Internet Appendix

for

Retail Trading and Return Predictability in China

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

Appendix A. Boehmer et al. (2021) Two-stage decomposition

Following Boehmer et al. (2021), we estimate the first stage of the two-stage decomposition. For each day *d*, we estimate the cross-sectional specification: [1]

$$Oib_{i,d} = h0_d + h1_d Oib_{i,d-1} + h2'_d Ret_{i,d-1} + h3_d Overconf_{i,d-1} + h4_d Gamble_{i,d-1} + u8_{i,d}. \quad \text{(A1)}$$

After we obtain the time-series of coefficients, $\{\widehat{h0}_d, \widehat{h1}_d, \widehat{h2}'_d, \widehat{h3}_d, \widehat{h4}_d\}$, we conduct statistical inference using the time-series means of the above coefficients, $\{\widehat{h0}, \widehat{h1}, \widehat{h2}', \widehat{h3}, \widehat{h4}\}$, and standard errors, which are adjusted using Newey and West (1987) method with the optimal bandwidth chosen by using Newey and West (1994) approach, to understand how each of the four hypotheses contributes to retail order flows. The first-stage estimation naturally decomposes $Oib_{i,d}$ into five components:

$$Oib_{i,d} = \widehat{Oib}_{i,d}^{persistence} + \widehat{Oib}_{i,d}^{liquidity} + \widehat{Oib}_{i,d}^{overconf} + \widehat{Oib}_{i,d}^{gamble} + \widehat{Oib}_{i,d}^{other}, \quad \text{(A2)}$$

With $\widehat{Oib}_{i,d}^{persistence} = \widehat{h1} Oib_{i,d}$, $\widehat{Oib}_{i,d}^{liquidity} = \widehat{h2}' Ret_{i,d-1}$, $\widehat{Oib}_{i,d}^{overconf} = \widehat{h3} Overconf_{i,d-1}$, $\widehat{Oib}_{i,d}^{gamble} = \widehat{h4} Gamble_{i,d-1}$, and $\widehat{Oib}_{i,d}^{other} = Oib_{i,d} - \widehat{Oib}_{i,d}^{persistence} - \widehat{Oib}_{i,d}^{liquidity} - \widehat{Oib}_{i,d}^{overconf} - \widehat{Oib}_{i,d}^{gamble}$. The "other" component is the residual component, which potentially contains other relevant information about future returns.

For the second stage of the decomposition, we relate future returns to the five components of order flow by the following specification using Fama and MacBeth's (1973) methodology:

$$Ret_{i,d+1} = m0_{d+1} + m1_{d+1}\widehat{Oib}_{i,d}^{persistence} + m2_{d+1}\widehat{Oib}_{i,d}^{liquidity} + m3_{d+1}\widehat{Oib}_{i,d}^{overconf} + m4_{d+1}\widehat{Oib}_{i,d}^{gamble} + m5_{d+1}\widehat{Oib}_{i,d}^{other} + m6'_{d+1} Controls_{i,d} + u9_{i,d+1}. \quad \text{(A3)}$$

The coefficient estimates in equation (A3) show how each component of order flows contributes to the predictive power of order flows for future stock returns. According to Boehmer et al. (2021), the advantage of the two-stage decomposition approach is that it includes various components of $Oib_{i,d}$ from the alternative hypotheses in a unified and internally consistent empirical framework. One caveat of this approach is that it is necessary to make empirical assumptions when choosing proxies for different hypotheses. Although these assumptions seem reasonable, we must be cautious as the results still depend on the validity of the empirical assumptions.

[1] All estimations in this study are estimated within each investor group. Beginning in this section, we omit the subscripts G to make the formula more readable.

Appendix B. Barber et al. (2009) Methods for Trading Performances

With access to all trading records from the Taiwan Stock Exchange, Barber, Lee, Liu, and Odean (2009, BLLO hereafter) designs a simple and intuitive method to compute the trading performances of different types of investors in four steps. First, BLLO separate all investors into "individuals", "corporations", "dealers", "foreigners", and "mutual funds", using identities provided by the exchange. Second, within each investor group, BLLO compute the aggregate daily net buy and net sell positions and create daily matching trading portfolios, the "net buy" and "net sell" portfolios. Specifically, the net buy volumes and sell volumes for stock *i* on day *d* of group G are summed over investor *j* in group *G*:

$$NetBuyVol_{i,d,G} = max\left(\sum_{j\in G} BuyVol_{i,d,j} - \sum_{j\in G} SellVol_{i,d,j}, 0\right), \tag{A4}$$
$$NetSellVol_{i,d,G} = max\left(\sum_{j\in G} SellVol_{i,d,j} - \sum_{j\in G} BuyVol_{i,d,j}, 0\right),$$

where the $Buyvol_{i,d,j}$ and $Sellvol_{i,d,j}$ are defined in equation (1). If net buy volume for stock *i* on day *d* in investor group *G* is positive, it belongs to the net buy portfolio, $NetBuy_{d,G}$; if the net buy volume is negative (thus the net sell volume is positive), it belongs to the net sell portfolio, $NetSell_{d,G}$. Third, BLLO track these "net buy" and "net sell" portfolios over a holding horizon of *n* days, and compute cumulative cash flows and returns from this tracking strategy, and treat them as proxies for trading performances of different investor groups. Specifically, BLLO computes the total performance of the net buy and net sell portfolios as:

$$Total_{d,G} = \sum_{i\in NetBuy_{d,G}} NetBuyVol_{i,G,d} \times AvgBuyPrice_{i,G,d} \times (ret_{i,d+1,d+n} - rf_{d+1,d+n}) - \sum_{i\in NetSell_{d,G}} NetSellVol_{i,G,d} \times AvgSellPrice_{i,G,d} \times \left(ret_{i,d+1,d+n} - rf_{d+1,d+n}\right) - TrdCost_{d,G}. \tag{A5}$$

The tracking net buy and sell portfolios are rebalanced once a day, so we follow BLLO and use the average prices, defined as

$$AvgBuyPrice_{i,G,d} = \frac{\max\left(\sum_{j\in G} BuyVolCNY_{i,d,j} - \sum_{j\in G} SellVolCNY_{i,d,j}, 0\right)}{NetBuyVol_{i,d,G}}, \tag{A6}$$
$$AvgSellPrice_{i,G,d} = \frac{\max\left(\sum_{j\in G} SellvolCNY_{i,d,j} - \sum_{j\in G} BuyvolCNY_{i,d,j}, 0\right)}{NetSellVol_{i,G,d}}.$$

Notice the $BuyVolCNY_{i,d,j}$ and $SellVolCNY_{i,d,j}$ in the numerator are cash volumes, products of trading share volumes and trading prices, rather than share volumes. Variable $ret_{i,d+1,d+n}$ and $rf_{d+1,d+n}$ are the cumulative return of stock *i* and the risk-free rate from day $d+1$ to day $d+n$, respectively. Trading cost is computed as follows:

$$TrdCost_{d,G} = \sum_{i}\sum_{j\in G} \tau_1 \times BuyVolCNY_{i,d,j} + \sum_{i}\sum_{j\in G} \tau_2 \times SellVolCNY_{i,d,j}. \tag{A7}$$

Variables τ_1 and τ_2 capture the proportional trading costs relative to cash volumes. According to China's policies and practices, there are commissions of 0.05% of cash volumes on both the buy and sell sides, a stamp tax of 0.10% on cash volumes of sales, and a transfer fee of 0.002% on cash volumes on both the buy and sell sides during our sample period. In other words, $\tau_1 = 0.05\% + 0.002\%$, $\tau_2 = 0.05\% + 0.10\% + 0.002\%$.

Three additional considerations are added in the third step. First, BLLO assumes that the net buy and sell portfolios are held for *n* days and sets *n* to different horizons. As the main discussion of BLLO focuses on the holding horizon of 140 days (*n*=140), we choose the same horizon for ease of comparison. Second, BLLO assumes that the net buy (sell) portfolios are independent for each day *d*, and annualizes the performance by aggregating the daily net buy and sell portfolios over each year.[2] Third, to have a perspective of total performance as a percentage of the total investment of these tracking portfolios, which is similar to the idea of return on investments, BLLO measures total investment as the aggregate holding value for each group of investors. For comparison, we choose the aggregate holding values in Panel A of Table I and present the return on investment as the total performance over aggregate holding.

After computing the total performance, BLLO decomposes the total into three intuitive components,

$$Total_{d,G} = StkSelect_{d,G} + MktTiming_{d,G} - TrdCost_{d,G}. \tag{A8}$$

Here, the stock selection component on day *d* for investor group *G* is computed as:

$$StkSelect_{d,G} = \sum_{i \in NetBuy_{d,G}} NetBuyVol_{i,G,d} \times AvgBuyPrice_{i,G,d} \times (ret_{i,d+1,d+n} - rmkt_{d+1,d+n}) - \sum_{i \in NetSell_{d,G}} NetSellVol_{i,G,d} \times AvgSellprice_{i,G,d} \times (ret_{i,d+1,d+n} - rmkt_{mkt,d+1,d+n}). \tag{A9}$$

In other words, BLLO compares each stock's return, $ret_{i,d+1,d+n}$, with the contemporaneous market return, $rmkt_{d+1,d+n}$, to determine whether investors are capable of choosing stocks that beat the market. The second component, the market timing component, is computed as:

$$MktTiming_{d,G} = \sum_{i \in NetBuy_{d,G}} NetBuyVol_{i,G,d} \times AvgBuyPrice_{i,G,d} \times (rmkt_{d+1,d+n} - rf_{d+1,d+n}) - \sum_{i \in NetSell_{d,G}} NetSellVol_{i,G,d} \times AvgSellprice_{i,G,d} \times (rmkt_{d+1,d+n} - rf_{d+1,d+n}). \tag{A10}$$

With the difference between the market portfolio and interest rate, the BLLO method aims to capture whether the investor can choose the allocation between stocks (market portfolio) and bonds (interest rate), and thus, time the market. The trading cost term is defined in equation (A7).

[2] We compute the standard error estimates using the Newey and West (1987) method, with the optimal bandwidth chosen by using Newey and West (1994) procedure. We conduct robustness check using optimal lag numbers selected by the Bayesian Information Criterion (BIC). The results are similar and available on request.

Appendix Table I. Previous Studies on Retail Investors in Different Markets

Publication Details	Data	Research Questions and Main Findings
Barber and Odean, Journal of Finance, 2000. "Trading is hazardous to your wealth: The common stock investment performance of individual investors?"	U.S. data, 66,465 households over 1991 to 1996 from a large discount brokerage	Research Question: Return performance of equities held directly by households. Findings: (1) Individual investors who hold common stocks directly pay a tremendous performance penalty for active trading. (2) Overconfidence can explain high trading levels and the resulting poor performance of individual investors.
Kaniel, Saar, and Titman, Journal of Finance, 2008. "Individual investor trading and stock returns."	U.S. data, all retail orders from NYSE CAUD file from year 2000 to 2003	Research Question: The relation between net individual investor trading and short-horizon cross-sectional stock returns. Findings: (1) Individuals buy stocks following declines in the previous month and sell following price increases. (2) Positive excess returns in the month following intense buying by individuals and negative excess returns after individuals sell.
Barber, Odean, and Zhu, Review of Financial Studies, 2008. "Do retail trades move markets?"	U.S. data, tick-by-tick trades from TAQ and ISSM over 1983 to 2001, small trades proxy for retail	Research Question: The trading of individual investors and stock returns. Findings: (1) Small trade order imbalance correlates well with order imbalance based on trades from retail brokers. (2) Retail investors herd. (3) Small trade order imbalance negatively forecasts future annual returns. (4) Small trade order imbalance positively predicts future week returns.
Kaniel, Liu, Saar, and Titman, Journal of Finance, 2012. "Individual investor trading and return patterns around earnings announcements."	U.S. data, all retail orders from NYSE CAUD file from year 2000 to 2003	Research Question: The informed trading by individual investors around earnings announcements Findings: (1) The intense aggregate individual investor buying (selling) predicts large positive (negative) abnormal returns on and after earnings announcement dates. (2) Decompose abnormal returns following the event into information and liquidity provision components, and show that about half of the returns can be attributed to private information.

Publication Details	Data	Research Questions and Main Findings
Kelley and Tetlock, Journal of Finance, 2013. "How wise are crowds? Insights from retail orders and stock returns."	U.S. data, retail orders from dozens of retail brokerages over year 2003 to 2007	Research question: The role of retail investors in stock pricing by separately examining aggressive (market) and passive (limit) orders. Findings: (1) Both market and limit order imbalance positively predict firms' monthly stock returns. (2) Market orders correctly predict firm news, including earnings surprises. (3) Limit orders following negative returns, consistent with traders providing liquidity.
Boehmer, Jones, Zhang and Zhang, Journal of Finance, 2021. "Tracking retail investor activity."	U.S. data, an algorithm to identify marketable retail trades from TAQ over year 2010 to 2015	Research Question: Provide an algorithm to identify marketable retail purchases and sales using TAQ. Findings: (1) Provide and validate the algorithm. (2) Marketable retail order imbalance positive predict future weekly stock return. (3) Predictive power of marketable retail order imbalance could be attributable to order flow persistence, contrarian trading, public news sentiment and unexplained part.
Barber, Lin, and Odean, Journal of Financial and Quantitative Analysis, forthcoming . "Resolving a paradox: Retail trades positively predict returns but are not profitable."	U.S. data, retail trades identified by Boehmer et al. (2021) from TAQ over year 2010-2019	Research Question: Retail order imbalance positively predicts returns, but in aggregate retail investor trades lose money. Findings: (1) Order imbalance tests equally weight stocks, but retail purchases concentrate in stocks that subsequently underperform. (2) Trades by retail investors with less knowledge, experience, and wealth are more likely to underperform.
Welch, Journal of Finance, 2022. "The wisdom of the Robinhood crowd."	U.S. data, Robinhood investors holding from May 2018 to August 2020	Research Question: Robinhood investors trading behavior. Findings: (1) Robinhood investors increased their holdings in the March 2020 COVID bear market. (2) Robinhood investors tend to buy stocks with high past share volume and dollar-trading volume. (3) From mid-2018 to mid-2020, an aggregated Robinhood portfolio had both good timing and good alpha.

Publication Details	Data	Research Questions and Main Findings
Ozik, Sadka, and Shen, Journal of Financial and Quantitative Analysis, 2021. "Flattening the illiquidity curve: Retail trading during the COVID-19 lockdown."	U.S. data, Robinhood investors holding from May 2018 to August 2020	Research Question: The impact of retail investors on stock liquidity during the Coronavirus pandemic lockdown. Findings: (1) Retail trading exhibits a sharp increase during Covid. (2) Retail trading attenuated the rise in illiquidity by roughly 40%, but less so for high-media-attention stocks.
Barber, Huang, Odean, and Schwarz, Journal of Finance, 2022. "Attention-Induced Trading and Returns: Evidence from Robinhood Users."	U.S. data, Robinhood investors holding from May 2018 to August 2020	Research Question: The influence of financial innovation by fintech brokerages on individual investors' trading and stock prices. Findings: (1) Robinhood investors engage in more attention-induced trading than other retail investors. (2) Intense buying by Robinhood users forecasts negative returns. Average 20-day abnormal returns are −4.7% for the top stocks purchased each day.
Grinblatt and Keloharju, Journal of Financial Economics, 2000. "The investment behavior and performance of various investor types: a study of Finland's unique data set."	Finnish data, daily retail trades and holdings from Finnish Central Securities Depository (FCSD) over 1994 to 1996	Research Question: The past-return-based behavior and the performance of various investor types. Findings: (1) Foreign investors tend to be momentum, while domestic investors, particularly households, tend to be contrarians. (2) The portfolios of foreign investors seem to outperform the portfolios of households, even after controlling for behavior differences.
Linnainmaa, Journal of Finance, 2011. "Do limit orders alter inferences about investor performance and behavior?"	Finnish data, daily retail trades and holdings from Finnish Central Securities Depository (FCSD) over 1995 to 2002	Research Question: Individual investors' trading behaviors could be explained by investors' use of limit orders. Findings: These patterns arise because limit orders are price-contingent and suffer from adverse selection.

Publication Details	Data	Research Questions and Main Findings
Grinblatt, Keloharju, Linnainmaa, Journal of Financial Economics, 2012. "IQ, trading behavior, and performance."	Finnish data, daily retail trades and holdings from Finnish Central Securities Depository (FCSD) over 1995 to 2002 and intelligence (IQ) test administered to Finnish male	Research Question: Whether IQ influences trading behavior, performance, and transaction costs. Findings: (1) High-IQ investors are less subject to disposition effect, more aggressive about tax-loss trading, and more likely to supply liquidity when stock experience a one-month high. (2) High-IQ investors exhibit superior market timing, stock-picking skill, and trade execution.
Bach, Calvet and Sodini, American Economic Review, 2020. "Rich pickings? Risk, return, and skill in household wealth."	Sweden data, annual administrative panel that reports the full balance sheet of Swedish residents between 2000 and 2007	Research Question: Examine the risk and return characteristics of household wealth. Findings: (1) The expected return on household net wealth increases with net worth. (2) The expected wealth return is driven by systematic risk-taking and exhibits strong persistence. Idiosyncratic risk is transitory but sufficiently large to generate substantial long-term dispersion in returns. (3) Wealth returns explain most of the historical increase in top wealth shares.
Dorn, Huberman and Sengmueller, Journal of Finance, 2008. "Correlated trading and returns."	German data, 37,000 retail clients at one of the three largest German discount brokers from 1998 to 2000	Research Question: Investors correlated trading and stock returns. Findings: (1) Investors tend to be on the same side of the market. (2) Correlated market orders lead returns due to persistent speculative price pressure. (3) Correlated limit orders also predict subsequent returns, consistent with liquidity demands.
Barrot, Kaniel, and Sraer, Journal of Financial Economics, 2016. "Are retail traders compensated for providing liquidity?"	French data, from a leading European online broker between 2002 and 2010	Research Question: Whether individual investors provide liquidity to the stock market and whether they are compensated for doing so. Findings: (1) The ability of aggregate retail order imbalances to predict short-term future returns is significantly enhanced during times of market stress. (2) Individual investors do not reap the rewards from liquidity provision because they experience a negative return on the trading day and reverse their trades for too long after the liquidity provision dissipated.

Publication Details	Data	Research Questions and Main Findings
Fong, Gallagher, and Lee, Journal of Financial and Quantitative Analysis, 2014. "Individual investors and broker types"	Australian data, transaction data from Australian Securities Exchange (ASX) SIRCA with identification of the broker between 1995 and 2007	Research Question: Examine the informativeness of trades via discount and full-service retail brokers. Findings: (1) Trades via full-service retail brokers are more informative than are trades via discount retail brokers. (2) Past returns, volatility, and news announcements could explain the net volume of discount retail brokers, but could not explain the net volume of full-service retail brokers.
Barber, Lee, Liu, and Odean, Review of Financial Studies, 2009. "Just how much do individual investors lose by trading?"	Chinese Taiwan data, transaction data in the Taiwan stock exchange between 1995 and 1999	Research Question: How much do individual investors lose by trading? Findings: (1) The aggregate portfolio of individuals suffers an annual performance penalty of 3.8%. (2) Nearly all individual trading losses can be traced to their aggressive orders.
Barber, Lee, Liu, and Odean, Journal of Financial Markets, 2014. "The cross-section of speculator skill: Evidence from day trading"	Chinese Taiwan data, day traders' transaction data in the Taiwan stock exchange between 1992 and 2006	Research Question: Examine the cross-sectional differences of returns earned by speculative day traders. Findings: Less than 1% of the day trader population can predictably and reliably earn positive abnormal returns net of fees.
Balasubramaniam, Campbell, Ramadorai and Ranish, Journal of Finance, 2023. "Who owns what? A factor model for direct stock holding"	Indian data, 10 million retail investors holdings data on August 2011from Indian stock market	Research Question: Build a cross-sectional factor model for retail investors' direct stock holdings. Findings: (1) Stock characteristics such as firm age and share price have strong investor clienteles. (2) Account attributes such as account age, account size, and extreme under diversification are associated with particular characteristic preferences. (2) Coheld stocks have higher return covariance.

Publication Details	Data	Research Questions and Main Findings
Anagol, Balasubramaniam, and Ramadorai, Journal of Financial Economics, 2021. "Learning from noise: Evidence from India's IPO lotteries"	Indian data, 1.5 million investors participate in allocation lotteries for 54 IPO stocks between 2007 and 2012	Research Question: Retail investors participate in allocation lotteries for Indian IPO stocks. Findings: Investors who wind the IPO lottery and obtain IPO stocks that rise in value increase portfolio trading volume in non-IPO stocks relative to lottery losers. A learning model could explain the results.
Titman, Wei, and Zhao, Journal of Financial Economics, 2022. "Corporate actions and the manipulation of retail investors in China: An analysis of stock splits"	Chinese data, account trading data from a major stock exchange in China between 1999 and 2015	Research Question: Corporate actions and the manipulation of retail investors in the stock splits events. Findings: (1) Share prices temporarily increase after stock splits, and subsequently decline below their presplit levels. (2) Small retail investors buy shares in firms initiating suspicious splits, while more sophisticated investors buy before suspicious split announcements and sell in the postsplit period. (3) Insiders sell large blocks of shares and obtain loans using company stock as collateral before the suspicious splits.
An, Lou, and Shi, Journal of Monetary Economics, 2022. "Wealth redistribution in bubbles and crashes."	Chinese data, account trading from a major stock exchange in China, 2014-15 bubbles and crashes episode	Research Question: What are the social-economic consequences of financial market bubbles and crashes? Findings: The largest 0.5% households in the equity market gain, while the bottom 85% lose, 250B RMB through active trading in this period.
Liao, Peng, Zhu, Review of Financial Studies, 2021. "Extrapolative bubbles and trading volume."	Chinese data, account-level transaction data from one of the largest brokerage firms between 2014 and 2015	Research Questions: Propose an extrapolative model to explain the sharp rise in prices *and* volume during financial bubbles. Findings: (1) The model proposes a novel mechanism for volume: because of extrapolative beliefs and disposition effects, investors are quick to buy assets with positive past returns and sell them if good returns continue. (2) Use Chinese account-level data to confirm the model's predictions.

Publication Details	Data	Research Questions and Main Findings
Chen, Gao, He, Jiang, and Xiong, Journal of Econometrics, 2019. "Daily price limits and destructive market behavior."	**Chinese data**, account trading from a major stock exchange in China from between 2012 and 2015	**Research Question:** Daily price limits and investors trading behavior. **Findings**: Large investors tend to buy on the day when a stock hits the 10% upper price limit and then sell on the next day; and their net buying on the limit-hitting day predicts stronger long-run price reversal.
Li, Geng, Subrahmanyam and Yu, Journal of Empirical Finance, 2017. "Do wealthy investors have an informational advantage? Evidence based on account classifications of individual investors"	**Chinese data,** a brokerage firm providing one million investors' trading records between January 2007 and 2009	**Research Question:** Do wealthy investors have an informational advantage? **Findings:** (1) Wealthy investors with portfolio values above the 99.5th percentile ("super" investors) outperform all other investors. (2) Part of their excess returns could be explained by informational advantages.
Liu, Peng, Xiong, and Xiong, Journal of Financial Economics, 2022. "Taming the bias zoo."	**Chinese data,** Combine subjective survey responses in 2018 with account-level transaction data from a major stock exchange in China	**Research Question:** Many biases offer observationally similar predictions for a targeted financial anomaly. This study combines subjective survey responses with observational data to tame the bias zoo. **Findings:** In cross-sectional regressions of respondents' actual turnover on survey-based trading motives, perceived information advantage, and gambling preference dominate other motives.

Appendix Table II. Comparing Main Findings of This Study with the Results from Previous Literature

Our main findings	**Findings from previous literature**
Return Predictability: Retail investors with smaller account sizes negatively predict future returns, whereas retail investors with larger account sizes positively predict future returns.	Retail trading negatively predicts future returns: Barber and Odean (2000), Grinblatt and Keloharju (2000), Barber, Huang, Odean, and Schwarz (2022) Retail trading positively predicts future returns: Kaniel, Saar, and Titman (2008), Barber, Odean, and Zhu (2008), Dorn, Huberman and Sengmueller (2008), Grinblatt, Keloharju, Linnainmaa (2012), Kelley and Tetlock (2013), Fong, Gallagher, and Lee (2014), Barber, Lee, Liu, and Odean (2014), Barrot, Kaniel, and Sraer (2016), Li, Geng, Subrahmanyam, and Yu (2017), Boehmer, Jones, Zhang, and Zhang (2021), Welch (2022)
Momentum/Contrarian: Retail investors with smaller account sizes display daily momentum and weekly contrarian patterns, whereas retail investors with larger account sizes display contrarian patterns.	Retail investors display momentum trading patterns: Kelley and Tetlock (2013), Boehmer, Jones, Zhang, and Zhang (2021) Retail investors display contrarian trading patterns: Grinblatt and Keloharju (2000), Kaniel, Saar, and Titman (2008), Linnainmaa (2011), Grinblatt, Keloharju, Linnainmaa (2012), Barrot, Kaniel, and Sraer (2016), Welch (2022)
Process Information: Retail investors with smaller account sizes fail to process public news, whereas retail investors with larger account sizes incorporate public news in trading.	Retail investors can process information: Kaniel, Liu, Saar, and Titman (2012), Kelley and Tetlock (2013), Boehmer, Jones, Zhang, and Zhang (2021) Retail investors fail to process information: Li, Geng, Subrahmanyam, and Yu (2017), Titman, Wei, and Zhao (2022)
Behavior Biases: Retail investors with smaller account sizes show behavioral patterns such as overconfidence and gambling preferences, while retail investors with larger account sizes could better predict returns in stocks more attractive to investors with behavioral biases.	Retail investors display behavioral biases: Barber and Odean (2000), Liu, Peng, Xiong, and Xiong (2022), Barber, Huang, Odean, and Schwarz (2022) Some retail investors display behavioral biases, while others trade against behavioral biases: Chen, Gao, He, Jiang, and Xiong (2019)

Appendix Table III. Distributions of Investor Trading and Holding

This table reports the summary statistics for the trading and holding of different investor groups across different stock characteristics. Our sample period covers January 2016 to June 2019. Our sample firms are A-share stocks listed on a major stock exchange with at least 15 days of trade during the previous month. The trading and holding for different investor groups across different stock characteristics are reported in Panels A and B, which are the time-series averages of cross-sectional means. The bottom two and top two sectors, in terms of trading volume and holding within each investor group, are reported in Panels C and D. The sectors are classified according to the Datastream Level 4 sector classification. Panel E reports the trading volumes of different investor groups across different trade sizes, which are the time-series averages of cross-sectional means.

Panel A. Trading volumes across different stock characteristics

	RT1	RT2	RT3	RT4	RT5	INST	CORP
Small	6.3%	23.3%	32.5%	13.4%	14.0%	9.9%	0.7%
Medium	5.8%	21.2%	31.3%	13.3%	13.9%	13.7%	0.8%
Large	4.3%	16.3%	26.7%	12.9%	15.7%	22.4%	1.8%
Low EP	6.3%	22.0%	31.4%	13.7%	15.3%	10.5%	0.8%
Medium EP	5.6%	21.2%	30.7%	12.9%	13.8%	14.9%	0.9%
High EP	4.3%	17.1%	28.0%	13.1%	14.7%	21.2%	1.6%
Low Turnover	4.7%	17.5%	28.0%	12.9%	14.5%	20.7%	1.7%
Medium Turnover	5.2%	19.4%	29.9%	13.5%	15.0%	15.8%	1.1%
High Turnover	6.3%	23.4%	32.1%	13.2%	14.2%	10.1%	0.6%

Panel B. Holding shares across different stock characteristics

	RT1	RT2	RT3	RT4	RT5	INST	CORP
Small	4.1%	13.4%	20.2%	9.9%	17.4%	7.1%	27.9%
Medium	3.6%	10.8%	16.5%	8.0%	15.4%	11.1%	34.6%
Large	2.2%	6.6%	10.6%	5.3%	10.1%	16.7%	48.6%
Low EP	3.9%	11.2%	16.9%	8.5%	15.5%	8.2%	35.7%
Medium EP	3.4%	11.3%	17.3%	8.2%	15.6%	12.6%	31.6%
High EP	2.4%	7.3%	12.0%	6.2%	11.8%	14.5%	45.8%
Low Turnover	2.2%	5.8%	9.3%	4.9%	11.2%	11.9%	54.6%
Medium Turnover	3.1%	9.0%	14.7%	7.8%	15.5%	12.9%	37.0%
High Turnover	4.3%	15.0%	22.2%	10.1%	16.2%	10.5%	21.6%

Panel C. Sectors with lowest and highest trading volumes

	RT1	RT2	RT3	RT4	RT5	INST	CORP
Bottom1	**Banks & Life Insurance** 2.4%	**Banks & Life Insurance** 10.5%	**Banks & Life Insurance** 20.4%	**Banks & Life Insurance** 11.9%	**Alternative Energy** 12.2%	**Alternative Energy** 6.5%	**Alternative Energy** 0.4%
Bottom2	Financial Services 3.6%	Financial Services 15.8%	Beverages 27.4%	Health Care Equipment & Services 11.9%	Gas, Water and Multiutilities 12.3%	Industrial Metals & Mining 11.6%	General Industrials 0.7%
Top2	Industrial Metals & Mining 6.5%	Industrial Engineering 22.3%	Gas, Water and Multiutilities 32.0%	Technology Hardware & Equipment 14.5%	Software & Computer Services 17.7%	Travel & Leisure 21.3%	Financial Services 2.9%
Top1	**Alternative Energy** 7.6%	**Alternative Energy** 25.5%	**Alternative Energy** 34.2%	**Financial Services** 14.5%	**Banks & Life Insurance** 18.0%	**Banks & Life Insurance** 32.8%	**Banks & Life Insurance** 4.1%

Panel D. Sectors with lowest and highest holding shares

	RT1	RT2	RT3	RT4	RT5	INST	CORP
Bottom1	**Banks & Life Insurance** 1.1%	**Banks & Life Insurance** 4.0%	**Banks & Life Insurance** 7.6%	**Banks & Life Insurance** 4.2%	**Banks & Life Insurance** 5.6%	**Alternative Energy** 2.9%	**Household Goods & Home Construction** 19.2%
Bottom2	Aerospace & Defense 1.9%	Aerospace & Defense 7.2%	Food & Drug Retailers 12.1%	Industrial Transportation 5.6%	Electricity 7.3%	Mobile Telecommunications 4.8%	Health Care Equipment & Services 23.2%
Top2	Forestry & Paper 4.5%	Household Goods & Home Construction 12.8%	Support Services 19.2%	Support Services 9.7%	Support Services 19.5%	Banks & Life Insurance 23.3%	Banks & Life Insurance 54.3%
Top1	**Alternative Energy** 6.1%	**Alternative Energy** 15.5%	**Alternative Energy** 20.9%	**Forestry & Paper** 9.8%	**Software & Computer Services** 23.8%	**Health Care Equipment & Services** 23.5%	**Mobile Telecommunications** 55.1%

Panel E. Trading volumes across different trade sizes

Trade Size	RT1	RT2	RT3	RT4	RT5	INST	CORP
<40,000 CNY	4.78%	10.86%	7.44%	1.22%	0.74%	5.35%	0.10%
40,000-200,000 CNY	0.69%	8.47%	14.54%	5.04%	3.65%	4.36%	0.27%
>200,000 CNY	0.00%	1.03%	8.19%	6.96%	10.30%	5.48%	0.73%

Appendix Table IV. Long-short Portfolios

This table reports an alternative method for Tables III and IV using long-short portfolios. On each day, we first group stocks into different counterparty bins in Panel A and into different days-to-cover ratios in Panel B. Within each bin, we then sort stocks into three groups according to the order imbalances of different investor groups. We form a long-short portfolio by longing the stock with the highest 1/3 order imbalance measures and shorting those with the lowest 1/3 order imbalance measures. The long-short portfolios are held on the next day and the portfolio returns are adjusted by Liu, Stambaugh and Yuan (2019) three-factor model. To account for serial correlation in the coefficients, the standard errors of the time series are adjusted using Newey and West (1987) method, with optimal bandwidths chosen by following Newey and West (1994). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Daily alphas for long-short portfolios with different counterparties

RT1-4	RT5	INST	Sample Coverage		RT1-RT4	RT5	INST
Buy	Buy	Sell	17%	Alpha	-0.0004**	0.0003**	0.0019***
				[*t*-stat]	[-2.16]	[2.32]	[11.05]
Buy	Sell	Buy	11%	Alpha	-0.0006**	0.0002	-0.0003*
				[*t*-stat]	[-2.33]	[0.83]	[-1.85]
Buy	Sell	Sell	21%	Alpha	-0.0005***	0.0011***	0.0020***
				[*t*-stat]	[-2.75]	[8.36]	[11.30]
Sell	Buy	Buy	22%	Alpha	-0.0028***	0.0011***	0.0009***
				[*t*-stat]	[-14.81]	[6.74]	[6.11]
Sell	Buy	Sell	13%	Alpha	-0.0033***	0.0023***	-0.0004**
				[*t*-stat]	[-9.10]	[10.88]	[-1.99]
Sell	Sell	Buy	17%	Alpha	-0.0011***	-0.0003*	0.0007***
				[*t*-stat]	[-6.60]	[-1.86]	[4.33]

Panel B. Daily alphas for long-short portfolios with different counterparties

Holding Horizon		OibRT1	OibRT2	OibRT3	OibRT4	OibRT5	INST
(0,10]Days	Alpha	-0.0054***	-0.0048***	-0.0034***	-0.0003*	0.0039***	0.0027***
	[*t*-stat]	[-19.22]	[-17.82]	[-14.72]	[-1.67]	[12.40]	[13.30]
(10,20]Days	Alpha	-0.0028***	-0.0023***	-0.0016***	-0.0006***	0.0014***	0.0021***
	[*t*-stat]	[-14.54]	[-11.45]	[-8.39]	[-4.63]	[7.80]	[9.85]
(20,60]Days	Alpha	-0.0020***	-0.0014***	-0.0011***	-0.0003***	0.0006***	0.0019***
	[*t*-stat]	[-14.70]	[-8.19]	[-7.25]	[-4.05]	[5.79]	[14.08]
Above60Days	Alpha	-0.0025***	-0.0049***	-0.0056***	-0.0017**	0.0003***	0.0015***
	[*t*-stat]	[-4.56]	[-5.47]	[-4.49]	[-2.52]	[3.07]	[12.84]

Appendix Table V. Predict Returns over the Next 12 Weeks Using Previous Week Order Imbalances
This table reports the alternative results for Table V using previous week order imbalances to predict the cross-section of returns over the next 12 weeks. The independent variables are the previous week's order imbalance, *Oib*(*w*-1). The control variables are the same as those in Table II, and are not reported for brevity. To account for serial correlation in the coefficients, the standard errors of the time series are adjusted using Newey and West (1987) method, with optimal bandwidths chosen by following Newey and West (1994). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

*w*th weeks		OibRT1(*w*-1)	OibRT2(*w*-1)	OibRT3(*w*-1)	OibRT4(*w*-1)	OibRT5(*w*-1)	OibINST(*w*-1)
1	Estimate	-0.0449***	-0.0428***	-0.0280***	-0.0041***	0.0064***	0.0081***
	[*t*-stat]	[-12.37]	[-11.26]	[-8.70]	[-3.42]	[8.05]	[12.14]
2	Estimate	-0.0176***	-0.0162***	-0.0095***	0.0004	0.0034***	0.0031***
	[*t*-stat]	[-7.64]	[-7.18]	[-5.09]	[0.39]	[5.17]	[5.15]
3	Estimate	-0.0099***	-0.0082***	-0.0050**	-0.0003	0.0031***	0.0021***
	[*t*-stat]	[-3.83]	[-2.93]	[-2.05]	[-0.33]	[5.64]	[2.93]
4	Estimate	-0.0072***	-0.0065***	-0.0036*	-0.0002	0.0023***	0.0008
	[*t*-stat]	[-3.48]	[-2.93]	[-1.71]	[-0.21]	[3.54]	[1.42]
5	Estimate	-0.0059***	-0.0053***	-0.0050***	-0.0025**	0.0009	0.0011**
	[*t*-stat]	[-3.60]	[-2.93]	[-2.71]	[-2.58]	[1.57]	[2.14]
6	Estimate	-0.0098***	-0.0089***	-0.0062***	-0.0010	0.0003	0.0021***
	[*t*-stat]	[-4.73]	[-4.24]	[-3.42]	[-1.04]	[0.65]	[4.14]
7	Estimate	-0.0070***	-0.0060***	-0.0035**	-0.0004	0.0009	0.0018***
	[*t*-stat]	[-4.11]	[-3.44]	[-2.21]	[-0.33]	[1.53]	[4.46]
8	Estimate	-0.0048**	-0.0033	-0.0027	-0.0002	0.0013**	0.0010*
	[*t*-stat]	[-2.49]	[-1.53]	[-1.39]	[-0.16]	[2.02]	[1.80]
9	Estimate	-0.0030	-0.0015	-0.0011	0.0006	0.0010*	0.0002
	[*t*-stat]	[-1.46]	[-0.75]	[-0.62]	[0.80]	[1.81]	[0.39]
10	Estimate	-0.0035*	-0.0030	-0.0024	0.0001	-0.0007	0.0010*
	[*t*-stat]	[-1.66]	[-1.38]	[-1.22]	[0.11]	[-1.25]	[1.72]
11	Estimate	-0.0011	-0.0007	0.0009	0.0008	0.0001	0.0001
	[*t*-stat]	[-0.54]	[-0.29]	[0.41]	[0.77]	[0.19]	[0.10]
12	Estimate	-0.0027	-0.0027	-0.0004	-0.0005	0.0000	0.0001
	[*t*-stat]	[-1.22]	[-1.20]	[-0.19]	[-0.63]	[0.05]	[0.20]

Appendix Table VI. Two Stage Decomposition with Additional Lags of Order Imbalances

This table presents the alternative two-stage decomposition results for Table VI with additional lags of order imbalances. Panel A reports the first-stage estimation results controlling additional two weeks order imbalance. The order imbalances are decomposed into five components as specified in the following equation:

$$Oib_{i,d} = p0_d + \left[p1_d^1 Oib_{i,d-1} + p1_d^2 Oib_{i,d-2,d-6} + p1_d^3 Oib_{i,d-7,d-11}\right] + p2'_d Ret_{i,d-1} + p3_d Overconf_{i,d-1} + p4_d Gamble_{i,d-1} + u10_{i,d}, \tag{A11}$$

where the order persistence components contain additional two weeks order imbalances, and other components are same as in Table VI. Panel B reports the second-stage decomposition of the order imbalance's predictive power, as specified in the following equation:

$$Oib_{i,d} = \widehat{Oib}_{i,d}^{persistence^{MoreLags}} + \widehat{Oib}_{i,d}^{liquidity} + \widehat{Oib}_{i,d}^{overconf} + \widehat{Oib}_{i,d}^{gamble} + \widehat{Oib}_{i,d}^{other}, \tag{A12}$$

where the order imbalance component, *Oib*(-1,*Persistence*^{MoreLags}), is constructed using the previous one day and previous two weeks order imbalance in first stage, and the rest are the same as those in Table VI. To account for serial correlation in the coefficients, the standard errors of the time series are adjusted using Newey and West (1987) method, with optimal bandwidths chosen by following Newey and West (1994). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A. First stage of projecting the order imbalance on persistence, past returns, overconfidence, and gambling proxies

		OibRT1	OibRT2	OibRT3	OibRT4	OibRT5
Oib(-1)	Estimate	0.1517***	0.1675***	0.1496***	0.0456***	0.0917***
	[*t*-stat]	[24.43]	[36.07]	[37.47]	[15.01]	[28.14]
Oib(-6,-2)	Estimate	0.2434***	0.2246***	0.1931***	0.0950***	0.1524***
	[*t*-stat]	[38.71]	[37.27]	[33.78]	[18.14]	[30.75]
Oib(-11,-7)	Estimate	0.0691***	0.0541***	0.0533***	0.0475***	0.0691***
	[*t*-stat]	[12.43]	[10.33]	[11.53]	[13.47]	[22.22]
Ret(-1)	Estimate	0.5303***	0.6989***	0.4174***	-0.2229***	-1.3117***
	[*t*-stat]	[9.16]	[15.30]	[13.53]	[-7.67]	[-30.37]
Ret(-6,-2)	Estimate	-0.2261***	-0.0271**	0.0435***	-0.0285	-0.0799***
	[*t*-stat]	[-12.25]	[-2.54]	[3.85]	[-1.60]	[-4.95]
Ret(-27,-7)	Estimate	0.0077*	0.0074**	-0.0033	-0.0282***	-0.0170***
	[*t*-stat]	[1.95]	[2.28]	[-1.19]	[-7.51]	[-3.99]
Overconf(-1)	Estimate	0.0391***	0.0204**	0.0218***	0.0167	-0.0533***
	[*t*-stat]	[3.11]	[2.17]	[2.90]	[1.63]	[-5.05]
Gamble(-1)	Estimate	0.0172	0.0580***	0.1430***	0.2229***	-0.0425*
	[*t*-stat]	[1.03]	[4.49]	[11.54]	[12.02]	[-1.94]

Intercept	Estimate	-0.0131***	-0.0076***	-0.0080***	-0.0062***	0.0174***
	[*t*-stat]	[-4.14]	[-3.60]	[-6.05]	[-3.75]	[8.57]
Adj.R2		9.68%	7.64%	5.37%	1.21%	2.99%

Panel B. Second stage decomposition of order imbalance's predictive power

Dep.var		Ret	Ret	Ret	Ret	Ret
Oib.var		OibRT1	OibRT2	OibRT3	OibRT4	OibRT5
Oib(-1,Persistence^{MoreLags})	Estimate	-0.0284***	-0.0274***	-0.0202***	-0.0062***	0.0059***
	[*t*-stat]	[-13.95]	[-13.52]	[-11.19]	[-3.95]	[8.36]
Oib(-1,Liquidity)	Estimate	-0.0085**	-0.0174***	-0.0255***	0.0204	0.0037
	[*t*-stat]	[-2.16]	[-3.87]	[-3.03]	[1.29]	[1.50]
Oib(-1,Overconf)	Estimate	-0.1022	-0.4609***	-0.6067***	-0.9331***	0.1815***
	[*t*-stat]	[-1.63]	[-3.79]	[-4.29]	[-5.06]	[5.14]
Oib(-1,Gamble)	Estimate	-0.5580***	-0.1310***	-0.0538***	-0.0453***	0.3008***
	[*t*-stat]	[-4.87]	[-4.00]	[-4.05]	[-5.32]	[6.35]
Oib(-1,Other)	Estimate	-0.0084***	-0.0082***	-0.0058***	-0.0008***	0.0011***
	[*t*-stat]	[-15.19]	[-15.17]	[-13.90]	[-6.06]	[9.40]
Adj.R2		10.54%	10.37%	10.06%	9.61%	9.48%
Interquartile return						
Oib(-1,Persistence^{MoreLags})		-0.1387%	-0.1103%	-0.0656%	-0.0123%	0.0335%
Oib(-1,Liquidity)		-0.0124%	-0.0223%	-0.0195%	0.0102%	0.0089%
Oib(-1,Overconf)		-0.0150%	-0.0338%	-0.0404%	-0.0428%	0.0252%
Oib(-1,Gamble)		-0.0317%	-0.0251%	-0.0254%	-0.0334%	0.0422%
Oib(-1,Other)		-0.1694%	-0.1433%	-0.0980%	-0.0227%	0.0475%

Appendix Table VII. Annual Performance Using Tracking Portfolios of Retail Trading, from Liquidity Provision and Liquidity Demand

This table reports the annual performance for different retail investor groups using the similar framework as in Table VIII and from the liquidity provision and demand perspective. The annual total performance could be decomposed into liquidity provision, liquidity demand, and trading cost, as specified in the following equation:

$$Total_{d,G} = LiquidityProvision_{d,G} + LiquidityDemand_{d,G} - TrdCost_{d,G}. \tag{A13}$$

We assume that if investors trade direction is opposite to direction of previous stock price movement, then investors provide liquidity; and if investors trade in the same direction of previous stock price movement, then investor demand liquidity. The total performance from liquidity provision and liquidity demand are defined in the following equations:

$$\begin{aligned} LiquidityProvision_{d,G} \\ = \sum\nolimits_{i \in NetBuy_{d,G} \text{ } and \text{ } ret_{i,d-1} \leq 0} NetBuyVol_{i,G,d} \times AvgBuyPrice_{i,G,d} \times (ret_{i,d+1,d+n} - rf_{d+1,d+n}) \\ - \sum\nolimits_{i \in NetSell_{d,G} \text{ } and \text{ } ret_{i,d-1} \geq 0} NetSellVol_{i,G,d} \times AvgSellPrice_{i,G,d} \times (ret_{i,d+1,d+n} - rf_{d+1,d+n}). \end{aligned} \tag{A14}$$

$$\begin{aligned} LiquidityDemand_{d,G} \\ = \sum\nolimits_{i \in NetBuy_{d,G} \text{ } and \text{ } ret_{i,d-1} > 0} NetBuyVol_{i,G,d} \times AvgBuyPrice_{i,G,d} \times (ret_{i,d+1,d+n} - rf_{d+1,d+n}) \\ - \sum\nolimits_{i \in NetSell_{d,G} \text{ } and \text{ } ret_{i,d-1} < 0} NetSellVol_{i,G,d} \times AvgSellPrice_{i,G,d} \times (ret_{i,d+1,d+n} - rf_{d+1,d+n}). \end{aligned} \tag{A15}$$

Panel A presents the total annual trading performance by different investor groups, from liquidity provision and demand perspective. Panel B further decompose the annual trading performance into stock selection and market timing components. Panel C reports the annual trading performance around earnings announcements. The earnings announcement events include the stock on announcement day *d* and the previous day *d*-1. To account for serial correlation in the coefficients, the standard errors of the time series are adjusted using Newey and West (1987) method, with optimal bandwidths chosen by following Newey and West (1994). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Annual trading performance using tracking portfolios

		Total	Liquidity Provision	Liquidity Demand	Trading Cost
RT1	Performance	-5.61%***	-0.34%	-3.89%***	-1.38%***
	[*t*-stat]	[-10.29]	[-0.82]	[-7.65]	[-23.47]
RT2	Performance	-4.61%***	-0.04%	-2.78%***	-1.80%***
	[*t*-stat]	[-8.96]	[-0.09]	[-6.45]	[23.07]
RT3	Performance	-3.60%***	-0.25%	-1.63%***	-1.72%***
	[*t*-stat]	[-9.49]	[-0.90]	[-6.06]	[-20.89]
RT4	Performance	-2.80%***	-0.11%	-1.11%***	-1.58%***
	[*t*-stat]	[-8.51]	[-0.49]	[-4.63]	[-19.53]
RT5	Performance	-0.29%	1.89%***	-1.14%***	-1.03%***
	[*t*-stat]	[-0.92]	[5.12]	[-3.05]	[-18.84]
INST	Performance	1.15%***	0.84%***	0.73%***	-0.41%***
	[*t*-stat]	[3.57]	[5.70]	[2.91]	[-20.49]

Panel B. Annual trading performance using tracking portfolios: separate stock selection and market timing

		Total	Liquidity Provision		Liquidity Demand		TradingCost
			Stock Selection	Market Timing	Stock Selection	Market Timing	
RT1	Performance	-5.61%***	0.07%	-0.41%	-4.10%***	0.22%	-1.38%***
	[*t*-stat]	[-10.29]	[0.25]	[-1.39]	[-8.51]	[1.03]	[-23.47]
RT2	Performance	-4.61%***	0.43%	-0.46%*	-2.97%***	0.19%	-1.80%***
	[*t*-stat]	[-8.96]	[1.46]	[-1.79]	[-7.63]	[1.12]	[23.07]
RT3	Performance	-3.60%***	0.17%	-0.42%**	-1.77%***	0.14%	-1.72%***
	[*t*-stat]	[-9.49]	[0.85]	[-2.47]	[-7.48]	[1.02]	[-20.89]
RT4	Performance	-2.80%***	0.33%**	-0.44%***	-1.23%***	0.12%	-1.58%***
	[*t*-stat]	[-8.51]	[2.01]	[-3.32]	[-6.66]	[0.65]	[-19.53]
RT5	Performance	-0.29%	2.45%***	-0.56%***	-1.40%***	0.26%	-1.03%***
	[*t*-stat]	[-0.92]	[7.96]	[-3.03]	[-5.32]	[1.18]	[-18.84]
INST	Performance	1.15%***	0.89%***	-0.05%	0.49%***	0.23%	-0.41%***
	[*t*-stat]	[3.57]	[6.71]	[-0.62]	[3.04]	[1.32]	[-20.49]

Panel C. Annual trading performance using tracking portfolios: around earnings announcements

		Total		Liquidity Provision		Liquidity Demand		Trading Cost	
		EA	EA/all days	EA	EA/all days	EA	EA/all days	EA	EA/all days
RT1	Performance	-0.39%***	6.89%	-0.05%	14.43%	-0.30%***	7.60%	-0.04%***	3.06%
	[*t*-stat]	[-3.68]		[-0.86]		[-4.10]		[-5.91]	
RT2	Performance	-0.36%***	7.71%	-0.03%	93.81%	-0.27%***	9.54%	-0.06%***	3.12%
	[*t*-stat]	[-4.08]		[-0.64]		[-3.89]		[-5.90]	
RT3	Performance	-0.24%***	6.61%	-0.03%	11.05%	-0.16%***	9.52%	-0.05%***	3.18%
	[*t*-stat]	[-4.24]		[-0.75]		[-3.44]		[-5.86]	
RT4	Performance	-0.14%***	5.02%	0.00%	0.24%	-0.09%**	8.01%	-0.05%***	3.26%
	[*t*-stat]	[-2.70]		[-0.01]		[-2.06]		[-5.86]	
RT5	Performance	0.10%	-33.28%	0.19%***	9.90%	-0.06%	4.96%	-0.03%***	3.37%
	[*t*-stat]	[0.91]		[3.13]		[-1.18]		[-5.87]	
INST	Performance	0.11%*	9.80%	0.10%***	11.57%	0.03%	4.23%	-0.01%***	3.56%
	[*t*-stat]	[1.81]		[3.48]		[0.67]		[-5.86]	

Appendix Table VIII. Gender and Age

This table examines retail investors by gender and age. The sample period covers January 2019 to March 2019. The sample firms are A-share stocks with at least 15 days of trade during the previous month. Panel A reports the summary statistics of trading volumes across age and gender groups. Panel B reports the return predictions across genders and age groups. To account for serial correlation in the coefficients, the standard errors of the time series are adjusted using Newey and West (1987) method, with optimal bandwidths chosen by following Newey and West (1994). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Summary Statistics of gender and age groups

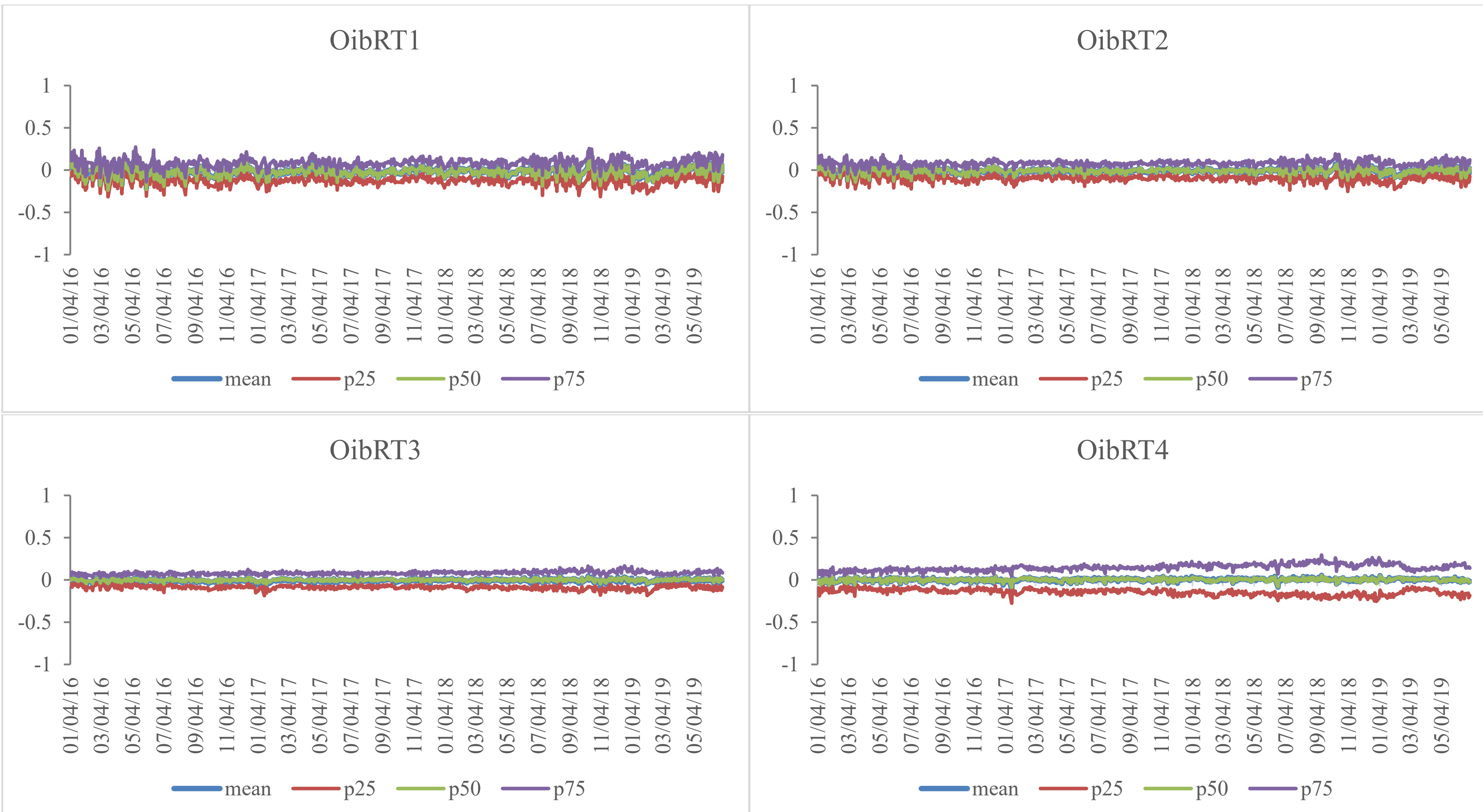
Gender	Trading Volume (% of total)	
Age	<45	>=45
Male	29%	38%
Female	13%	20%

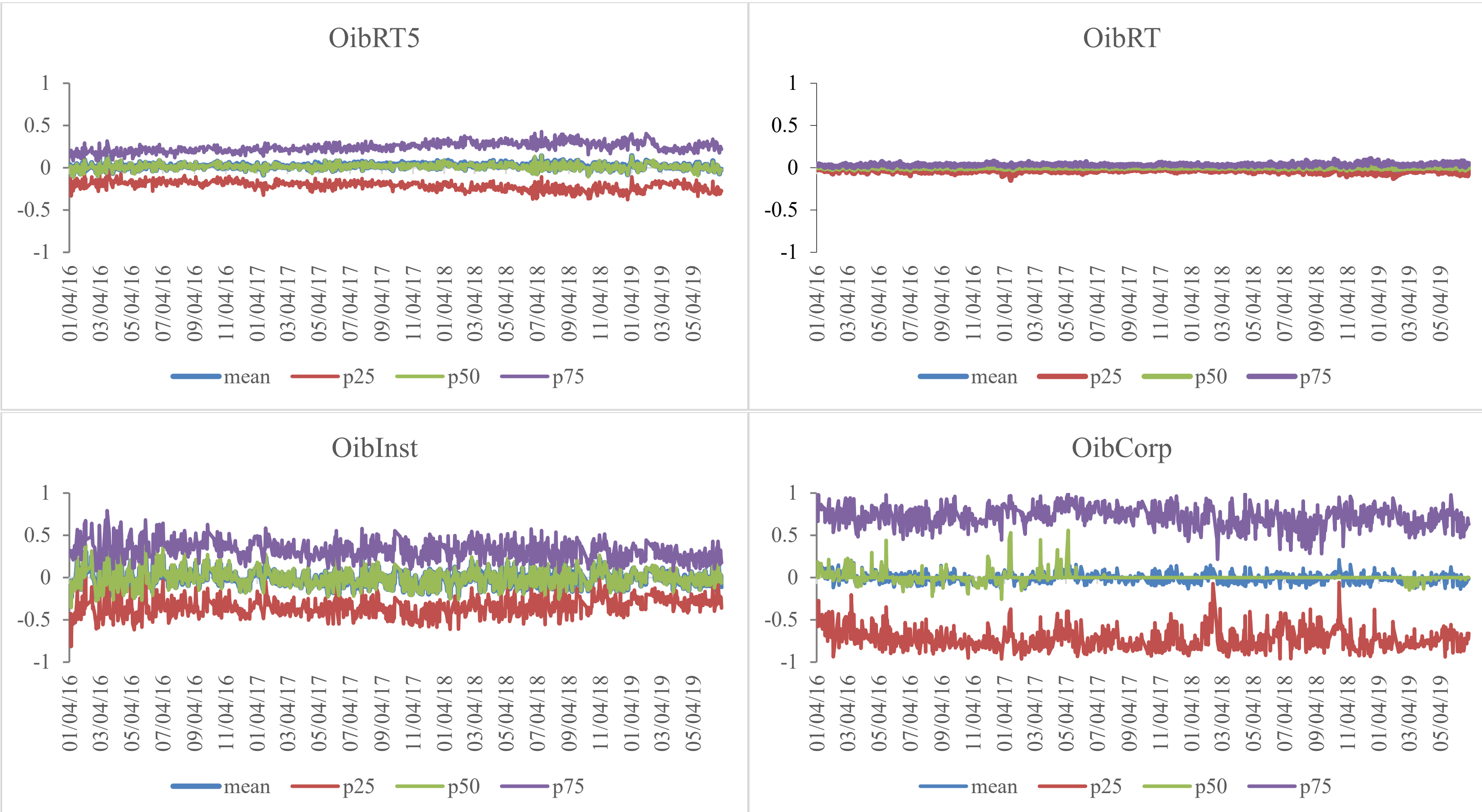
Panel B. Cross-sectional return predictions, by different gender and age groups

Dep.var		Ret	Ret	Ret	Ret
Gender		Male	Male	Female	Female
Age		<45	>45	<45	>45
Oib(-1)	Estimate	-0.0026***	-0.0060***	0.0007	-0.0002
	[*t*-stat]	[-5.45]	[-3.48]	[1.46]	[-0.32]
	Interquartile return	-0.05%	-0.11%	0.02%	0.00%

Appendix Figure I. Time Series of Different Types of Investor Order Imbalance

These figures report the time series of the different types of investor trading activities. The sample period covers January 2019 to March 2019. The sample firms are A-share stocks with at least 15 days of trade during the previous month. We present the cross-sectional mean, median, 25th percentiles, and 75th percentiles of scaled daily order imbalances for each investor group each day. Order imbalance is computed as the buy share volume minus the sell share volume divided by the buy share volume plus sell share volume for each investor group, as specified in Equation (1).



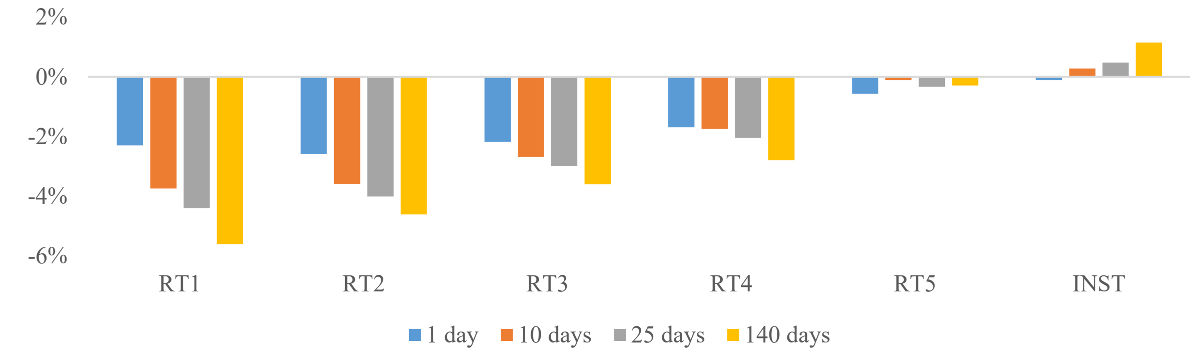


OibRT5
OibRT
OibInst
OibCorp
1
0.5
0
-0.5
-1
01/04/16
03/04/16
05/04/16
07/04/16
09/04/16
11/04/16
01/04/17
03/04/17
05/04/17
07/04/17
09/04/17
11/04/17
01/04/18
03/04/18
05/04/18
07/04/18
09/04/18
11/04/18
01/04/19
03/04/19
05/04/19
mean
p25
p50
p75

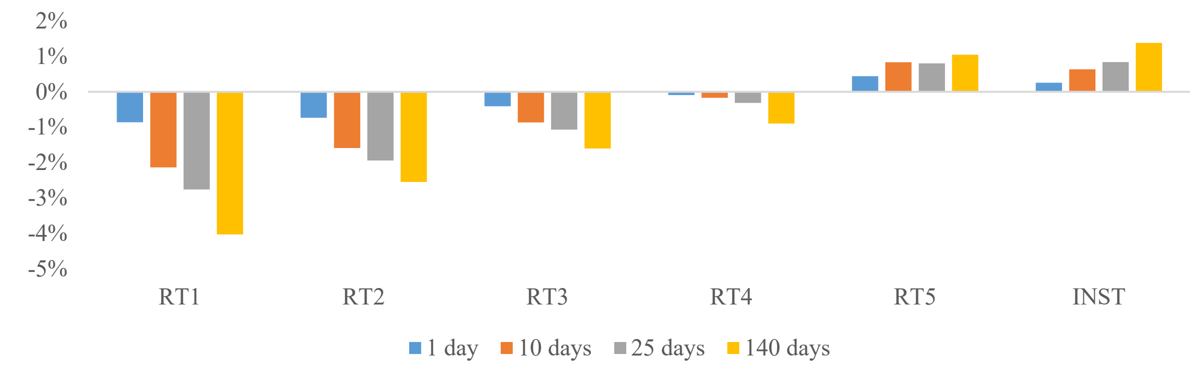
Appendix Figure II. Annual Performances over Different Portfolio Holding Horizon

This table reports the alternative performance for different retail investor groups over different portfolio holding days. Following Barber et al. (2009) method, as specified in equation (A4) to (A10), we construct the portfolios tracking aggregate retail trading and present the annual performance for holding *n* days (*n* = 1 day, 10 days, 25 days, and 140 days). Panel A presents the annual total performance and Panel B and C present stock selection and market timing.

Panel A. Total trading performances over different portfolio holding days



Panel B. Stock selection over different portfolio holding days



Panel C. Market timing over different portfolio holding days