# Heterogeneity in Retail Investors: Evidence from Comprehensive Account-Level Trading and Holdings Data

Charles M. Jones, Donghui Shi, Xiaoyan Zhang and Xinran Zhang*

February 01, 2021

## Abstract

Retail investors are heterogeneous, with vast differences in wealth and skills. Using comprehensive proprietary account-level data on trading and holdings from the Shanghai Stock Exchange from 2016 to 2019, we separate tens of millions of retail investors into five groups by their account sizes, and we examine their trading behavior and return performance in Chinese equities. Retail investors with account sizes less than three million CNY follow momentum trading strategies, and the prices of stocks they buy experience negative returns next day, while the ones they sell experience positive returns. In contrast, retail investors with larger account balances follow contrarian strategies, and they buy and sell stocks in directions consistent with future price movements. In addition, retail investors with smaller account sizes fail to process public news and they incur losses from trading, while retail investors with larger account sizes incorporate public news in their trading and experience trading gains.

Keywords: Retail investors, Chinese stock market, return predictability, gain and loss
JEL code: G12, G14, G15

* Jones is at Columbia University, Shi is at Fudan University, Fanhai International School of Finance, and Zhang and Zhang are at Tsinghua University, PBC School of Finance. The authors thank the Shanghai Stock Exchange for providing data and assistance. Corresponding author: Xinran Zhang, PBC School of Finance, Tsinghua University, 43 Chengfu Road, Beijing, 100083, zhangxr.15@pbcsf.tsinghua.edu.cn.

**Heterogeneity in Retail Investors: Evidence from Comprehensive Account-Level Trading and Holdings Data**


## Abstract

Retail investors are heterogeneous, with vast differences in wealth, skills and demographics. Using comprehensive proprietary account-level data on trading and holdings from the Shanghai Stock Exchange from 2016 to 2019, we separate tens of millions of retail investors into five groups by their account sizes as well as other demographic variables, and we examine their trading behavior and return performance in Chinese equities. Retail investors with account sizes less than three million CNY follow momentum trading strategies, and the prices of stocks they buy experience negative returns next day, while the ones they sell experience positive returns. In contrast, retail investors with larger account balances follow contrarian strategies, and they buy and sell stocks in directions consistent with future price movements. In addition, retail investors with smaller account sizes fail to process public news and they incur losses from trading, while retail investors with larger account sizes incorporate public news in their trading and experience trading gains. These patterns are stronger for young male retail investors.



Keywords: Retail investors, Chinese stock market, return predictability, gain and loss
JEL code: G12, G14, G15

Many observers suspect that retail investors, in comparison with institutional investors, are generally uninformed and, if anything, make systematic mistakes when selecting equity investments. For instance, Barber and Odean (2000, 2002, 2008) document some behavioral biases exhibited by retail investors, and they show that retail investors are overconfident and tend to trade too much. However, some more recent evidence, including Kaniel, Saar, and Titman (2008), Kaniel, Liu, Saar, and Titman (2012), Kelley and Tetlock (2013), Fong, Gallagher, and Lee (2014), Barrot, Kaniel, and Sraer (2016), and Boehmer, Jones, Zhang and Zhang (2020) suggests that retail investors can correctly predict future stock returns and trade accordingly, which indicates that retail investors might be informed about future stock price movements.

How is it possible to reconcile the conflicting results from previous studies? It is possible that retail investors are heterogeneous, and the above-mentioned empirical results are dominated by subgroups of retail investors. Due to data limitations, few previous studies directly examine the heterogeneity of retail investors. We are fortunate to have access to a comprehensive, account-level database from the Shanghai Stock Exchange with all trading and holdings over the period 2016 to 2019. In this study, we focus our attention on the heterogeneity of retail investors and investigate whether different subgroups of retail investors behave drastically differently.

Retail trading in China has two prominent features. First, according to the annual report of the Shanghai Stock Exchange, 85% of daily trading volume on the exchange comes from retail investors, while only 15% is from institutional investors. That is, retail investors are active and

dominant in trading in the Chinese stock market. Second, to comply with regulatory requirements, all investors in the market are categorized by general identity and account size. For instance, retail investors are separated from institutional investors, and they are further separated into five groups based on their account balances: less than 100,000 CNY, between 100,000 and 500,000 CNY, between 500,000 and 3,000,000 CNY, between 3,000,000 and 10,000,000 CNY, and greater than 10,000,000 CNY. The separation of retail investors into different groups greatly helps us to understand the heterogeneity in trading patterns among retail investors.

Over the three-year horizon in our sample, we collect daily trading information for over 53 Million retail accounts, with 58.7%, 28.6%, 10.9%, 1.4% and 0.4% of the accounts coming from the [<100k CNY], (100k, 500k CNY], (500k, 3m CNY], (3m, 10m CNY] and (>10m CNY) groups. We further separate these accounts by gender and age. The largest group of retail investors are young and middle-aged males, with account sizes mostly below 500k CNY. Since wealth is the key information used by the exchanges themselves to categorize different retail investors, our empirical results also focus on the account size subgroups.

With this rich cross section of retail investors, we first examine their trading behavior, in terms of what determines their trading. In general, momentum strategies (buying winners and selling losers) demand liquidity, while contrarian strategies (buying losers and selling winners) supply liquidity. From previous studies on U.S. stocks, retail investors, in aggregate, tend to be momentum over daily horizons and contrarian over weekly horizons. From Chinese trade level data, we find

drastically different trading behaviors from different retail groups. Retail investors with account size less than three million CNY are momentum traders over a daily horizon, potentially demanding market liquidity, while retail investors with account sizes larger than three million CNY and institutions are contrarian traders, potentially supplying liquidity.

Given these trading patterns of retail investors, we next investigate whether the buy and sell activities from retail investors can predict future price movements. If the market is perfectly efficient, and if all investors have the same information, stock price movements would resemble random walks, and trading would not predict future returns. On the other hand, if the market is not perfectly efficient, or if some investors have information advantages over other investors, we should observe return predictability. We use daily retail order imbalances from each retail group to predict future stock returns at horizons ranging from one day to 60 days. We find that retail investors with account sizes less than ten million CNY buy and sell in the opposite directions of future price movements. The prices of stocks they buy experience negative returns the next day, while the ones they sell experience positive returns. In contrast, retail investors with larger account balances buy and sell stocks in directions consistent with future price movements. These patterns are robust for different subsets of stocks, sorted by size, turnover and value, and the patterns persist for at least one month.

One possibility for the differences in the predictive power of trading for returns is that investors differ in their skills of acquiring and processing news. For instance, Barber and Odean (2008) find

that retail investors pay more attention to non-essential information rather than fundamental information. Using pubic news data from Financial News Database of Chinese Listed Companies (CFND), we examine whether retail investors incorporate the previous day's positive or negative news into their next day trading. We find that trades for accounts with balances below 10 million CNY deviate are less correlated with future stock price movements after a public news release, regardless of the direction of the news; while the predictive power of retail investors with balances higher than 10 million CNY strengthens after a public news release. This finding indicates that less wealthy retail investors fail to process public news, while wealthier retail investors can successfully incorporate valuable news into their trading. Kelley and Tetlock (2013) find that in the U.S., aggressive market orders from retail investors can predict earnings news, while limit orders cannot. Their interpretation is that retail market orders are more informed than retail limit orders. In our setting, we find that less wealthy retail investors predict earnings surprises with the wrong and sign, while more wealthy retail investors and institutions predict earnings surprises with the correct sign. This finding, again, supports the hypothesis that wealthy retail investors and institutions might be more informed about earnings news than less wealthy retail investors.

Do retail investors' trading help them to collect gains or losses in the stock market? Previous literature on U.S. stocks shows that some retail investors are informed about future price movements, and might potentially make money, while most studies, such as Barber and Odean (2000, 2002), show that U.S. retail investors have many behavioral biases and end up losing money

by trading, especially those who trade fast and frequently. Given the large base of retail investors in the Chinese stock market, data on gains and losses would give regulators a clear picture about how to guide and educate retail investors to trade rationally. With granular intraday data for each account and for each trade, we compute the approximate gain or loss for each investor group by following each trade from each account. Combining all retail investors altogether, Chinese retail investors lose 497 billion CNY (or 71 billion USD) each year in our sample, which is consistent with Barber et al (2008) findings in Taiwan. More interestingly, we find that retail investors with higher balances lose less in aggregate than the retail investors with lower balances. Again, the heterogeneity of retail investors significantly affects their overall gains and losses in the stock market.

Finally, we investigate whether the differences in demographics affect retail investors' trading and their ability to anticipate the cross-section of future stock returns. We find that male investors across all ages are momentum investors, while female investors are mostly contrarian investors. Male investors across all ages negatively predict returns and the younger ones lose the most, while some female retail investors, with age less than 35 and above 55, can positively predict future returns. These findings are generally in line with Barber and Odean (2001).

Our study is closely related to the retail investor literature. Researchers have carefully examined retail investors in the U.S. and other developed countries. As mentioned earlier, using detailed data from a discount broker in the U.S., Barber and Odean (2000, 2001, 2002) examine

the trading and investment behavior of retail investors in the U.S. and document many behavioral biases. For return predictability, Kaniel, Saar and Titman (2008), Barber, Odean, Zhu (2009), Kelley and Tetlock (2013), and Boehmer, Jones, Zhang and Zhang (2020) use different datasets from the U.S., and all four papers find that aggregate retail trading can positively predict the cross-section of future returns. However, each study provides different explanations for the return predictability. Using data from France, Barrot, Kaniel and Sraer (2016) provide further evidence that retail investors provide liquidity, especially during market downturns. Our results on retail investor heterogeneity with different account sizes are quite different from these studies.

Recently, a few papers study the rapidly growing Chinese stock market. Liu, Stambaugh and Yuan (2019) construct refined size and value factors in China, and show that Chinese stock returns, to a large extent, can be captured by a small number of factors. An, Bian, Lou and Shi (2020) study the wealth redistribution role of financial bubbles and crashes over July 2014 and December 2015, and document a net transfer of 250 billion CNY from the poor to the ultra-wealthy over this period. Compared to these studies using Chinese data, our focus is on return predictability in the cross section, and we investigate sample periods when the market is relatively stable.

The rest of the paper is organized as follow. Section I introduces the data. Section II provides the empirical results. We conduct robustness check in Section III. Section IV concludes.

## I. Data

## A. Data from the Exchange

The Shanghai Stock Exchange (SHSE) generously offered us access to their historical trade level data, which includes all investors' trading and holding data of all A-share stocks from the exchange's main board. The original data is available from January 2014. The Chinese stock market experienced a crisis during 2014-2015, and so we exclude these two years. Our final sample period is January 2016 to June 2019. The Science and Technology Innovation Board (or STAR Market) was launched on the SHSE on July 22, 2019, and thus is not included in our study.

To have a general understanding of the SHSE and its listed stocks, here we briefly discuss a few key measures of the exchange. There are two stock exchanges in China, the SHSE and the Shenzhen Stock Exchange (hereafter SZSE). In June 2019, there are 1,471 A-share stocks listed on the SHSE, with a total market capitalization of $ 4.6 trillion. In comparison, 2,157 A-share stocks are listed on the SZSE, with a total market capitalization of $ 3 trillion. That is to say, SHSE accounts for 60% of the total market capitalization in China and is a good representation of the overall Chinese stock market.

The proprietary SHSE data that we use covers the entire investor population, with roughly 53 million accounts. The data are collected for compliance purposes and are kept on the exchange's internal servers. Our data have three unique advantages compared to previous studies. First, investors are categorized into many groups, which greatly facilitate our study of investor

heterogeneity, and we discuss the details below. Second, the investor ids are unique, and each intraday order can be traced back to a specific investor. Third, the SHSE also provides each account's holding at the firm level at the close of each trading day. The latter two features greatly facilitate our later computation of account level gains and losses.

All investors on the exchange are first grouped into three major categories: retail (RT), institutional (INST), and corporations (CORP). Retail investors are further stratified into five groups based on their account sizes, which is the average portfolio value (including equity holdings in both SHSE and SZSE-listed firms, plus cash) over the previous twelve months. There are five subgroups: below 100,000 CNY (RT1), 100,000-500,000 CNY (RT2), 500,000 - 3 million CNY (RT3), 3 million - 10 million CNY (RT4), and above 10 million CNY (RT5). Starting from March 31, 2010, margin buying and short selling are allowed in Chinese stock exchanges for subsets of stocks. Investor margin buys, short sells and collateral trades are also identified in our trading data. We include these leveraged trades in our main results, and provide additional analysis excluding leveraged trading in our robustness checks.

Table 1 Panel A presents the aggregate account summary statistics. During our sample period, the number of active accounts for retail investors, institutions and corporations are 53.4 million, 40,000 and 47,000, respectively. Within the retail investor category, there are 31.4 million accounts with balances lower than 100,000 CNY, 15.3 million accounts with balances between 100,000 and 500,000 CNY, 5.8 million accounts with balances between 500,000 CNY and three million CNY,

0.7 million accounts with balances between three and the million CNY, and 0.2 million accounts with balances above ten million CNY. Clearly, most of the retail investors have accounts less than 500,000 CNY. The overall trading volume on the SHSE averages 201 billion CNY per day, and retail investors, institutions and corporations account for 81%, 17% and 2% respectively of the total trading volume. Within the retail investor sector, the trading volume for the five subgroups (in increasing order of account size) are 5%, 17%, 27%, 13% and 19% of the total trading volume, which is more evenly distributed than the numbers of accounts. For stock holdings, out of the overall 25 trillion CNY tradable market capitalization, retail investors' holdings account for 22%, institutions 17% and corporations 62%. Within the retail investor sector, the account values for the five groups (in increasing order of account size) are 1%, 4%, 6%, 3% and 7% of the total tradable market cap. That is to say, in the Chinese stock market, retail investors dominate in terms of trading, while corporations dominate in holdings.

**B. Data on Stock Returns and Characteristics**

We obtain data on stock returns, volumes, and accounting information from Wind Information Inc. (WIND), the largest financial data provider in China. We compute daily stock returns using close prices, which are dividend and split adjusted. In the U.S., there is a large amount of high frequency trading, including establishing and closing positions on the same day or even within the same second. In China, the "*T+1* trading rule" requires that if stocks are bought on day *T*, they

cannot be sold on the same day. The reverse trade has to be executed on day *T+1* or later. That is, there is essentially a minimum holding period of one day.

Previous literature using the U.S. data shows that microstructure frictions can generate noise in daily return measures. For instance, Blume and Stambaugh (1986) show that daily returns computed from the end-of-day closing prices can have an upward bias due to bid-ask bounce. To assess the potential magnitude of the bias, they measure the bias as $\left(\frac{P_A - P_B}{P_A + P_B}\right)^2$, where $P_A$ and $P_B$ are closing ask and bid prices. Blume and Stambaugh (1986) find that the average bias for small stocks is 0.051%, and for large stocks, the bias is 0.001%. Therefore, they recommend using closing bid-ask average prices to compute daily returns. We compute this bias measure using the closing bid and ask prices for all A-share stocks listed on the SHSE. The average bias measure is below 0.0002% across all stocks, which is negligible compared to the bias computed in Blume and Stambaugh (1983). Therefore, we compute daily returns using daily close prices without the Blume and Stambaugh (1983) adjustments.

To better understand the cross-section of Chinese stocks, we present the distribution of four prominent firm-level characteristics. The first is market capitalization, *SIZE*, computed as the previous month's closing price times total A shares outstanding. The second is the earnings to price ratio, *EP*, which is computed as the ratio of the most recently reported quarterly net profit excluding non-recurrent gains/losses over the product of last month-end's close price and total shares. We choose the earnings to price ratio for a potential "value effect", instead of the book to

market ratio, because Liu, Stambaugh and Yuan (2019) find that the *EP* ratio can better explain stock returns in the Chinese stock market. The third variable is monthly turnover, calculated as monthly share trading volume divided by tradable shares outstanding at the end of the previous month. We use the tradable shares outstanding as the denominator of turnover because we are most interested in shares of the stock that are tradable in practice. Finally, we report the distribution of daily stock returns.

From Panel B of Table 1, the mean market capitalization is 20.1 billion CNY or 3 billion USD, about half of the cross-sectional mean in the U.S stock market during the same period, which is 6.9 billion USD. The average monthly turnover in China is 48.32%, which is much bigger than the monthly turnover of 22% in the U.S. during the same period. The average earnings to price ratio is 0.0075 in China, while the average quarterly earnings to price ratio is 0.0272 in U.S stock market. This difference is possibly driven by the high valuation ratios in China. Finally, the average daily stock return is -0.01%, while the average daily stock return is 0.04% in the U.S stock market over the same sample period.

### C. Summary Statistics on Retail Trading

We merge the data from SHSE and WIND by stock ticker. To get rid of outliers in the data, we mainly use the filters in Liu, Stambaugh and Yuan (2019). Following their study, we exclude stocks with less than 15 days of the trading records during the most recent month. On each day, we have

an average of around 1,200 firms included in the sample. Liu, Stambaugh and Yuan (2019) also eliminate stocks that have become public within the past six months, stocks with less than 120 days of trading records during the past 12 months, and the smallest 30% of total firms listed in SHSE and SZSE. We do not exclude these stocks because retail investors trade actively in small stocks and during the IPO period. We present the results with all Liu et al (2019) filters in our robustness checks, and our findings are robust to these stricter filters.

In Table I Panel C, we present summary statistics on different types of investor activity across stocks. We report the time-series average of the cross-sectional mean for trading volume in shares and holdings in shares across investor types each day. Our sample covers over 1.1 million stock-day observations. For instance, on each day, the average stock-day buy volume in shares (*BuyVol*) for the five types of retail investors is 0.8 million, 2.7 million, 4.2 million, 2 million and 2.8 million; institutions buy 2.3 million and corporations buy 0.3 million shares per stock day. The sell volume in shares (*SellVol*) for different types of investors are similar. Consistent with what we observe in Panel A, most of the trading volume is from retail investors. As for the holding shares (*HoldShr*), for an average stock, corporations hold 36% of tradable shares on average, institutions hold 12% and the five groups of retail investors hold 52%. To have a rough idea about how long positions last, we assume that shareholders within same investor group have the same holding period. Then we compute the holding period for stock $i$, type $G$ investors as,

$$HP_{i,G} = 1/TO_{i,G}, \quad (1)$$

where $TO_{i,G}$ is the turnover (shares traded/shares held by this type of investor) of stock *i* for type *G* investors. For example, if 1% of the shares trade each day, then it takes 100 days for the entire stock of tradable shares held by this group to turn over, and the average holding period would be 100 days. In the last row of Panel C, the average holding period for the five groups of retail investors ranges from 35 days to 50 days, reflecting their active trading. In comparison, the market overall monthly turnover in the U.S. over the same period is 22%, indicating a holding period of 1/0.22 = 4.5 months, which is about 90 days. Institutional holding periods in our sample are much longer at 109 trading days. Corporations barely trade in our sample period, and their estimated holding period is 6,319 trading days. These dramatic differences in holding periods suggest that different types of Chinese investors might have very different trading patterns and trading goals.

To observe whether there are time trends in investors' trading, we plot the time series of cross sectional means of various investors' trading activity in Figure I. Panel A shows the average stock level share volume for different investor groups. During the sample period, the volumes are relatively stable across various investor groups, except for February 2019 to April 2019, during which the market return is positive and high, and more trading is observed. Panel B shows each group's trading volume as a percentage of total trading volume. The RT3 group has the most trading of shares, accounting for about 30% of the total trading volume. RT2 accounts for 20%.

RT4 and RT5 gradually decrease their trading percentages, while institutional trading increases, especially after 2018. Panel C displays the shares held percentage by each group. Held shares are quite stable and corporations hold the most, at 40% overall.

We measure different types of investors' directional trades by computing scaled order imbalance measures for each type of investor G trading each stock *i* on each day *t*,

$$oib(i,t,G) = \frac{buyvol(i,t,G) - sellvol(i,t,G)}{buyvol(i,t,G) + sellvol(i,t,G)}. \quad (2)$$

Table I Panel D reports the summary statistics of these scaled order imbalances. The mean order imbalance for the five retail groups are -0.021, -0.011, -0.006, 0.002 and 0.019, and the mean of order imbalances for institutions and corporations are -0.011 and -0.004. The small magnitude of these average order imbalance measures indicates that most buys and sells within each investor group cancel out each other. Sells are slightly more prevalent than buys in smaller retail investors and in institutions. Moreover, the standard deviations of order imbalances are larger for large retail investors and institutions compared to small and medium retail investors, which means there is more cross-stock variation in large retail investor and institutional trading activity. One-day autocorrelations of our scaled order imbalances are 0.243, 0.259, 0.216, 0.059 and 0.102 for the five retail groups, suggesting that small and medium retail order imbalances are more persistent than large retail imbalances. For contemporaneous cross-sectional correlations of the seven groups' order imbalances, small and medium retail investors' order imbalance (*OibRT1*, *OibRT2* and *OibRT3*) are highly correlated with a correlation over 0.6. Larger retail investors order imbalance

(*OibRT4*) is still positively correlated with *OibRT1-OibRT3*, but with a much lower correlation, around 0.2. The largest retail investors' order imbalance (*OibRT5*) is negatively correlated with all four other groups, with correlation around -0.15. Institutional order imbalances are negatively correlated with all five retail groups, with correlations ranging from -0.38 to -0.188. [1]

In Figure II, we plot the time-series of the cross-sectional mean, median and 25th and 75th percentiles of different types of investors' order imbalance over the three and half year sample period. Across all seven order imbalance measures, the means and medians are all close to zero. The order imbalances of small and medium retail investors (RT1, RT2 and RT3) are less volatile than those of large retail investors, institutions and corporations, which means most of the trades within small and medium retail investors offset within groups. There are no obvious time trends or structural breaks in the time series observations.

## II. Empirical Results

We start by investigating the properties of the order imbalance measures in Section II.A. In Section II.B, we examine whether past order imbalance measures can predict future stock returns using Fama-MacBeth regressions and long-short portfolios. In Section II.C, we study how retail trades are related to published news. We measure gains and losses for different types of investors

[1] In Appendix Table I, we compute correlations between our order imbalance data with past returns. Between returns and order imbalance measures, *OibRT1-OibRT4* are negatively with contemporaneous stock returns, and *OibRT5* and *OibInst* are positively with contemporaneous stock returns. We also include past daily, weekly and monthly stock returns (*Ret(-1)*, *Ret(-6,-2)* and *Ret(-27,-7*)). They are most negative.

in Section II.D. Finally, in Section II.E, we focus on the trading behavior of retail investors of different ages and genders.

### A. *What Explains Different Investor Type Order Imbalances?*

We start our empirical investigation by examining what drives the trading of different types of investors. Specifically, we examine how investors' order flow is related to past order flow and past returns. To allow maximal time-series flexibility and focus on cross-sectional patterns, we adopt the Fama and MacBeth (1973) two-stage estimation. At the first stage, for each day *d*, we estimate the following predictive regression in the cross section:

$$oib(i,d) = b0(d) + b1(d)'ret(i,d-1) + b2(d)'oib(i,d-1) \quad (3)$$
$$+b3(d)'controls(i,d-1) + u1(i,d).$$

The dependent variable, $Oib(i,d)$, is the scaled order imbalance measure for stock *i* on day *d*. We use past returns, past order imbalances, and past firm level controls (log market cap, earnings to price ratio, and turnover) to explain the cross sectional pattern of order imbalances. The first-stage estimation generates a daily time-series of coefficients, $\{b0(d), b1(d)', b2(d), b3(d)'\}$. At the second stage, we conduct statistical inference using the time-series average of the daily coefficients. The time-series coefficients can be serially correlated, so we compute standard errors using Newey-West (1987) with 5 lags.[2]

[2] The optimal lag number is chosen using BIC.

We estimate equation (3) for each group of investors separately and present the results in Table II. For past returns, we include the previous day return, previous week return and previous month return. The order imbalances of RT1, RT2, and RT3 load positively and significantly on the previous day return, indicating that these investors buy more if the previous day return is positive, and sell more if the previous day return is negative, which corresponds to a momentum trading strategy. For larger retail investors in RT4 and RT5, institutions, and corporations, order imbalances load negatively on returns from the previous day, indicating that they are contrarian investors. If we include past week returns and past month returns, then all retail investors load negatively on past returns, meaning they buy losers and sell winners, which is contrarian.

Our finding that some types of retail investors are contrarian and some are momentum traders is quite interesting and different from some previous studies. For instance, contrarian patterns have been documented in Kaniel, Saar and Titman (2008) using monthly horizons in the U.S., and Barrot, Kaniel and Sraer (2016) using daily and weekly horizons in France. Using U.S. data, Kelley and Tetlock (2013) and Boehmer et al (2020) both find that retail trades follow momentum over daily horizons, but are contrarian at weekly horizons. In our setting, we find the trading patterns from investors with smaller account sizes are similar to those in Kelley and Tetlock (2013) and Boehmer et al (2020), while the investors with the largest account sizes behave similarly to the patterns documented in Kaniel et al (2008) and Barrot et al (2016).

Previous literature finds that order imbalances are persistent. In Table II, the coefficients on previous day order imbalances are all positive and significant, which is consistent with other researchers. For the firm characteristics, order imbalances for smaller investors such as RT1 and RT2 are positively related to size, while larger investors load negatively on the size variable, indicating that smaller investors tend to buy larger stocks, while larger investors tend to buy smaller stocks. The coefficients on EP are quite mixed and do not show a clear pattern. For turnover, order imbalances from most groups load positively on turnover, indicating more buying on average for more actively traded stocks.

Our results in Table II reveal two important drivers affecting daily order imbalances. The first is past return; we show that small and medium retail investors are short-term momentum traders, while large retail investors and institutions are contrarian. The second is past order imbalance, which indicates that order imbalance measures are persistent and are more persistent for small and medium retail investors and less persistent for large retail investors. Our findings clearly indicate large trading heterogeneity in the retail investor space.

## *B. Predicting Future Stock Returns with Different Type of Investors' Order Imbalances*

### *B.1. Methodology and Overall Predictive Power*

Can retail investors' activity provide useful information for future Chinese stock returns? In this section, we examine the predictive power of our order imbalance measures using Fama-MacBeth regressions as follows:

$$ret(i,d) = c0(d) + c1(d)oib(i,d-1) + c2(d)'controls(i,d-1) + u2(i,d), \quad (4)$$

where we use the order imbalance measures from the previous day, $oib(i,d-1)$, and various control variables to predict the next day's stock return for firm *i* on day *d*, $ret(i,d)$. If a particular type of order imbalance can predict future returns in the right direction, we expect the associated coefficient *c1* to be significantly positive. If coefficient *c1* turns out to be significantly negative, then that group of investors is consistently wrong and might be making systematic trading mistakes. For control variables, we include past returns, log market cap, earnings-to-price ratio, and turnover, all from the previous month. The statistical inference is conducted on the time series of the regression coefficients, and we compute standard errors using Newey-West with 5 lags to adjust for autocorrelation, which is the optimal lag number using a Bayesian Information Criterion.

We report the estimation results in Table III, and again we find distinct heterogeneity in the cross section of retail investors. For the first retail investor group, we use order imbalances from RT1 (account balance < 100,000 CNY) to predict the next day return. The coefficient on *oib(-1)* is -0.0093, with a t-statistic of -24.98. The negative and strongly significant coefficient shows that, if retail investors RT1 buy more than they sell on a given day, the next day return on that stock is significantly negative, which indicates that RT1 as a group trades in the wrong direction vs. future

stock price movements. In terms of magnitude, we report at the bottom of the table that the inter-quartile range for the *oibRT1* measure is 0.2222 per day. Multiplying the interquartile difference by the regression coefficient of -0.0093 generates a daily return difference of -20.62 basis points (more than 50% annualized!) when we move from the 25th to the 75th percentile of the *OibRT1* variable.

For retail investors in groups RT2 to RT4, the patterns are qualitatively quite similar. The coefficients on the past day order imbalance are -0.0091, -0.0065, and -0.0009. All coefficients are negative and statistically significant. When we turn to the inter-quartile differences at the bottom of the table, the daily return differences are -16.68, -10.89, and - 2.47 basis points, respectively. That is to say, the first four groups of investors all trade in the wrong direction vs. future price movements. Interestingly, when we move from the smaller account size to the larger ones, the negative return differences become smaller, indicating that larger retail investors trade less incorrectly. Indeed, for the largest retail investor group, RT5, the coefficient on past day order imbalance actually becomes 0.0012, which is positive and highly significant with a t-statistic of 12.26. The interquartile return difference now becomes 5.23 basis points per day (over 12% per year). It seems that the largest retail investors' trading predicts the cross-section of future stock price movements in the correct direction.

As a comparison, the coefficient on the previous day order imbalance is 0.0016 for institutions, with a t-statistic of 20.34. That is to say, institutional trades also predict future stock price

movements in the right direction, and the interquartile return difference is 10.46 basis points, about twice the level of the RT5 coefficient. This is consistent with many previous studies that institutional investors are more informed, and their trades contain more information than retail investor trades. As noted earlier, corporations rarely trade, and thus the coefficient on the past day order imbalance of corporate trades is nearly zero and statistically insignificant.

For the control variables, we observe negative coefficients on the previous day, previous week return and previous month return, which indicates return reversals. Size is mostly insignificant. The earnings-to-price variable is positively and significantly related to future return, while turnover is negatively and significantly related to future return, which are consistent with findings in Liu, Stambaugh and Yuan (2019). This result also confirms that the predictability we find is not simply a manifestation of a size, value or turnover effect.

### *B.2. Subgroups in the Cross Section*

Our sample includes on average around 1,200 firms each day. Is the predictive power of different types of investor order imbalances restricted to a particular type of firm? We investigate this question by analyzing various firm subgroups in this section. We first sort all firms into three groups based on a firm or stock characteristic observed at the end of previous month. Then, we estimate equation (4) within each characteristic group. That is, we allow all coefficients in

equation (4) to be different within each group, which allows substantial flexibility in the possible predictive relationship across these different groups.

We first sort all of our sample stocks into three terciles based on market capitalization. We then estimate Fama-MacBeth regressions for each tercile and for each group of investors. The results are reported in Panel A of Table IV. In the top of the panel, we report coefficients on $oib(i, d-1)$ for different types of investors. Suppose we take RT1 as an example. When we move from the smallest tercile of SHSE firms by market cap to the largest tercile, the coefficient on $oib(i, d-1)$ decreases from -0.0114 to -0.0075, and the *t*-statistic decreases from -24.34 to -19.58. Economically, the interquartile difference in average daily returns for this investor group is -24.20 basis points for the smallest market-cap firms, and -18.25 basis points for the largest firms. That is to say, the negative predictive power of order imbalance from smaller investors is the strongest for the smallest 1/3 of firms, and it is smaller in magnitude for the largest 1/3 of firms, while still strongly negative and statistically significant. We observe similar patterns for RT2, RT3, and RT4.

When we move to the largest retail investor group, RT5, the pattern changes. When we move from the smallest tercile of firms by market cap to the largest tercile, the coefficient on $oib(i, d-1)$ decreases from 0.0019 to 0.0000, and the associated *t*-statistic decreases from 15.06 to 0.17. That is, the positive predictive power of order imbalance from larger investors is also the strongest for the smallest tercile of firms, and it diminishes to virtually zero for the largest tercile

of SHSE firms. Overall, the predictive power of retail order imbalance for future returns, no matter whether it is positive or negative, is the strongest for small firms, and weakest for large firms.

For the remaining two groups of investors, the trading from institutions always predicts future returns positively and significantly, and the positive coefficient becomes larger and more significant for the larger firms. This pattern suggests that institutional investors might be more informed about larger firms than smaller firms. In contrast, corporations do not have any predictive power for future returns for any firm size group.

In Panel B of Table IV, we sort all firms into three groups based on previous month turnover. We find that the negative predictive power from smaller retail investor order imbalances, and the positive predictive power from large retail investors and institutional order imbalances, to be the strongest for firms with higher turnover. This result is quite intriguing. Existing literature normally interprets turnover as a proxy for liquidity, and high liquidity normally implies high market efficiency. In contrast, our empirical results show that small retail investors make bigger mistakes for firms with high turnover, while the larger retail investors and institutional investors appear to be more informed about the firms with high turnover, despite their apparently high liquidity and efficiency. It is possible that large investors' trading contributes to the liquidity and efficiency, and it is also possible that turnover might also measure something other than liquidity and information efficiency, such as irrational herding behavior by some groups of investors.

Finally, we sort all firms based on the previous-month closing share price, and we present results of the Fama-MacBeth regressions in Panel C of Table IV. We find that small retail investors' negative predictability and institutions' positive predictability are the strongest in medium share price groups, though the economic differences between price terciles are smaller than the differences between size and turnover groups. This result is also quite interesting. Existing literature often considers low priced stocks to be lottery stocks preferred by retail investors. Since low price stocks may have greater limitations on arbitrage, they may be more mispriced. In contrast, our empirical results show that there is not much difference between share price terciles, particularly for small retail investors and institutions. It is possible that retail investors may not consider share price to be important when trading stocks in the Chinese stock market, and it is also possible that low price might not be a good lottery stock proxy because some high quality stocks like bank stocks are also low price (less than 5 CNY per share) in the Chinese stock market.

Overall, we find that the predictive power of various investors' order imbalances is significant and carries the same signs in different stock groups, which shows that the predictive power is not exclusive to subgroups of stocks. More importantly, we observe clear cross-sectional patterns for the predictive power. The negative predictive power of the small account retail order imbalance is much stronger for small cap, high turnover, and low earnings to price firms. The positive predictive power of the large account retail order imbalances comes mostly from small cap, high turnover

and low earnings to price firms. The positive predictive power of institutions exists among all subgroups, but is stronger in large cap, high turnover and high earnings to price firms.

### *B.2. Longer Horizons*

In the previous section, we show that different types of investor order imbalances can predict the next day's returns, negatively and significantly for small and medium retail investors, and positively and significantly for large retail investors and institutions. It is natural to now ask whether the predictive power is persistent. If the predictive power quickly vanishes or reverses, what we observe might be driven by short-term noise; if the predictive power continues at longer horizons, it is more likely the return predictability is linked to firm fundamentals or persistent biases. To answer this question, we extend equation (4) to longer horizons as follows:

$$ret(i,w) = d0(w) + d1(w)oib(i,d-1) + d2(w)'controls(i,d-1) + u3(i,w). \quad (5)$$

That is, we use previous day order imbalance, $oib(i,d-1)$, to predict k-week ahead returns. To precisely observe the decay rate of the predictive power of order imbalance, the return to be predicted is a weekly return over a specific five-day period, rather than a cumulative return over *n* weeks. If order imbalances have only short-lived predictive power for future returns that then vanishes, we should observe the coefficient $d1$ decrease to zero quickly. Alternatively, if the specified retail order imbalance has longer predictive power, the coefficient $d1$ should remain

statistically significant for a longer period. In our empirical estimation, $k$ ranges from two to 12 weeks.

We report the results in Table V. When we extend the window from one to 12 weeks, the coefficient on *oibRT1* monotonically decreases in magnitude, from -0.0226 at one week to -0.0005 at 12 weeks. The coefficients are statistically significant up to nine weeks ahead. The negative predictive power of *oibRT2*, *oibRT3*, and *oibRT4* becomes statistically insignificant by seven weeks. The *oibRT5* and *oibInst* positive predictive power decays more quickly, lasting for about four weeks. The *oibCorp* does not have significant predictive power for future returns in general, and the coefficients are economically close to zero regardless of the horizon.

Overall, the various order imbalance measures' predictive pattern for future returns varies substantially, and the most persistent ones are order imbalance measures from the smallest and largest retail investors, as well as order imbalances from institutional investors. The persistence of cross-sectional predictability over at least four weeks indicates that either the predictive power is rooted in information related to fundamentals or from persistent noise trading or behavioral biases.

### *B.3. Long-Short Portfolios*

In this section, we form quintile portfolios based on the previous day's average order imbalance from a particular investor group. That is, we aim to track a group of investors by buying the stocks in the highest order imbalance quintile for that investor group and shorting the stocks in the

lowest order imbalance quintile. We then hold this portfolio for the next day. Portfolio returns are value-weighted using the previous month-end market cap.

A portfolio approach is a natural way to measure cross-sectional differences and has several advantages over regression methods. First, it is easy to interpret, because it replicates the gross returns of a potential trading strategy, assuming (counterfactually) that one could observe all these trading flow data in real time and trade without frictions. Second, compared to a regression approach, the aggregation into portfolios can reduce the impact of outliers. Finally, portfolios are able to capture certain nonlinearities that might characterize the relationship between trading activity of one group of investors and future returns. If one type of investors on average can select the right stocks to buy and sell, then firms with higher or positive order imbalances would outperform firms with lower or negative order imbalances. Notice that this exercise merely uses order imbalance measures as a signal to predict future stock returns. It ignores trade frictions, shorting impediments, and transaction costs, and it provides no direct information on whether a group of investors can make profits from their own trades.

From Table I, typical holding periods for retail and institutional investors are mostly between 30 and 50 days, and we compute the long-short strategy holding period returns for horizons of one to 60 days. To be conservative, we compute the standard errors of the portfolio returns using the Newey-West (1987) approach with lags equal to twice the horizon in days. Since the return holding horizon extends to 60 days, compensation for exposures to systematic risks can be substantial, and

thus here we use the Liu, Stambaugh and Yuan (2019) three-factor model to compute risk-adjusted returns (alphas).

Table VI reports long-short portfolio alphas. The one-day long-short portfolio alpha, using the previous day order imbalance from RT1, is -0.0028, with a t-statistic of -10.28. The weekly (5-day) alpha for the long-short portfolio is -0.0043, with a *t*-statistic of -5.45. When we increase the holding horizon to 60 days, the average alpha becomes -1.31%, with a *t*-statistic of -4.21. The general pattern is that cumulative holding-period alphas and returns continue to grow in magnitude, at a decreasing rate, for up to 60 days. We observe no evidence of a reversal in return predictability. Similar patterns exist for RT2 and RT3. For larger retail investors in RT4, the one-day alpha is negative at -0.0008, with a significant t-statistic of -3.32, but it becomes insignificant for horizons longer than one day. For the largest retail investors in RT5, the one-day alpha is positive but insignificant, while the five-day and ten-day alphas are both positive and significant, indicating that the signals in RT5 trading are fairly weak but accumulate over time.

In contrast, the long-short strategy following institutional order imbalance generates positive and significant returns for the next one to 60 days, ranging between 36 basis points and 1.38%. We also observe no evidence of return reversals for these long-short portfolios. This shows that the information in institutional trading is persistent, consistent with findings from previous literature

that institutional trading is more informed. Finally, for the corporations, the long-short strategy mostly generates negative returns, and half of them are statistically significant for up to 60 days.[34]

To make sure that the statistical significance in return differences is not driven by particular sample periods, we present time-series plots of the return differences between quintiles one and five in Figure III, where the portfolios are sorted on order imbalances of different types of investors and the holding period is one day. Over our three and half-year sample period, we observe both time-variation in the return differences and positive and negative spikes. However, most data points are negative for small and medium retail investors, and the negative returns are not driven by particular sample sub-periods. Most data points are positive for institutional investors, and the positive returns are not driven by particular sample sub-periods. Unreported plots of alphas show the same pattern.

### *C. Heterogeneity in Information Processing Ability*

[3] We report the raw returns and subgroup results in Appendix Table II. The patterns for the raw returns are qualitatively similar to those in Table VI. The subgroup results for a holding period of 20 days, the general patterns are quite similar to those in Table IV.

[4] We conduct a rough calculation that includes transaction costs. Shanghai stock exchange market quality (2018) state that a reasonable estimate of one-way transaction cost on value-weighted Shanghai stock exchange stocks is about 11 basis points from 2015 to 2017. To be conservative, we assume that for each rebalance, we change 100% of the positions. That is, each rebalance we incur a 2*11 bps = 22 bps rebalance cost. After the transaction cost, the one-day alpha turns negative and significant for RT5, while institution's one-day alpha remains positive and significant.

The patterns in predictability that we observe may reflect differences in the information content of trades. Do different groups of retail investors have the same access to information? Do they pay attention to relevant information? Can they correctly process the information? Do their trades contain information related to the firm's fundamentals? These are all interesting and important questions. In this section, we first focus on publicly available news to investigate how different groups of investors trade on news days, and how this affects their predictive power for the cross-section of next day returns. Next, we turn to the most influential fundamental information, earnings news, and examine whether different investors can predict earnings announcement returns. By these investigations, we hope to understand whether different groups of retail investors pay attention to these important types of news, and more generally whether they can process news and incorporate it into their trading.

We obtain news data from the Financial News Database of Chinese Listed Companies (CFND), which includes news on all A-share stocks from both newspaper and internet sources. A brief description of the data is provided in Appendix III. Using their proprietary algorithm, CFND provides a sentiment score for each piece of news. If the news is positive, the score is +1; if negative, -1; and 0, if neutral. For each stock each day, we compute the net sentiment score by averaging the sentiment scores from all news on that firm that day. This average net sentiment score can be interpreted as the net positivity of news the firm receives during that day. Altogether, there are 369,217 firm-days (35% of the total sample) with public news.

Since news can contain negative or positive sentiment, the sentiment of a news event might affect different investors trading in different ways. Therefore, we define a positive news dummy and a negative news dummy. The positive news dummy *Pos(i, d-1)* is equal to one if the net news sentiment score is positive for this firm on day d-1, and zero otherwise; the negative news dummy *Neg(i, d-1)* is equal to one if the net news sentiment score is negative for this firm on day d-1, and zero otherwise.

We investigate how news affects retail trades with the following Fama MacBeth regression:

$$ret(i,d) = e0(d) + e1(d)Pos(i,d-1) + e2(d)Neg(i,d-1)$$
$$+[e3(d) + e4(d)Pos(i,d-1) + e5(d)Neg(i,d-1)]oib(i,d-1)$$
$$+e6'(d)controls(i,d-1) + u4(i,d). \quad (6)$$

If there is no news, the predictive power of retail trades is measured by coefficient e3; when there is only positive news, the predictive power is measured by e3+e4; when there is only negative news, the predictive power is measured by e3+e5. If coefficients e4 and e5 are significantly different from zero, that group of retail investors anticipates future stock returns differently on these news days.

The results are presented in Table VII Panel A. Here we take the smallest retail investors, RT1, as an example. First of all, the coefficient on the positive news dummy is 0.0005, with a significant t-statistic of 6.76, and the coefficient on the negative news dummy is -0.0004, with a significant t-

statistic of -3.53. That is, both positive news and negative news contain relevant information for future returns in the correct direction. Now the question is: can the retail investors incorporate the information in their trading?

For the coefficients on order imbalance, the coefficient c3 is -0.0077, which is quite similar to the -0.0095 in Table III, indicating that on average the trades from RT1 negatively predict future returns. When there is positive news, the coefficient on the interaction of positive news and the order imbalance is -0.0046, with a significant t-statistic of -9.35, implying that the wrong prediction of RT1 for future stock returns becomes significantly larger in magnitude when there is positive news on the firm. Similarly, the coefficient on the interaction of negative news and the order imbalance is -0.0053, which is again highly significant, implying that the wrong prediction of retail trades for future stock returns becomes significantly larger in magnitude when there is negative news on the firm. That is, when there is public news on the firm, no matter whether it is positive or negative, the RT1 investors do not seem to be able to correctly incorporate this news in their trading.

The pattern is similar for RT2, RT3 and RT4. But it is significantly different for RT5 and institutional investors. For instance, the coefficients c3, c4 and c5 for RT5 are 0.0008, 0.0014 and 0.0015, respectively. That is to say, the RT5 group can on average predict future returns in the correct direction, and their prediction becomes much stronger on both positive and negative news days, indicating that RT5 as a group is able to incorporate valuable information in public news into

their trading, thereby enhancing their ability to anticipate the cross-section of future stock returns. The pattern for institutional investors is quite similar to that of RT5.

Can the retail trades directly predict earnings news? To answer this question, we conduct one more test, following Kelley and Tetlock (2013), to examine whether trades from different retail groups contain information about future earnings surprises. For each earnings announcement day for firm i on day d, we compute the cumulative abnormal return over day d-1, d, and d+1, and relate this to the previous order imbalance using the following specification:

$$CAR(i, d-1, d+1) \tag{7}$$
$$= f0(d) + f1(d)oib(i, d-2) + f2(d)'controls(i, d-2) + u5(i, d).$$

Here we use order imbalance measures from day d-2 to avoid overlapping with the CAR calculation. The above specification is estimated using one Fama-MacBeth regression per calendar quarter. If a particular type of retail order imbalance can predict earnings surprises in the right direction, we expect the associated coefficient *f1* to be significantly positive. If coefficient *f1* turns out to be significantly negative, then that group of investors incorrectly predicts future earnings news. For control variables, we include past returns, log market cap, earnings-to-price ratio, and turnover, all from the previous month. Statistical inference is conducted on the time series of the

regression coefficients, and we compute standard errors using Newey-West with 4 lags to adjust for auto-correlations.[5]

We present the estimation results in Panel B of Table VII. For retail investors RT1-RT3, the coefficients f1 are -0.0247, -0.0226, and -0.0156, respectively, all with highly significant t-statistics. These negative and significant coefficients indicate that as a group, these investors incorrectly predict earnings surprises. In contrast, the coefficients f1 for RT5 and institutional investors are 0.0019, and 0.0039, both statistically significant, implying that these investors are able to correctly predict future earnings surprises. In between these two extreme cases, the coefficient f1 for RT4 is close to zero and insignificant.

Our results in this subsection reveal interesting heterogeneous patterns of how retail investors process public information and predict fundamental news. On one hand, the smaller retail investors seem to lack skills to process public news and predict fundamental news, while the largest retail investors and institutions are able to correctly process public news and predict fundamental news.

### *D. Gains and Losses of Different Type of Investors*

Given the large heterogeneity in trading behavior across different retail investor groups in China, do they actually gain or lose from their trading activities? To answer this question, we compute the gains and losses for each investor group. Previous literature on U.S. stocks shows that

[5] The optimal lag number is chosen using BIC.

some retail investors are informed about future price movements, and might potentially make money, while most studies show that retail investors overall have many behavioral biases and end up losing money by trading, especially those who trade frequently. Given the large base of retail investors in the Chinese stock market, a measure of gains and losses would provide a clearer picture for regulators, companies, exchanges, and the investors themselves.

For calculations of gains and losses from trading, previous studies mostly follow the approach in Barber, Lee, Liu and Odean (2008), who estimate the total gains and losses for investors in the Taiwan stock market from 1995 to 1999. Like us, they have a dataset that details every transaction in every Taiwanese stock, and they are able to map each of these trades to an account ID. Since they do not have investors' holding data, Barber et al (2008) calculate the gain/loss from trading each day for each investor group by assuming that investors have fixed holding periods of sixty days. The Barber et al (2008) measure of gains and losses is innovative, but it has two limitations. First, the assumption of a fixed holding period for all investors is at odds with the observed trading data. Second, their gain/loss calculation does not include positions that are unchanged, so their gain/loss measures only reflect the gain/loss from new positions, which makes the measured overall gain/loss for these investors imprecise. With access to both holdings and trading data, we are able to improve the Barber et al (2008) approach and compute precise measures for the gain/loss of each investor group.

On each trading day, we first compute the total gain/loss for each account by measuring the gain/loss on the holdings and the day's trading, both marked to market. Then we sum up the gains/losses for all investors within the same investor group. We also decompose the gain/loss into three parts: market timing, stock selection and transaction costs. Exact formulas for each component are provided in Appendix IV. Here we discuss the main intuition. At the end of each day, the gain/loss from market timing captures the comovement of the account balances with the market portfolio, while the stock selection component captures abnormal returns from each stock's risk-adjusted returns (alphas). For the transaction cost component, we use parameters directly from the SHSE, which include the commission cost (0.05% imposed on both the buy and sell side, with a minimum of 5 CNY for each trade), the stamp tax (0.10% of the sales amount during our sample period), and the transfer fee (0.002% imposed on both sides).

We present the gain/loss estimates in Table VIII. In Panel A, all five retail investor groups' total gain/loss estimates are negative, and are 77, 101, 177, 121 and 21 billion CNY per year, respectively. The third group RT3 loses the most at 177 billion CNY, while the fifth group RT5 loses the least at 21 billion CNY. In comparison, the total annualized gains for all institutional investors and corporations are 532 and 1,099 billion CNY, respectively.

If we add all five retail groups together, retail investors in SHSE-listed firms lose an aggregate of 497 billion CNY each year on average (71 billion USD per year), or 1.8% of the Shanghai Stock Exchange stock tradable market capitalization. Compared to the investor gain/loss structure in

Taiwan as in Barber et al (2008) with the total loss of individual investor in Taiwan stock market from 1995 to 1999 being 6.4 billion USD per year, the magnitude of losses by retail investors in Shanghai Stock Exchange is 11 times the dollar loss in Taiwan, possibly due to the large investor base in China. But as a percentage of total capitalization, the relative magnitude of the aggregate loss is smaller than that of Taiwan.

Where does the money go? From the second to fourth column, we decompose the gain/loss into three components: market timing, stock selection and transaction costs. The aggregate gain/loss of the market timing part is -24, -46, -91, -59 and -97 billion CNY per year for each of the five retail investor groups, while the stock selection part is -48, -37, -59, -48, and 95 billion per year. All five groups are unable to gain from market timing, which might not be surprising given the volatile nature of an emerging stock market. More interestingly, for RT1-RT4, the losses from stock selection are generally similar to the losses from market timing, which indicates that the poor stock selection from these investors, perhaps due to a lack of skills in information processing, is as hazardous as their market (mis)timing. Maybe it is not surprising to find that the largest retail investors in RT5, with better information processing skills, actually gain from stock selection, with annual gains of 95 billion CNY. Finally, the transaction costs are provided in the second column, and they are 5, 17, 27, 14, and 19 billion CNY per year for each group, respectively. We would like to mention that for the largest retail group, RT5, the annual total loss of 21 billion CNY is actually very close to the transaction cost of 19 billion CNY. In other words, if there were no

transaction costs, the largest retail investors as a group would have roughly broken even from stock trading over our sample period. In the last column, we report the gain/loss as a percentage of the beginning of period account value. For RT1-RT5, the loss is between -1.62% to -20.53% of the account values from beginning of the year.

In Panel B of Table VIII, we compute average gain/loss at the account level by dividing the aggregate gains/losses by the number of accounts for each investor group. For RT1-RT4 investors, the average account level loss is between 2,457 and 164,503 CNY per year. For RT5 investors, the loss per account is only 89,890 CNY. Despite the large aggregate losses and percentage losses, these average account level losses are not large, and might not be devastating for retail investors in general. This finding probably explains why retail investors have not quit investing in the stock market: the loss from trading stocks remains within a tolerable range. Given that gambling is illegal in China, investing in the stock market might be an alternative to gambling for some of the retail investors, and this explanation is also consistent with the lack of information processing skills and the presence of high-risk, speculative trading by smaller retail investors.[6]

An, Bian, Lou and Shi (2019) examine the wealth redistribution role of financial bubbles and crashes over July 2014 and December 2015, using a similar dataset over the stock market crisis

[6] In Appendix Table V, we compute the gains/losses following the original algorithm in Barber and Odean (2008) by assuming a holding period of 60 days. The patterns of gains/losses are generally consistent with our findings in Table VIII, but with smaller magnitudes. The reason is that our calculation in Table VIII also contains holding information.

period in 2015. They document a net transfer of 250 billion CNY (25 billion USD annually) from the poor to the ultra-wealthy from 2014 to 2015.

To understand whether their patterns persist over the latter period over 2016 to 2019, here we investigate whether the gains/losses are similar over calm vs. volatile conditions. We separate our sample into four different sets of market conditions according to the distribution of market returns. During our sample period, the mean of the market return is -0.01% and the standard deviation is 1.18%. There are 28 days when the market daily return is below the mean by two standard deviations or more, from -7.04% to -2.37%, and there are 24 days when the market daily return is above the mean by at least two standard deviations (from 2.35% to 5.60%).

In Panel C of Table VIII, we examine the gain/loss for each investor groups under these different market conditions. For the 28 days when market is below the mean by more than two standard deviations, all groups lose money, especially the larger retail investors, institutional and corporation investors. For the 24 days when the market is above its mean by more than two standard deviations, all groups make money, especially the larger ones. But the amount they make is generally 10-50% smaller than the amount they lose when the market is down. For the calm market conditions in the second and third columns, the gains and losses are less extreme. Overall, this pattern indicates a negative skew for the gain/loss measures, which possibly contributes to the aggregate losses for the retail investor groups.

## E. Ages and Genders

In this section, we examine heterogeneity through demographic differences, such as gender and age, of retail investors. According to Barber and Odean (2001), male investors could be more susceptible to behavioral biases, such as overconfidence and lack of attention.

Due to the limited access to data, we only have a three-month sample period from January 2019 to March 2019 on investor gender and age. We first present summary statistics on age and gender in Table IX Panel A. Male investors contribute 56% of trading volume on average, and females account for 44%. Within the male group, the trading volume (%) across age groups below 35, between 35 to 45, between 45 to 55 and above 55 is 10%, 19%, 24% and 14% (summing to the 56% male total), while the trading volume (%) for the same age groups for females is 5%, 9%, 11% and 9% (summing to the 44% of volume traded by females). That is, across all gender-age groups, younger male investors trade the most.

Next, we examine the determinants of order imbalance and return prediction for each gender-age group specified in equation (2) and (3). The results are reported in Table IX Panel B and Panel C. We find that male investors across all ages are on average momentum investors, while female investor groups are mostly contrarian investors. For return predictions, male investors across all ages negatively predict returns, with middle-aged males losing the most, while the youngest and oldest female retail investors (age less than 35 or above 55) can positively predict future returns.

These interesting patterns across age and gender provide further evidence regarding heterogeneity of retail investors.

## III. Robustness

### A. Applying stricter filters from Liu, Stambaugh and Yuan (2019)

In this study, we apply a filter from Liu et al (2019) and discard stocks with less than 15 days of the trading records during the most recent month. In addition, Liu et al (2019) also eliminate stocks that have become public within the past six months, stocks with less than 120 days of trading records during the past 12 months, and the smallest 30% of firms listed in SHSE and SZSE. We add all these additional filters and check the robustness of our results.

In Table X Panel A, the order imbalance determinants are similar to the results in Table II. The small and medium retail investors are momentum, and large retail investors and institutions are contrarian at a daily horizon. The coefficients on previous day order imbalances are still all positive and significant, consistent with order imbalance persistence. In Panel B, the order imbalance prediction directions are similar to the results in Table III. The first four groups of retail investors tend to trade in the wrong direction for future price movements, while the largest retail investor group RT5 and institutions trade in the same direction as the cross-section of future stock returns. The economic magnitudes for the first four type of retail investors are quantitatively similar, while RT5's economic magnitude is only half as large when adding these additional filters, perhaps

because RT5's positive return mainly comes from small stocks. The economic magnitude for institutions is still large. In conclusion, our main results are robust to the stricter filters from Liu, Stambaugh and Yuan (2019).

**B. Leveraged positions**

Our trade level data also identify investors' margin buys, short sells and collateral trades. Leveraged trading may be different from non-leverage trading. On each day, margin buys account for 10% of the trading volume, short sales account for 0.2% and collateral trading accounts for 15% during our sample period. We exclude the leverage trades and redo Table II and Table III.

Results are reported in Table X Panel C and Panel D. In Panel C, the order imbalance determinants results are similar to results in Table II; the small and medium retail investors are short-term momentum traders, and large retail investors and institutions are contrarian. The coefficients on previous day order imbalances are still all positive and significant, consistent with order imbalance persistence. In Panel D, the order imbalance prediction directions are similar to the results in Table III. The first four groups of retail investors trade in the wrong direction of future price movements, while the largest retail investor group RT5 and institutions trade in a way that positively predicts the cross-section of future stock returns. The economic magnitudes are quantitatively similar. In conclusion, our results are robust to whether or not we include these leverage trades.

## IV. Conclusion

Using proprietary data from the Shanghai Stock Exchange over 2016 to 2018, we separate tens of millions of retail investors into five groups by their account sizes, and we examine how their trading activities are related to future stock returns. We identify substantial heterogeneity in retail investors' trading behavior, information processing skills and their gains/losses from the stock market. Retail investors with account sizes less than 3mil CNY buy and sell in the wrong directions. The prices of stocks they buy experience negative returns the next day, while the ones they sell experience positive returns. For retail investors with large account balances, their aggregate trading predicts returns in the correct direction. In terms of news processing skills, we find the less wealthy retail investors fail to correctly process public news, while wealthier retail investors can successfully incorporate news into their trading and thereby benefit. Finally, we find that all groups of retail investors lose money, though retail investors with larger account sizes lose significantly less on average.

Our results on heterogeneity of retail investors help to understand the conflicting empirical results in the previous literature regarding retail investors. In addition, it is interesting that the exchange itself acknowledges the heterogeneity in retail investors, and is focused on adopting policies on suitability that restrict some kinds of trading for the smallest accounts. For example, the Shanghai Stock Exchange requires a retail investor to have at least 500k CNY holding of stocks for at least 20 trading days to open a leverage trading account or to trade on the riskier Science and

Technology Innovation Board (or STAR Market). These policies effectively exclude the smallest retail investors from leverage trading and trading on riskier start-ups, which could help protect these small retail investors from even worse losses.

## References


An, Li., Bian, Jiangze, Lou, Dong, and Shi, Donghui, 2019. Wealth Redistribution in Bubbles and Crashes. *Available at SSRN 3402254*.

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., and Terrance Odean, 2001, Boys will be boys: Gender, overconfidence, and common stock investment, *The Quarterly Journal of Economics*, *116*(1), 261-292.

Barber, Brad M., and Terrance Odean, 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies* 21, 785-818.

Barber, Brad M., Terrance Odean, and Ning Zhu, 2009, Do retail trades move markets? *Review of Financial Studies* 22, 151−186.

Barber, B. M., Lee, Y. T., Liu, Y. J., & Odean, T., 2008. Just how much do individual investors lose by trading?. *The Review of Financial Studies*, *22*(2), 609-632.

Barrot, Jean-Noel, Ron Kaniel and David Alexandre Sraer, 2016, Are Retail Traders Compensated for Providing Liquidity? *Journal of Financial Economics* 120, 146-168.

Blume, Marshall E. and Robert F. Stambaugh, 1983, Biases in computed returns: An application to the size effect, *Journal of Financial Economics*, 12, 387-404.

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

Boehmer, Ekkehart, Charles M. Jones, Juan Wu, and Xiaoyan Zhang, 2016. What Do Short Sellers Know? Working paper, Columbia.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang and Xinran Zhang, 2019. Tracking retail investor activity. *Available at SSRN 2822105*

Campbell, John Y., Tarun Ramadorai, and Allie Schwartz, 2009, Caught on tape: Institutional trading, stock returns, and earnings announcements, *Journal of Financial Economics*, 92(1), 66-91.

Chordia, Tarun, and Avanidhar Subrahmanyam, 2004, Order imbalance and stock returns: Theory and evidence, *Journal of Financial Economics* 72, 485−518.

Chordia, Tarun, Sahn-Wook Huh, and Avanidhar Subrahmanyam, 2006, The cross-section of expected trading activity, *Review of Financial Studies*, 20, 709-740.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607−636.

Fong, Kingsley YL, David R. Gallagher, and Adrian D. Lee, 2014, Individual investors and broker types, *Journal of Financial and Quantitative Analysis* 49.2 : 431-451.

Hendershott, Terrence, Dmitry Livdan, and Norman Schürhoff, 2015. Are institutions informed about news?. *Journal of Financial Economics*, 117(2), 249-287.

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

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

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.

Lee, Charles M. and Balkrishna Radhakrishna, 2000, Inferring investor behavior: Evidence from TORQ data, *Journal of Financial Markets*, 3(2), 83-111.

Lee, Charles M.C. and Mark J. Ready, 1991, Inferring investor behavior from intraday data, *Journal of Finance*, 46(2), 733-746.

Lee, Yi-Tsung, Liu, Yu-Jane, Roll, Richard, and Subrahmanyam, Avanidhar, 2004, Order imbalances and market efficiency: Evidence from the Taiwan Stock Exchange. *Journal of Financial and Quantitative Analysis*, 39, 327-341.

Liu, Jianan, Robert F. Stambaugh, and Yu Yuan, 2019. "Size and value in China." *Journal of Financial Economics*.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703−708.

You, Jiaxing, Bohui Zhang, and Le Zhang, 2017. "Who captures the power of the pen?." *The Review of Financial Studies* 31.1, 43-96.

**Table I. Summary statistics**

This Table reports summary statistics for trading by different investor groups. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on the Shanghai Stock Exchange with at least 15 days with trades during the previous month. Panel A reports the time series average of the number of accounts trading each year and aggregate daily trading activity. Panel B reports the time series average of cross sectional mean, standard deviation, 25$^{th}$ and 75$^{th}$ percentiles of stock characteristics and return distributions. Panel C reports the time series average of cross sectional mean buy (*Buyvol*) and sell (*Sellvol*) share volume and holdings in shares (*HoldShr*) across different investor groups. Panel D reports the time series average of cross sectional mean, standard deviation, autocorrelation and cross correlation of scaled daily order imbalances by each investor group. Order imbalance is computed as the buy share volume minus sell share volume divided by buy share volume plus sell share volume for each investor group.

Panel A. Trading and Holdings by Different Types of Investors in the Aggregate

| | RT1 | RT2 | RT3 | RT4 | RT5 | INST | CORP |
|---|---|---|---|---|---|---|---|
| Account value | <100K CNY | (100K,500K) CNY | (500K,3M) CNY | (3M,10M) CNY | >10M CNY | | |
| Number of Accounts | 31,409,950 | 15,282,005 | 5,826,800 | 734,705 | 235,420 | 39,547 | 46,882 |
| Trading Volume (Bil. CNY) | 9 | 35 | 54 | 27 | 37 | 35 | 3 |
| Trading Volume (% of total) | 5% | 17% | 27% | 13% | 19% | 17% | 2% |
| Holdings Value (Bil CNY) | 336 | 951 | 1,566 | 840 | 1,794 | 4,201 | 15,547 |
| Holdings Value (% of total) | 1% | 4% | 6% | 3% | 7% | 17% | 62% |

Panel B. Stock Characteristics and Return Distribution

| | Mean | Std | P25 | P50 | P75 |
|---|---|---|---|---|---|
| Market Capitalization (Billion CNY) | 20.1 | 80.3 | 2.9 | 5.6 | 12.1 |
| Monthly Turnover (of Tradable A shares) | 48.32% | 72.48% | 14.09% | 25.40% | 49.97% |
| Earnings to Price Ratio | 0.0075 | 0.0155 | 0.0018 | 0.0060 | 0.0122 |
| Daily Stock Return | -0.01% | 2.17% | -1.09% | -0.22% | 0.77% |

Panel C. Trading and Holdings by Different Types of Investors in the Cross Section

| | RT1 | RT2 | RT3 | RT4 | RT5 | INST | CORP |
|---|---|---|---|---|---|---|---|
| BuyVol | 801,220 | 2,708,679 | 4,189,939 | 2,052,381 | 2,776,871 | 2,344,479 | 255,426 |
| SellVol | 799,445 | 2,698,182 | 4,179,190 | 2,050,914 | 2,776,958 | 2,366,058 | 259,280 |
| BuyVol(%) | 5.4% | 20.1% | 30.0% | 13.2% | 14.9% | 15.2% | 1.2% |
| SellVol(%) | 5.6% | 20.6% | 30.2% | 13.2% | 14.4% | 15.1% | 1.0% |
| HoldShr | 39,970,679 | 97,128,784 | 144,784,216 | 72,360,705 | 135,012,468 | 256,301,535 | 1,626,316,536 |
| HoldShr(%) | 3.3% | 10.4% | 16.0% | 7.8% | 14.5% | 11.5% | 36.4% |
| Holding Period (Days) | 50 | 36 | 35 | 35 | 49 | 109 | 6,319 |

Panel D. Order Imbalance in the Cross Section by Investor Group

| | Mean | Std | AR1 | Correlations | OibRT1 | OibRT2 | OibRT3 | OibRT4 | OibRT5 | OibInst | OibCorp |
|---|---|---|---|---|---|---|---|---|---|---|---|
| OibRT1 | -0.021 | 0.187 | 0.243 | OibRT1 | 1 | | | | | | |
| OibRT2 | -0.011 | 0.171 | 0.259 | OibRT2 | 0.802 | 1 | | | | | |
| OibRT3 | -0.006 | 0.166 | 0.216 | OibRT3 | 0.610 | 0.710 | 1 | | | | |
| OibRT4 | 0.002 | 0.25 | 0.059 | OibRT4 | 0.194 | 0.244 | 0.256 | 1 | | | |
| OibRT5 | 0.019 | 0.352 | 0.102 | OibRT5 | -0.151 | -0.158 | -0.163 | -0.091 | 1 | | |
| OibInst | -0.011 | 0.455 | 0.205 | OibInst | -0.315 | -0.365 | -0.380 | -0.263 | -0.188 | 1 | |
| OibCorp | -0.004 | 0.72 | 0.088 | OibCorp | 0.022 | 0.029 | 0.021 | -0.007 | -0.043 | -0.044 | 1 |

**Table II. Determinants of Order Imbalance by Different Investor Groups**

This table reports determinants of different investor group trading activity. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on the Shanghai Stock Exchange with at least 15 days with trades during the previous month. We estimate Fama MacBeth regressions as specified in equation (3). The dependent variables are scaled order imbalance measures on day d, defined as buy share volume minus sell share volume divided by buy share volume plus sell share volume for each investor type. As independent variables, we include the previous day return *Ret(-1),* the previous week return *Ret(-6,-2)* the previous month return *Ret(-27,-7),* and the previous day order imbalance *Oib(-1)*. As control variables, we include log market cap (*Size*) from the end of the previous month, the earnings to price ratio (*EP*) as of the end of the previous month, and last month's monthly turnover (*Turnover*). To account for serial correlation in the coefficients, the standard errors of the time-series are adjusted using Newey-West (1987) with 5 lags.

| Dep.var | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |
| Ret(-1) | 0.5484 | 16.52 | 0.7688 | 30.76 | 0.4837 | 23.59 | -0.2155 | -9.59 | -1.2644 | -38.89 | -2.2759 | -33.78 | -0.6050 | -11.75 |
| Ret(-6,-2) | -0.4173 | -26.88 | -0.2087 | -18.58 | -0.0820 | -8.69 | -0.0517 | -4.65 | -0.0598 | -4.90 | 0.0377 | 1.20 | -0.3553 | -14.08 |
| Ret(-27,-7) | -0.0395 | -6.50 | -0.0140 | -2.99 | -0.0065 | -1.77 | -0.0190 | -4.57 | -0.0398 | -6.97 | -0.0182 | -1.26 | -0.0667 | -5.64 |
| Oib(-1) | 0.1836 | 39.22 | 0.1947 | 48.07 | 0.1712 | 49.34 | 0.0499 | 21.54 | 0.1036 | 39.17 | 0.2254 | 66.17 | 0.0859 | 33.08 |
| Size | 0.0054 | 5.77 | 0.0036 | 4.47 | 0.0011 | 1.70 | -0.0038 | -5.96 | -0.0105 | -13.54 | -0.0026 | -1.49 | -0.0340 | -12.68 |
| EP | 0.0010 | 0.03 | 0.0072 | 0.29 | -0.0182 | -0.86 | -0.1024 | -3.77 | -0.1557 | -4.70 | 0.3129 | 4.56 | 0.8470 | 11.40 |
| Turnover | 0.0055 | 5.40 | 0.0026 | 3.93 | 0.0022 | 4.91 | 0.0013 | 2.25 | -0.0075 | -8.57 | -0.0016 | -0.79 | 0.0061 | 2.98 |
| Intercept | -0.1424 | -6.75 | -0.0913 | -4.91 | -0.0282 | -1.89 | 0.0922 | 6.43 | 0.2667 | 15.16 | 0.0411 | 0.97 | 0.7716 | 12.88 |
| Adj. R2 | 8.41% | | 7.03% | | 4.83% | | 1.06% | | 2.36% | | 8.19% | | 2.32% | |

**Table III. Predicting Future Returns Using Order Imbalances by Different Investor Groups**

This table reports estimation results on whether trading activity by different investor groups can predict the cross section of one-day-ahead returns. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on the Shanghai Stock Exchange with at least 15 days with trading during the previous month. We estimate Fama-MacBeth regressions as specified in equation (4). The dependent variable is the return on day d. As independent variables, we include the previous day order imbalance *Oib(-1)*, the previous day return *Ret(-1)*, the previous week return *Ret(-6,-2)* and the previous month return *Ret(-27,-7)*. In addition, we include previous month log market cap(*Size*), earnings to price ratio (*EP*) and monthly turnover (*Turnover*) as control variables. For each regression, we also provide the interquartile range for the relevant explanatory order imbalance to compute the difference in predicted day-ahead returns for observations at the two ends of the interquartile range. To account for serial correlation in the coefficients, the standard errors of the time-series are adjusted using Newey-West (1987) with 5 lags.

| Dep.var | Ret | | Ret | | Ret | | Ret | | Ret | | Ret | | Ret | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Oib.var | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
| | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |
| Oib(-1) | -0.0093 | -24.98 | -0.0091 | -22.58 | -0.0065 | -18.50 | -0.0009 | -7.21 | 0.0012 | 12.26 | 0.0016 | 20.34 | 0.0000 | -1.51 |
| Ret(-1) | -0.0027 | -0.62 | -0.0091 | -2.07 | 0.0006 | 0.13 | 0.0189 | 4.06 | 0.0190 | 4.13 | 0.0132 | 2.79 | 0.0206 | 4.42 |
| Ret(-6,-2) | -0.0149 | -8.06 | -0.0132 | -7.07 | -0.0124 | -6.62 | -0.0120 | -6.37 | -0.0115 | -6.13 | -0.0113 | -6.04 | -0.0118 | -6.29 |
| Ret(-27,-7) | -0.0039 | -4.36 | -0.0036 | -4.04 | -0.0034 | -3.86 | -0.0033 | -3.72 | -0.0032 | -3.62 | -0.0034 | -3.85 | -0.0033 | -3.70 |
| Size | 0.0000 | 0.36 | 0.0000 | 0.17 | 0.0000 | -0.16 | 0.0000 | -0.32 | 0.0000 | -0.18 | 0.0000 | -0.21 | 0.0000 | -0.33 |
| EP | 0.0147 | 3.54 | 0.0150 | 3.57 | 0.0145 | 3.41 | 0.0144 | 3.42 | 0.0146 | 3.47 | 0.0140 | 3.34 | 0.0146 | 3.45 |
| Turnover | -0.0007 | -3.47 | -0.0007 | -3.63 | -0.0007 | -3.69 | -0.0007 | -3.83 | -0.0007 | -3.83 | -0.0007 | -3.83 | -0.0007 | -3.87 |
| Intercept | -0.0012 | -0.48 | -0.0006 | -0.26 | 0.0002 | 0.06 | 0.0005 | 0.19 | 0.0001 | 0.06 | 0.0002 | 0.10 | 0.0005 | 0.20 |
| Adj R2 | 8.83% | | 8.68% | | 8.43% | | 8.10% | | 8.11% | | 8.25% | | 8.08% | |
| Interquartile | 0.2222 | | 0.1827 | | 0.1678 | | 0.2868 | | 0.4536 | | 0.6740 | | 1.4844 | |
| IQ Ret Diff | -0.2062% | | -0.1668% | | -0.1089% | | -0.0247% | | 0.0523% | | 0.1046% | | -0.0059% | |

**Table IV. Return Predictability within Subgroups**

This table reports whether order imbalances by different investor groups can predict the cross section of returns for a subset of stocks. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on the Shanghai Stock Exchange with at least 15 days with trading during the previous month. We first sort all firms into 3 groups based on previous month-end characteristics. Then we estimate Fama-MacBeth regressions as specified in equation (4) for each subgroup. The dependent variable is the return on day d. As independent variables, we include the previous day order imbalance *Oib(-1)*, the previous day return *Ret(-1),* the previous week return *Ret(-6,-2)* and previous month return *Ret(-27,-7)*. In addition, we include last month's log market cap(*Size*), earnings to price ratio (*EP*) and previous monthly turnover (*Turnover*) as control variables. For each regression, we also provide the interquartile range for the relevant explanatory order imbalance to compute the difference in predicted day-ahead returns for observations at the two ends of the interquartile range. To account for serial correlation in the coefficients, the standard errors of the time-series are adjusted using Newey-West (1987) with 5 lags.

Panel A. Market cap groups

| | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |
| Small | -0.0114 | -24.34 | -0.0123 | -22.93 | -0.0083 | -17.61 | -0.0006 | -3.47 | 0.0019 | 15.06 | 0.0012 | 13.10 | -0.0001 | -1.10 |
| Medium | -0.0095 | -21.14 | -0.0093 | -18.81 | -0.0066 | -15.13 | -0.0010 | -5.92 | 0.0009 | 7.92 | 0.0014 | 14.57 | 0.0000 | -0.18 |
| Large | -0.0075 | -19.58 | -0.0070 | -17.62 | -0.0053 | -14.48 | -0.0015 | -7.33 | 0.0000 | 0.17 | 0.0024 | 15.94 | -0.0001 | -1.40 |
| | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | |
| Small | -0.2420% | | -0.2069% | | -0.1346% | | -0.0196% | | 0.1000% | | 0.0965% | | -0.0076% | |
| Medium | -0.2016% | | -0.1608% | | -0.1045% | | -0.0297% | | 0.0438% | | 0.1013% | | -0.0015% | |
| Large | -0.1825% | | -0.1476% | | -0.0981% | | -0.0366% | | 0.0009% | | 0.1295% | | -0.0112% | |

Panel B. Turnover groups

| | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |
| Low | -0.0063 | -19.68 | -0.0061 | -18.20 | -0.0047 | -15.67 | -0.0010 | -6.78 | 0.0003 | 3.01 | 0.0015 | 17.40 | -0.0001 | -1.03 |
| Medium | -0.0087 | -20.80 | -0.0087 | -17.87 | -0.0064 | -14.61 | -0.0010 | -6.21 | 0.0011 | 8.26 | 0.0014 | 15.30 | 0.0000 | -0.93 |
| High | -0.0156 | -26.68 | -0.0171 | -25.60 | -0.0119 | -19.23 | -0.0007 | -2.81 | 0.0025 | 13.29 | 0.0017 | 14.11 | 0.0000 | -0.41 |
| | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | |
| Low | -0.1711% | | -0.1451% | | -0.1033% | | -0.0335% | | 0.0129% | | 0.0955% | | -0.0078% | |
| Medium | -0.2007% | | -0.1673% | | -0.1127% | | -0.0295% | | 0.0495% | | 0.0928% | | -0.0061% | |
| High | -0.2783% | | -0.2338% | | -0.1459% | | -0.0169% | | 0.1001% | | 0.1239% | | -0.0027% | |

Panel C. Share price groups

| | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |
| Low | -0.0085 | -17.61 | -0.0081 | -15.86 | -0.0065 | -14.27 | -0.0013 | -9.20 | 0.0008 | 6.52 | 0.0013 | 15.17 | 0.0000 | -0.84 |
| Medium | -0.0098 | -22.17 | -0.0096 | -20.03 | -0.0070 | -16.52 | -0.0007 | -4.42 | 0.0013 | 9.91 | 0.0016 | 14.50 | 0.0000 | -0.65 |
| High | -0.0092 | -22.92 | -0.0093 | -20.99 | -0.0059 | -14.62 | -0.0006 | -3.28 | 0.0014 | 10.49 | 0.0017 | 13.70 | -0.0001 | -1.19 |
| | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | | Daily Ret Diff | |
| Low | -0.1952% | | -0.1548% | | -0.1098% | | -0.0390% | | 0.0373% | | 0.0947% | | -0.0063% | |
| Medium | -0.2167% | | -0.1747% | | -0.1168% | | -0.0210% | | 0.0561% | | 0.1086% | | -0.0045% | |
| High | -0.1968% | | -0.1642% | | -0.0979% | | -0.0171% | | 0.0616% | | 0.1064% | | -0.0084% | |

**Table V. Predicting K-weeks Ahead**

This table reports estimation results on whether different investor groups' trading activity can predict the cross section of future returns at more distant horizons. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on Shanghai Stock Exchange with at least 15 trading days in the previous month. We estimate Fama-MacBeth regressions, as specified in equation (5). The dependent variable is the k weeks ahead weekly future return. The independent variable is the previous day order imbalance *Oib(-1)*, and other variables are the same as those in Table III; those coefficients are not reported. To account for serial correlation in the coefficients, the standard errors of the time-series are adjusted using Newey-West (1987) with 5 lags.

| Oib.var | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| weeks Ahead | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |
| 1 | -0.0226 | -25.04 | -0.0220 | -22.11 | -0.0144 | -16.62 | -0.0019 | -6.70 | 0.0027 | 12.43 | 0.0044 | 21.02 | 0.0000 | -0.58 |
| 2 | -0.0065 | -10.65 | -0.0060 | -9.07 | -0.0037 | -6.11 | 0.0001 | 0.46 | 0.0010 | 6.11 | 0.0012 | 6.05 | -0.0001 | -0.95 |
| 3 | -0.0038 | -6.32 | -0.0031 | -4.77 | -0.0015 | -2.55 | 0.0001 | 0.54 | 0.0007 | 3.99 | 0.0007 | 3.52 | 0.0000 | -0.02 |
| 4 | -0.0024 | -4.00 | -0.0021 | -2.97 | -0.0012 | -1.81 | -0.0001 | -0.37 | 0.0007 | 4.18 | 0.0005 | 2.35 | -0.0002 | -2.08 |
| 5 | -0.0014 | -2.89 | -0.0014 | -2.59 | -0.0011 | -2.16 | -0.0005 | -1.98 | 0.0004 | 2.37 | 0.0002 | 1.42 | -0.0001 | -0.65 |
| 6 | -0.0029 | -5.68 | -0.0024 | -4.12 | -0.0016 | -3.04 | -0.0003 | -1.24 | 0.0001 | 0.74 | 0.0005 | 2.92 | 0.0000 | -0.26 |
| 7 | -0.0027 | -5.49 | -0.0025 | -4.68 | -0.0018 | -3.77 | -0.0002 | -0.70 | 0.0001 | 0.40 | 0.0007 | 4.42 | -0.0001 | -0.89 |
| 8 | -0.0015 | -2.92 | -0.0010 | -1.66 | -0.0007 | -1.29 | -0.0002 | -0.88 | 0.0003 | 2.11 | 0.0004 | 2.66 | 0.0002 | 2.29 |
| 9 | -0.0010 | -2.09 | -0.0005 | -1.04 | -0.0004 | -0.76 | 0.0002 | 0.95 | 0.0004 | 2.47 | 0.0001 | 0.44 | 0.0000 | 0.56 |
| 10 | -0.0007 | -1.37 | -0.0004 | -0.69 | -0.0004 | -0.84 | 0.0002 | 0.99 | -0.0001 | -0.45 | 0.0002 | 0.99 | -0.0001 | -1.49 |
| 11 | -0.0006 | -1.14 | -0.0007 | -1.16 | -0.0001 | -0.14 | 0.0000 | -0.04 | -0.0002 | -1.00 | 0.0002 | 1.18 | -0.0001 | -1.16 |
| 12 | -0.0005 | -1.06 | -0.0001 | -0.14 | 0.0005 | 0.86 | 0.0001 | 0.41 | 0.0001 | 0.39 | 0.0000 | 0.06 | 0.0000 | 0.00 |

**Table VI. Long-short Strategy Returns Based on Order Imbalances**

This table reports portfolio returns using a long-short strategy wherein we buy the stocks in the highest quintile of the relevant group's order imbalance, and we short the stocks in the lowest order imbalance quintile. The order imbalance is computed as the buy share volume minus sell share volume divided by buy share volume plus sell share volume for the specified investor group from the previous day (day 0). Portfolio returns are value-weighted by previous month-end market cap. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on the Shanghai Stock Exchange with at least 15 days with trades in the previous month. We report risk-adjusted returns using the Liu Stambaugh and Yuan (2019) three-factor model. As our data are overlapping, we adjust the standard errors of the portfolio return time-series using Newey-West (1987) with twice the horizon of the longest overlapping lags.

| Holding | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Period (days) | Alpha | t-stat | Alpha | t-stat | Alpha | t-stat | Alpha | t-stat | Alpha | t-stat | Alpha | t-stat | Alpha | t-stat |
| 1 | -0.0028 | -10.28 | -0.0023 | -8.17 | -0.0018 | -7.16 | -0.0008 | -3.32 | 0.0003 | 1.51 | 0.0036 | 12.84 | -0.0008 | -3.57 |
| 5 | -0.0043 | -5.45 | -0.0034 | -4.81 | -0.0020 | -2.87 | 0.0002 | 0.26 | 0.0013 | 2.05 | 0.0064 | 7.52 | -0.0010 | -1.65 |
| 10 | -0.0060 | -5.02 | -0.0051 | -4.86 | -0.0031 | -3.33 | 0.0001 | 0.07 | 0.0020 | 2.55 | 0.0102 | 6.84 | -0.0026 | -2.80 |
| 20 | -0.0082 | -3.52 | -0.0073 | -3.82 | -0.0050 | -3.45 | -0.0011 | -0.88 | 0.0018 | 1.27 | 0.0114 | 5.62 | -0.0027 | -2.24 |
| 30 | -0.0094 | -4.44 | -0.0080 | -4.41 | -0.0052 | -3.52 | 0.0007 | 0.35 | 0.0033 | 1.36 | 0.0116 | 5.23 | -0.0034 | -2.96 |
| 40 | -0.0110 | -5.84 | -0.0095 | -5.90 | -0.0063 | -3.75 | -0.0014 | -0.58 | 0.0025 | 1.18 | 0.0158 | 6.44 | -0.0033 | -1.64 |
| 50 | -0.0120 | -4.99 | -0.0101 | -4.07 | -0.0064 | -2.50 | -0.0025 | -0.78 | 0.0021 | 0.68 | 0.0158 | 4.21 | -0.0042 | -2.12 |
| 60 | -0.0131 | -4.21 | -0.0112 | -3.54 | -0.0071 | -2.38 | -0.0048 | -1.43 | -0.0020 | -0.55 | 0.0138 | 4.95 | -0.0032 | -1.27 |

**Table VII. Information Processing Ability by Different Investor Groups**

This table reports estimation results on the information processing ability of different investor groups. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on the Shanghai Stock Exchange with at least 15 days with trades in the previous month. Panel A reports how news days affect the return predictability of different investor group trades. We estimate Fama MacBeth regressions, as specified in equation (6). The dependent variable is the return on day d. The independent variables are the previous day's order imbalance *Oib(-1)*, the news dummies *Pos(-1)* and *Neg(-1)* and the interaction terms *Oib(-1)*Pos(-1)* and *Oib(-1)*Neg(-1)*. The *Pos(-1)* dummy is equal to 1 if the news sentiment score is positive for that firm-day and zero otherwise; *Neg(-1)* dummy is equal to 1 if the news sentiment score is negative for that firm-day and zero otherwise. Other control variables are the same as those in Table III; those coefficients are not reported. To account for serial correlation in the coefficients, the standard errors of the time-series are adjusted using Newey-West (1987) with 5 lags. Panel B reports whether different investor groups' trading activity can predict earnings surprises. For each quarter, we estimate Fama-MacBeth regressions as specified in equation (7) to measure stock returns around the earnings announcement for firm i on day d. The dependent variable, earnings surprise, is proxied by the cumulative abnormal return from day d-1 to day d+1, *CAR[-1,1]*. As independent variables, we use order imbalance measures from day d-2, *Oib(-2)*, to avoid overlapping with the CAR calculation. Other control variables are same as those in Table III. To account for serial correlation in the coefficients, the standard errors of the time-series are adjusted using Newey-West (1987) with 4 lags.

Panel A. Order Imbalance Prediction, with Positive/Negative News

| Dep.var | Ret | | Ret | | Ret | | Ret | | Ret | | Ret | | Ret | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Oib.var | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
| | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |
| Pos(-1) | 0.0005 | 6.76 | 0.0005 | 6.59 | 0.0005 | 6.27 | 0.0004 | 5.61 | 0.0004 | 5.06 | 0.0005 | 5.98 | 0.0004 | 5.35 |
| Neg(-1) | -0.0004 | -3.53 | -0.0004 | -3.95 | -0.0005 | -4.47 | -0.0005 | -4.40 | -0.0005 | -4.60 | -0.0005 | -4.12 | -0.0005 | -4.67 |
| Oib(-1) | -0.0077 | -24.34 | -0.0073 | -21.72 | -0.0049 | -16.92 | -0.0005 | -4.27 | 0.0008 | 10.27 | 0.0014 | 17.49 | 0.0000 | 1.41 |
| Oib(-1)Pos(-1) | -0.0046 | -9.35 | -0.0054 | -10.72 | -0.0052 | -10.20 | -0.0016 | -5.73 | 0.0014 | 6.33 | 0.0008 | 5.85 | -0.0004 | -4.86 |
| Oib(-1)Neg(-1) | -0.0053 | -8.67 | -0.0059 | -8.29 | -0.0052 | -7.31 | -0.0012 | -3.11 | 0.0015 | 5.66 | 0.0010 | 4.65 | -0.0003 | -2.65 |

Panel B. Predicting Earnings Surprise Using Order Imbalances by Different Investor Groups

| Dep.var | CAR[-1,1] | | CAR[-1,1] | | CAR[-1,1] | | CAR[-1,1] | | CAR[-1,1] | | CAR[-1,1] | | CAR[-1,1] | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Oib.var | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst. | | OibCorp. | |
| | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |
| Oib(-2) | -0.0247 | -9.53 | -0.0226 | -5.84 | -0.0156 | -4.86 | 0.0000 | 0.01 | 0.0019 | 2.22 | 0.0039 | 3.44 | -0.0005 | -1.04 |
| Ret(-2) | -0.1876 | -3.37 | -0.1874 | -3.27 | -0.1629 | -2.63 | -0.1288 | -2.33 | -0.1284 | -2.29 | -0.1493 | -2.88 | -0.1299 | -2.36 |
| Ret(-7,-3) | -0.1127 | -5.78 | -0.1079 | -5.57 | -0.1051 | -5.55 | -0.1034 | -5.45 | -0.1029 | -5.40 | -0.1032 | -5.41 | -0.1037 | -5.53 |
| Ret(-28,-8) | -0.0091 | -0.96 | -0.0083 | -0.87 | -0.0079 | -0.84 | -0.0075 | -0.79 | -0.0074 | -0.77 | -0.0081 | -0.86 | -0.0074 | -0.78 |
| Size | 0.0010 | 0.86 | 0.0009 | 0.80 | 0.0008 | 0.74 | 0.0008 | 0.70 | 0.0008 | 0.70 | 0.0008 | 0.71 | 0.0008 | 0.70 |
| EP | -0.0951 | -1.59 | -0.0968 | -1.55 | -0.0966 | -1.55 | -0.0986 | -1.63 | -0.1005 | -1.63 | -0.1002 | -1.67 | -0.0984 | -1.60 |
| Turnover | -0.0034 | -1.25 | -0.0036 | -1.31 | -0.0035 | -1.29 | -0.0035 | -1.27 | -0.0035 | -1.27 | -0.0037 | -1.36 | -0.0035 | -1.25 |
| Intercept | -0.0170 | -0.62 | -0.0154 | -0.55 | -0.0141 | -0.49 | -0.0132 | -0.46 | -0.0131 | -0.46 | -0.0130 | -0.46 | -0.0132 | -0.46 |
| Adj.R2 | 6.32% | | 6.04% | | 5.76% | | 5.52% | | 5.49% | | 5.73% | | 5.48% | |

**Table VIII. Gain/Loss for Various Investor Groups**

This table reports the gain and loss for various investor groups from January 2016 to June 2019. Panel A reports the gain/loss using holding and trading data. On each trading day, we first compute the total gain/loss for each account, by following the gain/loss on both the holdings and the day's trading marked to market. Then we sum up the gains/losses for all investors within the same investor group. We also decompose the gain/loss into three parts: market timing, stock selection and transaction costs. Exact formulas for each component are provided in Appendix IV. At the end of each day, the gain/loss from market timing captures the comovement of the account balances with the market portfolio, while the stock selection component captures abnormal returns from each stock's risk-adjusted returns (or alphas). For the transaction cost component, we use parameters directly from the SHSE, which include commission cost (0.05% imposed on both the buy and sell side, with a minimum of 5 CNY for each trade), the stamp tax (0.10% of the sales amount), and the transfer fee (0.002% imposed on both sides). We also compute average gain/loss at the account level and the percentage of gain/loss over investor holding position. Panel B reports the gain/loss pattern during market ups and downs. We separate our sample into four different regions, using the mean (-0.01%) and standard deviation (1.18%) of daily market portfolio returns, and report the average daily gain/loss during different market conditions.

Panel A. Annualized gain/loss

| Annualized gain/loss | Total (mil CNY) | Market timing (mil CNY) | Stock selection (mil CNY) | Transaction costs (mil CNY) | Gain/loss percentage |
|---|---|---|---|---|---|
| RT1 | -77,186 | -24,303 | -48,136 | -4,747 | -20.53% |
| RT2 | -100,877 | -46,116 | -37,339 | -17,422 | -9.98% |
| RT3 | -177,388 | -90,654 | -59,267 | -27,467 | -10.79% |
| RT4 | -120,861 | -59,340 | -47,954 | -13,567 | -13.20% |
| RT5 | -21,162 | -96,895 | 94,703 | -18,969 | -1.62% |
| Inst. | 531,814 | -165,183 | 714,804 | -17,807 | 11.22% |
| Corp. | 1,098,956 | -692,276 | 1,792,984 | -1,752 | 6.68% |

Panel B. Annualized gain/loss per account

| Annualized gain/loss per account | Total gain/loss per account (CNY) | Market timing per account (CNY) | Stock selection per account (CNY) | Transaction costs per account (CNY) |
|---|---|---|---|---|
| RT1 | -2,457 | -774 | -1,532 | -151 |
| RT2 | -6,601 | -3,018 | -2,443 | -1,140 |
| RT3 | -30,443 | -15,558 | -10,171 | -4,714 |
| RT4 | -164,503 | -80,767 | -65,269 | -18,466 |
| RT5 | -89,890 | -411,584 | 402,271 | -80,577 |
| Inst. | 13,447,655 | -4,176,872 | 18,074,792 | -450,265 |
| Corp. | 23,440,904 | -14,766,355 | 38,244,620 | -37,361 |

Panel C. Daily gain/loss for each investor group under different market conditions, million CNY.

| Market Ret | Below mean by more than two stdev. | Below mean within two stdev | Above mean within two stdev | Above mean by more than two stdev. |
|---|---|---|---|---|
| N days | 28 | 367 | 429 | 24 |
| RT1 | -19,898 | -2,451 | 2,267 | 9,750 |
| RT2 | -39,636 | -6,935 | 6,529 | 27,792 |
| RT3 | -70,140 | -11,134 | 10,545 | 46,022 |
| RT4 | -41,193 | -5,868 | 5,465 | 24,608 |
| RT5 | -86,067 | -11,468 | 12,245 | 54,849 |
| Inst. | -144,358 | -24,289 | 26,911 | 135,421 |
| Corp | -504,827 | -81,016 | 87,944 | 411,397 |

**Table IX Gender and Age**

This table reports trading behavior and return prediction across different age and gender investor groups. The sample period covers January 2019 to March 2019. The sample firms are A-share stocks listed on the Shanghai Stock Exchange with at least 15 days with trades during the previous month. Since age and gender are applicable only for retail investors, we only include retail investors in this table. Panel A reports the summary statistics of trading volume across different age and gender group. Panel B reports the determinants of order imbalance by different investor groups. We estimate Fama-MacBeth regressions as specified in equation (3). The dependent variables are scaled order imbalance measures on day d, defined as buy share volume minus sell share volume divided by buy share volume plus sell share volume for each investor type. As independent variables, we include the previous day return *Ret(-1),* the previous week return *Ret(-6,-2)* the previous month return *Ret(-27,-7),* and the previous day order imbalance *Oib(-1)*. As additional control variables, we include log market cap (*Size*) from the end of the previous month, the earnings to price ratio (*EP*) as of the end of the previous month, and last month's monthly turnover (*Turnover*). Panel C reports the return prediction using order imbalance by different investor groups. We estimate Fama-MacBeth regressions as specified in equation (4). The dependent variable is the return on day d. The independent variable and control variables are same as above. To account for serial correlation in the coefficients, the standard errors of the time-series are adjusted using Newey-West (1987) with 5 lags.

Panel A. Summary Statistics

| Gender | Trading Volume (% of total for that gender) | | | |
|---|---|---|---|---|
| Age | <35 | 35-45 | 45-55 | >55 |
| Male | 10% | 19% | 24% | 14% |
| Female | 5% | 9% | 11% | 9% |

Panel B. Determinants of Order Imbalance by Different Investor Groups

| Dep.var | Oib | | Oib | | Oib | | Oib | | Oib | | Oib | | Oib | | Oib | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gender | Male | | Male | | Male | | Male | | Female | | Female | | Female | | Female | |
| Age | <35 | | 35-45 | | 45-55 | | >55 | | <35 | | 35-45 | | 45-55 | | >55 | |
| | Coef | t-stat | Coef | t-stat | Coef | t-stat | Coef | t-stat | Coef | t-stat | Coef | t-stat | Coef | t-stat | Coef | t-stat |
| Ret(-1) | 0.7762 | 9.22 | 0.4560 | 5.50 | 0.2173 | 1.98 | 0.3340 | 3.19 | 0.0018 | 0.02 | -0.1438 | -2.33 | -0.0594 | -0.63 | -0.4831 | -6.95 |
| Ret(-6,-2) | -0.2848 | -5.06 | -0.2340 | -4.78 | -0.1595 | -4.21 | -0.1031 | -2.37 | -0.2242 | -3.53 | -0.1231 | -2.46 | -0.1077 | -1.95 | -0.1005 | -2.18 |
| Ret(-27,-7) | -0.0334 | -2.63 | -0.0358 | -3.49 | -0.0180 | -1.10 | -0.0178 | -0.90 | -0.0209 | -1.41 | -0.0185 | -1.99 | -0.0358 | -1.88 | -0.0198 | -1.23 |
| Oib(-1) | 0.0073 | 0.83 | 0.0501 | 5.92 | 0.0909 | 11.60 | 0.0638 | 18.91 | -0.0096 | -2.32 | 0.0337 | 3.54 | 0.0700 | 11.52 | 0.0432 | 5.04 |
| Size | 0.0030 | 1.18 | 0.0055 | 1.86 | 0.0039 | 1.29 | 0.0022 | 0.56 | -0.0005 | -0.12 | 0.0031 | 0.91 | 0.0022 | 0.64 | -0.0007 | -0.17 |
| EP | -0.0459 | -1.03 | -0.0271 | -0.84 | 0.0326 | 0.86 | 0.0254 | 0.42 | -0.0209 | -0.39 | -0.0359 | -0.70 | 0.0275 | 0.59 | 0.0009 | 0.02 |
| Turnover | 0.0032 | 1.54 | 0.0062 | 2.11 | 0.0064 | 3.76 | 0.0088 | 2.88 | 0.0013 | 0.40 | 0.0076 | 2.02 | 0.0117 | 4.20 | 0.0045 | 1.09 |
| Intercept | -0.0727 | -1.35 | -0.1316 | -2.12 | -0.0916 | -1.39 | -0.0557 | -0.66 | 0.0102 | 0.13 | -0.0796 | -1.03 | -0.0549 | -0.72 | 0.0122 | 0.14 |

Panel C. Predicting Future Returns Using Order Imbalances by Different Investor Groups

| Dep.var | Ret | | Ret | | Ret | | Ret | | Ret | | Ret | | Ret | | Ret | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gender | Male | | Male | | Male | | Male | | Female | | Female | | Female | | Female | |
| Age | <35 | | 35-45 | | 45-55 | | >55 | | <35 | | 35-45 | | 45-55 | | >55 | |
| | Coef | t-stat | Coef | t-stat | Coef | t-stat | Coef | t-stat | Coef | t-stat | Coef | t-stat | Coef | t-stat | Coef | t-stat |
| Oib(-1) | -0.0015 | -3.31 | -0.0032 | -4.82 | -0.0054 | -5.11 | -0.0026 | -3.20 | 0.0007 | 2.16 | -0.0004 | -0.80 | -0.0021 | -2.57 | 0.0007 | 1.74 |
| Ret(-1) | 0.0504 | 2.53 | 0.0501 | 2.51 | 0.0422 | 2.11 | 0.0412 | 2.05 | 0.0502 | 2.51 | 0.0492 | 2.45 | 0.0440 | 2.17 | 0.0509 | 2.50 |
| Ret(-2,-6) | -0.0153 | -1.98 | -0.0154 | -2.01 | -0.0157 | -2.04 | -0.0153 | -1.99 | -0.0152 | -1.97 | -0.0152 | -1.97 | -0.0153 | -1.99 | -0.0151 | -1.97 |
| Ret(-27,-7) | 0.0017 | 0.46 | 0.0016 | 0.43 | 0.0015 | 0.40 | 0.0017 | 0.44 | 0.0018 | 0.47 | 0.0017 | 0.46 | 0.0016 | 0.42 | 0.0017 | 0.45 |
| Size | 0.0000 | -0.13 | 0.0000 | -0.04 | 0.0000 | 0.03 | 0.0000 | -0.06 | -0.0001 | -0.17 | 0.0000 | -0.09 | 0.0000 | -0.07 | -0.0001 | -0.18 |
| EP | 0.0008 | 0.06 | 0.0008 | 0.06 | 0.0009 | 0.07 | 0.0016 | 0.12 | 0.0011 | 0.08 | 0.0009 | 0.07 | 0.0010 | 0.07 | 0.0012 | 0.09 |
| Turnover | -0.0011 | -1.70 | -0.0010 | -1.65 | -0.0010 | -1.61 | -0.0010 | -1.66 | -0.0011 | -1.71 | -0.0010 | -1.68 | -0.0010 | -1.66 | -0.0011 | -1.72 |
| Intercept | 0.0048 | 0.64 | 0.0041 | 0.55 | 0.0037 | 0.51 | 0.0043 | 0.58 | 0.0051 | 0.67 | 0.0045 | 0.61 | 0.0045 | 0.60 | 0.0051 | 0.68 |
| Interquartile | 0.2851 | | 0.2288 | | 0.2136 | | 0.2939 | | 0.3825 | | 0.3024 | | 0.2904 | | 0.3280 | |
| ReturnDiff | -0.0425% | | -0.0732% | | -0.1156% | | -0.0761% | | 0.0251% | | -0.0130% | | -0.0619% | | 0.0229% | |

**Table X. Robustness and Predicting Aggregate Market Returns**

This Table reports robustness results by adding more filters. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on the Shanghai Stock Exchange with at least 15 trading days in the previous month. Panel A and Panel B report order imbalance determinants and return prediction by including all filters in Liu Stambaugh and Yuan (2019). Panel C and Panel D report order imbalance determinants and return prediction by excluding leveraged trading, consisting of margin buys, short sales and collateral trading. We report the key independent variables. Other variables are the same as those in Table II and Table III; those coefficients are not reported. To account for serial correlation in the coefficients, the standard errors of the time-series are adjusted using Newey-West (1987) with 5 lags.

Panel A. Order Imbalance Determinants, including all filters from Liu Stambaugh and Yuan (2019)

| Dep.var | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |
| Ret(-1) | 0.4227 | 11.31 | 0.7823 | 23.78 | 0.5479 | 18.66 | -0.2450 | -7.83 | -1.3345 | -34.61 | -1.8080 | -21.12 | -0.6514 | -8.85 |
| Ret(-6,-2) | -0.5207 | -26.78 | -0.2857 | -18.59 | -0.1379 | -10.28 | -0.1046 | -6.76 | -0.1041 | -6.45 | 0.2090 | 5.75 | -0.4100 | -11.64 |
| Ret(-27,-7) | -0.0464 | -6.22 | -0.0195 | -3.29 | -0.0104 | -2.18 | -0.0310 | -5.67 | -0.0529 | -7.22 | 0.0191 | 1.15 | -0.0996 | -5.49 |
| Oib(-1) | 0.2173 | 43.99 | 0.2176 | 48.25 | 0.1914 | 48.10 | 0.0612 | 22.53 | 0.1131 | 37.29 | 0.2551 | 70.96 | 0.1102 | 31.02 |

Panel B. Cross-sectional Return Predictions, including all filters from Liu Stambaugh and Yuan (2019)

| Dep.var | Ret | | Ret | | Ret | | Ret | | Ret | | Ret | | Ret | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Oib.var | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
| | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |
| Oib(-1) | -0.0075 | -24.16 | -0.0071 | -20.81 | -0.0052 | -16.40 | -0.0011 | -7.91 | 0.0004 | 4.49 | 0.0017 | 17.88 | 0.0000 | -0.05 |
| Ret Diff | -0.1720% | | -0.1373% | | -0.0912% | | -0.0299% | | 0.0183% | | 0.1078% | | -0.0003% | |

Panel C. Order Imbalance Determinants, excluding Leveraged Trading: Margin trading, Short Sells and Collateral Stock Trading

| Dep.var | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ret(-1) | 0.5533 | 16.70 | 0.7984 | 31.78 | 0.5268 | 24.35 | -0.3157 | -13.66 | -1.3479 | -35.20 | -2.2570 | -29.44 | -0.7094 | -13.56 |
| Ret(-6,-2) | -0.4178 | -27.32 | -0.2163 | -19.09 | -0.1157 | -11.11 | -0.1206 | -9.05 | -0.1914 | -12.16 | 0.0254 | 0.81 | -0.3756 | -14.17 |
| Ret(-27,-7) | -0.0426 | -6.96 | -0.0173 | -3.62 | -0.0101 | -2.47 | -0.0308 | -6.08 | -0.0724 | -9.11 | -0.0126 | -0.89 | -0.0576 | -4.61 |
| Oib(-1) | 0.1839 | 39.28 | 0.1942 | 47.93 | 0.1571 | 44.51 | 0.0335 | 12.92 | 0.0632 | 10.89 | 0.2228 | 63.95 | 0.0853 | 31.17 |

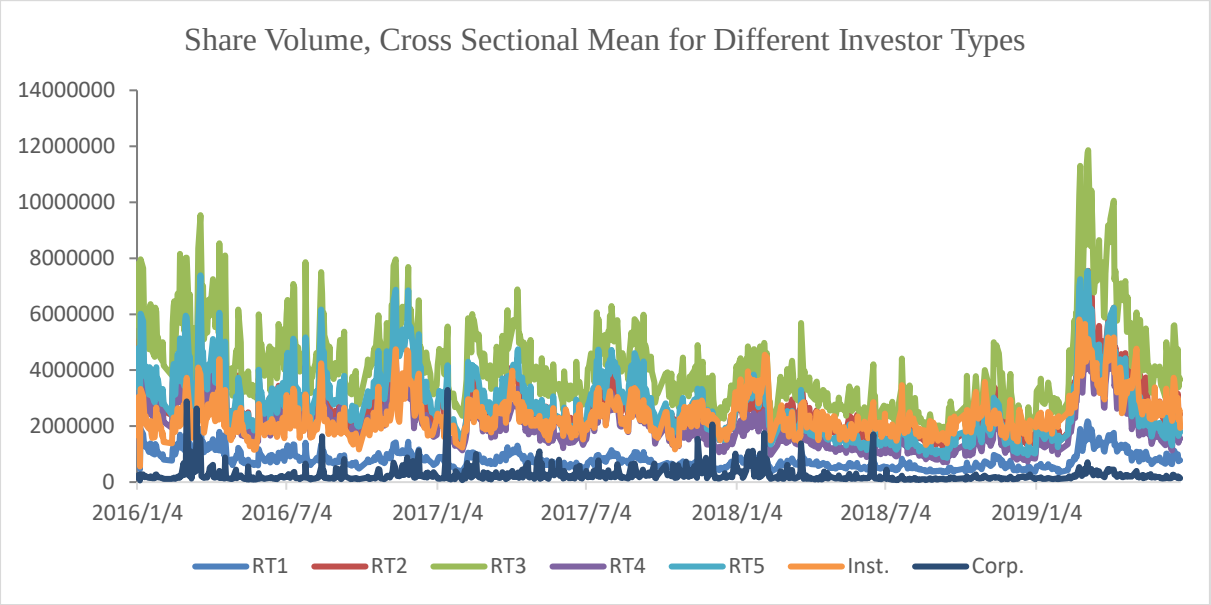
Panel D. Cross-sectional Return Predictions, excluding Leveraged Trading: Margin trading, Short Sells and Collateral Stock Trading

| Dep.var | Ret | | Ret | | Ret | | Ret | | Ret | | Ret | | Ret | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Oib.var | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
| | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat | Coef. | t-stat |
| Oib(-1) | -0.0092 | -24.62 | -0.0088 | -21.97 | -0.0056 | -17.20 | -0.0003 | -3.43 | 0.0008 | 11.25 | 0.0016 | 20.21 | 0.0000 | -0.83 |
| Ret Diff | -0.2037% | | -0.1644% | | -0.1064% | | -0.0116% | | 0.0481% | | 0.1058% | | -0.0030% | |

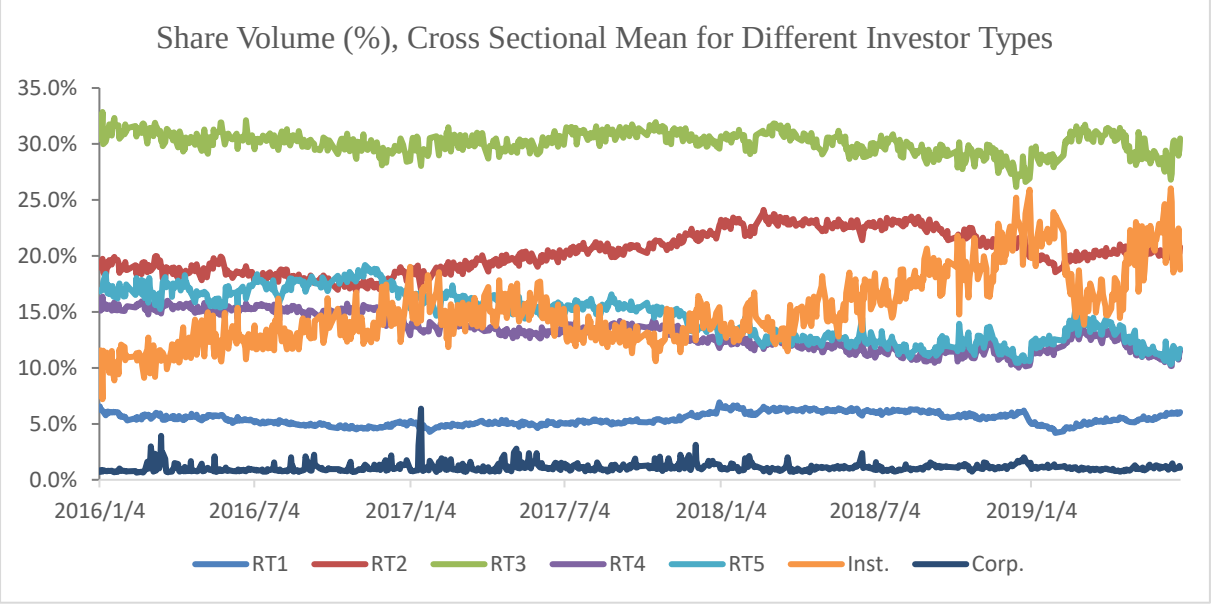
**Figure I. Different Investor Type Order Flows between Jan 2016 and Jun 2019**

There figures report the time series plot of the cross sectional mean for different types of investor trading activity from January 2016 to Jun 2019. Our sample firms are A-share stocks listed on the Shanghai Stock Exchange. In Panel A, we show time series plots of the average buy volume and sell volume by each type of investor. In Panel B, we present the volume percentage by each type of investor. In Panel C, we show the shares held by each type of investor.

Panel A. Share Volume




Panel B. Share Volume (%)



Panel C. Shares Held (%)

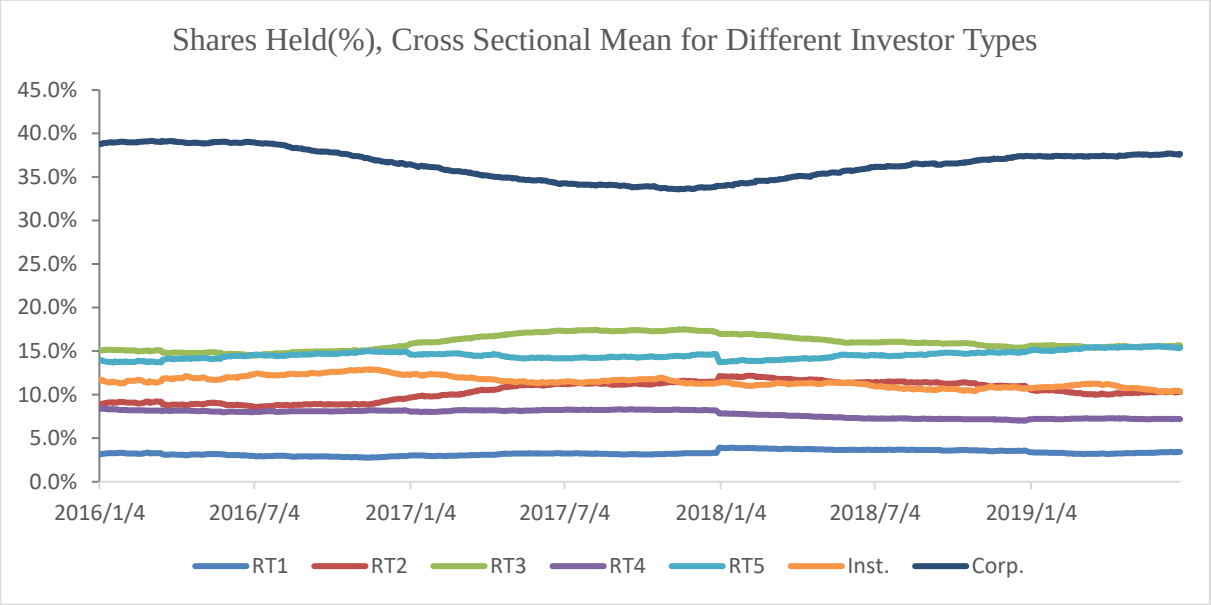
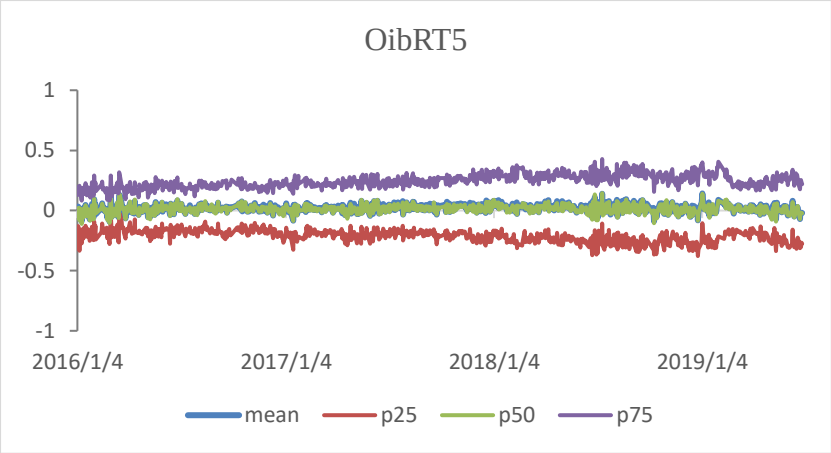
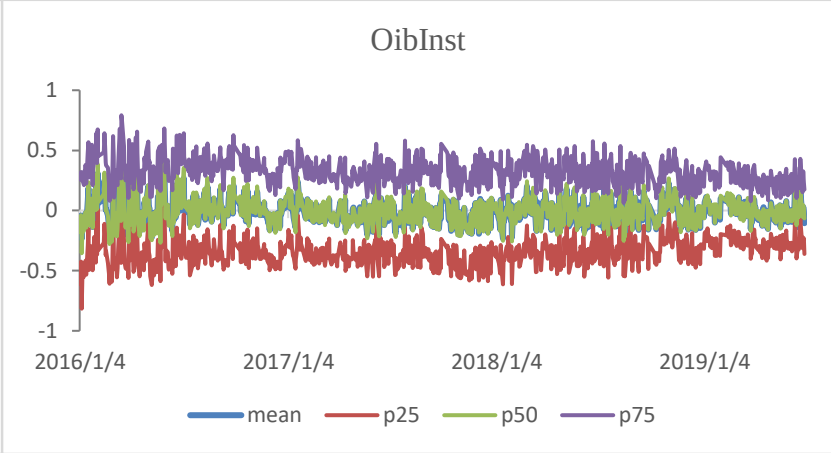


**Figure II. Time Series of Different Types of Investor Order Imbalance**

These figures reports time series of different types of investor trading activity. Our sample period covers January 2016 to June 2019, and our firms are A-share stocks listed on the Shanghai Stock Exchange. We present the cross-sectional mean, median, 25$^{th}$ percentiles and 75$^{th}$ percentiles of scaled daily order imbalances by each investor group each day. Order imbalance is computed as the buy share volume minus sell share volume divided by buy share volume plus sell share volume for each investor group.


OibRT5
1
0.5
0
-0.5
-1
2016/1/4
2017/1/4
2018/1/4
2019/1/4
mean
p25
p50
p75


OibInst
1
0.5
0
-0.5
-1
2016/1/4
2017/1/4
2018/1/4
2019/1/4
mean
p25
p50
p75

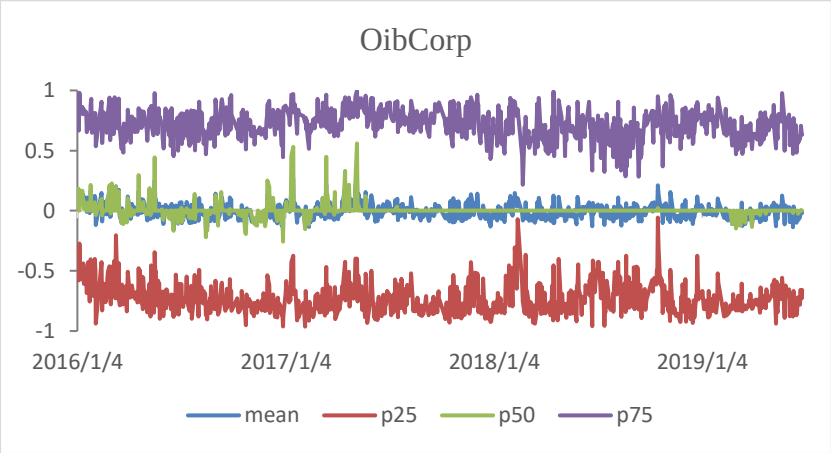
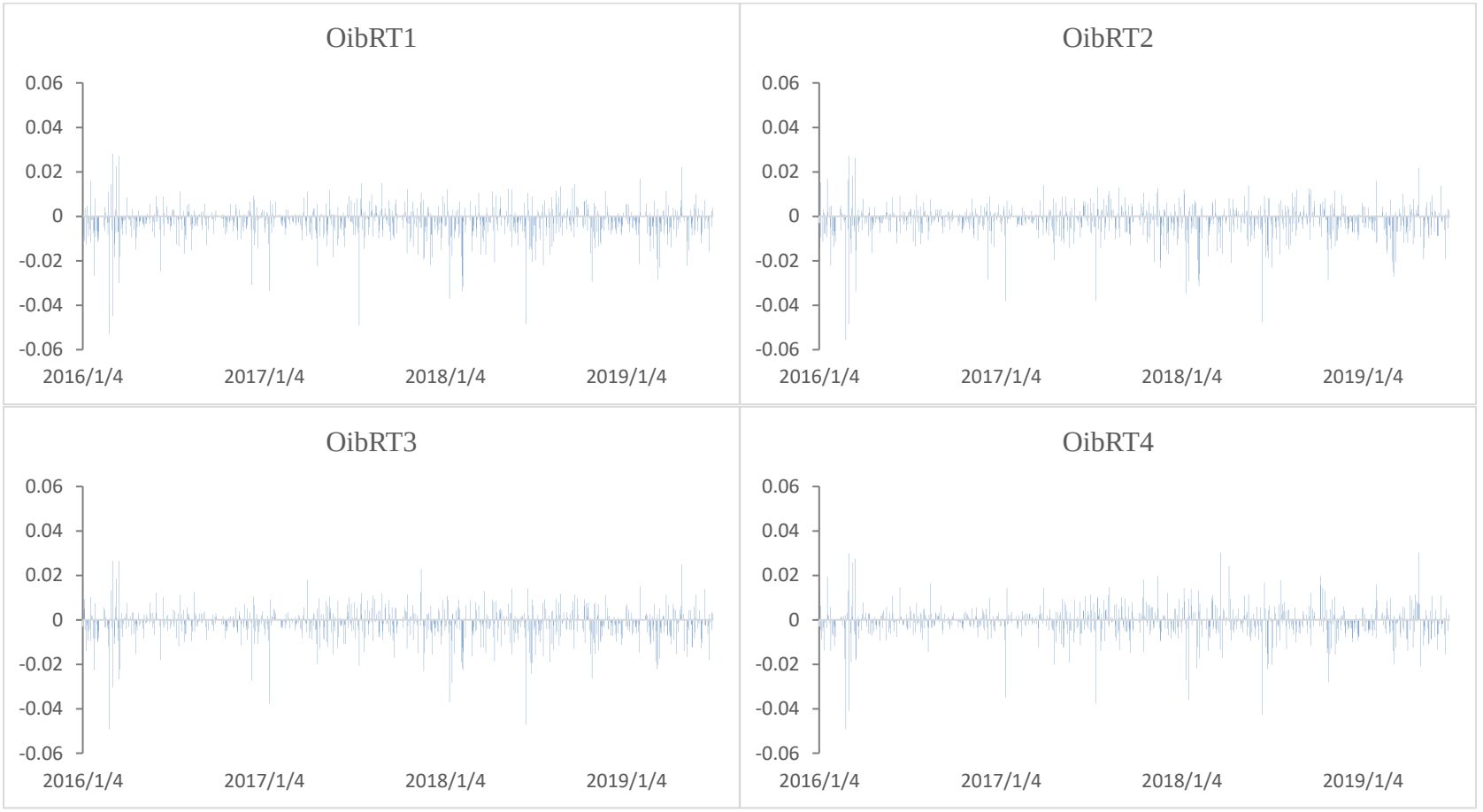
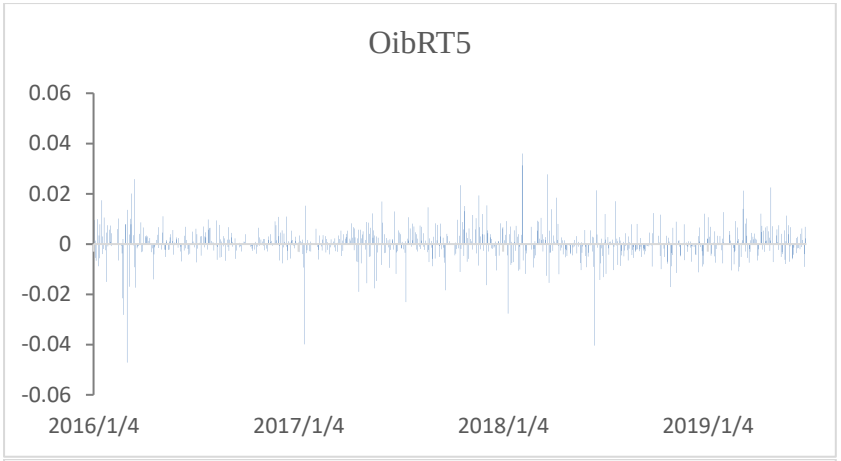
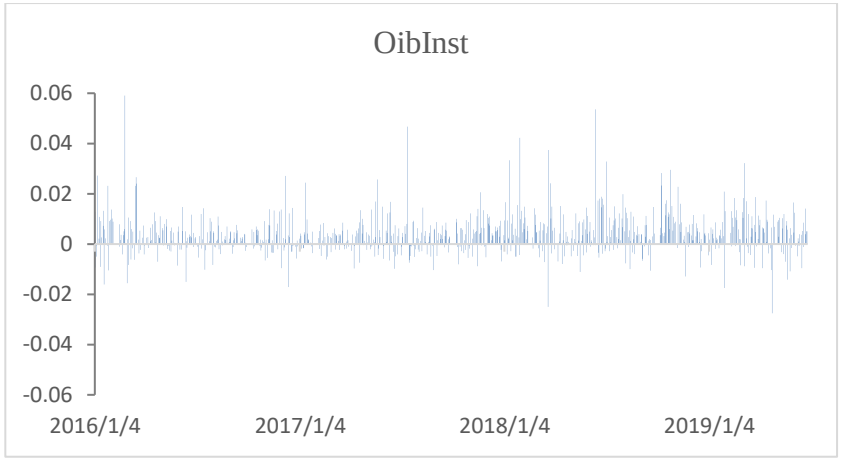

OibCorp
1
0.5
0
-0.5
-1
2016/1/4
2017/1/4
2018/1/4
2019/1/4
mean
p25
p50
p75

**Figure III. Portfolio Return Difference Using Previous Day Order Imbalances**

These figures plot daily value-weighted portfolio return differences between quintile 5 and quintile 1, where stocks are sorted on the previous-day order imbalance measures within each investor groups. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on Shanghai Stock Exchange with at least 15 trading days in the previous month. Order imbalance is computed as the buy share volume minus sell share volume divided by buy share volume plus sell share volume for each investor group.




OibRT5
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
2016/1/4
2017/1/4
2018/1/4
2019/1/4


OibInst
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
2016/1/4
2017/1/4
2018/1/4
2019/1/4

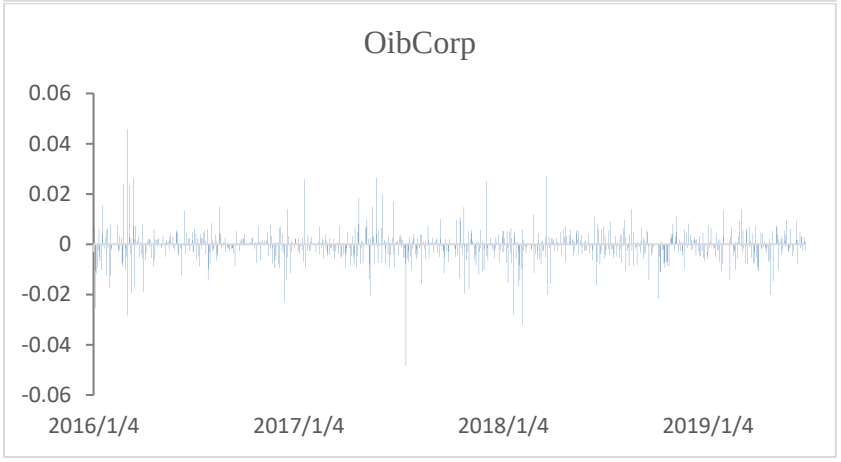
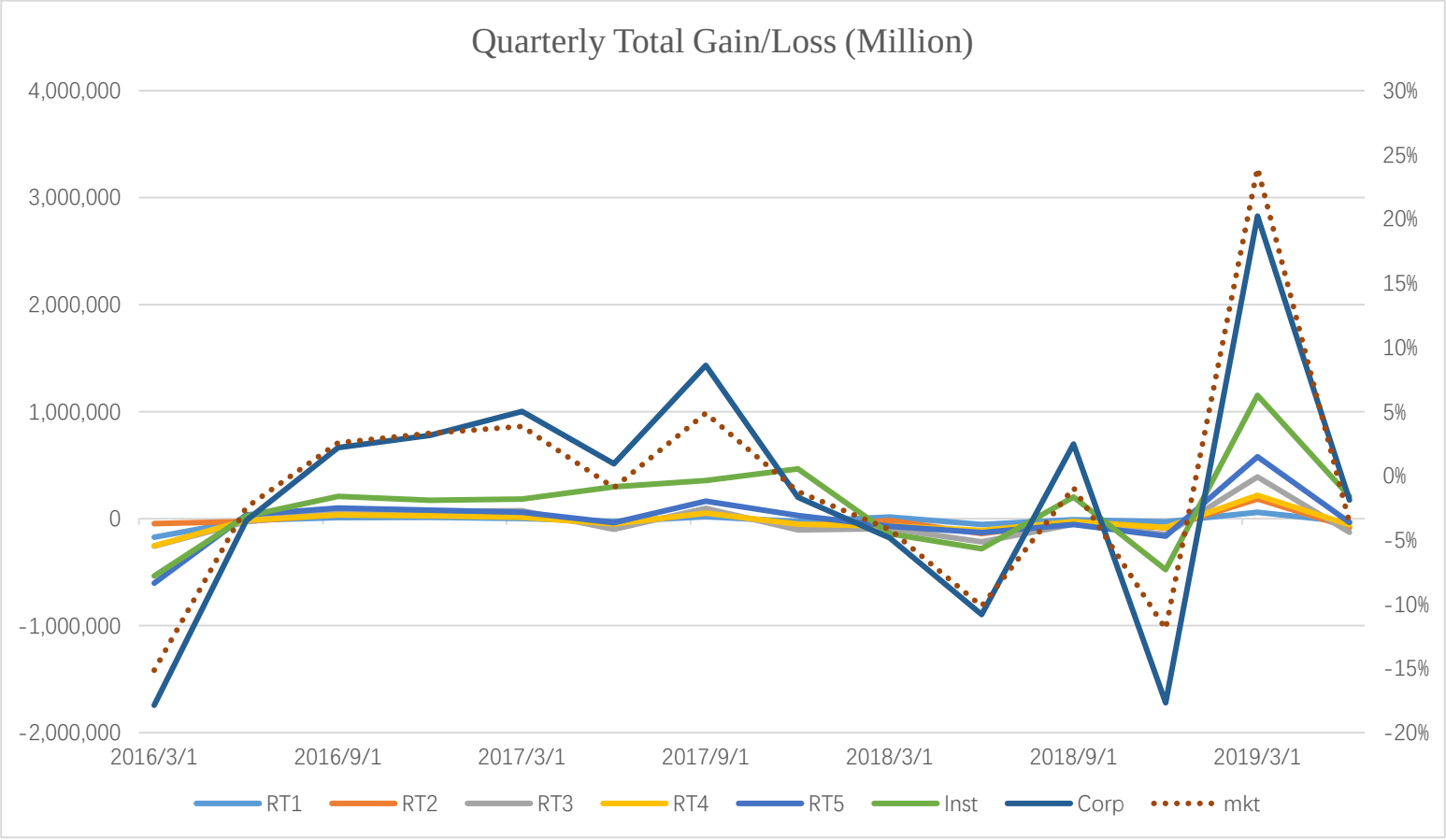

OibCorp
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
2016/1/4
2017/1/4
2018/1/4
2019/1/4

**Figure IV. Quarterly Total Gain/loss by Different Investor Groups**

This figure reports quarterly gain/loss of different types of investor. Our sample period covers January 2016 to June 2019, and our firms are A-share stocks listed on the Shanghai Stock Exchange.



**Appendix Table I. Order Imbalance Correlation with Returns**

This Table reports time series averages of daily cross sectional order imbalance and stock return correlations. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on the Shanghai Stock Exchange with at least 15 days with trading in the previous month. Order imbalance is computed as the buy share volume minus sell share volume divided by buy share volume plus sell share volume for the specified investor group.

| | OibRT1 | OibRT2 | OibRT3 | OibRT4 | OibRT5 | OibInst | OibCorp |
|---|---|---|---|---|---|---|---|
| Ret | -0.327 | -0.449 | -0.422 | -0.152 | 0.096 | 0.158 | -0.059 |
| Lag Ret(-1) | -0.068 | -0.096 | -0.098 | -0.049 | -0.039 | -0.065 | -0.024 |
| Lag Ret(-6,-2) | -0.149 | -0.150 | -0.124 | -0.040 | -0.003 | -0.063 | -0.034 |
| Lag Ret(-27,-7) | -0.014 | -0.007 | -0.004 | -0.006 | -0.012 | 0.003 | -0.005 |

**Appendix Table II. Long Short Portfolios**

This table reports portfolio returns using a long-short strategy wherein we buy the stocks in the highest quintile of the relevant group's order imbalance, and short the stocks in the lowest order imbalance quintile. The order imbalance is computed as the buy share volume minus sell share volume divided by buy share volume plus sell share volume for the specified investor group from the previous day (day 0). Portfolio returns are value-weighted by previous month-end market cap. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on the Shanghai Stock Exchange with at least 15 days with trades in the previous month. We report the raw returns in Panel A and the risk-adjusted returns of subgroup results in Panel B using the Liu Stambaugh and Yuan (2019) three factor model, with a holding period of 20 days in Panel B. As our data are overlapping, we adjust the standard errors of the portfolio return time-series using Newey-West (1987) with twice the horizon of the longest overlapping lags.

Panel A. Raw Return

| Holding | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Period (days) | Mean | t-stat | Mean | t-stat | Mean | t-stat | Mean | t-stat | Mean | t-stat | Mean | t-stat | Mean | t-stat |
| 1 | -0.0030 | -11.22 | -0.0024 | -8.92 | -0.0020 | -7.83 | -0.0010 | -4.28 | 0.0002 | 1.00 | 0.0036 | 13.74 | -0.0008 | -3.76 |
| 5 | -0.0053 | -7.24 | -0.0040 | -6.08 | -0.0026 | -4.19 | -0.0006 | -1.02 | 0.0008 | 1.50 | 0.0068 | 9.06 | -0.0012 | -2.17 |
| 10 | -0.0075 | -7.20 | -0.0060 | -6.63 | -0.0041 | -4.86 | -0.0015 | -1.75 | 0.0009 | 1.21 | 0.0102 | 10.06 | -0.0031 | -3.38 |
| 20 | -0.0081 | -5.22 | -0.0060 | -4.42 | -0.0043 | -3.49 | -0.0018 | -1.51 | 0.0014 | 1.28 | 0.0108 | 7.12 | -0.0038 | -2.84 |
| 30 | -0.0085 | -4.45 | -0.0060 | -3.54 | -0.0040 | -2.61 | -0.0015 | -1.05 | 0.0011 | 0.81 | 0.0111 | 6.07 | -0.0041 | -2.72 |
| 40 | -0.0087 | -3.97 | -0.0055 | -2.83 | -0.0031 | -1.75 | -0.0010 | -0.59 | 0.0015 | 0.95 | 0.0135 | 6.17 | -0.0050 | -2.99 |
| 50 | -0.0086 | -3.73 | -0.0049 | -2.29 | -0.0027 | -1.35 | -0.0006 | -0.32 | 0.0024 | 1.44 | 0.0129 | 5.42 | -0.0069 | -3.52 |
| 60 | -0.0104 | -4.12 | -0.0054 | -2.28 | -0.0028 | -1.31 | -0.0014 | -0.66 | 0.0025 | 1.31 | 0.0131 | 5.42 | -0.0086 | -4.05 |

Panel B. alphas for different stock subgroups, holding period of 20 days

| | OibRT1 | | OibRT2 | | OibRT3 | | OibRT4 | | OibRT5 | | OibInst | | OibCorp | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Alpha | t-stat | Alpha | t-stat | Alpha | t-stat | Alpha | t-stat | Alpha | t-stat | Alpha | t-stat | Alpha | t-stat |
| Size groups | | | | | | | | | | | | | | |
| Small | -0.0131 | -8.87 | -0.0114 | -7.02 | -0.0042 | -2.78 | 0.0005 | 0.58 | 0.0076 | 6.76 | 0.0081 | 8.20 | 0.0000 | -0.04 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Medium | -0.0143 | -8.18 | -0.0109 | -7.10 | -0.0059 | -5.05 | -0.0002 | -0.29 | 0.0046 | 6.70 | 0.0097 | 6.22 | -0.0005 | -0.82 |
| Large | -0.0077 | -3.50 | -0.0071 | -3.72 | -0.0053 | -3.52 | -0.0019 | -1.54 | 0.0003 | 0.19 | 0.0108 | 5.09 | -0.0033 | -2.49 |
| Turnover Groups | | | | | | | | | | | | | | |
| Low | -0.0073 | -3.45 | -0.0063 | -3.62 | -0.0046 | -3.31 | -0.0016 | -1.45 | -0.0005 | -0.33 | 0.0106 | 5.17 | -0.0012 | -0.81 |
| Medium | -0.0088 | -3.03 | -0.0077 | -3.31 | -0.0034 | -1.96 | 0.0001 | 0.03 | 0.0040 | 2.56 | 0.0085 | 4.06 | -0.0003 | -0.26 |
| High | -0.0161 | -4.71 | -0.0148 | -5.18 | -0.0108 | -4.52 | -0.0012 | -0.66 | 0.0102 | 5.21 | 0.0132 | 5.22 | -0.0052 | -2.87 |
| Share Price Groups | | | | | | | | | | | | | | |
| Low | -0.0058 | -1.75 | -0.0045 | -1.75 | -0.0033 | -1.37 | -0.0021 | -1.34 | 0.0016 | 1.03 | 0.0062 | 3.03 | 0.0002 | 0.17 |
| Medium | -0.0086 | -3.86 | -0.0079 | -3.88 | -0.0047 | -2.68 | 0.0010 | 0.58 | 0.0039 | 2.42 | 0.0108 | 3.99 | -0.0028 | -1.59 |
| High | -0.0085 | -3.07 | -0.0067 | -3.05 | -0.0043 | -2.43 | -0.0006 | -0.42 | 0.0036 | 1.79 | 0.0149 | 4.97 | -0.0061 | -3.42 |

## Appendix III. CFND news dataset

Our news data is from the Financial News Database of Chinese Listed Companies (CFND), which includes news on all A-share stocks from over 400 internet media and over 600 newspapers. The internet sources trace back to 2002 and have over 8.3 million pieces of A-share stock news. The newspaper sources trace back to 1994 and have over 2.6 million pieces of A-share stock news. The main sources of the database are listed in Appendix Table V Panel A. CFND uses a supervised machine learning algorithm (Support Vector Machine) to label each piece of news as positive (+1), negative (-1), or neutral (0). The accuracy of their algorithm on the test sample of 23,970 pieces of news is over 85%.

We include news from both internet and newspaper sources to aggregate the public information and merge the news data with our trading data. Summary statistics are reported in Panel B. There are 369,217 firm-days (35% of the total sample) with public news. On a typical firm-news day, there are total of 8 pieces of news, with 3.53, 2.37, and 2.29 pieces of positive, negative and neutral news, respectively. For each stock each day, we compute the net sentiment score by averaging the sentiment scores from all news on that firm that day. This average net sentiment score can be interpreted as a scaled measure of net positive news the firm receives during that day. In Panel C, we estimate the correlation between the net sentiment score, the order imbalance of each investor group and the daily stock return. The less wealthy retail investor group trades, RT1-RT4, are negatively correlated with public news, while the largest retail group, RT5, and institutional trades are positively correlated with public news. The net sentiment score is positively correlated with the stock return,.

**Appendix Table III. CFND News Dataset Summary**

This table reports the summary statistics of the CFND news dataset. Panel A reports the main internet and newspaper sources used in CFND. Panel B reports the summary statistics of the CFND news dataset from newspapers, the internet and aggregated across both sources. *Newsday* is the number of firm-days with public news, *Coverage (%)* is the percentage of News-days scaled by the total number of firm-days. *#News*, *#Positive*, *#Negative*, *#Neutral* is the number of total, positive, negative and neutral news on a typical firm-day with public news. The net sentiment score (*Sent*) is computed by averaging the sentiment scores from all news on a firm-day with news. Panel C reports the time series average of cross sectional order imbalance and news correlations. Our sample period covers January 2016 to June 2019, and our sample firms are A-share stocks listed on the Shanghai Stock Exchange with at least 15 days with trading in the previous month. Order imbalance is computed as the buy share volume minus sell share volume divided by buy share volume plus sell share volume for the specified investor group.

Panel A. Main sources of internet media and newspaper used in CFND

| Internet | Translated name | Website | Newspaper | Translated name |
|---|---|---|---|---|
| 和讯网 | Hexun Net | www.hexun.com | 中国证券报 | China Securities Journal |
| 新浪财经 | Sina Finance | finance.sina.com.cn | 上海证券报 | Shanghai Securities News |
| 东方财富网 | East Money Net | www.eastmoney.com | 第一财经日报 | China Business News |
| 腾讯财经 | Tencent Finance | finance.qq.com | 中国经营报 | China Economic Times |
| 网易财经 | NetEase Finance | money.163.com | 经济观察报 | Economic Observer |
| 凤凰财经 | Phoenix Finance | finance.ifeng.com | 证券日报 | Security Daily |
| 中国经济网 | China Economic Net | www.ce.cn | 证券时报 | Security Times |
| 搜狐财经 | Sohu Finance | business.sohu.com | 21 世纪经济报道 | 21st Century Business Herald |
| 华讯财经 | Huaxun Finance | www.591hx.com | | |
| FT 中文网 | Financial Times Chinese | m.ftchinese.com | | |
| 全景网 | Panorama Net | www.p5w.net | | |
| 中国证券网 | Shanghai Securities Net | www.cnstock.com | | |

| | | |
|---|---|---|
| 证券之星 | Stock Star | www.stockstar.com |
| 财新网 | Caixin Net | www.caixin.com |
| 澎湃新闻网 | The Paper | www.thepaper.cn |
| 第一财经 | China Business Net | m.yicai.com |
| 财经网 | Caijing Net | www.caijing.com |
| 金融界 | Chinan Finance Online | www.jrj.com.cn |

Panel B. Summary Statistics

| | News day | Coverage(%) | #News | #Positive | #Negative | #Neutral | Sent |
|---|---|---|---|---|---|---|---|
| Newspaper | 177,150 | 16.8% | 3.96 | 1.79 | 0.84 | 1.35 | 0.32 |
| Internet | 312,157 | 29.6% | 7.32 | 3.15 | 2.33 | 1.95 | 0.18 |
| Both | 369,217 | 35.1% | 8.09 | 3.53 | 2.37 | 2.29 | 0.22 |

Panel C. Order Imbalance Correlation with news sentiment

| | OibRT1 | OibRT2 | OibRT3 | OibRT4 | OibRT5 | OibInst | OibCorp | Ret |
|---|---|---|---|---|---|---|---|---|
| Newspaper Sent | -0.0064 | -0.0061 | -0.0029 | -0.0004 | 0.0003 | 0.0012 | -0.0024 | 0.0050 |
| Internet Sent | -0.0439 | -0.0573 | -0.0536 | -0.0241 | 0.0150 | 0.0206 | -0.0177 | 0.1587 |
| Both sent | -0.0401 | -0.0504 | -0.0457 | -0.0192 | 0.0138 | 0.0182 | -0.0142 | 0.1394 |

## Appendix IV. Gain and Loss for different investor groups

### 1. Gain and Loss per investor, per stock and per day

At the end of day *t*, investor *i* holds stock *s*: $shares_{i,s,t}^{H}$, the closing price is $p_{s,t}^{C}$.

At the end of day *t+1*, investor *i* holds stock *s*: $shares_{i,s,t+1}^{H}$, the closing price is $p_{s,t+1}^{C}$.

**Condition 0 (Benchmark)**: On day *t+1*, the investor *i* does not trade.

At the end of day *t+1*, investor *i* holds stock *s*: $shares_{i,s,t+1}^{H} = shares_{i,s,t}^{H}$

The daily gain of investor *i* on stock *s* from day *t* to day *t+1* is:

$$\text{Gain}_{i,s,t,t+1} = p_{s,t+1}^{C} shares_{i,s,t+1}^{H} - p_{s,t}^{C} shares_{i,s,t}^{H} \quad \text{(A.1)}$$

We can decompose the gain/loss into a market timing part and a stock selection part. The expected return of stock *s* on day *t* from the capital asset pricing model is $\text{R}_{\text{st}}^{expected} = [\text{R}_{ft} + \beta_{\text{s}}(R_{mt} - R_{ft})]$, where $\text{R}_{ft}$ and $R_{mt}$ is the risk free rate and market return of day *t*, and $\beta_{\text{s}}$ is estimated from the previous month for stock *s*. The gain/loss from market timing and stock selection can be represented as:

$$\text{Gain}_{i,s,t,t+1}^{market\ timing} = p_{s,t}^{C} shares_{i,s,t}^{H} \text{R}_{\text{st}}^{expected} \quad \text{(A.2)}$$

$$\text{Gain}_{\text{i,s,t,t+1}}^{stock\ selection} = p_{s,t}^{C} shares_{i,s,t}^{H} \left( \frac{p_{s,t+1}^{C}}{p_{s,t}^{C}} - 1 - \text{R}_{\text{st}}^{expected} \right) \quad \text{(A.3)}$$

There are four trading types. For a long side trade, the investor buys to open the position and then sells to close the position. For a short side trade, the investor sells to open the position and then buys to close the position. We discuss the daily gain for each type of trade from Condition 1 to Condition 4.

**Condition 1:** On day *t+1*, the investor *i* buys to open or increases a long position in stock *s* using $N_1$ trades. Each buy to open trade is to buy $shares_{i,s,t+1,n}^{BO}$ at price $p_{i,s,t+1,n}^{BO}$, where $n \in \{1, \dots, N_1\}$

At the end of day *t+1*, investor *i* holds $shares_{i,s,t+1}^{H}$ of stock *s*, which can be represented as:

$$shares_{i,s,t+1}^{H} = shares_{i,s,t}^{H} + \sum_{N_1} shares_{i,s,t+1,n}^{BO} \tag{A.4}$$

The daily gain of investor *i* on stock *s* from day *t* to day *t+1* is:

$$\text{Gain}_{i,s,t,t+1} = p_{s,t+1}^{C} shares_{i,s,t+1}^{H} - p_{s,t}^{C} shares_{i,s,t}^{H} - \sum_{N_1} p_{i,s,t+1,n}^{BO} shares_{i,s,t+1,n}^{BO} \tag{A.5}$$

If we put (A.4) into (A.5), the daily gain can also be represented as:

$$\begin{aligned} &\text{Gain}_{i,s,t,t+1} \\ &= p_{s,t+1}^{C}\left(shares_{i,s,t}^{H} + \sum_{N_1} shares_{i,s,t+1,n}^{BO}\right) - p_{s,t}^{C} shares_{i,s,t}^{H} \\ &- \sum_{N_1} p_{i,s,t+1,n}^{BO} shares_{i,s,t+1,n}^{BO} \\ &= \left(p_{s,t+1}^{C} - p_{s,t}^{C}\right) shares_{i,s,t}^{H} + \left(p_{s,t+1}^{C} \sum_{N_1} shares_{i,s,t+1,n}^{BO} - \sum_{N_1} p_{i,s,t+1,n}^{BO} shares_{i,s,t+1,n}^{BO}\right) \end{aligned} \tag{A.6}$$

We can decompose the gain/loss into a market timing part and a stock selection part. The expected return of stock *s* on day *t*, from the capital asset pricing model is $\text{R}_{\text{st}}^{expected} = [\text{R}_{ft} + \beta_{\text{s}}(R_{mt} - R_{ft})]$, where $\text{R}_{ft}$ and $R_{mt}$ is the risk free rate and market return of day *t* and $\beta_{\text{s}}$ is estimated from the previous month for stock *s*. The gain/loss from market timing and stock selection can be represented as:

$$\text{Gain}_{i,s,t,t+1}^{market\ timing} = \left(p_{s,t}^{C} shares_{i,s,t}^{H} + \sum_{N_1} p_{i,s,t+1,n}^{BO} shares_{i,s,t+1,n}^{BO}\right) \times \text{R}_{\text{st}}^{expected} \tag{A.7}$$

$$\begin{aligned}&\mathrm{Gain}_{\mathrm{i,s,t,t+1}}^{stock\ \ selection} \\ &= p_{s,t}^{C} shares_{i,s,t}^{H}\left(\frac{p_{s,t+1}^{C}}{p_{s,t}^{C}} - 1 - \mathrm{R}_{\mathrm{st}}^{expected}\right) \\ &+ \sum_{N_1} p_{i,s,t+1,n}^{BO} shares_{i,s,t+1,n}^{BO}\left(\frac{p_{s,t+1}^{C}}{p_{i,s,t+1,n}^{BO}} - 1 - \mathrm{R}_{\mathrm{st}}^{expected}\right)\end{aligned} \tag{A.8}$$

**Condition 2:** On day *t+1*, investor *i* sells to close or reduces a long position in stock *s* using $N_2$ trades. Each sale trade sells $shares_{i,s,t+1,n}^{SC}$ at price $p_{i,s,t+1,n}^{SC}$, where $n \in \{1, \dots, N_2\}$

At the end of day *t+1*, investor *i* holds stock *s*: $shares_{i,s,t+1}^{H}$, can be represented as:

$$shares_{i,s,t+1}^{H} = shares_{i,s,t}^{H} - \sum_{N_2} shares_{i,s,t+1,n}^{SC} \tag{A.9}$$

The daily gain of investor *i* on stock *s* from day *t* to day *t+1* in this case is:

$$\mathrm{Gain}_{i,s,t,t+1} = p_{s,t+1}^{C} shares_{i,s,t+1}^{H} - p_{s,t}^{C} shares_{i,s,t}^{H} + \sum_{N_2} p_{i,s,t+1,n}^{SC} shares_{i,s,t+1,n}^{SC} \tag{A.10}$$

If we put (A.5) into (A.6), the daily gain can also be represented as:

$$\begin{aligned}&\mathrm{Gain}_{i,s,t,t+1} \\ &= p_{s,t+1}^{C}\left(shares_{i,s,t}^{H} - \sum_{N_2} shares_{i,s,t+1,n}^{SC}\right) - p_{s,t}^{C} shares_{i,s,t}^{H} \\ &+ \sum_{N_2} p_{i,s,t+1,n}^{SC} shares_{i,s,t+1,n}^{SC} \\ &= \left(p_{s,t+1}^{C} - p_{s,t}^{C}\right)\left(shares_{i,s,t}^{H} - \sum_{N_2} shares_{i,s,t+1,n}^{SC}\right) \\ &+ \left(\sum_{N_2} p_{i,s,t+1,n}^{SC} shares_{i,s,t+1,n}^{SC} - p_{s,t}^{C} \sum_{N_2} shares_{i,s,t+1,n}^{SC}\right)\end{aligned} \tag{A.11}$$

Again we can decompose the gain/loss into a market timing part and a stock selection part. The expected return of stock *s* on day *t*, from the capital asset pricing model is ($\mathrm{R}_{\mathrm{st}}^{expected} = [\mathrm{R}_{ft} +$

$\beta_s(R_{mt} - R_{ft})]$), where $R_{ft}$ and $R_{mt}$ is the risk free rate and market return on day *t* and $\beta_s$ is estimated from the previous month for stock *s*. The gain/loss from market timing and stock selection can be represented as:

$$\text{Gain}_{i,s,t,t+1}^{market\ timing} = (p_{s,t}^{C} shares_{i,s,t}^{H} - \sum_{N_2} p_{i,s,t+1,n}^{SC} shares_{i,s,t+1,n}^{SC}) \times \text{R}_{\text{st}}^{expected} \quad \text{(A.12)}$$

$$\text{Gain}_{\text{i,s,t,t+1}}^{stock\ selection} \quad \text{(A.13)}$$

$$= p_{s,t}^{C} shares_{i,s,t}^{H} \left( \frac{p_{s,t+1}^{C}}{p_{s,t}^{C}} - 1 - \text{R}_{\text{st}}^{expected} \right)$$

$$- \sum_{N_2} p_{i,s,t+1,n}^{SC} shares_{i,s,t+1,n}^{SC} \left( \frac{p_{s,t+1}^{C}}{p_{i,s,t+1,n}^{SC}} - 1 - \text{R}_{\text{st}}^{expected} \right)$$

**Condition 3:** On day *t+1*, investor *i* sells to open or increases a short position in stock *s* using $N_3$ trades. Each sale trade sells $shares_{i,s,t+1,n}^{SO}$ at price $p_{i,s,t+1,n}^{SO}$, where $n \in \{1, \dots, N_3\}$

At the end of day *t+1*, investor *i* holds a short position of $shares_{i,s,t+1}^{H}$ in stock *s*, which can be represented as:

$$shares_{i,s,t+1}^{H} = shares_{i,s,t}^{H} - \sum_{N_3} shares_{i,s,t+1,n}^{SO} \quad \text{(A.14)}$$

The daily gain of investor *i* on stock *s* from day *t* to day *t+1* is:

$$Gain_{i,s,t,t+1} = p_{s,t+1}^{C} shares_{i,s,t+1}^{H} - p_{s,t}^{C} shares_{i,s,t}^{H} + \sum_{N_3} p_{i,s,t+1,n}^{SO} shares_{i,s,t+1,n}^{SO} \quad \text{(A.15)}$$

If we put (A.8) into (A.9), the daily gain can also be represented as:

$$
\begin{aligned}
&Gain_{i,s,t,t+1} \\
&= p^{C}_{s,t+1}\left(shares^{H}_{i,s,t} - \sum_{N_3} shares^{SO}_{i,s,t+1,n}\right) - p^{C}_{s,t} shares^{H}_{i,s,t} \\
&+ \sum_{N_3} p^{SO}_{i,s,t+1,n} shares^{SO}_{i,s,t+1,n} \\
&= \left(p^{C}_{s,t+1} - p^{C}_{s,t}\right)\left(shares^{H}_{i,s,t} - \sum_{N_3} shares^{SO}_{i,s,t+1,n}\right) \\
&+ \left(\sum_{N_3} p^{SO}_{i,s,t+1,n} shares^{SO}_{i,s,t+1,n} - p^{C}_{s,t} \sum_{N_3} shares^{SO}_{i,s,t+1,n}\right)
\end{aligned}
\tag{A.16}
$$

We can decompose the gain/loss into market timing part and stock selection part. The expected return of stock *s* on day *t*, from the capital asset pricing model is ($\mathrm{R}_{\mathrm{st}}^{expected} = [\mathrm{R}_{ft} + \beta_{\mathrm{s}}(R_{mt} - R_{ft})]$), where $\mathrm{R}_{ft}$ and $R_{mt}$ is the risk free rate and market return of day *t*, $\beta_{\mathrm{s}}$ is estimated form previous month for stock *s*. The gain/loss from market timing and stock selection can be represented as:

$$
\mathrm{Gain}^{market\ timing}_{i,s,t,t+1} = \left(p^{C}_{s,t} shares^{H}_{i,s,t} - \sum_{N_3} p^{SO}_{i,s,t+1,n} shares^{SO}_{i,s,t+1,n}\right) \times \mathrm{R}^{expected}_{\mathrm{st}} \tag{A.17}
$$

$$
\begin{aligned}
&\mathrm{Gain}^{stock\ selection}_{\mathrm{i,s,t,t+1}} \\
&= p^{C}_{s,t} shares^{H}_{i,s,t}\left(\frac{p^{C}_{s,t+1}}{p^{C}_{s,t}} - 1 - \mathrm{R}^{expected}_{\mathrm{st}}\right) \\
&- \sum_{N_3} p^{SO}_{i,s,t+1,n} shares^{SO}_{i,s,t+1,n}\left(\frac{p^{C}_{s,t+1}}{p^{SO}_{i,s,t+1,n}} - 1 - \mathrm{R}^{expected}_{\mathrm{st}}\right)
\end{aligned}
\tag{A.18}
$$

**Condition 4:** On day *t+1*, investor *i* buys to close or reduces a short position in stock *s* using $\mathrm{N}_4$ trades. Each buy-to-close trade is $\mathrm{shares}^{BC}_{i,s,t+1,n}$ at price $p^{BC}_{i,s,t+1,n}$, where $n \in \{1, \ldots, N_4\}$. At the end of day t+1, investor *i* has a position of $shares^{H}_{i,s,t+1}$ in stock s, which can be represented as

$$shares_{i,s,t+1}^{H} = shares_{i,s,t}^{H} + \sum_{N_4} shares_{i,s,t+1,n}^{BC} \tag{A.19}$$

The daily gain of investor *i* on stock *s* from day *t* to day *t+1* is:

$$Gain_{i,s,t,t+1} = p_{s,t+1}^{C} shares_{i,s,t+1}^{H} - p_{s,t}^{C} shares_{i,s,t}^{H} - \sum_{N_4} p_{i,s,t+1,n}^{BC} shares_{i,s,t+1,n}^{BC} \tag{A.20}$$

If we put (A.11) into (A.12), the daily gain can also be represented as:

$$\begin{aligned} & Gain_{i,s,t,t+1} \\ & = p_{s,t+1}^{C} \left( shares_{i,s,t}^{H} + \sum_{N_4} shares_{i,s,t+1,n}^{BC} \right) - p_{s,t}^{C} shares_{i,s,t}^{H} \\ & - \sum_{N_4} p_{i,s,t+1,n}^{BC} shares_{i,s,t+1,n}^{BC} \\ & = \left(p_{s,t+1}^{C} - p_{s,t}^{C}\right) shares_{i,s,t}^{H} + (p_{s,t+1}^{C} \sum_{N_4} shares_{i,s,t+1,n}^{BC} \\ & - \sum_{N_4} p_{i,s,t+1,n}^{BC} shares_{i,s,t+1,n}^{BC}) \end{aligned} \tag{A.21}$$

We can decompose the gain/loss into market timing part and stock selection part. The expected return of stock *s* on day *t*, from the capital asset pricing model is ($\mathrm{R}_{\mathrm{st}}^{expected} = [\mathrm{R}_{ft} + \beta_{\mathrm{s}}(R_{mt} - R_{ft})]$), where $\mathrm{R}_{ft}$ and $R_{mt}$ is the risk free rate and market return of day *t*, $\beta_{\mathrm{s}}$ is estimated form previous month for stock *s*. The gain/loss from market timing and stock selection can be represented as:

$$\mathrm{Gain}_{i,s,t,t+1}^{market\ timing} = (p_{s,t}^{C} shares_{i,s,t}^{H} + \sum_{N_4} p_{i,s,t+1,n}^{BC} shares_{i,s,t+1,n}^{BC}) \times \mathrm{R}_{\mathrm{st}}^{expected} \tag{A.22}$$

$$\text{Gain}_{\text{i,s,t,t+1}}^{stock\ \ selection} \tag{A.23}$$

$$= p_{s,t}^{C} shares_{i,s,t}^{H} \left( \frac{p_{s,t+1}^{C}}{p_{s,t}^{C}} - 1 - \text{R}_{\text{st}}^{expected} \right)$$

$$+ \sum_{N_4} p_{i,s,t+1,n}^{BC} shares_{i,s,t+1,n}^{BC} \left( \frac{p_{s,t+1}^{C}}{p_{i,s,t+1,n}^{BC}} - 1 - \text{R}_{\text{st}}^{expected} \right)$$

There are three special cases: (1) For a day where the stock goes ex-dividend, we add the amount of the cash dividend directly into the daily gain. (2) For a stock dividend or stock split, the shares and price are both adjusted, and the gain does not change purely due to the stock dividend or split. (3) If investor *i*'s position in stock *s* changes via another type of transaction, including an SEO, converting restricted shares into tradable shares, a block trade, gifted stock and stock delisting, since we don't have the exact cost of these transactions, we conservatively assume the holding cost is the closing price $\text{p}_{\text{s,t+1}}^{\text{C}}$. For an IPO, the cost is the initial shareholders' issue price.

Finally, the trading and holding data we get from the Shanghai Stock Exchange do not record short sell trades and the short positions of investors who initiate the shorts. Instead, these trades are recorded under the investor's broker company account due to regulatory requirements. So we cannot trace back the profits or losses from shorting to the investors who did these shorts. Short sales are still infrequent in Chinese stock markets, with short sale trades accounting for about 0.20% of the daily trading volume, and aggregate short interest is only about 0.02% of tradable shares. So this limitation due to our data is negligible to our gain and loss results.

## 2. Gain and Loss at aggregate level

(1) Gain and Loss for investor group G, per day

Suppose investor *i* holds *S* stocks and investor *i* belongs to group *G*. Since the buy to open and buy to close trade are in the same direction and sell to close and sell to open are in the same direction, we combine them together. The daily gain of investor group *G* from day *t* to day *t+1* is

$$Gain_{G,t,t+1} \tag{A.24}$$
$$= \sum_{G}\sum_{S} p^{C}_{s,t+1} shares^{H}_{i,s,t+1} - \sum_{G}\sum_{S} p^{C}_{s,t} shares^{H}_{i,s,t}$$
$$- \sum_{G}\sum_{S}\sum_{N} p^{B}_{i,s,t+1,n} shares^{B}_{i,s,t+1,n} + \sum_{G}\sum_{S}\sum_{N} p^{S}_{i,s,t+1,n} shares^{S}_{i,s,t+1,n}$$

We can decompose the gain/loss into a market timing part and stock selection part. The expected return of stock *s* on day *t*, from the capital asset pricing model is ($\mathrm{R}_{\mathrm{st}}^{expected} = [\mathrm{R}_{ft} + \beta_{\mathrm{s}}(R_{mt} - R_{ft})]$), where $\mathrm{R}_{ft}$ and $R_{mt}$ is the risk free rate and market return of day *t*, $\beta_{\mathrm{s}}$ is estimated from the previous month for stock *s*. The gain/loss from market timing and stock selection can be represented as:

$$\mathrm{Gain}^{market\ timing}_{i,s,t,t+1} \tag{A.25}$$
$$= (\sum_{G}\sum_{S} p^{C}_{s,t} shares^{H}_{i,s,t} + \sum_{G}\sum_{S}\sum_{N_1} p^{B}_{i,s,t+1,n} shares^{B}_{i,s,t+1,n}$$
$$- \sum_{G}\sum_{S}\sum_{N_1} p^{S}_{i,s,t+1,n} shares^{S}_{i,s,t+1,n}) \times \mathrm{R}_{\mathrm{st}}^{expected}$$

$$\mathrm{Gain}^{stock\ selection}_{\mathrm{i,s,t,t+1}} \tag{A.26}$$
$$= \sum_{G}\sum_{S} p^{C}_{s,t} shares^{H}_{i,s,t} \left( \frac{p^{C}_{s,t+1}}{p^{C}_{s,t}} - 1 - \mathrm{R}_{\mathrm{st}}^{expected} \right)$$
$$+ \sum_{G}\sum_{S}\sum_{N_1} p^{B}_{i,s,t+1,n} shares^{B}_{i,s,t+1,n} \left( \frac{p^{C}_{s,t+1}}{p^{B}_{i,s,t+1,n}} - 1 - \mathrm{R}_{\mathrm{st}}^{expected} \right)$$
$$- \sum_{G}\sum_{S}\sum_{N_2} p^{S}_{i,s,t+1,n} shares^{S}_{i,s,t+1,n} \left( \frac{p^{C}_{s,t+1}}{p^{S}_{i,s,t+1,n}} - 1 - \mathrm{R}_{\mathrm{st}}^{expected} \right)$$

(2) Gain and Loss for investor category G, from day $\tau$ to day $\tau + \mathrm{n}$

$$Gain_{G,\tau,\tau+n} \tag{A.27}$$
$$= \sum_{G}\sum_{S} p^{C}_{s,\tau+n} shares^{H}_{i,s,\tau+n} - \sum_{G}\sum_{S} p^{C}_{s,\tau} shares^{H}_{i,s,\tau}$$
$$- \sum_{t=\tau}^{\tau+n-1} \sum_{G}\sum_{S}\sum_{N} p^{B}_{i,s,t+1,n} shares^{B}_{i,s,t+1,n} + \sum_{t=\tau}^{\tau+n-1} \sum_{G}\sum_{S}\sum_{N} p^{S}_{i,s,t+1,n} shares^{S}_{i,s,t+1,n}$$

**Appendix Table V. Gain/Loss for Various Investor Groups using Barber, Lee, Liu and Odean (2008)'s approach**

This table reports the gain and loss for various investor groups following the Barber, Lee, Liu and Odean (2008) approach from January 2016 to June 2019. Barber et al (2008) assumes that each day represents an independent observation of the total profits earned by a specific group. On each day, they construct two portfolios: the net buy portfolio includes stocks whose net buy shares is positive for specific group and the net sell portfolio includes stocks whose net sell shares is positive for specific group. We assume buy and sell portfolios held by a fixed holding period of 60 days. The gain/loss from stock from selection is the buy portfolio minus sell portfolio times portfolio return minus the cumulative market return. For the gain/loss from market timing, everything is the same as above, except the return is the expected return from the capital asset pricing model, $([\mathrm{R_{ft}} + \beta_s(R_{mt} - R_{ft})])$, $\beta_s$ is estimated form previous month for stock *s*. For the gain/loss from transaction costs, the commission costs on average are 0.05% of the dollar trading shares, imposed on both the buy and sell side, and the minimum commission cost is 5 CNY for each trade. Transaction taxes include the stamp tax, 0.10% of the value of sales, and the transfer fee, which is 0.002% of the value of trade and is imposed on both sides. Finally, the total gain/loss is aggregated from stock selection, market timing and trading costs.

| Annualized gain/loss | Total (mil CNY) | Market timing (mil CNY) | Stock selection (mil CNY) | Transaction costs (mil CNY) |
|---|---|---|---|---|
| RT1 | -16,454.55 | -183.24 | -11,474.60 | 4,796.71 |
| RT2 | -40,652.36 | -1,719.54 | -21,304.69 | 17,628.13 |
| RT3 | -53,032.91 | -4,632.72 | -20,687.42 | 27,712.77 |
| RT4 | -22,279.88 | -3,419.28 | -5,204.03 | 13,656.57 |
| RT5 | -8,351.83 | -4,996.44 | 15,739.56 | 19,094.95 |
| Inst. | 33,617.91 | 5,238.65 | 46,214.76 | 17,835.50 |
| Corp. | 4,674.07 | 9,715.17 | -3,286.18 | 1,754.92 |