.02	-0.02	-0.02
	(-1.88)	(-1.89)	(-1.57)	(-1.56)	(-0.99)	(-0.78)	(-0.68)	(-0.65)
MAX	0.05	0.05	0.05	0.05	-0.02	-0.02	-0.02	-0.02
	(0.97)	(0.95)	(0.93)	(0.85)	(-0.42)	(-0.43)	(-0.49)	(-0.48)
ILLIQ	0.07	0.07	0.09	0.05	0.07**	0.06**	0.14***	0.14***
	(1.34)	(1.42)	(1.46)	(1.03)	(2.13)	(2.15)	(3.82)	(3.69)
OPBaspread	-0.22***	-0.22***	-0.23***	-0.24***	-0.07*	-0.08*	-0.08*	-0.08*
	(-8.13)	(-8.08)	(-8.79)	(-8.65)	(-1.74)	(-1.84)	(-1.90)	(-1.91)
Ln(Mep)	0.05	0.05	0.10*	0.10**	-0.05	-0.05	-0.03	-0.03
	(0.97)	(1.02)	(1.96)	(2.13)	(-1.38)	(-1.25)	(-0.66)	(-0.73)

Table 1.10. Noise Trader Risk Hypothesis Test: The Impact of Arbitrage Cost on the Pricing Effect of Retail Trading Volume

This table tests the impact of arbitrage cost on the pricing effect of retail trading volume. Each month, I regress the delta-hedged call option returns (in percentage) or put option returns (in percentage) on the total retail trading volume and the interaction terms with the arbitrage cost variables. All control variables are standardized (N~(0,1)) to make the results comparable. Column (1) to (4) show the results for call options, and Column (5) and (8) report the results for put options. I use two measures of the arbitrage cost: the idiosyncratic volatility (*IVOL*) and the short interest (*SIR*). The variables of interest are the total retail trading volume (*Rtrd*) and the interaction terms. The interaction terms are based on both indicator variables and continuous variables, including high IVol dummy (*IVol>Q80%*) and high short interest dummy (*SIR%>Q80%*). The *t*-statistics are shown in parentheses using the Newey and West (1987) corrected standard errors. ***, **, and, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

(Table 1.10 continued)

	Call option				Put option			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rtrd	-0.20***	-0.28***	-0.28***	-0.33***	-0.26***	-0.20***	-0.31***	-0.29***
	(-4.14)	(-5.61)	(-7.00)	(-7.41)	(-8.19)	(-5.03)	(-11.98)	(-7.75)
Rtrd × IVol>Q80%	-0.35***				-0.25***			
	(-7.76)				(-3.35)			
Rtrd × SIR%>Q80%		-0.25***				-0.49***		
		(-2.98)				(-10.33)		
Rtrd × IVol			-0.11***				-0.10***	
			(-6.12)				(-4.15)	
Rtrd × SIR				-0.09***				-0.20***
				(-2.77)				(-10.71)
SIR		0.03		0.03		-0.09***		-0.07**
		(1.08)		(1.53)		(-2.75)		(-2.20)
IVol	-0.35***	-0.39***	-0.33***	-0.39***	-0.21***	-0.23***	-0.19***	-0.22***
	(-6.26)	(-7.25)	(-5.93)	(-7.20)	(-4.93)	(-5.19)	(-4.41)	(-5.11)
EIS	-0.24***	-0.22***	-0.24***	-0.22***	-0.09	-0.05	-0.09	-0.05
	(-3.48)	(-3.36)	(-3.52)	(-3.45)	(-1.21)	(-0.67)	(-1.27)	(-0.63)
MIS	-0.06*	-0.05	-0.06*	-0.05	-0.02	-0.00	-0.02	-0.00
	(-1.86)	(-1.50)	(-1.94)	(-1.53)	(-0.66)	(-0.07)	(-0.81)	(-0.12)
MAX	0.05	0.05	0.06	0.05	-0.02	-0.02	-0.01	-0.02
	(0.90)	(0.92)	(1.13)	(0.93)	(-0.55)	(-0.45)	(-0.30)	(-0.45)
ILLIQ	0.03	-0.01	0.03	-0.01	0.03	-0.07**	0.03	-0.08**
	(0.73)	(-0.15)	(0.72)	(-0.15)	(0.94)	(-2.27)	(0.98)	(-2.51)
OPBaspread	-0.22***	-0.22***	-0.22***	-0.22***	-0.08*	-0.10**	-0.07*	-0.10**
	(-8.29)	(-7.48)	(-8.31)	(-7.39)	(-1.80)	(-2.29)	(-1.72)	(-2.24)
Ln(Mep)	0.06	0.07	0.07	0.07	-0.03	-0.06*	-0.02	-0.06
	(1.28)	(1.29)	(1.52)	(1.40)	(-0.70)	(-1.95)	(-0.44)	(-1.64)

Table 1.11. The Impact of Abel Noser Institutional Trading Activities on Option Returns

This table reports the regress results of the impact of institutional trading activities on the cross-sectional option returns. Panel A shows the summary statistics of the institutional trading activities, and Panel B reports the regression results. Each month, I regress the delta-hedged call option returns (in percentage) or put option returns (in percentage) on different institutional trading activities measures estimated from the Abel Noser dataset. I construct four institutional trading activities, including the institutional trading volume (Itrd), the net institutional buying volume (Inby), and the volatility of the two measures (Std(Itrd) and Std(Inby)). All control variables are standardized (N~(0,1)) to make the results comparable. Column (1) to (4) show the results for call options, and Column (5) and (8) report the results for put options. The *t*-statistics are shown in parentheses using the Newey and West (1987) corrected standard errors. ***, **, and, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

Panel A: Summary of the Abel Noser institutional trading measures

Variables	Mean	Media	Std	P10	P25	P75	P90
Itrd (%)	6.30	5.41	4.09	1.94	3.30	8.36	11.84
Inby (%)	0.21	0.14	3.24	-3.37	-1.30	1.69	3.89

(Table 1.11 continued)

Panel B: Fama-MacBeth regression

	Call option				Put option			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Itrd	-0.03				0.05			
	(-0.74)				(1.39)			
Std(Itrd)		-0.04				0.02		
		(-1.03)				(0.64)		
Inby			0.01				0.02	
			(0.38)				(1.13)	
Std(Inby)				-0.03				0.01
				(-0.82)				(0.32)
IVol	-0.39***	-0.40***	-0.39***	-0.40***	-0.20**	-0.20**	-0.21**	-0.21**
	(-4.44)	(-4.54)	(-4.14)	(-4.56)	(-2.37)	(-2.39)	(-2.30)	(-2.43)
EIS	-0.03	-0.03	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03
	(-1.44)	(-1.44)	(-1.45)	(-1.44)	(-1.45)	(-1.46)	(-1.47)	(-1.46)
MIS	-0.12**	-0.12**	-0.12*	-0.12**	-0.01	-0.01	-0.01	-0.01
	(-2.01)	(-2.01)	(-1.97)	(-2.00)	(-0.14)	(-0.16)	(-0.17)	(-0.16)
MAX	0.29***	0.29***	0.29***	0.29***	0.09**	0.09**	0.09**	0.09**
	(5.68)	(5.63)	(5.51)	(5.67)	(2.45)	(2.39)	(2.34)	(2.36)
ILLIQ	-0.12***	-0.12***	-0.13***	-0.12***	-0.11**	-0.11**	-0.11**	-0.11**
	(-3.03)	(-2.99)	(-3.00)	(-2.98)	(-2.28)	(-2.28)	(-2.29)	(-2.31)
OPBaspread	-0.06	-0.05	-0.06	-0.05	0.17**	0.17**	0.17**	0.17**
	(-1.36)	(-1.30)	(-1.44)	(-1.30)	(2.24)	(2.25)	(2.34)	(2.27)
Ln(Mep)	0.28***	0.27***	0.27***	0.27***	0.01	0.02	0.01	0.02
	(2.97)	(3.07)	(2.93)	(3.10)	(0.30)	(0.42)	(0.27)	(0.39)

Table 1.12. Evidence from out-of-money options

This table examines the impact of different retail trading activities on the cross-sectional delta-hedged out-of-money option returns. Panels A and B show the sample distribution of the out-of-money options, and Panel C reports the monthly Fama-MacBeth regression results. I examine four retail trading activities, namely the retail trading volume (Rtrd), the net retail trading volume (Rnby), and the volatility of the two measures (Std(Rtrd) and Std(Rnby)). All control variables are standardized (N~(0,1)) to make the results comparable. Column (1) to (4) in Panel C show the results for call options, and column (5) and (8) report the results for put options. The *t*-statistics are shown in parentheses using the Newey and West (1987) corrected standard errors. ***, **, and, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

Panel A: Summary statistics of the Delta-hedged call option sample (77,516 obs)

Variables	Mean	Media	Std	P10	P25	P75	P90
Oret_ex (%)	-4.82	-5.32	13.55	-14.75	-9.78	-1.52	3.60
Oret_ed (%)	-3.27	-3.69	10.16	-10.82	-7.00	-0.54	3.49
Maturity (days)	50	50	3	46	49	51	53
Moneyness (%)	113.2	112.6	2.50	110.4	111.2	114.7	117.1
Vega	0.10	0.10	0.03	0.06	0.08	0.12	0.13
OPBaspread (%)	55.64	40.00	44.77	9.23	20.69	80.00	130.4

Panel B: Summary statistics of the Delta-hedged put option sample (85,140 obs)

Variables	Mean	Media	Std	P10	P25	P75	P90
Oret_ex (%)	-3.92	-4.40	9.82	-12.16	-7.88	-1.50	3.40
Oret_ed (%)	-2.99	-3.40	7.77	-9.45	-6.17	-0.85	2.79
Maturity (days)	50	50	3	46	49	51	52
Moneyness (%)	86.89	87.53	2.45	83.07	85.46	88.85	89.55
Vega	0.09	0.09	0.02	0.06	0.07	0.10	0.12
OPBaspread (%)	51.05	40.00	42.96	7.93	18.18	66.67	120.0

(Table 1.12 continued)

Panel C: Fama-MacBeth regression

	Call option				Put option			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rtrd	-0.79***				-0.70***			
	(-7.45)				(-11.69)			
Rnby		0.02				0.13**		
		(0.43)				(2.49)		
Std(Rtrd)			-0.30***				-0.40***	
			(-4.05)				(-9.70)	
Std(Rnby)				-0.45***				-0.47***
				(-6.39)				(-12.44)
IVol	-0.28***	-0.53***	-0.51***	-0.55***	-0.20**	-0.21**	-0.20**	-0.21**
	(-3.70)	(-6.78)	(-6.44)	(-7.17)	(-2.37)	(-2.30)	(-2.39)	(-2.43)
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2.1: Summary Statistics

This table reports the summary statistic of my sample data. Panel A reports the pooled distribution of retail investor trading activities. Panel B to Panel D report the time-series average of the mean values of stock characteristics on decile portfolios sorted by aggregate retail selling orders, aggregate retail buying orders, and an order imbalance measure. Panel E reports the time-series average correlation between different variables. In Panel B to E, I rebalance the decile portfolio at the daily frequency. $DVol$ is the average daily trading volume of individual stock, reported in dollars. $SVol$ is the average daily trading volume of individual stock, reported in shares. $NTrd$ is the average of total number of trading orders on each day. $DVolR$, $SVolR$, and $NTrdR$ are defined similar for the retail trading activities. $Rbuy\_DVol$, $Rbuy\_SVol$, $Rbuy\_NTrd$, $Rsel\_DVol$, $Rsel\_SVol$, $Rsel\_NTrd$ are defined similar for retail buying orders and retail selling orders. R_{buy} is the total daily buying volume from retail investors scaled by the total trading volume on that day. R_{sel} is the total daily selling volume from retail investors scaled by the total trading volume on that day. R_{imb} is the difference between $Rbuy$ and $Rsel$. OI_{bjzz} is the retail order imbalance measure proposed by Boehmer et al. (2022), which is the total retail buying orders minus the total retail selling orders scaled by the total retail trading volume, equivalent to the ratio between retail buying orders and retail selling orders. Return t+1 is the equal-weighted average return (in bps) for each portfolio in day t+1. Return t is the equal-weighted average return (in bps) for each portfolio on the formation day (day t). STR21 is the short term reversal factor, defined as the buy and hold return in previous 21 trading days. MAX21 is the maximum daily return during previous 21 trading days. MOM252 is the buy and hold return during previous 252 trading days except the recent 21 trading days. IVOL21 is the idiosyncratic volatility during the previous 21 trading days adjusted by the Fama-French three factor model. Mep is the market capitalization on the portfolio formation day (i.e. day t), in $Billions. B/M is the most recent available book-to-market equity. OP is the most recent available operating profitability. Spread is the average daily bid-ask spread scaled by the average bid-ask prices during the previous 21 trading days. ILLIQ is the Amihud (2002) illiquidity measure at day t, estimated using the previous 21 trading day data. To t is the stock turnover on day t. To3 is the average daily stock turnover in previous 63 days. IO% is the percentage of shares held by institutional investors in previous quarter. Analysts are the number of analysts covering the stocks in previous month. Dispersion is the standard deviation of the analyst forecasts. All continuous variables are winsorized at 1% and 99% levels. My sample covers 1762 trading days during 2010 and 2016.

Panel A: Summary statistics of retail trading activities

		N	Mean	STD	Median	Q1	Q3
1	DVol ($)	4,904,080	38,141,222	94,434,810	4,577,305	695,409	26,921,172
2	SVol	4,904,080	1,069,951	2,448,620	232,859	62,005	864,335
3	NTrd	4,904,080	6,338	11,441	1,968	530	6,563
4	DVolR($)	4,904,080	2,310,808	6,757,418	263,288	60,044	1,253,591
5	SvolR	4,904,080	77,067	200,378	14,540	4,442	52,180
6	NTrdR	4,904,080	228	480	65	21	201
7	Rbuy_Dvol	4,904,080	1,146,970	3,401,315	122,351	25,718	603,972
8	Rbuy_Svol	4,904,080	38,312	101,670	6,788	1,900	25,275
9	Rbuy_NTrd	4,904,080	115	249	31	9	98
10	Rsel_Dvol	4,904,080	1,158,276	3,386,778	131,576	28,604	627,101
11	Rsel_Svol	4,904,080	38,547	99,759	7,297	2,100	26,381
12	Rsel_NTrd	4,904,080	113	236	32	10	101
13	R_{buy}(%)	4,904,080	4.11	4.69	2.46	1.33	4.89
14	R_{sel}(%)	4,904,080	4.28	4.79	2.62	1.46	5.06
15	R_{imb}(%)	4,904,080	-0.18	4.56	-0.08	-1.21	0.96
16	OI_{bjzz}	4,904,080	-0.03	0.37	-0.02	-0.25	0.19

Panel B: Summary statistics of stock characteristics (sorted by aggregated retail selling orders, Rsel)

	1	2	3	4	5	6	7	8	9	10
Rsel (%)	0.57	1.09	1.49	1.90	2.37	2.96	3.78	5.09	7.64	15.90
Rbuy (%)	2.11	2.01	2.19	2.45	2.80	3.28	4.00	5.14	7.14	9.90
OIbjzz	0.32	0.11	0.04	-0.01	-0.04	-0.07	-0.09	-0.11	-0.14	-0.29
Return *t*+1	10.03	8.51	7.33	7.19	6.95	7.18	7.24	7.05	4.73	-0.38
Return *t*	7.97	7.25	7.63	8.09	8.83	8.79	9.84	13.85	23.95	-8.67
STR21 (%)	1.04	1.20	1.29	1.34	1.42	1.41	1.48	1.60	1.95	1.72
Max21 (%)	4.44	4.33	4.38	4.49	4.66	4.90	5.35	6.17	7.45	8.57
MOM252	15.52	16.84	17.52	17.91	18.21	18.65	18.87	18.79	16.79	11.63
IVOL21	1.46	1.42	1.45	1.49	1.56	1.66	1.84	2.16	2.65	3.10
Mep ($B)	2.15	3.77	4.91	6.11	7.62	9.46	10.00	6.62	2.42	0.36
B/M	0.66	0.61	0.60	0.60	0.61	0.62	0.65	0.69	0.73	0.85
OP	0.12	0.13	0.14	0.14	0.14	0.13	0.12	0.11	0.07	0.03
Spread	2.92	2.82	2.83	2.88	2.98	3.12	3.38	3.87	4.61	5.20
ILLIQ (10^{-}	0.09	0.04	0.04	0.04	0.05	0.06	0.08	0.12	0.20	0.55
To *t* (%)	0.60	0.72	0.74	0.75	0.75	0.75	0.76	0.79	0.79	0.52
To3 (%)	0.57	0.72	0.75	0.77	0.77	0.77	0.78	0.81	0.80	0.61
DVol ($M)	15.59	30.09	38.49	46.48	56.09	68.70	76.00	60.31	36.65	4.41
IO%	0.72	0.76	0.76	0.75	0.73	0.70	0.65	0.58	0.45	0.27
Analysts	7.63	9.83	10.54	10.81	10.80	10.53	9.81	8.22	6.06	3.39
Dispersion	0.21	0.19	0.19	0.19	0.20	0.22	0.25	0.30	0.36	0.38

Panel C: Summary statistics of stock characteristics (sorted by aggregated retail buying orders, Rbuy)

	1	2	3	4	5	6	7	8	9	10
Rbuy (%)	0.48	0.97	1.36	1.76	2.22	2.80	3.62	4.93	7.46	15.45
Rsel (%)	2.36	2.20	2.36	2.61	2.95	3.41	4.11	5.26	7.28	10.25
OIbjzz	-0.43	-0.20	-0.11	-0.05	-0.01	0.02	0.05	0.08	0.11	0.25
Return *t*+1	6.28	6.31	6.38	6.08	6.18	6.58	6.05	6.37	4.29	11.31
Return *t*	7.01	5.50	5.27	6.07	6.98	6.76	6.27	2.68	-0.07	41.16
STR21 (%)	1.77	1.54	1.45	1.41	1.39	1.37	1.28	1.21	1.28	1.72
Max21 (%)	4.58	4.36	4.39	4.49	4.66	4.90	5.31	6.08	7.37	8.59
MOM252	15.89	16.81	17.48	18.00	18.40	18.59	18.95	18.67	16.43	11.50
IVOL21 (%)	1.48	1.42	1.44	1.49	1.56	1.66	1.83	2.14	2.64	3.11
Mep (B)	2.07	3.67	4.79	5.97	7.42	9.02	9.70	7.34	2.93	0.50
B/M	0.68	0.62	0.61	0.60	0.61	0.62	0.64	0.69	0.73	0.84
OP	0.12	0.13	0.14	0.14	0.14	0.13	0.13	0.11	0.07	0.03
Spread	2.89	2.81	2.83	2.89	2.99	3.13	3.39	3.86	4.60	5.22
ILLIQ (10^{-5})	0.08	0.04	0.04	0.04	0.05	0.06	0.08	0.11	0.20	0.56
To *t* (%)	0.58	0.69	0.72	0.74	0.74	0.75	0.76	0.79	0.82	0.58
To3 (%)	0.57	0.70	0.74	0.75	0.76	0.77	0.78	0.81	0.82	0.63
DVol ($M)	14.79	28.19	36.71	44.61	54.58	66.00	72.97	65.44	42.00	7.53
IO%	0.71	0.76	0.76	0.75	0.73	0.70	0.66	0.58	0.46	0.28
Analysts	7.43	9.61	10.39	10.73	10.75	10.50	9.85	8.49	6.41	3.63
Dispersion	0.21	0.19	0.19	0.19	0.20	0.22	0.25	0.30	0.36	0.38

Panel D: Summary statistics of stock characteristics (sorted by aggregated retail orders imbalance, OIbjzz)

	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10
OIbjzz	-0.79	-0.58	-0.44	-0.34	-0.27	-0.21	-0.15	-0.10	-0.06	-0.02	0.02	0.06	0.09	0.14	0.18	0.24	0.31	0.40	0.53	0.76
Rbuy (%)	1.07	1.91	2.31	2.59	2.83	3.07	3.28	3.50	3.74	3.92	4.18	4.33	4.47	4.63	4.79	5.04	5.40	5.96	6.93	9.73
Rsel (%)	9.37	6.82	5.81	5.22	4.86	4.61	4.42	4.29	4.21	4.06	4.03	3.89	3.72	3.53	3.32	3.12	2.89	2.61	2.21	1.32
Return *t*+1	1.34	3.51	3.46	2.77	3.42	3.76	4.48	5.53	5.02	5.16	5.68	5.38	6.37	7.76	8.73	9.98	11.47	11.18	14.64	15.91
Return *t*	-9.51	-4.70	-3.15	-0.47	2.49	3.08	6.32	6.22	9.72	12.15	16.34	19.35	17.91	16.17	13.77	14.34	12.82	14.40	16.63	20.17
STR21 (%)	1.69	1.69	1.68	1.69	1.64	1.66	1.68	1.70	1.67	1.67	1.73	1.64	1.54	1.41	1.29	1.15	0.99	0.84	0.70	0.47
Max21 (%)	5.67	5.43	5.37	5.34	5.34	5.38	5.42	5.46	5.51	5.55	5.63	5.64	5.60	5.56	5.50	5.44	5.39	5.37	5.39	5.55
MOM252	13.71	14.97	15.88	16.54	17.16	17.71	18.06	18.69	19.30	19.21	20.02	19.70	19.45	18.40	17.85	17.07	16.14	15.37	14.00	12.03
IVOL21	1.96	1.85	1.83	1.82	1.82	1.83	1.84	1.86	1.88	1.89	1.92	1.92	1.92	1.91	1.89	1.88	1.87	1.87	1.89	1.99
Mep	0.84	1.67	2.41	3.27	4.41	5.56	6.92	8.16	9.31	9.58	10.36	9.79	8.66	7.17	5.79	4.44	3.30	2.41	1.62	0.78
B/M	0.81	0.73	0.69	0.67	0.65	0.63	0.62	0.62	0.61	0.63	0.61	0.61	0.62	0.62	0.64	0.64	0.66	0.68	0.73	0.82
OP	0.09	0.10	0.11	0.11	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.11	0.11	0.11	0.11	0.10	0.09
Spread	3.56	3.45	3.41	3.39	3.38	3.38	3.38	3.39	3.42	3.43	3.46	3.48	3.48	3.48	3.48	3.48	3.48	3.50	3.55	3.69
ILLIQ (10^{-5})	0.34	0.18	0.13	0.11	0.08	0.08	0.07	0.06	0.06	0.10	0.05	0.06	0.06	0.07	0.08	0.09	0.12	0.14	0.21	0.43
To *t* (%)	0.33	0.45	0.54	0.61	0.69	0.74	0.81	0.86	0.92	0.92	1.01	0.99	0.96	0.90	0.83	0.76	0.67	0.59	0.47	0.32
To3 (%)	0.39	0.51	0.60	0.67	0.74	0.78	0.83	0.88	0.92	0.90	0.97	0.96	0.92	0.88	0.82	0.76	0.68	0.61	0.50	0.35
DVol ($M)	5.09	11.47	17.06	23.33	32.30	41.02	51.51	62.78	75.08	84.96	96.36	87.48	74.08	59.42	46.69	35.59	25.93	18.55	11.49	4.48
IO%	0.52	0.60	0.63	0.65	0.66	0.67	0.67	0.67	0.68	0.66	0.68	0.67	0.67	0.67	0.66	0.66	0.64	0.63	0.59	0.50
Analysts	4.80	6.29	7.36	8.14	8.99	9.60	10.25	10.71	11.15	11.04	11.50	11.26	10.87	10.31	9.66	9.01	8.17	7.33	6.23	4.68
Dispersion	0.29	0.26	0.24	0.23	0.23	0.23	0.22	0.23	0.22	0.23	0.23	0.23	0.23	0.23	0.23	0.24	0.24	0.24	0.26	0.28

Panel E: Correlation matrix between different retail trading activities and some typical stock characteristics

No:	Variable	OI_{bjzz}	R_{imb}	R_{buy}	R_{sel}	R_{vol}	Return *t*	STR21	Max21	MOM252	IVOL21	Ln(Mep)	Ln(B/M)
(1)	OI_{bjzz}	1.00	0.94	0.45	-0.38	0.03	0.03	-0.03	0.01	-0.01	0.02	0.00	-0.01
(2)	R_{imb}	0.75	1.00	0.39	-0.38	0.00	0.04	-0.03	0.00	0.00	0.01	0.02	-0.01
(3)	R_{buy}	0.39	0.47	1.00	0.57	0.86	0.00	-0.03	0.28	-0.09	0.37	-0.37	0.02
(4)	R_{sel}	-0.35	-0.51	0.51	1.00	0.87	-0.03	-0.01	0.29	-0.09	0.36	-0.39	0.04
(5)	R_{vol}	0.02	-0.03	0.86	0.86	1.00	-0.02	-0.03	0.32	-0.11	0.41	-0.47	0.06
(6)	Return *t*	0.03	0.05	0.03	-0.02	0.01	1.00	0.19	0.02	0.01	-0.02	0.03	0.00
(7)	STR21	-0.03	-0.01	0.00	0.01	0.00	0.21	1.00	0.25	0.02	-0.01	0.07	0.01
(8)	Max21	0.00	-0.01	0.28	0.28	0.31	0.10	0.38	1.00	-0.15	0.81	-0.44	0.01
(9)	MOM252	0.00	0.00	-0.06	-0.06	-0.07	0.00	0.00	-0.10	1.00	-0.18	0.21	-0.03
(10)	IVOL21	0.01	-0.01	0.37	0.37	0.42	0.04	0.11	0.84	-0.11	1.00	-0.56	0.00
(11)	Ln(Mep)	0.00	0.02	-0.45	-0.46	-0.53	0.00	0.03	-0.38	0.14	-0.50	1.00	-0.24
(12)	Ln(B/M)	-0.01	-0.01	0.06	0.07	0.08	0.01	0.01	0.01	-0.02	0.00	-0.21	1.00

Table 2.2. Daily Portfolio Analysis Based on Aggregate Retail Orders

This table reports the time-series average returns or alphas (in bps) of daily rebalanced decile portfolios sorted by their aggregate retail orders. Panel A and Panel B report the results based on aggregate retail selling orders (Rsel) and aggregate retail buying orders (Rbuy) respectively. Excess return is the average returns in excess of the risk-free rate. Alpha is the intercept from the regression of daily excess returns on risk factors specified by an asset pricing model. The factor models include: the CAPM, the Fama-French three-factor model, and a four-factor model that includes Fama-French three factors and Carhart momentum factor. Long-Short return or Alpha is the return or alpha of a zero-cost portfolio that longs the corresponding high retail measure decile portfolio and shorts the low retail measure decile portfolio (i.e., H – L). I also report the long-short hedge portfolio return based on the 2nd and 9th decile portfolio. The t-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ** and * indicate statistical significance at the 1% and 5% levels, respectively, using two-tailed tests. The sample period is from January 2010 to December 2016, covering total 1762 trading days.

Panel A: decile portfolio based on aggregate retail selling orders (Rsel)

	Equal-weighted returns				Value-weighted returns			
	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha
1 (Low)	10.03	4.11	4.66	4.72	8.60	2.88	3.15	3.19
	(3.92)	(4.42)	(11.18)	(11.74)	(3.63)	(4.03)	(5.81)	(5.91)
2	8.51	2.44	2.88	2.97	7.04	1.31	1.46	1.48
	(3.28)	(2.89)	(7.13)	(7.80)	(3.00)	(2.36)	(3.10)	(3.13)
3	7.33	1.23	1.65	1.75	6.41	0.75	0.86	0.87
	(2.85)	(1.58)	(4.11)	(4.59)	(2.80)	(1.58)	(1.91)	(1.94)
4	7.19	1.04	1.46	1.55	6.51	0.92	1.00	1.01
	(2.76)	(1.32)	(3.93)	(4.27)	(2.90)	(2.10)	(2.36)	(2.40)
5	6.95	0.77	1.19	1.29	5.41	-0.02	0.00	0.03
	(2.68)	(0.97)	(3.26)	(3.68)	(2.52)	(-0.05)	(0.00)	(0.08)
6	7.18	0.97	1.41	1.51	5.87	0.67	0.64	0.65
	(2.65)	(1.12)	(3.56)	(3.93)	(2.83)	(1.69)	(1.63)	(1.65)
7	7.24	0.89	1.38	1.47	3.94	-1.27	-1.32	-1.31
	(2.58)	(0.92)	(2.87)	(3.22)	(1.87)	(-2.98)	(-3.18)	(-3.16)
8	7.05	0.57	1.14	1.31	4.35	-1.17	-1.17	-1.20
	(2.35)	(0.45)	(1.68)	(2.06)	(1.98)	(-1.92)	(-1.93)	(-1.97)
9	4.73	-1.74	-1.10	-0.86	3.88	-2.34	-2.27	-2.32
	(1.40)	(-1.05)	(-0.99)	(-0.81)	(1.41)	(-1.87)	(-1.92)	(-1.95)
10	-0.38	-5.41	-4.92	-4.77	-0.55	-6.27	-5.89	-5.85
	(-0.13)	(-3.02)	(-3.43)	(-3.38)	(-0.18)	(-3.86)	(-4.23)	(-4.22)
H – L	-10.42**	-9.52**	-9.58**	-9.50**	-9.14**	-9.14**	-9.04**	-9.04**
(*t*-stat)	(-7.10)	(-6.01)	(-6.11)	(-6.06)	(-5.77)	(-5.71)	(-5.94)	(-5.94)
(9)-(2)	-3.78**	-4.18**	-3.98**	-3.83**	-3.16*	-3.66**	-3.74**	-3.80**
(*t*-stat)	(-2.83)	(-3.22)	(-3.33)	(-3.23)	(-2.31)	(-2.67)	(-2.83)	(-2.86)

(Table 2.2, continued)

Panel B: decile portfolio based on aggregate retail buying orders (Rbuy)

	Equal-weighted returns				Value-weighted returns			
	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha
1 (Low)	6.28	0.51	1.05	1.11	5.78	0.29	0.56	0.59
	(2.50)	(0.54)	(2.68)	(2.87)	(2.52)	(0.42)	(1.14)	(1.22)
2	6.31	0.25	0.71	0.80	6.30	0.65	0.82	0.84
	(2.46)	(0.31)	(2.14)	(2.43)	(2.77)	(1.33)	(1.99)	(2.04)
3	6.38	0.26	0.69	0.78	6.05	0.43	0.53	0.55
	(2.49)	(0.34)	(1.95)	(2.34)	(2.72)	(0.89)	(1.21)	(1.26)
4	6.08	-0.08	0.34	0.43	6.01	0.44	0.51	0.52
	(2.33)	(-0.10)	(0.91)	(1.18)	(2.66)	(0.98)	(1.20)	(1.21)
5	6.18	-0.04	0.39	0.49	4.93	-0.54	-0.52	-0.50
	(2.32)	(-0.04)	(1.04)	(1.34)	(2.25)	(-1.34)	(-1.30)	(-1.24)
6	6.58	0.32	0.76	0.85	5.56	0.25	0.23	0.24
	(2.45)	(0.38)	(1.86)	(2.26)	(2.62)	(0.64)	(0.58)	(0.60)
7	6.05	-0.25	0.24	0.35	5.30	0.08	0.05	0.05
	(2.16)	(-0.26)	(0.49)	(0.79)	(2.56)	(0.23)	(0.13)	(0.14)
8	6.37	-0.12	0.44	0.60	4.51	-0.91	-0.91	-0.92
	(2.14)	(-0.10)	(0.65)	(0.94)	(2.04)	(-1.55)	(-1.57)	(-1.57)
9	4.29	-2.19	-1.55	-1.31	6.73	0.77	0.80	0.80
	(1.31)	(-1.39)	(-1.52)	(-1.37)	(2.62)	(0.69)	(0.75)	(0.75)
10	11.31	6.21	6.69	6.85	7.42	1.39	1.66	1.61
	(3.52)	(3.28)	(4.28)	(4.46)	(2.30)	(0.74)	(0.93)	(0.91)
H – L	5.03**	5.70**	5.65**	5.75**	1.64	1.11	1.10	1.01
(*t*-stat)	(3.21)	(3.39)	(3.39)	(3.47)	(0.84)	(0.56)	(0.57)	(0.53)
(9)-(2)	-2.03	-2.44*	-2.26*	-2.11*	0.43	0.12	-0.02	-0.03
(*t*-stat)	(-1.63)	(-2.03)	(-2.05)	(-1.97)	(0.35)	(0.09)	(-0.01)	(-0.03)

Table 2.3. 10×10 Double Sorted Portfolios Based on Aggregate Retail Selling and Retail Buying Orders

This table reports the 10×10 double sorted portfolio analysis results based on the retail buying measure and retail selling measure. In Panel A, I first sort stocks into decile groups based on the aggregate retail buying orders. To construct the 10×10 portfolios, I further sort the stocks within each decile group into decile groups based on the aggregate retail selling orders. In Panel B, I first sort stocks by retail selling orders and then sort stock by retail buying orders. The time series average returns or Fama-French four-factor alphas (in bps) of the decile portfolios are reported. The t-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ** and * indicate statistical significance at the 1% and 5% levels, respectively, using two-tailed tests. To save space, I only report the results for the 1st, 2nd, 5th, 6th, 9th, and 10th decile portfolios.

Panel A: 10×10 double sorted portfolios first by aggregate retail buying orders (Rbuy)

	Equal-weighted returns						Value-weighted returns					
	Low Buy	2	5	6	9	High Buy	Low Buy	2	5	6	(9)	High Buy
1 (Low sell)	10.79	9.06	9.81	11.10	12.83	18.34	8.18	9.11	7.40	10.16	8.65	11.67
2	8.60	8.28	7.69	9.31	13.01	20.40	8.13	7.96	6.86	6.88	5.77	12.57
5	8.16	6.70	4.94	7.73	5.95	12.82	7.08	5.24	5.15	5.49	8.30	7.56
6	3.95	5.52	6.23	7.12	4.16	7.38	3.39	6.70	4.92	7.22	7.17	-0.38
9	5.13	4.92	5.23	3.27	-1.97	1.16	4.59	5.44	2.86	5.14	-1.99	-0.27
10 (High sell)	1.48	3.67	2.76	-0.20	-9.15	1.69	2.13	4.51	2.65	3.34	-8.96	-3.13
H – L	-9.30**	-5.39**	-7.05**	-11.30**	-21.97**	-16.65**	-6.05**	-4.60**	-4.75*	-6.81**	-17.61**	-14.79**
(*t*-stat)	(-5.69)	(-3.56)	(-4.02)	(-5.97)	(-8.81)	(-5.30)	(-3.19)	(-2.80)	(-2.46)	(-3.05)	(-6.55)	(-4.94)
FF4 Alpha	-8.95**	-4.98**	-6.64**	-10.78**	-21.05**	-15.93**	-6.15**	-4.41**	-4.71*	-6.52**	-17.25**	-13.23**
(*t*-stat)	(-5.47)	(-3.36)	(-3.76)	(-5.75)	(-8.04)	(-4.97)	(-3.20)	(-2.68)	(-2.41)	(-2.94)	(-6.38)	(-4.23)

Panel B: 10×10 double sorted portfolios first by aggregate retail selling orders (Rsel)

	Equal-weighted returns						Value-weighted returns					
	Low Sell	2	5	6	9	High Sell	Low Sell	2	5	6	(9)	High Sell
1 (Low buy)	9.10	6.30	5.63	6.38	2.40	-1.74	7.13	4.79	4.92	7.72	5.12	3.30
2	9.52	6.68	4.89	4.39	2.60	-2.75	7.20	6.41	5.07	4.31	3.65	0.37
5	8.43	7.80	6.93	6.24	0.65	-5.63	7.42	6.90	4.62	6.34	4.92	-3.47
6	9.95	7.14	5.36	4.92	2.33	-6.70	11.54	7.53	5.62	6.23	3.41	-4.89
9	11.88	10.44	8.19	11.02	10.46	8.93	10.45	8.07	7.80	7.62	9.67	5.71
10 (High buy)	15.29	13.71	13.05	16.64	23.26	17.94	9.86	11.12	6.94	10.31	11.59	10.59
H – L	6.19**	7.41**	7.42**	10.26**	20.86**	19.68**	2.72	6.33**	2.03	2.59	6.47*	7.29*
(*t*-stat)	(3.75)	(4.45)	(4.46)	(5.42)	(6.86)	(6.99)	(1.47)	(3.56)	(0.95)	(1.22)	(2.08)	(2.38)
FF4 Alpha	6.33**	7.60**	7.50**	10.60**	21.82**	20.62**	2.12	6.10**	2.12	2.63	6.85*	8.42**
(*t*-stat)	(3.80)	(4.56)	(4.50)	(5.56)	(6.97)	(7.10)	(1.14)	(3.45)	(1.00)	(1.25)	(2.22)	(2.60)

Table 2.4. Fama-MacBeth Cross-sectional Return Regressions

This table reports the regression results. All stock characteristics are standardized (N~(0,1)) to make the results comparable. Panel A and B show the results based on equal-weighted least square (EWLS) and value-weight least square (VWLS) regressions. The dependent variable is a firm's daily stock return (in bps). The explanatory variable of interest is the aggregate retail selling orders (Rsel), aggregate retail buying orders (Rbuy), and aggregate retail order imbalance (OIbjzz). The control variables include: the Lee and Ready (1991) total trading order imbalance (TOI_LR), the stock returns in previous 3 days (Return(t), Return(t-1), and Return(t-2)), the daily turnover in previous day (To(t)), the logarithm of market capitalization in previous day (Ln(SIZE)), market beta (Beta252), the logarithm of book-to-market equity (Ln(B/M)), annual operating profitability (OP), total asset growth (TAG), momentum (MOM252), Short-term reversal (STR21), idiosyncratic volatility (IVOL21), maximum daily return (MAX21), and average daily turnover ratio during previous 3 months (TO3). The t-statistics are shown in parentheses using the Newey and West (1987) corrected standard errors with up to twelve lags. ** and * indicate statistical significance at the 1% and 5% levels, respectively, using two-tailed tests. The sample period is from January 2010 to December 2016, covering total 1762 trading days.

Panel A: EW results

	(1)	(2)	(3)	(4)	(5)	(6)
Rsel	-3.39**			-4.82**	-2.28**	
	(-10.45)			(-16.36)	(-6.22)	
Rbuy		2.27**		3.98**		0.78*
		(7.16)		(13.87)		(2.07)
OIbjzz			3.19**		2.43**	2.91**
			(22.04)		(13.55)	(14.81)
TOI_LR	1.63**	1.85**	1.53**	0.94**	1.32**	1.46**
	(8.07)	(9.08)	(7.59)	(4.65)	(6.54)	(7.18)
Return(t)	-6.30**	-6.35**	-6.29**	-6.35**	-6.31**	-6.33**
	(-11.59)	(-11.70)	(-11.58)	(-11.71)	(-11.61)	(-11.67)
Return(t-	-1.81**	-1.84**	-1.80**	-1.80**	-1.81**	-1.82**
	(-5.09)	(-5.18)	(-5.06)	(-5.11)	(-5.12)	(-5.15)
Return(t-	-1.86**	-1.87**	-1.83**	-1.82**	-1.83**	-1.83**
	(-5.45)	(-5.50)	(-5.38)	(-5.35)	(-5.38)	(-5.38)
To(t)	0.61	1.03*	0.78	0.68	0.60	0.82
	(1.33)	(2.25)	(1.70)	(1.49)	(1.32)	(1.79)
Ln(SIZE	-2.30**	-0.46	-1.16*	-1.47**	-1.92**	-0.95
	(-4.46)	(-0.90)	(-2.28)	(-2.82)	(-3.72)	(-1.84)
Beta252	-1.24	-0.37	-0.67	-0.82	-1.04	-0.56
	(-1.45)	(-0.43)	(-0.79)	(-0.96)	(-1.21)	(-0.66)
Ln(B/M)	0.37	0.54	0.50	0.46	0.43	0.52
	(0.79)	(1.18)	(1.09)	(1.00)	(0.94)	(1.12)
OP	0.65	1.10**	0.95*	0.87*	0.75	1.00**
	(1.69)	(2.87)	(2.45)	(2.26)	(1.95)	(2.61)
TAG	-0.62	-0.70*	-0.69*	-0.65*	-0.64	-0.68*
	(-1.86)	(-2.12)	(-2.09)	(-1.97)	(-1.95)	(-2.05)
MOM25	0.75	0.72	0.73	0.76	0.75	0.76
	(1.40)	(1.34)	(1.37)	(1.43)	(1.41)	(1.42)
STR21	-1.93**	-1.95**	-1.86**	-1.82**	-1.83**	-1.83**
	(-3.63)	(-3.68)	(-3.51)	(-3.44)	(-3.47)	(-3.45)
IVOL21	-1.68**	-2.94**	-2.44**	-2.27**	-1.99**	-2.64**
	(-2.64)	(-4.57)	(-3.72)	(-3.61)	(-3.14)	(-4.14)
MAX21	1.28*	1.49*	1.43*	1.34*	1.33*	1.42*
	(2.18)	(2.51)	(2.42)	(2.28)	(2.26)	(2.40)
TO3	-1.97**	-2.47**	-2.23**	-2.12**	-2.03**	-2.28**
	(-3.93)	(-4.84)	(-4.36)	(-4.23)	(-4.04)	(-4.50)

(Table 2.4, continued)

Panel B: VW results

	(1)	(2)	(3)	(4)	(5)	(6)
Rsel	-2.60**			-3.79**	-1.97**	
	(-4.59)			(-7.42)	(-3.23)	
Rbuy		-0.15		1.88**		-1.50*
		(-0.29)		(3.81)		(-2.47)
OIbjzz			1.56**		1.22**	2.03**
			(7.62)		(5.31)	(7.94)
TOI_LR	0.63*	0.80**	0.56*	0.49	0.46	0.51
	(2.29)	(2.95)	(2.02)	(1.76)	(1.66)	(1.84)
Return(t)	-2.17**	-2.23**	-2.15**	-2.18**	-2.15**	-2.16**
	(-3.93)	(-4.04)	(-3.91)	(-3.94)	(-3.90)	(-3.91)
Return(t-	-1.09*	-1.05*	-1.04*	-1.09*	-1.09*	-1.07*
	(-2.31)	(-2.23)	(-2.22)	(-2.32)	(-2.32)	(-2.28)
Return(t-	-0.59	-0.57	-0.60	-0.56	-0.57	-0.55
	(-1.25)	(-1.22)	(-1.27)	(-1.19)	(-1.20)	(-1.17)
To(t)	-0.05	-0.02	-0.06	-0.09	-0.09	-0.11
	(-0.11)	(-0.05)	(-0.14)	(-0.19)	(-0.20)	(-0.24)
Ln(SIZE	-0.45	-0.73	-0.83	-0.55	-0.56	-0.60
	(-0.98)	(-1.59)	(-1.77)	(-1.21)	(-1.22)	(-1.30)
Beta252	-0.21	-0.13	-0.10	-0.17	-0.18	-0.17
	(-0.20)	(-0.12)	(-0.09)	(-0.16)	(-0.17)	(-0.16)
Ln(B/M)	-0.11	-0.04	-0.02	-0.08	-0.08	-0.06
	(-0.33)	(-0.12)	(-0.06)	(-0.24)	(-0.23)	(-0.19)
OP	0.34	0.42	0.45	0.37	0.38	0.39
	(0.81)	(1.03)	(1.11)	(0.90)	(0.91)	(0.94)
TAG	-0.01	-0.10	-0.15	-0.05	-0.05	-0.06
	(-0.04)	(-0.32)	(-0.48)	(-0.15)	(-0.17)	(-0.21)
MOM25	0.68	0.64	0.65	0.68	0.68	0.67
	(0.99)	(0.93)	(0.93)	(0.97)	(0.97)	(0.96)
STR21	-2.09**	-2.12**	-1.96**	-2.06**	-2.03**	-2.03**
	(-3.31)	(-3.35)	(-3.10)	(-3.25)	(-3.21)	(-3.20)
IVOL21	-2.76**	-3.18**	-3.18**	-2.89**	-2.89**	-2.97**
	(-3.67)	(-4.20)	(-4.14)	(-3.82)	(-3.84)	(-3.93)
MAX21	2.73**	2.81**	2.77**	2.75**	2.76**	2.77**
	(4.11)	(4.21)	(4.12)	(4.14)	(4.15)	(4.16)
TO3	-0.31	-0.42	-0.42	-0.34	-0.33	-0.33
	(-0.64)	(-0.86)	(-0.85)	(-0.70)	(-0.68)	(-0.66)

Table 2.5. Weekly Rebalanced Portfolio Analysis Based on Aggregate Retail Orders

This table reports the average returns or alphas (in bps) of weekly rebalanced portfolios sorted by their aggregate retail orders in previous week. Panel A and B report the results based on aggregate selling orders (Rsel) and aggregate buying orders (Rbuy). Excess return is the average returns in excess of the risk-free rate. Alpha is the intercept from the regression of weekly excess returns on risk factors. The factor models include: the CAPM, the Fama-French three-factor model, and a four-factor model that includes Fama-French three factors and Carhart momentum factor. Long-Short return or Alpha is the return or alpha of a zero-cost portfolio that longs the high retail measure decile portfolio and shorts the low retail measure decile portfolio (i.e., H – L). The t-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ** and * indicate statistical significance at the 1% and 5% levels, respectively, using two-tailed tests. The sample period is from January 2010 to December 2016, covering total 365 trading weeks. To save space, I only report the results for the 1st, 2nd, 5th, 6th, 9th, and 10th decile portfolios.

Panel A: weekly decile portfolio based on aggregate retail selling orders (Rsel)

	Equal-weighted returns				Value-weighted returns			
	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha
1 (Low)	39.25	10.95	13.87	14.34	36.06	8.67	10.05	10.18
	(3.43)	(2.83)	(6.23)	(6.19)	(3.53)	(3.02)	(3.96)	(3.97)
2	31.55	2.72	5.15	5.55	30.97	4.08	4.90	4.86
	(2.78)	(0.82)	(2.54)	(2.70)	(3.00)	(1.54)	(2.04)	(2.00)
5	32.42	2.99	5.55	5.98	26.89	1.47	1.57	1.61
	(2.96)	(0.94)	(3.43)	(3.67)	(3.08)	(0.95)	(1.01)	(1.03)
6	32.38	2.45	5.32	5.73	26.26	1.82	1.54	1.54
	(2.77)	(0.73)	(3.10)	(3.36)	(3.07)	(1.08)	(0.96)	(0.97)
9	17.05	-13.52	-9.33	-7.94	32.23	0.13	0.74	0.12
	(1.22)	(-2.10)	(-1.98)	(-1.66)	(2.30)	(0.01)	(0.08)	(0.01)
10	6.18	-19.08	-15.86	-15.07	-8.79	-37.82	-34.97	-34.41
	(0.43)	(-2.16)	(-2.10)	(-2.03)	(-0.56)	(-3.82)	(-3.82)	(-3.81)
H – L	-33.07**	-30.03**	-29.73**	-29.41**	-44.85**	-46.50**	-45.02**	-44.59**
(t-stat)	(-4.19)	(-3.73)	(-3.79)	(-3.73)	(-4.81)	(-4.88)	(-4.87)	(-4.83)

Panel B: weekly decile portfolio based on aggregate retail buying orders (Rbuy)

	Equal-weighted returns				Value-weighted returns			
	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha
1 (Low)	27.21	0.41	3.51	3.89	29.01	2.85	4.39	4.55
	(2.58)	(0.11)	(1.95)	(2.07)	(2.92)	(0.95)	(1.80)	(1.85)
2	27.81	-1.31	1.24	1.58	29.35	2.54	3.30	3.18
	(2.50)	(-0.41)	(0.86)	(1.07)	(2.99)	(1.17)	(1.68)	(1.57)
5	31.77	2.16	4.72	5.19	27.02	1.18	1.23	1.31
	(2.82)	(0.65)	(2.68)	(2.90)	(2.97)	(0.61)	(0.63)	(0.67)
6	29.27	-0.34	2.36	2.71	23.75	-1.30	-1.59	-1.61
	(2.50)	(-0.09)	(1.22)	(1.45)	(2.88)	(-0.79)	(-0.99)	(-0.99)
9	26.04	-4.91	-0.79	0.72	26.93	-2.95	-2.53	-2.46
	(1.78)	(-0.67)	(-0.15)	(0.15)	(2.26)	(-0.40)	(-0.36)	(-0.35)
10	19.10	-7.10	-3.83	-2.84	7.28	-23.71	-21.28	-20.94
	(1.35)	(-0.87)	(-0.55)	(-0.41)	(0.50)	(-2.26)	(-2.11)	(-2.08)
H – L	-8.11	-7.51	-7.34	-6.73	-21.72*	-26.55*	-25.67*	-25.49*
(t-stat)	(-1.01)	(-0.94)	(-0.93)	(-0.85)	(-1.99)	(-2.42)	(-2.43)	(-2.39)

Table 2.6. Buy and Hold Returns in the Following 21 Trading Day

This table reports the average buy and hold returns or alphas (in percentage) of daily rebalanced portfolios sorted by their aggregate retail orders. Panel A and B report the results based on aggregate selling orders (Rsel) and aggregate buying orders (Rbuy). On each day, I establish the decile portfolios, and hold them for 21 trading days. Excess return is the average buy and hold return in excess of the 21-day cumulative risk-free rate. Alpha is the intercept from the regression of excess returns on risk factors. I compound the daily risk factors for 21 trading days. The factor models include: the CAPM, the Fama-French three-factor model, and a four-factor model that includes Fama-French three factors and Carhart momentum factor. Long-Short return or Alpha is the return or alpha of a zero-cost portfolio that longs the corresponding high retail measure decile portfolio and shorts the low retail measure decile portfolio (i.e., H – L). The t-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ** and * indicate statistical significance at the 1% and 5% levels, respectively, using two-tailed tests. To save space, I only report the results for the 1st, 2nd, 5th, 6th, 9th, and 10th decile portfolios.

Panel A: decile portfolio based on aggregate retail selling orders (Rsel)

	Equal-weighted returns				Value-weighted returns			
	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha
1 (Low)	1.55	0.26	0.45	0.47	1.44	0.22	0.32	0.33
	(4.26)	(1.94)	(10.74)	(11.12)	(4.39)	(2.62)	(5.64)	(5.71)
2	1.45	0.13	0.29	0.31	1.38	0.17	0.24	0.24
	(4.00)	(1.15)	(6.95)	(7.60)	(4.32)	(2.38)	(4.13)	(4.15)
5	1.29	-0.05	0.09	0.12	1.15	0.01	0.02	0.02
	(3.54)	(-0.48)	(2.40)	(3.62)	(3.89)	(0.34)	(0.50)	(0.69)
6	1.28	-0.08	0.07	0.10	1.12	0.04	0.02	0.02
	(3.45)	(-0.73)	(1.90)	(3.04)	(3.98)	(1.07)	(0.66)	(0.61)
9	0.79	-0.74	-0.52	-0.44	1.27	-0.01	0.02	0.03
	(1.75)	(-3.53)	(-3.95)	(-3.44)	(3.52)	(-0.04)	(0.12)	(0.17)
10 (High)	0.43	-0.92	-0.73	-0.65	0.66	-0.66	-0.56	-0.53
	(1.01)	(-3.76)	(-3.84)	(-3.48)	(1.76)	(-3.98)	(-4.08)	(-3.91)
H – L	-1.11**	-1.18**	-1.18**	-1.12**	-0.77**	-0.88**	-0.88**	-0.86**
(*t*-stat)	(-5.31)	(-5.73)	(-5.86)	(-5.54)	(-4.91)	(-5.28)	(-5.57)	(-5.40)

Panel B: decile portfolio based on aggregate retail buying orders (Rbuy)

	Equal-weighted returns				Value-weighted returns			
	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha	Excess Return	CAPM Alpha	3-factor Alpha	4-factor Alpha
1 (Low)	1.25	-0.00	0.19	0.20	1.25	0.08	0.18	0.19
	(3.54)	(-0.02)	(4.79)	(4.89)	(3.93)	(0.90)	(3.44)	(3.50)
2	1.29	-0.02	0.14	0.17	1.28	0.09	0.15	0.16
	(3.59)	(-0.16)	(3.71)	(4.70)	(4.10)	(1.45)	(3.06)	(3.21)
5	1.30	-0.05	0.09	0.12	1.17	0.02	0.03	0.04
	(3.54)	(-0.45)	(2.49)	(3.28)	(3.92)	(0.56)	(0.76)	(0.87)
6	1.30	-0.06	0.09	0.12	1.13	0.03	0.02	0.02
	(3.49)	(-0.54)	(2.43)	(3.63)	(3.94)	(0.85)	(0.55)	(0.59)
9	0.88	-0.65	-0.44	-0.35	1.22	-0.01	-0.00	0.03
	(1.94)	(-3.17)	(-3.28)	(-2.74)	(3.59)	(-0.07)	(-0.00)	(0.25)
10 (High)	0.71	-0.66	-0.47	-0.38	0.77	-0.56	-0.50	-0.47
	(1.63)	(-2.63)	(-2.39)	(-1.99)	(1.96)	(-2.83)	(-2.89)	(-2.73)
H – L	-0.54*	-0.66**	-0.66**	-0.58**	-0.48*	-0.64**	-0.68**	-0.66**
(*t*-stat)	(-2.47)	(-3.06)	(-3.18)	(-2.82)	(-2.48)	(-3.17)	(-3.76)	(-3.61)

Table 2.7. Portfolio Order Imbalance in Each 10×10 Double Sorted Portfolios

This table reports the portfolio's time-series average retail order imbalance (in %) in the 10×10 double sorted portfolios based on the retail buying measure and retail selling measure. Both the *ex ante* (as of day t) and *ex post* (as of day t+1) retail order imbalance measures are reported. Retail order imbalance (Rimb) is defined as the net retail buying orders scaled by the total trading volume on that day, i.e. (Rbuy-Rsel). In Panel A, I first sort stocks into decile groups based on the aggregate retail buying orders. To construct the 10×10 portfolios, I further sort the stocks within each decile group into finer decile groups based on the aggregate retail selling orders. In Panel B, I reverse the sorting order. The percentage numbers below the portfolio index number are the average retail buying order or the average retail selling orders in that decile portfolio. The t-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ** and * indicate statistical significance at the 1% and 5% levels, respectively, using two-tailed tests. To save space, I only report the results for the 1st, 2nd, 5th, 6th, 9th, and 10th decile portfolios.

Panel A: Portfolio order imbalance in the 10×10 double sorted portfolios first by aggregate retail buying orders (Rbuy)

	Rimb as of formation day t						Rimb as of day t+1					
	Low Buy	2	5	6	9	High Buy	Low Buy	2	5	6	9	High Buy
	0.47%	0.97%	2.22%	2.80%	7.46%	15.72%	2.32%	2.16%	2.88%	3.35%	7.06%	10.22%
1 (Low sell)	0.14	0.52	1.50	1.99	6.19	13.62	-0.18	-0.10	0.00	0.03	0.13	0.21
2	-0.07	0.19	0.96	1.34	4.64	11.61	-0.20	-0.09	-0.01	0.04	0.25	0.13
5	-0.66	-0.41	0.09	0.27	1.57	6.04	-0.27	-0.17	-0.09	-0.04	0.08	-0.09
6	-0.94	-0.64	-0.20	-0.05	0.77	4.64	-0.35	-0.21	-0.14	-0.08	0.06	-0.23
9	-3.28	-2.27	-1.97	-2.07	-3.96	-0.73	-0.63	-0.47	-0.41	-0.44	-0.62	-0.87
10 (High sell)	-9.48	-6.95	-6.86	-7.44	-11.81	-8.47	-1.17	-0.84	-0.86	-0.82	-1.10	-1.18
H – L	-9.62**	-7.47**	-8.36**	-9.43**	-18.00**	-22.09**	-0.99**	-0.75**	-0.86**	-0.85**	-1.23**	-1.40**
(*t*-stat)	(-48.72)	(-109.61)	(-117.36)	(-116.20)	(-83.74)	(-65.92)	(-17.53)	(-17.92)	(-19.09)	(-16.54)	(-14.65)	(-11.88)

Panel B: Portfolio order imbalance in the 10×10 double sorted portfolios first by aggregate retail selling orders (Rsel)

	Rimb as of formation day t						Rimb as of day t+1					
	Low Sell	2	5	6	9	High Sell	Low Sell	2	5	6	9	High Sell
	0.55%	1.09%	2.37%	2.96%	7.65%	16.19%	2.59%	2.36%	3.08%	3.54%	7.31%	10.78%
1 (Low buy)	-0.25	-0.70	-1.75	-2.27	-6.60	-14.24	-0.26	-0.31	-0.45	-0.50	-0.89	-1.33
2	-0.08	-0.39	-1.24	-1.64	-5.21	-12.57	-0.16	-0.22	-0.32	-0.37	-0.74	-1.19
5	0.43	0.18	-0.32	-0.50	-1.78	-7.17	-0.08	-0.09	-0.15	-0.17	-0.30	-0.81
6	0.67	0.39	-0.03	-0.15	-0.79	-5.43	-0.06	-0.06	-0.09	-0.10	-0.22	-0.72
9	2.70	1.83	1.65	1.80	3.68	0.21	0.05	0.05	0.07	0.16	0.10	-0.69
10 (High buy)	8.63	6.14	6.11	6.76	10.93	7.59	0.14	0.19	0.17	0.20	-0.27	-0.82
H – L	8.87**	6.83**	7.86**	9.03**	17.54**	21.83**	0.40**	0.50**	0.62**	0.70**	0.61**	0.51**
(*t*-stat)	(46.83)	(99.38)	(102.07)	(104.67)	(81.14)	(64.83)	(6.90)	(12.30)	(13.33)	(14.08)	(8.17)	(3.88)

Table 2.8. Portfolio Performance and Retail Trading Activities Around Intensive Individual Trading Activities

This table presents the portfolio performance around intensive individual trading measures. Panel A and B show the results based on retail selling order and retail buying orders. On each trading day, I rank stocks into decile groups according to the two retail trading measures. I then report average daily portfolio return and the retail trading actitivities during the (-20, 20) window around the formation day (i.e. Day 0). The reported returns and other variables are the time series equal-weighted average within each portfolio. Dvol is the total dollar amount trading volume. Rsel is the percentage of retail orders. Rbuy is the percentage of retail buying orders. Rimb is the difference between Rbuy and Rsel. The t-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ** and * indicate statistical significance at the 1% and 5% levels, respectively, using two-tailed tests.

Panel A: Decile portfolio based on aggregate retail selling orders (Rsel)

		Day (-20,-16)	Day (-15,-11)	Day (-10,-6)	Day (-5,-3)	Day (-2)	Day (-1)	Day (0)	Day (1)	Day (2)	Day (3,5)	Day (6,10)	Day (11,15)	Day (15,20)
Decile 1 Low sell	Dvol($M)	15.07**	15.13**	15.15**	15.14**	15.13**	15.17**	16.25**	15.59**	15.68**	15.53**	15.44**	15.35**	15.33**
	Rsel(%)	2.47**	2.45**	2.42**	2.36**	2.30**	2.22**	0.57**	2.23**	2.31**	2.37**	2.42**	2.45**	2.48**
	Rbuy (%)	2.30**	2.29**	2.26**	2.23**	2.19**	2.15**	2.11**	2.19**	2.23**	2.27**	2.30**	2.32**	2.34**
	Rimb (%)	-0.17**	-0.16**	-0.15**	-0.13**	-0.11**	-0.07**	1.49**	-0.04**	-0.08**	-0.10**	-0.12**	-0.13**	-0.14**
	Return (‱)	6.80**	6.21**	5.71*	3.78	1.48	1.84	7.78**	10.01**	8.35**	7.74**	7.62**	7.40**	7.44**
Decile 10 High Sell	Dvol($M)	4.86**	4.83**	4.80**	4.85**	4.87**	4.84**	4.53**	4.41**	4.44**	4.48**	4.51**	4.55**	4.59**
	Rsel(%)	9.82**	9.89**	9.99**	10.13**	10.28**	10.45**	15.88**	10.50**	10.33**	10.19**	10.05**	9.95**	9.87**
	Rbuy (%)	9.41**	9.46**	9.55**	9.66**	9.75**	9.88**	9.89**	9.80**	9.69**	9.59**	9.49**	9.42**	9.35**
	Rimb (%)	-0.40**	-0.42**	-0.44**	-0.47**	-0.52**	-0.56**	-5.73**	-0.69**	-0.64**	-0.59**	-0.55**	-0.52**	-0.52**
	Return (‱)	9.54**	10.86**	12.40**	14.43**	17.76**	10.83**	-11.73**	-0.41	2.99	3.34	3.39	3.32	3.61
H-L	Rimb (%)	-0.24**	-0.26**	-0.28**	-0.34**	-0.41**	-0.49**	-7.22**	-0.65**	-0.56**	-0.49**	-0.43**	-0.39**	-0.38**
	(*t*-stat)	(-9.01)	(-9.70)	(-10.85)	(-12.14)	(-13.99)	(-17.48)	(-85.50)	(-22.58)	(-19.22)	(-18.10)	(-16.34)	(-14.73)	(-14.31)
	Return(‱)	2.75	4.65**	6.69**	10.64**	16.27**	8.99**	-19.51**	-10.42**	-5.36**	-4.40**	-4.23**	-4.09**	-3.83**
	(*t*-stat)	(1.92)	(3.24)	(4.79)	(7.14)	(9.74)	(5.25)	(-10.82)	(-7.10)	(-3.73)	(-3.42)	(-3.18)	(-3.14)	(-2.99)

Panel B: Decile portfolio based on aggregate retail buying orders (Rbuy)

		Day (-20,-16)	Day (-15,-11)	Day (-10,-6)	Day (-5,-3)	Day (-2)	Day (-1)	Day (0)	Day (1)	Day (2)	Day (3,5)	Day (6,10)	Day (11,15)	Day (15,20)
Decile 1 Low Buy	Dvol($M)	14.71**	14.81**	14.85**	14.82**	14.80**	14.72**	15.49**	14.79**	14.78**	14.70**	14.62**	14.55**	14.52**
	Rsel(%)	2.58**	2.56**	2.54**	2.50**	2.46**	2.42**	2.36**	2.39**	2.44**	2.47**	2.50**	2.53**	2.54**
	Rbuy (%)	2.31**	2.28**	2.23**	2.17**	2.10**	2.02**	0.48**	2.04**	2.11**	2.17**	2.24**	2.27**	2.30**
	Rimb (%)	-0.27**	-0.28**	-0.30**	-0.33**	-0.36**	-0.40**	-1.82**	-0.35**	-0.32**	-0.30**	-0.27**	-0.25**	-0.24**
	Return (‱)	9.09**	9.42**	9.73**	10.13**	10.17**	9.06**	6.74**	6.26*	5.79*	6.01*	6.29**	6.36**	6.51**
Decile 10 High Buy	Dvol($M)	7.87**	7.84**	7.80**	7.85**	7.95**	7.89**	7.68**	7.53**	7.51**	7.48**	7.49**	7.57**	7.58**
	Rsel(%)	9.68**	9.74**	9.81**	9.91**	10.00**	10.10**	10.23**	10.27**	10.15**	10.05**	9.94**	9.86**	9.79**
	Rbuy (%)	9.46**	9.53**	9.64**	9.79**	9.95**	10.14**	15.42**	10.15**	9.96**	9.80**	9.66**	9.54**	9.46**
	Rimb (%)	-0.21**	-0.20**	-0.17**	-0.12**	-0.06*	0.03	4.93**	-0.12**	-0.19**	-0.24**	-0.29**	-0.31**	-0.33**
	Return (‱)	7.87**	8.58**	9.42**	9.19**	8.96**	14.16**	36.33**	11.29**	5.40	5.04	4.45	3.84	4.07
H-L	Rimb (%)	0.06*	0.08**	0.13**	0.22**	0.30**	0.43**	6.75**	0.23**	0.13**	0.05	-0.02	-0.06*	-0.09**
	(*t*-stat)	(2.16)	(3.12)	(4.97)	(7.90)	(10.36)	(14.25)	(73.73)	(7.56)	(4.35)	(1.95)	(-0.71)	(-2.21)	(-3.31)
	Return(‱)	-1.22	-0.85	-0.31	-0.94	-1.22	5.10*	29.59**	5.03**	-0.39	-0.97	-1.84	-2.52	-2.44
	(*t*-stat)	(-0.84)	(-0.59)	(-0.21)	(-0.61)	(-0.65)	(2.52)	(13.20)	(3.21)	(-0.26)	(-0.71)	(-1.37)	(-1.88)	(-1.86)

Table 2.9. Additional Analysis on Fear or Greed Sentiment

This table reports the regression results. All stock characteristics are standardized (N~(0,1)) to make the results comparable. Panel A and B show the results based on equal-weighted least square (EWLS) and value-weight least square (VWLS) regressions. The dependent variable is the stock's daily return (in bps). Besides the control variables in Table 2.4, I include additional interaction terms, including interactions with small size dummy (Size<Q20%), high idiosyncratic volatility dummy (IVol>Q80%), high VIX dummy (VIX_high), and low market return dummy (MKT_low). The t-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ** and * indicate statistical significance at the 1% and 5% levels, respectively, using two-tailed tests. The sample period is from January 2010 to December 2016, covering total 1762 trading days.

Panel A: EWLS Fama-MacBeth regression results

	(1)	(2)	(3)	(4)
Rsel	-2.84**	-2.54**	-1.39**	-0.81**
	(-6.64)	(-8.00)	(-7.10)	(-4.22)
Rsel×Size<Q20%	-1.13*			
	(-2.18)			
Rsel×IVol>Q80%		-2.20**		
		(-4.55)		
Rsel×VIX_high			-2.00**	
			(-6.68)	
Rsel×MKT_low				-2.58**
				(-10.63)
Control variables as in Table 2.4	yes	yes	yes	yes

Panel B: VWLS Fama-MacBeth regression results

	(1)	(2)	(3)	(4)
Rsel	-2.52**	-2.38**	-1.00**	-0.55
	(-4.13)	(-4.04)	(-2.93)	(-1.49)
Rsel×Size<Q20%	-2.40**			
	(-3.54)			
Rsel×IVol>Q80%		-2.44*		
		(-2.12)		
Rsel×VIX_high			-1.60**	
			(-3.66)	
Rsel×MKT_low				-2.05**
				(-4.94)
Control variables as in Table 2.4	yes	yes	yes	yes

Table 2.10. Investor Mood: Weekday Effects

This table reports the weekday performance of the weekly rebalanced portfolios sorted by their aggregate retail selling orders (Rsel). Panel A and B report the results based on equal weighted portfolio and value weighted portfolio. At the end of each week, I estimate the aggregate retail selling orders for each stock during the week, and establish the decile portfolios. I then report the portfolio performance in each weekday in the following week. To save space, I only report the results for the 1st, 2nd, 5th, 6th, 9th, and 10th decile portfolios. Long-Short excess return (H-L Exret) or 4-factor Alpha is the return or alpha of a zero-cost portfolio that longs the corresponding high retail measure decile portfolio and shorts the low retail measure decile portfolio (i.e., H – L). The t-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ** and * indicate statistical significance at the 1% and 5% levels, respectively, using two-tailed tests.

Panel A: EW portfolio results based on aggregate retail selling orders (Rsel)

	Low Rsel	2	5	6	9	High Rsel	H – L Exret	4-factor Alpha
Mon	0.73	-2.09	-2.16	-2.08	-2.58	-8.15	-8.88*	-9.07*
	(0.12)	(-0.33)	(-0.33)	(-0.31)	(-0.35)	(-1.08)	(-2.04)	(-2.05)
Tue	14.11	11.52	11.54	12.89	5.99	-0.66	-14.77**	-11.71**
	(2.36)	(1.84)	(1.79)	(1.92)	(0.91)	(-0.13)	(-4.26)	(-3.76)
Wed	7.76	6.19	7.46	6.86	3.58	4.69	-3.07	-0.42
	(1.62)	(1.30)	(1.55)	(1.39)	(0.71)	(1.23)	(-0.95)	(-0.15)
Thu	10.70	10.61	9.87	9.73	5.19	2.01	-8.70*	-6.29*
	(1.85)	(1.82)	(1.69)	(1.67)	(0.88)	(0.41)	(-2.36)	(-2.15)
Fri	7.40	6.28	6.95	6.26	8.35	14.31	6.91	6.64*
	(1.25)	(1.02)	(1.16)	(0.98)	(1.30)	(3.09)	(1.90)	(2.27)

Panel B: VW portfolio results based on aggregate retail selling orders (Rsel)

	Low Rsel	2	5	6	9	High Rsel	H – L Exret	4-factor Alpha
Mon	-3.14	-3.53	-2.52	-1.68	1.34	-9.13	-5.99	-5.98
	(-0.53)	(-0.60)	(-0.51)	(-0.35)	(0.17)	(-1.06)	(-1.07)	(-1.08)
Tue	12.55	10.58	9.87	10.83	16.69	0.09	-12.46**	-12.25**
	(2.23)	(1.80)	(1.81)	(2.05)	(2.06)	(0.01)	(-2.66)	(-2.74)
Wed	8.33	7.95	6.94	6.93	6.95	-1.76	-10.09**	-6.85*
	(1.87)	(1.75)	(1.62)	(1.59)	(1.13)	(-0.39)	(-3.11)	(-2.10)
Thu	11.59	10.97	9.58	7.16	9.50	-4.36	-15.96**	-14.76**
	(2.23)	(1.95)	(1.88)	(1.50)	(1.16)	(-0.67)	(-3.99)	(-4.34)
Fri	7.49	5.76	3.90	3.49	-1.00	7.29	-0.20	-2.73
	(1.34)	(1.00)	(0.72)	(0.67)	(-0.15)	(1.16)	(-0.05)	(-0.74)

Table 2.11. Retail Trading Activities Around S&P 500 Index Addition and Deletion

This table shows the time-series average of daily stock returns and retail trading activities around the S&P 500 index addition or deletion. Panel A and B present the results using addition and deletion respectively. EDay is defined as the effective addition/deletion day, which is Day 0. I focus on a window of 5 weeks (25 trading days) before and 4 week (20 trading days) after the effective day. Daily stock returns and the retail trading activates around the EDay are reported. Dvol is the average daily dollar trading volume. Rbuy is the daily aggregate retail buying orders. Rsel is the aggregate retail selling orders. Rimb is the difference between Rbuy and Rsel. Both the raw returns and market-adjusted returns are reported. The t-statistics shown in parentheses are computed based on bootstrap method with 10,000 iterations.

Panel A: Retail trading activities around the S&P500 index addition day

	(-25,-21)	(-20,-16)	(-15,-11)	(-10,-6)	(-5,-3)	-2	-1	EDay	1	2	(3,5)	(6,10)	(11,15)	(16,20)
Dvol($B)	0.15**	0.17**	0.18**	0.21**	0.30**	2.16**	0.32**	0.22**	0.22**	0.23**	0.21**	0.20**	0.16**	0.16**
Rsel(%)	2.73**	2.67**	2.54**	2.69**	2.46**	2.17**	2.03**	2.16**	2.35**	2.24**	2.39**	2.46**	2.41**	2.46**
Rbuy(%)	2.70**	2.78**	2.69**	2.71**	2.54**	2.24**	1.65**	2.24**	2.46**	2.37**	2.48**	2.43**	2.46**	2.48**
Rimb(%)	-0.01	0.11	0.16*	0.01	0.08	0.08	-0.30	0.08	0.11	0.13	0.09	-0.03	0.05	0.02
(*t*-stat)	(-0.10)	(1.21)	(2.45)	(0.12)	(0.92)	(0.81)	(-1.03)	(0.86)	(0.68)	(1.10)	(1.14)	(-0.46)	(0.90)	(0.32)
Return(‱)	14.61	16.69	14.15	17.03*	15.52	-7.51	-33.14*	3.69	-3.48	9.58	-29.21*	15.86	1.55	21.38**
(*t*-stat)	(1.69)	(1.52)	(1.37)	(1.99)	(0.96)	(-0.38)	(-1.97)	(0.13)	(-0.17)	(0.49)	(-2.38)	(1.70)	(0.19)	(2.85)
ExRet(‱)	6.66	21.06	8.37	12.37	12.07	-13.23	-53.95**	-1.19	1.46	-2.54	-23.63*	-0.05	-6.42	9.00
(*t*-stat)	(0.89)	(1.94)	(0.93)	(1.64)	(0.85)	(-0.80)	(-3.08)	(-0.05)	(0.09)	(-0.16)	(-2.35)	(-0.01)	(-0.90)	(1.37)

Panel B: Retail trading activities around the S&P500 index deletion day

	(-25,-21)	(-20,-16)	(-15,-11)	(-10,-6)	(-5,-3)	-2	-1	EDay	1	2	(3,5)	(6,10)	(11,15)	(16,20)
Dvol($B)	0.07**	0.07**	0.07**	0.08**	0.08**	0.12**	0.55**	0.11**	0.09**	0.09**	0.07**	0.07**	0.07**	0.08**
Rsel(%)	3.16**	3.02**	3.14**	3.22**	2.87**	2.73**	2.90**	2.29**	2.96**	3.17**	3.03**	3.24**	2.86**	2.98**
Rbuy(%)	3.06**	3.01**	3.34**	3.42**	2.90**	2.61**	2.78**	2.08**	3.02**	3.22**	3.08**	3.23**	3.19**	3.08**
Rimb(%)	-0.09	-0.01	0.20	0.15	0.03	-0.13	-0.12	-0.21	0.06	0.05	0.02	-0.01	0.32**	0.10
(*t*-stat)	(-0.75)	(-0.11)	(1.79)	(0.78)	(0.26)	(-0.49)	(-0.57)	(-1.57)	(0.38)	(0.21)	(0.12)	(-0.10)	(3.49)	(0.77)
Return(‱)	-11.99	-21.41	45.83*	2.18	-20.78	55.74	17.48	55.75	-9.73	82.80*	2.56	15.82	-16.51	37.68*
(*t*-stat)	(-0.74)	(-1.20)	(2.47)	(0.14)	(-1.16)	(1.42)	(0.59)	(1.66)	(-0.23)	(1.97)	(0.12)	(0.99)	(-0.98)	(2.30)
ExRet(‱)	-16.19	-14.95	25.42	-5.75	-21.12	41.20	18.19	44.09	-10.43	59.84	-6.01	5.49	-20.66	30.79
(*t*-stat)	(-1.05)	(-1.08)	(1.53)	(-0.43)	(-1.28)	(1.14)	(0.63)	(1.40)	(-0.26)	(1.68)	(-0.31)	(0.36)	(-1.36)	(1.84)

Table 2.12. Portfolio Return and Individual Trading Around the Earnings Announcement Day

This table presents the portfolio returns and individual trading activities around the earnings announcement days. In each quarter, I estimate the stocks' standard earning announcement surprise (SUEs) based on the actual EPS and the analyst census EPS in previous month, and then rank stocks into 1 of the 5 quintiles according to their SUE measure. Panel A to C report the results for the highest SUE group, middle SUE group, and lowest SUE group. I report the time-series average of the portfolio return and the associated retail trading measures. Both raw returns and the market adjusted returns are reported. The t-statistics shown in parentheses are computed based on standard errors with Newey-West corrections.

Panel A: Earnings announcement with extreme positive earnings surprise, ranked in Q5 (Top 20%)

	Day (-20,-16)	Day (-15,-11)	Day (-10,-6)	Day (-5,-3)	Day (-2)	Day (-1)	EADay (0)	Day (1)	Day (-2)	Day (3,5)	Day (6,10)	Day (11,15)	Day (15,20)
Dvol($M)	23.60**	23.45**	23.69**	23.94**	25.20**	31.93**	63.62**	37.59**	30.62**	27.91**	25.62**	24.80**	24.66**
Rsel(%)	5.44**	5.43**	5.43**	5.40**	5.30**	5.39**	5.54**	5.81**	5.69**	5.61**	5.55**	5.54**	5.57**
Rbuy(%)	5.22**	5.26**	5.27**	5.30**	5.47**	5.91**	5.43**	5.40**	5.26**	5.26**	5.31**	5.31**	5.28**
Rimb(%)	-0.22**	-0.17**	-0.16**	-0.10*	0.16**	0.52**	-0.11*	-0.41**	-0.43**	-0.35**	-0.24**	-0.23**	-0.29**
(*t*-stat)	(-7.45)	(-5.32)	(-10.29)	(-2.37)	(3.42)	(8.86)	(-2.32)	(-10.37)	(-10.04)	(-8.81)	(-10.20)	(-7.43)	(-11.12)
Return(‱	16.03**	13.69**	11.48*	7.35	7.93	35.22**	321.93**	19.35*	10.05	10.57	3.27	9.96*	9.53**
(*t*-stat)	(3.08)	(3.27)	(2.48)	(1.35)	(1.46)	(6.00)	(25.36)	(2.16)	(1.41)	(1.64)	(0.39)	(2.08)	(3.39)
ExRet(‱)	6.01	3.84	2.31	1.52	6.61*	34.56**	320.21**	15.21*	2.98	10.03*	3.69	5.77	4.75
(*t*-stat)	(1.94)	(1.12)	(0.54)	(0.30)	(2.09)	(6.07)	(25.17)	(2.71)	(0.74)	(2.67)	(1.06)	(1.86)	(1.94)

Panel B: Earnings announcement with moderate earnings surprise, ranked in Q3 (middle 20%)

	Day (-20,-16)	Day (-15,-11)	Day (-10,-6)	Day (-5,-3)	Day (-2)	Day (-1)	EADay (0)	Day (1)	Day (-2)	Day (3,5)	Day (6,10)	Day (11,15)	Day (15,20)
Dvol($M)	83.11**	84.03**	84.40**	88.19**	92.08**	116.70**	231.23**	131.00**	107.59**	95.75**	88.11**	87.44**	84.83**
Rsel(%)	2.60**	2.56**	2.55**	2.50**	2.49**	2.60**	2.85**	2.77**	2.69**	2.62**	2.61**	2.59**	2.58**
Rbuy(%)	2.50**	2.48**	2.47**	2.45**	2.53**	2.86**	2.75**	2.62**	2.51**	2.48**	2.48**	2.47**	2.47**
Rimb(%)	-0.11**	-0.09**	-0.08*	-0.06	0.04	0.26**	-0.10**	-0.15**	-0.18**	-0.14**	-0.13**	-0.12**	-0.11**
(*t*-stat)	(-5.46)	(-3.24)	(-2.75)	(-1.85)	(1.45)	(8.59)	(-5.74)	(-6.68)	(-5.36)	(-5.52)	(-5.07)	(-4.28)	(-5.29)
Return(‱	6.52	8.39**	8.80**	10.14**	3.89	11.22**	38.60**	7.99**	3.91	0.56	4.38	5.77	7.15**
(*t*-stat)	(1.44)	(2.79)	(5.45)	(3.55)	(1.39)	(3.25)	(5.36)	(2.90)	(1.06)	(0.14)	(0.88)	(1.24)	(3.56)
ExRet(‱)	0.39	-0.10	0.23	2.48	1.29	8.77**	32.47**	2.05	-3.54*	0.86	3.52**	2.34	2.94*
(*t*-stat)	(0.28)	(-0.09)	(0.18)	(1.14)	(0.54)	(5.35)	(5.92)	(1.26)	(-2.40)	(0.77)	(3.50)	(1.75)	(2.48)

Panel C: Earnings announcement with extreme negative earnings surprise, ranked in Q1 (Bottom 20%)

	Day (-20,-16)	Day (-15,-11)	Day (-10,-6)	Day (-5,-3)	Day (-2)	Day (-1)	EADay (0)	Day (1)	Day (-2)	Day (3,5)	Day (6,10)	Day (11,15)	Day (15,20)
Dvol($M)	17.35**	17.21**	17.65**	17.79**	18.61**	22.27**	46.24**	26.17**	21.90**	19.42**	18.40**	17.73**	17.06**
Rsel(%)	5.69**	5.67**	5.70**	5.63**	5.63**	5.67**	5.63**	6.00**	5.76**	5.65**	5.72**	5.69**	5.69**
Rbuy(%)	5.43**	5.52**	5.51**	5.53**	5.67**	5.94**	5.47**	5.57**	5.47**	5.43**	5.45**	5.47**	5.42**
Rimb(%)	-0.26**	-0.15**	-0.19**	-0.10**	0.04	0.27**	-0.17**	-0.42**	-0.29**	-0.22**	-0.26**	-0.21**	-0.27**
(*t*-stat)	(-10.88)	(-4.50)	(-5.77)	(-3.11)	(1.16)	(3.83)	(-3.85)	(-6.30)	(-5.79)	(-14.33)	(-7.84)	(-12.86)	(-8.62)
Return(‱	9.10	5.90	0.29	-4.53	-13.28	-23.27**	-352.90**	-43.06**	-21.01**	-1.35	2.75	9.74*	4.37
(*t*-stat)	(1.74)	(1.29)	(0.06)	(-1.09)	(-2.03)	(-2.86)	(-33.69)	(-3.28)	(-3.73)	(-0.19)	(0.35)	(2.73)	(1.39)
ExRet(‱)	-1.07	-4.34	-8.43	-9.21*	-12.88**	-23.65**	-353.52**	-47.22**	-27.03**	-1.41	2.79	4.97*	-1.03
(*t*-stat)	(-0.32)	(-1.18)	(-1.96)	(-2.48)	(-4.56)	(-5.39)	(-41.34)	(-5.22)	(-10.66)	(-0.33)	(0.76)	(2.09)	(-0.34)

Table 2.13. Retail Trading and Earnings Announcements: Fama-MacBeth Regression

This table presents the coefficients estimated from quarterly Fama-MacBeth regressions of cumulative returns (CAR) on different retail trading activities over different period. RET[0] is the stock return on the earnings announcement day. CAR[X,Y] refers to the cumulative return over the period t+X to t+Y. Rsel, Rbuy, and OIbjzz are the corresponding retail trading measures on the previous day of the earnings announcement day. Rsel5, Rbuy5, and OIbjzz5 are the corresponding average retail trading measures on the previous 5 days before the earnings announcement day. Rsel10, Rbuy10, and OIbjzz10 are the corresponding average retail trading measures on the previous 10 days before the earnings announcement day. I regress CAR on different retail trading measures separately, but report them together to save space. All the control variables in Table 4 are also included in the regression, but the coefficients are not reported. The t-statistics shown in parentheses are computed based on standard errors with Newey-West corrections. ** and * indicate statistical significance at the 1% and 5% levels, respectively, using two-tailed tests.

Panel A: Retail trading activities on day (-1)

	RET[0]	CAR[-1,1]	CAR[0,2]	CAR[2,5]	CAR[2,10]
Rsel	-0.30**	-0.38**	-0.43**	-0.13**	-0.24**
	(-10.87)	(-11.43)	(-14.93)	(-8.70)	(-4.92)
Rbuy	-0.18**	-0.27**	-0.32**	-0.10**	-0.14**
	(-7.10)	(-15.56)	(-13.37)	(-6.99)	(-3.92)
OIbjzz	0.15**	0.15**	0.16**	0.01	0.07*
	(6.77)	(5.59)	(6.84)	(0.76)	(2.40)
Control	Yes	Yes	Yes	Yes	Yes

Panel B: Average retail trading activities during day (-1) to day (-5)

	RET[0]	CAR[-1,1]	CAR[0,2]	CAR[2,5]	CAR[2,10]
Rsel5	-0.34**	-0.47**	-0.56**	-0.19**	-0.33**
	(-11.67)	(-14.47)	(-17.86)	(-6.41)	(-5.01)
Rbuy5	-0.29**	-0.41**	-0.48**	-0.15**	-0.26**
	(-10.29)	(-9.32)	(-14.24)	(-4.77)	(-4.13)
OIbjzz	0.07**	0.08*	0.11**	0.04**	0.08**
	(3.07)	(2.51)	(3.79)	(2.72)	(4.77)
Control	Yes	Yes	Yes	Yes	Yes

Panel C: Average retail trading activities during day (-1) to day (-10)

	RET[0]	CAR[-1,1]	CAR[0,2]	CAR[2,5]	CAR[2,10]
Rsel10	-0.38**	-0.52**	-0.62**	-0.19**	-0.33**
	(-14.13)	(-19.37)	(-25.43)	(-6.08)	(-4.39)
Rbuy10	-0.32**	-0.45**	-0.53**	-0.17**	-0.29**
	(-15.02)	(-13.39)	(-23.61)	(-5.18)	(-3.86)
OIbjzz1	0.09**	0.10**	0.13**	0.04**	0.08**
	(4.65)	(3.79)	(4.70)	(4.22)	(6.39)
Control	Yes	Yes	Yes	Yes	Yes

References

Akepanidtaworn, Klakow, Rick Di Mascio, Alex Imas, and Lawrence Schmidt, 2022, Selling fast and buying slow: Heuristics and trading performance of institutional investors. *Journal of Finance*, forthcoming.

Bakshi, Gurdip, and Nikunj Kapadia, 2003, Delta-hedged gains and the negative market volatility risk premium, *Review of Financial Studies* 16, 527-566.

Barber, Brad M., Xing Huang, Terrance Odean, and Christopher Schwarz, 2022, Attention induced trading and returns: Evidence from Robinhood users, *Journal of Finance*, forthcoming.

Barber, Brad M., Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean, 2009, Just how much do individual investors lose by trading? *Review of Financial Studies* 22, 609-632.

Barber, Brad M., and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773-806.

Barber, Brad M., Terrance Odean, and Ning Zhu, 2008, Do retail trades move markets? *Review of Financial Studies* 22, 151-186.

Barrot, Jean-Noel, Ron Kaniel, and David Sraer, 2016, Are retail traders compensated for providing liquidity? *Journal of Financial Economics* 120, 146-168.

Birru, Justin, 2018, Day of the week and the cross-section of returns. *Journal of financial economics* 130, 182-214.

Black, Fischer, 1986, Noise, *Journal of Finance* 41, 528-543.

Brennan, Michael J., Tarun Chordia, Avanidhar Subrahmanyam, and Qing Tong, 2012, Sell-order liquidity and the cross-section of expected stock returns. *Journal of Financial Economics* 105, 523-541.

Brennan, Michael J., Sahn-Wook Huh, and Avanidhar Subrahmanyam, 2016, Asymmetric effects of informed trading on the cost of equity capital. *Management Science* 62, 2460-2480.

Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang, 2008, Which shorts are informed? *Journal of Finance* 63, 491-527.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang, and Xinran Zhang, 2022, Tracking retail investor activity, *Journal of Finance*, forthcoming.

Boyer, Brian, Todd Mitton, and Keith Vorkink, 2010, Expected idiosyncratic skewness, *Review of Financial Studies* 23, 169-202.

Boyer, Brian, and Keith Vorkink, 2014, Stock options as lotteries, *Journal of Finance* 69, 1485-1527.

Campbell, John Y., Tarun Ramadorai, and Allie Schwartz, 2009, Caught on tape: Institutional trading, stock returns, and earnings announcements, *Journal of Financial Economics* 92, 66-91.

Cao, Jie, Tarun Chordia, and Xintong Zhan, 2020, The Calendar Effects of the Idiosyncratic-Volatility Puzzle: A Tale of Two Days? *Management Science* accepted.

Cao, Jie, and Bing Han, 2013, Cross section of option returns and idiosyncratic stock volatility, *Journal of Financial Economics* 108, 231-249.

Cao, Jie, Bing Han, Qing Tong, and Xintong Zhan, 2022, Option return predictability, *Review of Financial Studies*, forthcoming.

Cao, Jie, Aurelio Vasquez, Xiao Xiao, and Xintong Zhan, 2021, Volatility uncertainty and the cross-section of option returns, Working paper.

Chan, Yue-Cheong, Jennifer Conrad, Gang Hu, and Sunil Wahal, 2018, Trading by Crossing, ABFER working paper.

Choy, Siu-Kai, 2015, Retail clientele and option returns, *Journal of Banking and Finance* 51, 26-42.

Chung, Juliet, 2021, Melvin Capital Lost 53% in January, Hurt by GameStop and Other Bets, Retrieved from https://www.wsj.com/articles/melvin-capital-lost-53-in-january-hurt-by-gamestop-and-other-bets-11612103117

Coval, Joshua D., and Tyler Shumway, 2001, Expected option returns, *Journal of Finance* 56, 983-1009.

Cready, William, Abdullah Kumas, and Musa Subasi, 2014, Are trade Size-Based inferences about traders reliable? evidence from institutional Earnings-Related trading, *Journal of Accounting Research* 52, 877-909.

Eaton, Gregory W., T. Clifton Green, Brian Roseman, and Yanbin Wu, 2021, Zero-commission individual investors, high frequency traders, and stock market quality, Working Paper.

Eisele, Alexander, Tamara Nefedova, Gianpaolo Parise, and Kim Peijnenburg, 2020, Trading out of sight: An analysis of cross-trading in mutual fund families, *Journal of Financial Economics* 135, 359-378.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests. *Journal of political economy* 81, 607-636.

Figlewski, Stephen, 1989, Options arbitrage in imperfect markets, *Journal of Finance* 44, 1289-1311.

Garleanu, Nicolae, Lasse Heje Pedersen, and Allen M. Poteshman, 2009, Demand-based option pricing, *Review of Financial Studies* 22, 4259-4299.

Goyal, Amit, and Alessio Saretto, 2009, Cross-section of option returns and volatility, *Journal of Financial Economics* 94, 310-326.

Grinblatt, Mark, and Matti Keloharju, 2001, What makes investors trade? *Journal of Finance* 56, 589-616.

Grinblatt, Mark, and Matti Keloharju, 2009, Sensation seeking, overconfidence, and trading activity, *Journal of Finance* 64, 549-578.

Grossman, Sanford J., and Merton H. Miller, 1988, Liquidity and market structure. *Journal of Finance* 43, 617-633.

Han, Bing, and Alok Kumar, 2013, Speculative retail trading and asset prices, *Journal of Financial and Quantitative Analysis* 48, 377-404.

Hu, Gang, Koren M. Jo, Yi Alex Wang, and Jing Xie, 2018, Institutional trading and Abel Noser data, *Journal of Corporate Finance* 52, 143-167.

Huang, Darien, Christian Schlag, Ivan Shaliastovich, and Julian Thimme, 2019, Volatility-of-volatility risk, *Journal of Financial and Quantitative Analysis* 54, 2423-2452.

Kaniel, Ron, Shuming Liu, Gideon Saar, and Sheridan Titman, 2012, Individual investor trading and return patterns around earnings announcements. *Journal of Finance* 67, 639-680.

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2008, Individual investor trading and stock returns, *Journal of Finance* 63, 273-310.

Kelley, Eric K., and Paul C. Tetlock, 2013, How wise are crowds? Insights from retail orders and stock returns, *Journal of Finance* 68, 1229-1265.

Kelley, Eric K., and Paul C. Tetlock, 2017, Retail short selling and stock prices. *Review of Financial Studies* 30, 801-834.

Kumar, Alok, 2009, Who gambles in the stock market? *Journal of Finance* 64, 1889-1933.

Kumar, Alok, and Charles MC Lee, 2006, Retail investor sentiment and return comovements. *Journal of Finance* 61, 2451-2486.

Lee, Charles MC, and Balkrishna Radhakrishna, 2000, Inferring investor behavior: Evidence from TORQ data, *Journal of Financial Markets* 3, 83-111.

Lee, Charles MC, Andrei Shleifer, and Richard H. Thaler, 1991, Investor sentiment and the closed-end fund puzzle. *Journal of Finance* 46, 75-109.

Li, Yun, 2021, Melvin Capital, hedge fund targeted by Reddit board, closes out of GameStop short position, Retrieved from https://www.cnbc.com/2021/01/27/hedge-fund-targeted-by-reddit-board-melvin-capital-closed-out-of-gamestop-short-position-tuesday.html

McCabe, Caitlin, 2021, It Isn’t Just AMC. Retail Traders Increase Pull on the Stock Market, Retrieved from https://www.wsj.com/articles/it-isnt-just-amc-retail-traders-increase-pull-on-the-stock-market-11624008602

Puckett, Andy, and Xuemin Yan, 2011, The interim trading skills of institutional investors, *Journal of Finance* 66, 601-633.

Ruan, Xinfeng, 2020, Volatility-of-volatility and the cross-section of option returns, *Journal of Financial Markets* 48, 100492.

Ronalds-Hannon Eliza, and Jennifer Surane, 2021, Wall Street Reins In Hedge Funds’ Short Bets on Meme Stocks, Retrieved from https://www.bloomberg.com/news/articles/2021-06-04/wall-street-banks-rein-in-hedge-funds-short-bets-on-meme-stocks

Scuffham, Matt, and Saqib Iqbal, Ahmed, 2021, Some on Wall Street try options trade to bet against AMC without getting burned, Retrieved from https://www.reuters.com/business/exclusive-some-wall-street-try-options-trade-bet-against-amc-without-getting-2021-06-07

Shleifer, Andrei, and Lawrence H. Summers, 1990, The noise trader approach to finance. *Journal of Economic perspectives* 4, 19-33.

Stambaugh, Robert F., and Yu Yuan, 2017, Mispricing factors, *Review of Financial Studies* 30, 1270-1315.

Tetlock, Paul C, 2007, Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62, 1139-1168.

Vasquez, Aurelio, 2017, Equity volatility term structures and the cross section of option returns, *Journal of Financial and Quantitative Analysis* 52, 2727-2754.

Welch, Ivo, 2020, The Wisdom of the Robinhood crowd, NBER Working Paper.

Winck, Ben, 2020, Wall Street is being shaken to its core by a legion of Gen Z day traders. From a casual hobbyist to a 20-year-old running a 14,000-person platform, meet the new generation of retail investors, Retrieved from https://www.businessinsider.com/gen-z-millennial-day-retail-trader-interviews-wall-street-options-2020-7

Wright, William F., and Gordon H. Bower, 1992, Mood effects on subjective probability assessment. *Organizational behavior and human decision processes* 52, 276-291.