Retail Investors in the Pandemic

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

August 30, 2022

ABSTRACT

Using the Boehmer, Jones, Zhang, and Zhang (2021) algorithm, we identify a broad swath of marketable retail investor orders in the U.S. market between January 2020 and March 2022. During this pandemic period, the retail trading volumes we identify increase from 9% of total market volumes to 11%. The daily retail order flows significantly and positively predict returns for next day and for the next twelve weeks in the cross-section. Meanwhile, the heightened retail activities are associated with wider future effective spreads and higher future volatilities, as well as less activities from high frequency traders and short-sellers. Results using intraday and weekly data are similar.

Keywords: retail investors, order flows, liquidity, volatility, high frequency trading, short selling.
JEL codes: G11, G12, G14, G23

* Charles M. Jones is at Columbia Business School, Xiaoyan Zhang is at the PBC School of Finance at Tsinghua University, and Xinran Zhang is at Central University of Economics and Finance. We thank the Plato Partnership for funding. Xiaoyan Zhang acknowledges financial support from China NSF Project 71790605. We thank Anna Pavlova, Albert J. Menkveld, Sophie Moinas and conference participants at 2022 Plato Market Innovator (MI3) Conference and 2022 Transparency and Market Structure Conference for helpful comments and suggestions. All remaining errors are our own. Corresponding author: Xiaoyan Zhang, PBC School of Finance, 43 Chengfu Road, Beijing, China, 100083, zhangxiaoyan@pbcsf.tsinghua.edu.cn.

Retail Investors in the Pandemic

August 30, 2022

ABSTRACT

Using the Boehmer, Jones, Zhang, and Zhang (2021) algorithm, we identify a broad swath of marketable retail investor orders in the U.S. market between January 2020 and March 2022. During this pandemic period, the retail trading volumes we identify increase from 9% of total market volumes to 11%. The daily retail order flows significantly and positively predict returns for next day and for the next twelve weeks in the cross-section. Meanwhile, the heightened retail activities are associated with wider future effective spreads and higher future volatilities, as well as less activities from high frequency traders and short-sellers. Results using intraday and weekly data are similar.

Keywords: retail investors, order flows, liquidity, volatility, high frequency trading, short selling.
JEL codes: G11, G12, G14, G23

1. Introduction

Retail investors' involvement in the stock market significantly increases after the outbreak of the COVID-19 pandemic in the spring of 2020. Using an algorithm that identifies marketable retail order flow in U.S. stocks (Boehmer, Jones, Zhang and Zhang 2021, BJZZ hereafter), we compute the aggregate marketable retail trading volumes as $3.4 trillion in 2019, accounting for 9% of total market dollar volume. After pandemic starts, the aggregate marketable retail dollar volume doubles to $6.8 trillion in 2020 and steadily increases afterwards (accounting for around 11% of market total volume), with a couple of spikes around the Gamestop episodes in 2021 and Ukraine War in 2022. Many contribute the rise of retail trading to the introduction of retail-oriented investing platforms with zero trading commissions, such as Robinhood, which attract young and unexperienced individual investors, with relatively little wealth but are incentivized to develop investing expertise by actually trading.[1] For instance, Robinhood investors triple their trading and account for 20% of the retail trading volume during the second quarter of 2020, as in Welch (2022).

Given the heightened uncertainties during pandemic, the significant increase in aggregate retail trading and the new mix of retail investors, practitioners, regulators[2] and academic researchers are all interested in the increasingly important role that retail investors play in the stock market. To shed light on these issues, our study concentrates on three research questions: how the new mix of retail investors contributes to price discovery; how they are related to market quality measures during the pandemic; how other market participants, such as short-sellers and high frequency traders, respond to increased retail trading.

[1] Earlier studies on Robinhood investors include Welch (2022), Barber, Huang, Odean, and Schwarz (2022), Ozik, Sadka, and Shen (2022), and Eaton, Green, Roseman, and Wu (2022).
[2] For instance, the chairman of the U.S. SEC, Gary Gensler, publicly expresses concerns regarding retail investors (https://www.sec.gov/news/testimony/gensler-testimony-20210505).

We start by investigating how retail order flows during this period are related to future stock price movements in the cross section. BJZZ examine the U.S. equity market between 2010 and 2015 and find that overall signed retail order flows, or retail order imbalance *(buys minus sells)*, can predict future stock returns positively and significantly, indicating that retail marketable orders contain price-relevant information that is not yet incorporated into prices. After 2016, with the success of commission-free trading platforms, rise of social media as information gathering and distribution channels, and fierce competitions among trading venues, the overall trading environment has significantly changed. With the increased uncertainty and potential downward pressures brought by the pandemic, it is unclear whether overall retail order flow can still predict returns in the cross section. On the one hand, the retail traders might have increased savings from government salvage checks, and they might have more time and channels to acquire information than usual due to stay-at-home working practice and rising of social medias, which can potentially help them to invest. On the other hand, the new entrants of retail investors, mostly Robinhood-type, are less experienced and more susceptible to trading biases, and might make suboptimal investment choices.[3]

Using a sample period between January 2020 and March 2022, we show that overall retail order imbalances still positively and significantly predict future daily returns in the cross section. Economically, an interquartile increase in retail order imbalance is associated with 3.90 bps of higher return for the next day (9.75% annualized), which is highly statistically significant. The predictive pattern persists for the next twelve weeks and is robust across firms with different sizes and different turnover ratios. That is, the higher uncertainty during the pandemic and the arrival of

[3] For instance, Barber et al. (2022) find Robinhood investors engage in more attention-induced trading than other retail investors, and their behavior leads to negative future returns.

more inexperienced retail investors don't change the overall positive predictive pattern of retail trading for future stock returns. This might not be too surprising because Robinhood investors only account for a small part, around 21%, of the overall retail order flows.

We next examine how retail order activities are related to future market quality measures, such as liquidity and volatility, in the cross section. The pandemic brings a sharp market downturn to the U.S. equity market, together with higher volatility and lower liquidity in general. It is important to understand how retail investors, who many assume to be liquidity providers, affect future liquidity and volatility in the market, especially during this volatile period. Earlier studies provide mix evidence on this question. Ozik, Sadka, and Shen (2021) finds that retail trading, using Robinhood sample, significantly attenuated the rise of illiquidity during lockdown of spring 2020; while Eaton, Green, Roseman, and Wu (2022) uses platform outages as exogenous shocks and find decreases to Robinhood investor participation actually are associated with higher market liquidity. Our study includes many more retail investors than Robinhood investors in the market, and find that over our 2+ years of pandemic sample period, increases in overall retail trading activity *(buys plus sells)* are associated with higher effective spreads and higher volatility in the future. An interquartile increase in retail overall retail activity is associated with a 0.94 bps increase in effective spreads and a 0.64% increase in intraday volatility over the next day. The pattern extends over the next 12 weeks at least, indicating that overall retail activities might generally demand future liquidity and generate more uncertainty.

Given that more positive retail order imbalances predict higher future returns, and higher overall retail trading volumes are associated with lower liquidity and higher volatility in the future, how do other important market participants trade in response? The impacts the pandemic brings to professional investors probably significantly differ from those to retail investors. Presumably, the environment of high volatility and low market liquidity during the pandemic lead to decreasing funding liquidity, and it becomes more difficult for professional investors to acquire information with the quarantine practices and work-from-home routines. Here we focus on two groups of relatively sophisticated investors: high frequency traders (HFT) and short-sellers (SS). High frequency traders are generally believed to trade on arbitrage opportunities and are sensitive to short-term changes in prices and liquidity, while short-sellers are assumed to be informed pessimistic investors who collect and process information regarding future price movements. Our empirical results show that the overall activity levels of retail investors are associated with significantly lower activity levels of both HFTs and SS's for at least 12 weeks, as measured by cancel-to-trade ratios for HFT, and days-to-cover ratios for SS. It is possible that increased trading by retail investors makes it harder for HFTs and SS's to trade profitably, and it is also possible that the lower market quality associated with heightened retail trading activity makes it less attractive for HFTs and SS's to trade.

After we collect the predictive patterns of how retail trading is related to future returns, market quality measures and actions from other market participants, we investigate the economic rationale of these predictive patterns. BJZZ examine three alternatives for retail order imbalance's positive predictive power for future returns (order persistence, liquidity provision and information) and find both order persistence and information explanations contribute significantly to the predictive pattern. Other than these explanations, Da, Engelberg, and Gao (2011) proposes that retail investors' attention can drive their trading activities and lead to positive predictive power for future returns, while the attention itself might not necessarily contain firm fundamental information.[4] Following Da et al. (2011), we compute retail attention using Google search volume index, and find retail attention are significantly related to retail order flows, but its contribution to retail's predictive power for next-day returns is not significant. For longer horizons, Da et al. (2011) documents reversals in longer-term return prediction patterns, and interprets it as the predictive information embedded in retail attention might be temporary and not related to firm fundamentals. For our sample period, we find no evidence for prediction reversal over the longer term. Instead, we provide evidence that retail order imbalance significantly predicts earnings news over longer horizon, suggesting that the retail flow might carry information regarding firm fundamentals.

Finally, we conduct a battery of subperiod and subsample analysis. Given that our sample contains 27 eventful months, we divide the sample into subperiods to examine whether the predictive patterns differ over Covid-19 shocks, GME periods, and Ukraine war. All three events accompany increases in market illiquidity and volatility. Interestingly, the predictive power of retail order flow for future returns is significantly stronger for the outbreak of the pandemic, but weaker for the other two subperiods, suggesting that retail order flow probably contain more price relevant information for the outbreak of Covid-19. Meanwhile, retail trading is associated with much higher future effective spreads for the outbreak of the pandemic, but lower for the GME episode, implying that retail traders presumably demand liquidity at the outbreak of the Covid-19, but possibly provide liquidity for the GME event. We also study subsamples of stocks to separate firms with various sizes, and turnover ratios. Finally, our main results are based on daily data, so we examine the main patterns using intraday data and weekly data and find similar results.

[4] Hendershott et al. (2022) examine the asset price dynamics with limited attention, and find inattentive investors arrive stochastically to trade could also positively affect the stock price in subsequent periods.

Our paper is closely related to the vast retail investor behavior literature, and classical papers include Barber and Odean (2000, 2008), Kaniel, Saar, and Titman (2008), Kelley and Tetlock (2013), Barrot, Kaniel, and Sraer (2016) and others. The recent studies on retail investors mostly focus on Robinhood investors, which includes Pagano, Sedunov, and Velthuis (2021) and Welch (2022) on retail trading and price movements, Barber, Huang, Odean, and Schwarz (2022) on investor behaviors, Ozik, Sadka, and Shen (2021) and Eaton, Green, Roseman, and Wu (2022) on liquidity, and Hüfner, Strych, and Westerholm (2022) on crash risk. Other than Robinhood investors, Ortmann, Pelster, and Wengerek (2020) use study retail trading behaviors using transactional-level brokerage data from August 2019 to April 2020, and Bryzgalova, Pavlova, and Sikorskaya (2022) examine retail options trading during the pandemic and find them participating in trading frenzies.

Compared to the existing literature, our study is the first to examine the trading patterns of retail investors in general, instead of the Robinhood sample, during the pandemic and to provide an overall picture on how retail trading is related to future prices and market quality measures. Our unique and thorough findings, such as overall retail trading positively predicts the future stock return, higher retail activity is associated with wider future effective spreads and higher future volatilities, as well as lower activities of both HFTs and SS's, not only complement existing literature, but also provide significant references for future actions by practitioners, regulators and researchers.

2. Data and Empirical Methodology

2.1 Data and Measures

Our sample starts on January 1st of 2020, and ends on March 31st of 2022, a total of 567 trading days. Following existing literature, we collect trading data from TAQ and merge with stock returns and accounting data from CRSP and Compustat, respectively. We adopt the conventional filters by including only common stocks with share code 10 or 11 (which excludes mainly ETFs, ADRs, and REITs) that are listed on the NYSE, NYSE MKT (formerly the Amex), or NASDAQ, and removing low-priced stocks and guarantee a non-trivial minimum tick size by requiring the minimum stock price to be $1 on the previous trading day. For each day, we have an average of around 3,000 firms included in the sample. Overall, we have 1.2 million stock-day observations.

We identify a large subset of marketable retail order flows in the U.S. following the BJZZ algorithm, which relies on the special subpenny setup in the U.S. stock market. That is, transactions with a retail investor tend to be executed off-exchange and reported on a TRF at prices that are just above or below a round penny due to the small price improvement given by the trade counterparty. For all trades reported to a FINRA TRF (exchange code "D" in TAQ), let P_{it} be the transaction price for stock i at time t, and let $Z_{it} \equiv 100 * \text{mod}(P_{it}, 0.01)$ be the fraction of a penny associated with that transaction price. Variable Z_{it} can take any value in the unit interval [0,1). If Z_{it} is in the interval (0,0.4), we identify it as a retail sell transaction, because the transaction price is just slightly above the round penny, which potentially is a small price improvement for the retail seller. If Z_{it} is in the interval (0.6,1), then the transaction is coded as a retail buy transaction, because the transaction price is just slightly below the round penny, which potentially is a small price improvement for the retail buyer. To be conservative, transactions at a round penny ($Z_{it} = 0$) or near the half-penny ($0.4 \leqslant Z_{it} \leqslant 0.6$) are not assigned to the retail category.[5] With BJZZ

[5] The BJZZ retail identification algorithm captures most retail marketable orders, which normally are the more aggressive orders, while retail limited orders are excluded. According to the 606 filings by Charles Schwab, one of the largest retail broker, limit orders account for 32% of all orders, and marketable orders account for more than half. From a cross validation test using Nasdaq data, the BJZZ algorithm matches the NASDAQ TRF's correct buy/sell sign 98% of the time, which demonstrates its accuracy. Barardehi et al. (2022) argue that the BJZZ retail order imbalance also reflects the opposite of institutional order imbalance.

algorithm, on each day *t* for each stock *i*, we define retail buy and sell volumes as "$retailbuyvol_{it}$" and "$retailsellvol_{it}$". In this study, we measure retail activity from two perspectives. First, we compute signed retail order imbalances measures, $Oibvol_{\mathrm{it}}$ as follows:

$$Oibvol_{\mathrm{it}} = \frac{retailbuyvol_{it} - retailsellvol_{it}}{retailbuyvol_{it} + retailsellvol_{it}}. \qquad (1)$$

The order imbalance measures reflect net buy and sell directions for retail investors, and is mainly used for predicting future stock price up and down movements. Second, we measure total activity of retail investors, by summing up both retail buys and sells and comparing with the total buys and sells from all investors for that stock ($totalvol_{it}$), as follows:

$$Actvol_{\mathrm{it}} = \frac{retailbuyvol_{it} + retailsellvol_{it}}{totalvol_{it}}. \qquad (2)$$

The total retail activity measures reflect activeness of retail investors, and we use them mostly for predict market quality measures, such as liquidity and volatility, which are affected by both buys and sells rather than the net direction.[6]

Table I presents summary statistics on the retail order flows measures and other firm characteristics for our sample firms. We compute the mean, standard deviation, median, and 25th and 75th percentiles of the pooled stock-day sample. The retail order imbalance measure, *Oibvol*, has a mean of -0.0208, and a standard deviation of 0.3967. The small magnitude of mean and relatively large standard error are consistent with previous literature, suggesting that retail investors' trades mostly cancel each other on average, yet with a large dispersion in the cross section. For the retail activity measure, *Actvol*, the mean is 0.0904 with a standard deviation of

[6] We also consider using number of trades rather than share volumes in equation (1) and (2). Results are similar to those using share volumes and are available on request.

0.1002. That is, for an average stock on an average day, the retail investors using marketable orders contribute around 9% to the total trading activities, which is sizable.

To understand the aggregate magnitude and trend of marketable retail order flows in our sample period, in Figure I Panel A we plot the monthly time-series of aggregate retail trading volumes and compare with the market total trading volumes between January 2019 to March 2022. For 2019, monthly retail trading is mostly $283 billion (with an annual total of $3.4 trillion), and account for 9% of total market trading volume. After the Covid-19 outbreak in the U.S., monthly retail trading jumps to $568 billion for March 2020, and the annual total retail trading becomes $6.8 trillion for 2020, accounting for 11% of the total market trading volume. During the episode of GME in January 2021, we observe another surge of retail trading with a monthly total of $707 billion, accounting for 11% of market total volume for that month. Finally, when Ukraine War breaks out in February 2022, retail trading bounces again to $650 billion in February and $712 billion in March 2022. The time series of total retail trading clearly shows two patterns: retail trading generally increases and remains high during the pandemic period, and there are distinctive spikes around major market events.

We report the cross-sectional distribution of marketable retail order imbalances in Panel B, and retail activities in Panel C of Figure I. For each day of our sample, we compute the mean, 25th percentile, the median, and the 75th percentile of marketable retail order imbalances. For *Oibvol* in Panel B, the means and medians are close to zero, while the 25th percentiles are mostly around -0.2, and the 75th percentiles are mostly around 0.2. There is an obvious dip in March 2020 due to selling pressures for large market downturns, but no other obvious time trends or structural breaks are found. For the activity measures in Panels C, the ranges are mostly between 2% for the 25th percentile to 10% for the 75th percentile, while the time series for 75th percentile displays a slight upward trend, indicating that the surge in retail activity might be mostly driven by a smaller number of stocks preferred by retail investors. There's also a pattern of quarterly dips in the *Actvol* measures, which coincides with the quarterly witching days. It's likely that institutions rebalance significantly around these days, and the relative importance of retail investors drop accordingly.

For our main empirical results, we connect the retail trades future stock returns, liquidity and volatility, and trading from other market participant, and below we explain how we construct these measures in details. For returns, Blume and Stambaugh (1983) find that daily returns computed from end-of-day closing prices can generate an upward bias, due to bid-ask bounce, and recommend to use end-of-day bid-ask average prices to compute daily returns. Therefore, our study uses daily returns computed from end-of-day bid-ask average prices.

Previous studies use many alternative liquidity measures, such as effective spread, quoted spread, price impact and realized spreads. Given that trades can happen within the quoted bid and ask prices, the effective spread is more precise and we choose it as our main liquidity measure. Results using alternative measures are similar and available on request. To be specific, for the *k*-th trade for stock *i* on day *t* (out of a total of *N* trades for stock *i* on day *t*), the proportional effective spread is defined as , $EffSpd_{i,t} = \frac{1}{N}\sum_{k=1}^{N}\frac{2D_{i,k}(P_{i,k}-M_{i,k})}{M_{i,k}}$, where $D_{i,k}$ is set to +1 for buyer-initiated trades and -1 for seller-initiated trades using the Lee and Ready (1991) algorithm, $P_{i,k}$ is the transaction price, $M_{i,k}$ is the midpoint of the NBBO quote assigned to the *k*th trade. Higher effective spreads indicate lower liquidity. Our data on effective spread is obtained from WRDS Intraday Indicators, which utilizes intra-day trades and quotes data from TAQ, and applies filters and adjustments as in Holden and Jacobsen (2014).

For volatility, we follow Van Kervel and Menkveld (2019) and Boehmer, Fong, and Wu (2021) and adopt the algorithm in Andersen, Bollerslev, Diebold, and Ebens (2001). The annualized intraday volatility measure for stock i on day t is defined as $IntVol_{i,t} = \sqrt{250\sum_{m=1}^{T}\left(Ret_{i,m}\right)^2}$, which is the square root of the sum of squared end-of-minute returns, $Ret_{i,m} = Ln\frac{P_{i,m}}{P_{i,m-1}}$. Here m refers to minute, T is the number of minutes that stock i trades on day t, and we assume there are 250 trading days in one year.

We obtain high frequency trading data from WRDS SEC MIDAS. Hendershott, Jones, and Menkveld (2011), Hasbrouck and Saar (2013) and Weller (2018), propose that both "cancel to trade ratio" and "order to trade ratio" are valid proxies for high frequency trading. Since the two measures have a correlation of 0.80 in our sample, we mainly present results on "cancel to trade ratio". To be more specific, $HFTcancel_{i,t}$, the cancel to trade ratio, is calculated as the logarithm of number of full or partial cancellations divided by numbers of trades. Since high frequency traders tend to first place many orders to measure the depth of the market and then cancel them, higher cancel-to-trade ratios indicate higher level of activities from high frequency traders.

Finally, we obtain daily short-selling data from WRDS MARKIT. Our main results use the days-to-cover-ratio, as proposed by Hong, Li, Ni, Sheinkman, and Yan (2016), which is a standard measure capturing information from both supply and demand sides of equity loans. We compute shorts' days-to-cover-ratio, $SDTCR_{i,t}$, as the total number of shares on loan scaled by the daily trading volume. High *SDTCR* indicates more shares on loan, and high activity by short-sellers. Results using alternative measures, such as short supply and short flow, are available on request.

The summary statistics in Table I provide mean and standard deviations for all key variables. To illustrate the cross-sectional pattern over time, we also plot their cross-sectional distributions in Figure II, where we present the cross-sectional means, p25 (the 25th percentile), p50 (the 50th percentile), and p75 (the 75th percentile) for each day. In Panel A, the *EffSpd* is expressed in % and it spikes drastically from around 0.4% to over 1% in March 2020, and remains high until May 2020, which reflects the quick and lasting dry-up of stock market liquidity at the outbreak of the pandemic. For GME episode in 2021 and Ukraine War in 2022, we observe similar but relative smaller spikes in effective spread, from around 0.32% to 0.42% in GME episode, and from around 0.36% to 0.42% in Ukraine War. We present time-series of volatility in Panel B, and observe a sharp increase of *IntVol* during March and April of 2020. Volatilities are relatively higher around the GME episode and Ukraine War. For the high frequency cancel-to-order ratio in Panel C and shorting days-to-cover ratio in Panel D, the cross-sectional distributions are relative stable over time and there is no obvious time trend.

2.2 Empirical Method

Our study aims to understanding how retail trades predict future movements in stocks prices, liquidity, volatility and activities from other market participants. To establish the predictive relation, we mainly use current retail trades to predict next period movements in prices, and other variables. In a regression framework, retail measures are our main independent variables, while the prices, market quality measures, and activities from other investors are the dependent variables. Econometrical studies, such as Petersen (2009), suggest that for dependent variables with low time-series persistence, such as returns, Fama-MacBeth regression is a suitable choice; while for dependent variables with high persistence over time, such as liquidity, volatility, high frequency trading and short selling measures, [7] panel regression is preferred.

[7] The auto-correlation coefficients, AR(1), for effective spread, intraday volatility, cancel to trade ratio and days-to-cover-ratio are 0.86, 0.81, 0.70, and 0.80 respectively.

To be more specific, for predicting returns, we adopt the standard two-stage Fama-MacBeth regression approach, similar to the one in Boehmer, Jones, and Zhang (2008). That is, for each day *t*, we estimate the following cross-sectional specification:

$$Y(i,t) = a0(t) + a1(t)Retail(i,t-1) + a2(t)' \; Controls(i,t-1) + u1(i,t). \quad (3)$$

Here the dependent variable $Y(i,t)$ refers to returns of stock *i* at time *t*, and the independent variable, $Retail(i,t-1)$, refers to retail order imbalance measures from previous day, $Oibvol(i,t-1)$. We obtain the time-series of parameter estimates $\{a0(t), a1(t), a2(t)'\}$ from the cross-sectional regressions, and conduct inferences on the mean and standard errors of these parameter estimates, while the standard errors are adjusted using Newey-West (1987) approach with five lags for daily regressions.[8] If retail trading can predict future returns in the correct direction, we expect to a significant and positive coefficient of $a1$, the time-series average of $a1(t)$.

For predicting liquidity, volatility, HFT and SS measures, we follow the panel set up as in Eaton et al. (2022):

$$Y(i,t) = b0 + b1Retail(i,t-1) + b2' \; Controls(i,t-1) + \gamma_t + u2(i,t). \quad (4)$$

Here the dependent variable, $Y(i,t)$, represents the liquidity or volatility or HFT or SS measures for stock *i* at time *t*, and the independent variable, $Retail(i,t-1)$, represents retail trading activities from previous day, $Actvol(i,t-1)$. Notice we include time fixed effect, γ_t, to control for pure time variation across days. To control for firm effect, we directly include lagged depending variables to simultaneously capture the firm level effect and allow for time-series dynamics. Following previous literature, the standard errors are double clustered at day and firm

[8] Following Newey and West (1994), we use integer $[4(T/100)^{2/9}]$ to calculate the optimal lag. With number of days in our sample *T*=567, our optimal lag is five. Our results are robust if we use other algorithms to compute the optimal lag.

level. If retail activities are related to future liquidity, liquidity, HFT or SS activities, we expect to coefficient $b1$ to be significant.

We follow existing studies to select control variables that can potentially affect future stock prices, liquidity, volatility and other measures. Our control variables include the following: the log market capitalization from the previous month, *Lsize*; log book-to-market ratio at the most recent quarter end, *Lbm*; last month's consolidated trading volume as a fraction of outstanding shares *Lturnover,* and the previous month's daily return volatility following Ang, Hodrick, Xing, and Zhang (2006), *Lmvol*. We also include the lag of the dependent variables as controls. In the case of return prediction, we include previous day's return, *Ret*(-1), the return over the past week *Ret*(*w*-1), and the return over the past month *Ret*(*m*-1). For the other cases, we include lagged dependent variables as controls, and decide the lag length by AIC and SIC. The summary statistics of the control variables are presented in Table I, and they are consistent with previous literature.

The set-up in equation (3) and (4) is quite flexible and can be extended to various horizons. Our main results use daily data. When we examine return reversal patterns, we extend equation (3) and (4) to next 12 weeks. For robustness check, we also examine intraday retail order flows in different 30-minute time buckets using equation (3) and (4).

3. Empirical Results: Predictive Patterns

3.1 Retail Flows and the Cross-Section of Future Returns

We first use retail order imbalance to predict next day return, as in equation (3), and report the estimation results Table II Panel A. The coefficient on *Oibvol* is 0.0010, with a *t*-statistic of 9.85. The positive and significant coefficient indicates that, if retail investors buy more than they sell on a given day, the return on that stock on the next day is significantly higher. For economic magnitude, given that the daily interquartile of *Oibvol* is 0.3902, so when we move from the 25th

to the 75th percentile in the cross section of *Oibvol*, the daily return increases by 0.3902*0.0010 = 3.90bps (3.90*250=9.75% annualized). We compare the magnitude with the results in BJZZ, with the interquartile weekly return difference being 0.1089% (0.1089%*52=5.67% annualized). These magnitudes are comparable, and the magnitude is even larger during the pandemic than for the period of 2010-2015. That is, previous day marketable retail order imbalances significantly predict next-day stock price movements in the correct direction. For the control variables, we observe negative and significant coefficients on the previous day and previous week's return, which indicate daily as well as weekly return reversals. The coefficients on previous month return is insignificant. Size, book-to-market, turnover, and volatility all carry the expected signs, and most are not statistically significant. This confirms that the predictability we find is not a manifestation of size, book-to-market, turnover, or volatility anomalies. The average adjusted R2's from the first stage cross-sectional estimation are mostly around 7.86%.

Given the strong positive predictive pattern of retail order imbalance for next day return, it is natural to ask whether the predictive power persists over longer horizons. If the predictive pattern quickly disappears or reverses, retail investors may be capturing short-term information or driven by temporary fads; if the predictive pattern persists, then retail order imbalance might contain longer term information, such as firm fundamentals. Therefore, we extend equation (3) to the next 12 weeks. That is, we use previous day's retail order imbalance measures, $Retail(i, t-1)$, to predict cumulative returns over the next k weeks, $Ret(i, w+k)$, with k=1 to 12. To be specific, for k=1, $Ret(i, w+1)$ is the cumulative return over days *t*+1 to *t*+5; for k=12, $Ret(i, w+12)$ is the cumulative return over days *t*+1 to *t*+60. If marketable retail order imbalances have only short-lived predictive power for future returns, we might observe the coefficient $a1$ decrease to zero quickly. Alternatively, if the marketable retail order imbalance has longer predictive power, the coefficient $a1$ should remain statistically significant for a longer period.

We report the results in Panel B of Table II. To make sure the coefficients are comparable across different horizons, we scale each cumulative returns by number of days involved and the coefficients reflect the predictive power over daily horizons. When the estimation horizon is extended from one to 12 weeks, the coefficient on *Oibvol* gradually decreases from 0.00032 to 0.00009. If we compare the magnitude with the results in BJZZ, with first week interquartile return annualized difference being 5.67% in BJZZ, and is 3.25% in this study, the magnitude is smaller but comparable during the Pandemic. When we extend the window to 12 weeks, the coefficients gradually decrease over time, but there are no reversal patterns within 12 weeks, which is quite similar to findings in BJZZ. Given that the predictive power of the retail order imbalances persists for multiple weeks, they may potentially capture longer-term information.[9] [10]

The general finding of previous day retail order flow positively predicting next day and next 12 weeks of returns might be surprising to readers who have been paying attention to Robinhood investors, and their lack of trading experiences. However, even though Robinhood investors attract substantial attention from media and regulators, according to Welch (2022), they overall still only account for around 21% of the total retail order flows at the outbreak of the

[9] The BJZZ algorithm is based on the pay-for-order-flow (PFOF) practice, where wholesalers offer price improvement to retail investors to obtain their order flows. Readers might be concerned that if retail flows predict return positively, the counter-party would potentially lose money. BJZZ offers an economic explanation for why wholesalers would still like to pay for retail order flows on page 2299. In short, as long as the information content of marketable retail order flow is less than the bid-ask spread being charged, internalizers and wholesalers can still earn positive revenues by trading with these retail orders.

[10] We follow Barber, Lee, Liu, and Odean (2008)'s method to calculate the retail investors gain/loss over the sample period. That is, if we form net buy and sell portfolios based on retail order flow and assume a holding period of 140 days, the daily gain from this strategy is 0.45 million, or 257 million over our sample period.

pandemic.[11] Earlier studies, such as Da et al. (2011), Kelley and Tetlock (2013), and BJZZ, all find that retail order imbalance has strong positive predictive power for future returns. Both Kelley and Tetlock (2013) and BJZZ provide evidence that retail orders might contain relevant information regarding firm fundamentals. However, the predictive pattern doesn't necessarily mean that retail investors possess some private information or have advantages in information processing. For instance, Da et al. (2011) shows that retail attention-driven trades, rather than information-driven trades, can still predict future returns, but the predictive pattern would reverse over the longer term. In this study, we first establish the retail predictive patterns for returns, volatility, liquidity and other investors trading in Section 3, to have a more comprehensive view at the matter, and then we examine economic hypothesis for these predictive patterns in Section 4.

3.2 Retail Activities and the Cross-Section of Future Liquidity and Volatility

Are the activity of retail investors associated with future firm level liquidity and volatility? In this section, we examine whether retail activity measures can predict future liquidity and volatility in the cross section. Previous studies provide mixed evidence on this question. On the one hand, Barrot et al. (2016) argue that retail investors provide liquidity during financial crisis, when the VIX is high. Ozik et al. (2021) find that Robinhood investors significantly attenuate the rise of illiquidity during the Covid lockdown in early 2020. On the other hand, Eaton et al. (2022) uses Robinhood outages as exogenous shocks and find the decreased Robinhood participation are associated with higher market liquidity and lower volatility.

[11] Table I of Welch (2022) use payment-for-order-flow to impute the trading volume, and shows that during the second quarter of 2020, the implied Robinhood trading volume is 41 billion, while the implied trading volume from the four largest retail brokers (which covers the majority of the retail flows) is 199 billion. Based on these numbers, order flows from Robinhood account for 21% (41/199 = 21%) of total retail order flows. Figure A1 of Bryzgalova et al. (2022) also shows that Robinhood takes around 20% of all brokerages' payment for order flow from 2020 to 2021.

To answer this question, we estimate equation (4) in the panel setting, with the $\mathrm{Y}(i,t)$ variable being $EffSpd(i,t)$ or $IntVol(i,t)$, proxying for liquidity or volatility, respectively.[12] In terms of liquidity, if retail investors provide liquidity to the market, then higher retail activity would be associated with lower effective spread in the future, and the coefficient $b1$ would be significantly negative, and vice versa. Similarly, if retail trading stabilizes the market, we expect to see that higher retail activity to be associated with lower volatility for the future, and coefficient $b1$ would be significantly negative, and vice versa.

We report the estimation results in Table III Panel A. In regression I, we use *Actvol* and to predict future effective spread, and the coefficient on *Actvol* is 0.1017, with a significant *t*-statistic of 10.22. In terms of economic magnitude, given that the daily interquartile of *Actvol* is 0.0929, when we move from the 25^{th} to the 75^{th} percentile in the cross section of *Actvol*, the daily effective spread increases by 0.0929*0.1017 = 0.0094 percent or 0.94 bps, which is sizeable, given that the median effective spread is 16 bps for our sample in Table I. The positive coefficients indicate that, more activities from retail investors are associated with higher effective spread, or lower liquidity on the next day. It is possible that excessive retail activity dries out the liquidity in the market, which leads to higher effective spreads for the next day. In regression II, when we use *Oibvol* to predict next day effective spread, the coefficient is also positive and significant, but with smaller economic magnitude.

When we use our retail trading measures to predict next-day stock intraday volatility in regression III, the coefficient on *Actvol* is 0.0690, with a *t*-statistic of 5.46. Economically, an interquartile movement for *Actvol* (0.0929) is associated with an increase of intraday volatility of 0.0929*0.0690 = 0.0064 or 64 bps. The positive and significant coefficient shows that, if more

[12] We choose to include two lags of the dependent variable as controls, based on AIC or SIC.

activities from retail investors on a given day, the next day intraday volatility for that stock would significantly increase. When we use *Oibvol* to predict next day intraday volatility in regression IV, the coefficient is also positive and significant, but the magnitude is smaller. That is, for predicting future liquidity and volatility, retail overall activeness, rather than direction of retail orders, have stronger and larger predictive powers.

Do these daily positive predictive patterns persist in the long run? If the positive coefficients quickly reverses or diminishes to zero, then retail investors' trading activity only has temporary impact on market quality; if the positive coefficient continues to be significant over longer horizons, then the retail investors' trades affect the market in a lasting way. To examine this issue, we replace the dependent variables in equation (4) from next-day market quality measures to average weekly market quality measures for the next 12 weeks.

We report the results in Table III Panel B. In the first column, we predict future liquidity for the next one to 12 weeks using *Actvol*. The coefficients gradually decrease from 0.0810 to 0.00612 at the 12th week, while the *t*-statistics remain highly significant for all weeks, indicating that *Actvol* has long-lasting positive predictive power for future liquidity. In the second column, when we use *Oibvol* to predict liquidity, the coefficients are small and never significant for the next 12 weeks, indicating that the trading direction of retail investors probably only contain short term information about liquidity. In the third and fourth columns, we use *Actvol* and *Oibvol* to predict future intraday volatility, respectively. The patterns are quite similar to those in column I and II. That is, *Actvol* have lasting positive predictive power for future volatilities, while *Oibvol* only have short run predictive power. Overall, the above results show that while retail order imbalances have strong and long term predictive power for returns, they fail to predict long term market quality measures. In contrast, retail investors' activities are significantly associated with future illiquidity and uncertainties, over short or long horizons, with higher retail activities associated with higher illiquidity and higher volatility.

3.3 Retail Activities and the Cross-Section of Future High Frequency Trading and Short Selling

There are many participants in the stock market. Given the rise of retail investors during the pandemic, how do their activities affect the behavior of other participants? In this section, we focus on two important subsets of institutional investors, the high-frequency-traders and short-sellers, who tend to be quite sensitive to the market quality and information in the prices. On the one hand, high frequency traders are believed to be trade on arbitrage opportunities and improve the market quality (Hendershott et al. 2011), and short sellers are assumed to be informed pessimistic investors (Boehmer et al. 2008), they may trade against retail investors and reap profit (Barber et al. 2022); while on the other hand, the increased trading by retail investors, together with their associations with low liquidity and high volatility, might make it harder for HFTs and SS's to trade profitably, and they might reduce their participation in the market. We focus on the question how retail order flows and overall activities is related to next day activities from HFTs and short-sellers. In particular, we estimate our benchmark regression in equation (4), with the dependent variables being *HFTCancel* for high frequency trading, and *SDTCR* to for short-selling.

We report the estimation results in Table IV. We first investigate how retail trading is related to next day HFT and SS activities in Panel A. In column I, we predict next day *HFTCancel* with *Actvol*, and the coefficient is -0.2300 with a *t*-statistic of -17.12. Economically, an interquartile change in *Actvol* lead to a decrease in *HFTCancel* by -0.0214. The negative and significant coefficient implies that high retail activities are associated with lower activities from high frequency traders on the next day. It is possible that the lower market quality associated with heightened retail trading activity makes it less attractive to HFTs to trade, or large retail activity makes it harder for HFTs to trade profitably. In regression II, when we use *Oibvol* to predict next day high frequency trading, the coefficient is also negative and significant, but with smaller magnitude.

We next examine how next day short selling, *SDTCR*, is associated with prior retail measures in Panel A. In column III on Panel A, the coefficient on *Actvol* is -1.1141 with a *t*-statistic of -12.28. An interquartile movement of *Actvol* is associated with a decrease of -0.1035, in SDTCR. From the summary statistics in Table I, the mean of *SDTCR* is 4.26 days, and a decrease of -0.1035 day account for -0.1035/4.26 = -2.43% decrease in shorting activity. Similar to HFT result, when retail investors trading more on a given day, there is less shorting of that stock the next day. In regression IV, when we use *Oibvol* to predict next day short selling, the coefficient is negative and significant, but with smaller economic magnitude.

Do these negative coefficients last for the longer term? We present the long term predictions in Panel B, where we replace the dependent variables in equation (4) from one day ahead HFTs and SS's to weekly average HFTs and SS's for the next 12 weeks. In the first column, where we predict future HFTs for the next one to 12 weeks, the coefficient on *Actvol* gradually decreases from -0.1002 to -0.0756 at the 12^{th} week, and the *t*-statistics remain highly significant for all weeks. However, the coefficients of *Oibvol* in the second column is only significant for the first week. In the third column, we use retail activities to predict future short selling, and the coefficients of *Actvol* are always negative and significant. Again, in the fourth column, when we predict future volatility using *Oibvol,* it is only significant for week 1. To summarize, both retail order imbalances and retail activities negatively predict future activities by HFT and SS, while the predictive power is much stronger and long lasting for the activity measures. Given that higher retail trading activity is associated with wider effective spreads and higher volatility, they might make trading from HFT and SS more difficult or less profitable, and that's why we observe the decreases in their activities.

4. Economic Interpretations and Robustness

4.1 Retail trades and Attention

Results in Section 3.1 show that retail order flows positively significantly predict future returns or up to the next 12 weeks. There are multiple explanations for retail's positive predictive power. BJZZ examines three alternatives (order flow persistence, liquidity provision and information), and find the results are more consistent with the flow persistence and information hypotheses. In this study, we shift our focus to an alternative hypothesis, attention, which is not examined by BJZZ, but can also be consistent with the positive predictive pattern in the data. Barber and Odean (2008) first provide evidence that attention drives retail buys, and some of this attention is not related to firm fundamental news. Da et al. (2011) proposes a direct measure of retail investor attention, the Google Search Volume Index (*SVI*), and show that an increase in *SVI* predicts higher return for the short term such as 2 weeks. But since the *SVI* doesn't necessarily contain firm fundamental information, the return prediction reverses after 52 weeks. The most recent study by Barber et al.(2022) finds Robinhood investors engage in more attention-induced trading than other retail investors and the herding events predict negative return. In this study, we follow Da et al. (2011) and proxy retail attention by Google search volume index (*SVI*). To capture the changes in retail attention, we follow Da et al. (2011) and compute *ASVI*, which is the log *SVI* during the current week minus the log median *SVI* during the previous eight weeks.

We examine the relation between retail trades and attention in three steps. For the first step, we investigate whether retail order imbalances and total retail activities are related to retail attention. Previous studies find that retail order flows are related to firm characteristics, past returns, and its own lag. Here we add in retail attention from previous day and check whether it also contributes to retail trades. To be specific, we project retail order imbalance and retail activity measures onto the retail attention variable from previous day. Results are presented in Panel A of Table V. Column I is our benchmark without the *ASVI* variable. We find retail order imbalance is significantly related to its own lag, past returns, firm size and other firm characteristics, and the adjusted R2 is 0.0032. When we include *ASVI* in column II, the coefficient on *ASVI* is 0.0091 with a significant *t*-statistic of 10.44, but the R2 only slightly increases 0.0035. That is, higher retail attention significant contributes to higher order imbalance, but the explanatory power is not large. Results for *Actvol* in column III and IV are largely similar.

In our second step, we examine how much retail attention attributes to retail's predictive power for future returns and other variables. Take future returns as an example. We first decompose the retail order imbalance variable into attention-related component (*Att*) and attention-orthogonal (*AttOrth*) component, then re-estimate the predictive regression in equation (3). If retail attention drives the predictive power of retail order imbalance for future returns, then coefficient on *Att* would be significant; or if the predictive power is driven by elements other than attention, then coefficient on *AttOrth* would be significant.

Results are reported in Panel B of Table V. In column I, the coefficient on *Att* is 0.3337with a *t*-statistic of 1.46. That is, attention is related to the positive predictive power of retail order imbalance for next-day return, but it is not significant. Meanwhile, the coefficient on *AttOrth* is 00009 with a *t*-statistic of 9.32, which is similar to the statistics for the *Oibvol* in Table II Panel A, indicating that the part of order imbalance that's orthogonal to retail attention contribute significantly to its predictive power for returns. For the liquidity results in the second column of

Panel B, it is quite interesting to find that the coefficient on *Att* is -0.1022 (*t*-stat=-2.90), and the coefficient on *AttOrth* is 0.1077 (*t*-stat=10.79), indicating that attention-driven retail activity actually provides liquidity and leads to lower effective spread, while the retail activity driven by the orthogonal component demands liquidity and leads to higher effective spread. We present results on volatility, HFT and short-selling in the rest of the panel. Basically, higher *Att* and *AttOrth* both lead to higher volatility, lower HFT activity and short-selling activity, all with statistical significances.[13]

Our third step focuses on whether retail flow contains information regarding firm fundamentals, by directly examining whether retail trade can predict the most important firm level information, the earnings news. Following Kelley and Tetlock (2013), we use cumulative abnormal returns (*CAR*) over earnings announcement periods to estimate earnings news, where the abnormal returns are the returns in excess of expected return using the market model. If retail order flows contain information about earnings news, then the *Oibvol* should predict the earnings news positively and significantly. From results in Panel C of Table V, for various event windows around earnings announcements, the coefficients on *Oibvol* are always positive and mostly significant, which provide direct evidence that retail order imbalance might contain information related to firm fundamentals.

To summarize, we find evidence that retail attention contributes significantly to retail trading, and its prediction for future market quality measures. We also provide evidence that retail order flow contains information regarding firm earnings news.

[13] Following Da et al. (2011), we also examine the reversal pattern of the order imbalance measure over the next 52 weeks, using the *Att* and *AttOrth* components. Results are presented in Appendix Table A.1, and we fail to find significant reversal patterns for our sample.

4.2. Retail Trades over Pandemic, GME and Ukraine War

The 27 months in our sample is quite eventful. The initial shock of the outbreak of COVID-19 in spring of 2020 is substantial, and the capital market quickly react with large negative returns, and surges in market illiquidity and volatility. Presumably, the arrival of the pandemic contains systematic negative news, and affect the fundamentals of the economy. The VIX index, as an indicator or the market implied volatility, increases from lower than 20% in January 2020 to 82% on March 16, 2022. The episode of GME in January 2021 also attracts extensive attention from practitioners, researchers and regulators. Unlike the pandemic itself, the event concentrates on a handful of MEME stocks, mostly and heavily traded on Robinhood, rather than the general market. Maybe it is not surprising that the VIX only slightly increase to 33% during January 2021. The war between Ukraine and Russia, which breaks out in February 2022, potentially leads to shortage in food and energy supply in Europe, accounts for another major event. Unlike GME episode, the war clearly affects more than a handful of stocks. But due to its remote location, its impact on U.S. market is moderate, with VIX increasing to 35%.

Do retail investors trade differently for events with different natures? Barrot et al. (2016) find the ability of retail order imbalances to predict future returns is significantly enhanced during times of market stress indicated by VIX, and retail investors are likely compensated for liquidity provision during these times. Therefore, we first divide our sample period into high-VIX days and low-VIX days based the 75th percentile during our sample period. We define a dummy variable, *DVIX*, which takes the value of 1 when the day's VIX index is above the sample's 75th percentile and 0 otherwise. We also define individual event dummies to capture three events. The dummy *DCovid* takes a value of 1 for the month of March 2020, and zero otherwise; the dummy *DGME* takes a value of 1 between January 13 of 2021, the first day of large increase in GME, to February

12, one month after when trading calms down for the stock; and the dummy *DUKR* takes a value of 1 between February 21 of 2022, when Putin, the president of Russia, acknowledged independences of Ukraine lands, and to the end of our sample, and zero otherwise. For the Fama-MacBeth setup for return prediction, we directly project the coefficient $a1$ in equation (3) on a constant and the dummy variables. For the panel setup, we simply include the interaction term in the regression.

The empirical results using *DVIX* are reported in Panel A of Table VI. In regression I, the average retail coefficient is 0.0008 with a highly significant *t*-stat of 7.92, and the coefficient on *DVIX* is 0.0007 with a significant *t*-stat of 3.65, indicating that during period with higher VIX, the predictive coefficient becomes 0.0015 (0.0008+0.0007), much higher than other days, which is consistent with Barrot et al. (2016). For regression II, in which retail flows are used to predict future effective spread, the average retail coefficient is 0.0576 and significant, and the coefficient on *DVIX* is 0.1860 with *t*-stat of 6.50, which is not consistent with Barrot et al. (2016), in the sense that higher retail activity level actually leads to lower market liquidity. Similar patterns are observed for volatility prediction. For HFT activities, the patterns are interestingly different. On low-VIX days, the coefficient is -0.2599 (*t*-stat=-18.35), indicating higher retail activities are associated with lower activities from HFT. On high-VIX days, the coefficient on the interaction is 0.1227 (*t*-stat=3.94), meaning that even though the overall impact (-0.2599+0.1227=-0.1372) is still negative, the magnitude becomes significantly smaller. That is HFT seems to be more active on high VIX days, possibly because there are more arbitrage opportunities on these days. We find similar results for SS.

For individual episodes results in Panel B of Table VI, we observe several interesting patterns. First, the results are similar to high VIX days' results for the pandemic outbreak. Second, for GME episode, the coefficient of *DGME* for effective spread is -0.1021 with *t*-stat of -3.11, indicating that during the GME period, presumably less sophisticated retail investors join the market and provide liquidity. Finally, there is no significant differences for return and liquidity prediction during the war, but retail trading has higher association with volatility.

4.3 Retail Trades for firms with different characteristics

Our sample includes on average more than 3,000 firms each day. Is the predictive power of marketable retail order imbalances restricted to particular subsets of firm? Do retail investors have preferences for trading particular subsets of firms during the pandemic? We investigate these questions by analyzing various firm subgroups in this section. That is, we first sort all firms into groups based on firm characteristics and define group dummies, and we estimate equation (3) and (4) with interactions with group dummies. We then present the coefficient on the retail measures within each group and compare their magnitudes. Here we consider two essential firm characteristics, market capitalization and turnover.

From Table VII Panel A, we first separate firms by size. Moving from the smallest one-third of firms to the largest, the coefficient on *Oibvol* decreases from 0.0013 to 0.0007, while both statistically significant. Clearly, the predictive power of marketable retail order imbalances is much stronger for smaller firms than for larger firms, but the predictability remains reliably present in all three groups. When we sort firms by previous month turnover, moving from the lowest liquidity firms to the highest, the coefficient on *Oibvol* increases from 0.0006 to 0.0025, and all statistically significant. The pattern is clear: the predictive power of retail order imbalances for future returns is stronger for firms with higher turnovers. Results for effective spread, intraday volatility, HFT and SS are presented in the rest of the table.

Out of all stocks, Robinhood 50 list defines a small set of stocks that are mostly actively traded on Robinhood.[14] We define a Robinhood 50 dummy variable, *DRH50*, which takes the value of 1 if the stock belongs to Robinhood 50 list, and zero otherwise. We interact the *DRH50* with the retail measures for their predictive power for future variables. In Panel B of Table VII, we present the predictive coefficients for the retail measures for the Robinhood 50 stocks and other stocks. The coefficients of all other stocks are similar to those we observed in earlier sections. For Robinhood 50 stocks, it's interesting to find in the first column the coefficient on *Oibvol* is 0.0169, which is much larger than 0.0010 for non-Robinhood 50 stocks, implying the retail order imbalance has a much larger predictive power these stocks. There is no significant difference for effective spread prediction in column II, indicating that retail activities demand liquidity, even for the Robinhood 50 stocks, which is consistent with Eaton et al. (2022). In column III, the coefficient on *Actvol* is 0.1834 for Robinhood 50 stocks, which is significantly larger than 0.0671 for other stocks, suggesting that retail trading on these stocks potentially bring more uncertainty.

4.4 Retail Trades over intraday and weekly horizons

After seeing how daily retail trades are related to future returns, liquidity and volatility measures over daily and weekly horizons, in this section, we focus on the intraday dynamics of retail trades and its relation with returns and market quality measures. We divide the trading day into 13 intervals, starting from 09:30am to 16:00pm, with each interval lasting for 30 minutes, and present the each 30-minutes trading patterns in Figure III. That is, for each interval, we report the time-series mean of marketable retail investor trading share values and number of trades. Panel A and B of Figure III show that retail trading display a U-shape throughout the day. There are more shares traded by retail investors in the morning and near market close, and there are less trading

[14] We also design alternative retail concentration measures, and report results in Appendix Table A.2.

during the middle of the day. This U-shape is similar to the pattern of institutional trades, but it is slight different in the sense that there are more retail traders at the open rather than at the close, while the institutional traders trade more at the close than at the open.

Can intraday retail trades predict intraday returns? We report the estimation results for return prediction in Table VIII. In the first column, we use the *Oibvol* to predict the next half-hour of stock returns. For instance, the 09:30-10:00 interval retail *Oibvol* can predict the next half hour return with a coefficient of 0.000120 and a *t*-statistic of 6.76. Other than the last interval of 15:00-15:30, all intraday *Oibvol* predicts next 30-minute return positively and significantly. In the second column, we predict the future close-to-close daily returns using *Oibvol* from each 30-minute interval. All coefficients are positive and significant (except for the first three intervals), suggesting that the predictive power of retail order imbalance is embedded within each trading interval throughout the day. Overall, past interval marketable retail order imbalances can significantly impact future stock price movement, not only for the next interval but also for the next trading day.

We present results on volatility and liquidity in column III to VI in Table VIII.[15] We observe positive predictive power of retail activity for liquidity and volatility throughout the day and the next day, indicating that intraday retail trading activity consistently predicts higher illiquidity and high volatility. Overall, past interval marketable retail trading activity can significantly decrease future stock liquidity and increase volatility, not only for the next interval but also for the next trading day.[16]

5. Conclusions

---

[15] Since there is intraday data on HFT and short-selling, we can't perform intraday analysis on HFT and short-selling.
[16] BJZZ mainly focuses on weekly horizons in their prediction regressions. We present weekly analysis in the Appendix Table A.3. The main findings are consistent with daily and intraday results.

Using the BJZZ subpenny transaction prices algorithm, we identify retail investors' orders in the U.S. market between January 2020 and June 2021. During the pandemic period, the overall retail trading volumes increase from 9% of total market volumes to about 11%. With the heightened retail activity, retail order imbalances significantly and positively predict returns for next day and for the next twelve weeks in the cross-section. In terms of future liquidity and volatility, higher retail activities are associated with wider future effective spreads and greater future volatilities, along with significantly lower activities from high frequency traders and short-sellers.

Given the rising of retail-oriented investing platforms with zero trading commissions, and retail investors trading's impact on price discovery, market quality, and other market participants during the Pandemic, regulators may need to carefully consider updating policies on retail investors protection and their impact on market quality.[17] Besides, our study clearly leaves many interesting questions unsolved. For example, do retail investors trading a new form of risk? Should regulations be formed to protect retail investors during the pandemic? We leave these interesting and important questions to future research.

[17] For instance, the chairman of the U.S. SEC, Gary Gensler, publicly talked about investor protection in a digital age (https://www.sec.gov/news/speech/gensler-remarks-nasaa-spring-meeting-051722) and plans to update the regulation rules and drive greater efficiencies for retail investors (https://www.sec.gov/news/speech/gensler-remarks-piper-sandler-global-exchange-conference-060822).

REFERENCES

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H., 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61(1), 43-76.

Ang, A., Hodrick, R., Xing, Y., and Zhang, X., 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61: 259-299.

Barardehi, Y. H, Bernhardt, D., Da, Z., and Warachka, M., 2022. Internalized retail order imbalances and institutional liquidity demand. Unpublished working paper. Chapman University.

Barber, B. M., Huang, X., Odean, T., and Schwarz, C., 2022. Attention induced trading and returns: Evidence from Robinhood users. *Journal of Finance*, Forthcoming.

Barber, B. M., Lee, Y. T., Liu, Y. J., and Odean, T., 2009. Just how much do individual investors lose by trading?. *Review of Financial Studies* 22(2), 609-632.

Barber, B. M., and Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance* 55(2), 773-806.

Barber, B. M., and Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21(2), 785-818.

Barrot, J. N., Kaniel, R., and Sraer, D., 2016, Are retail traders compensated for providing liquidity?. *Journal of Financial Economics* 120, 146-168.

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

Boehmer, E., Fong, K., and Wu, J., 2021. Algorithmic trading and market quality: International evidence. *Journal of Financial and Quantitative Analysis* 56(8), 2659-2688.

Boehmer, E., Jones, C. M., and Zhang, X., 2008. Which shorts are informed? *Journal of Finance* 63(2), 491-527.

Boehmer, E., Jones, C. M., Zhang, X., and Zhang, X., 2021. Tracking retail investor activity. *Journal of Finance* 76(5), 2249-2305.

Bryzgalova, S., Pavlova, A., and Sikorskaya, T., 2022. Retail trading in options and the rise of the big three wholesalers. Unpublished working paper. London Business School.

Da, Z., Engelberg, J., and Gao, P., 2011. In search of attention. *Journal of Finance* 66(5), 1461-1499.

Eaton, G. W., Green, T. C., Roseman, B., and Wu, Y., 2022. Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics*, Forthcoming.

Fama, E. F., and MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81(3), 607-636.

Hasbrouck, J., and Saar, G., 2013. Low-latency trading. *Journal of Financial Markets* 16(4), 646-679.

Hendershott, T., Jones, C. M., and Menkveld, A. J., 2011. Does algorithmic trading improve liquidity?. *Journal of Finance* 66(1), 1-33.

Hendershott, T., Menkveld, A. J., Praz, R., and Seasholes, M., 2022. Asset price dynamics with limited attention. *Review of Financial Studies* 35(2), 962-1008.

Holden, C. W., and Jacobsen, S., 2014. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *Journal of Finance* 69(4), 1747-1785.

Hüfner, F., Strych, J., and Westerholm, J., 2022. The Impact of Retail Investors on Stock Liquidity and Crash Risk. Unpublished working paper. Karlsruhe Institute of Technology.

Hong, H., Li, W., Ni, S. X., Sheinkman, J., and Yan, P., 2016. Days to cover and stock returns. NBER working paper.

Kaniel, R., Saar, G., and Titman, S., 2008. Individual investor trading and stock returns. *Journal of Finance* 63(1): 273-310.

Kelley, E. K., and Tetlock, P. C., 2013. How wise are crowds? Insights from retail orders and stock returns. *Journal of Finance* 68(3), 1229-1265.

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

Newey, W.K. and West, K.D., 1987. Hypothesis testing with efficient method of moments estimation. *International Economic Review* 28: 777-787.

Newey, W.K. and West, K.D., 1994. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies* 61(4), pp.631-653.

Ortmann, R., Pelster, M., and Wengerek, S. T., 2020. COVID-19 and investor behavior. *Finance Research Letters* 37, 101717.

Ozik, G., Sadka, R., and Shen, S., 2021. Flattening the illiquidity curve: Retail trading during the COVID-19 lockdown. *Journal of Financial and Quantitative Analysis* 56(7), 2356-2388.

Pagano, M. S., Sedunov, J., and Velthuis, R., 2021. How did retail investors respond to the COVID-19 pandemic? The effect of Robinhood brokerage customers on market quality. *Finance Research Letters* 43, 101946.

Petersen, M. A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22(1), 435-480.

Van Kervel, V., and Menkveld, A. J. 2019. High frequency trading around large institutional orders. *Journal of Finance* 74(3), 1091-1137.

Welch, I., 2022. The wisdom of the Robinhood crowd. *Journal of Finance* 77(3), 1489-1527.

Weller, B. M., 2018. Does algorithmic trading reduce information acquisition? *Review of Financial Studies* 31(6), 2184-2226.

Table I. Summary Statistics

This table reports pool summary statistics for retail investor trading and stock characteristics. Our sample period is January 2020 to March 2022, and our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. For retail trading, we report the retail order imbalance measure (*Oibvol*) as defined in equation (1), and the percentage of retail share volume over total share volume (*Actvol*) as defined in equation (2). For liquidity and volatility, we report effective spread (*Effspr(%)*), and annualized intraday volatility (*Intvol*). For high frequency trading, we report the cancel to trade ratio (*HFTCancel*). For short selling, we report the days to cover ratio (*SDTCR*). For stock characteristics, we report daily stock return (*Ret*) in percentage, market capitalization (*Size*) in billions, book to market ratio (*Lbm*), monthly stock turnover (*Lturnover*) and monthly return volatility (*Lmvol*). The return is computed using bid-ask average prices.

Variables	Mean	Std	P25	P50	P75
Oibvol	-0.0208	0.3967	-0.2172	-0.0135	0.1725
Actvol	0.0904	0.1002	0.0274	0.0534	0.1203
Effspr(%)	0.3702	0.5568	0.0721	0.1615	0.4393
IntVol	0.8296	0.5769	0.4162	0.6772	1.0858
HFTCancel	2.98	0.51	2.64	2.92	3.25
SDTCR	4.26	5.61	0.75	2.20	5.47
Ret(%)	0.12	5.30	-1.65	0.00	1.62
Size($Bil.)	10.89	65.44	0.21	0.93	4.17
Lbm	0.65	1.78	0.16	0.40	0.83
Lturnover	0.02	0.28	0.00	0.01	0.01
Lmvol	0.57	0.65	0.29	0.45	0.69

Table II. Retail Trading Predicts Stock Returns

This table reports results on whether marketable retail investor order imbalance can predict future stock returns from one day to 12 weeks. Our sample period is January 2020 to March 2022, and our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. We estimate the Fama-MacBeth regressions, as specified in equation (3). The dependent variable is the next day return in Panel A, and cumulative return *n*-weeks ahead return, scaled by number of days involved in Panel B. The returns are computed using the end-of-day bid-ask average price. The independent variables are scaled marketable retail order imbalance measures, *Oibvol*, as defined in equation (1). Control variables include previous day return, *Ret*(-1), previous week return, *Ret*(w-1), previous month return, *Ret*(m-1), log market capitalization, *Lsize*, book to market ratio, *Lbm*, monthly stock turnover, *Lturnover*, and monthly return volatility *Lmvol*. To account for serial correlation in the coefficients, the standard deviations of the time-series are adjusted using Newey-West (1987) with five lags.

Panel A. Retail order flows predict next-day return

Dep.var	Ret	
Retail	Oibvol	
	Coef.	*t*-Stat
Retail(-1)	0.0010	9.85
Ret(-1)	-0.0181	-3.80
Ret(w-1)	-0.0230	-2.32
Ret(m-1)	0.0117	0.79
Lsize	-0.0002	-2.06
Lbm	0.0001	0.65
Lturnover	0.0040	1.58
Lmvol	-0.0066	-0.78
Intercept	0.0023	2.58
Adj.R2	0.0786	
Interquartile	0.3902	
Interquartile next-day return diff	0.0390%	

Panel B. Retail order flows predict return in the long run

Retail	Oibvol	
	Coef.	*t*-Stat
w=1	0.00032	7.28
w=2	0.00024	6.90
w=3	0.00020	7.49
w=4	0.00017	6.87
w=5	0.00015	6.88
w=6	0.00013	6.92
w=7	0.00010	5.47
w=8	0.00010	5.31
w=9	0.00009	6.06
w=10	0.00009	5.86
w=11	0.00008	6.00
w=12	0.00009	6.37

Table III. Retail Trading Predicts Stock Liquidity and Volatility

This table reports estimation results on whether retail trading activity can predict liquidity and intraday volatility from one day to 12 weeks. Our sample period is January 2020 to March 2022, and our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. We estimate panel regressions, as specified in equation (4). The dependent variables *Y* are next day liquidity and volatility in Panel A, and next *k*th week liquidity and volatility scaled by number of days involved in Panel B. Liquidity proxy is effective spread, *Effspr*(*%*), while volatility proxy is intraday volatility, *Intvol*. The independent variables are retail trading activity, *Actvol*, and retail order imbalance, *Oibvol*. Controls include previous day return, previous week return, previous month return, size, book-to-market ratio, previous month turnover and volatility of daily returns. We include second-order lagged dependent variable to control for its persistency and also include the day fixed effect. The standard errors are double clustered at day and stock level.

Panel A. Retail trading predicts next-day effective spread

Regression	I		II		III		IV	
Dep.var	Effspr		Effspr		Intvol		Intvol	
Retail	Actvol		Oibvol		Actvol		Oibvol	
	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat
Retail(-1)	0.1017	10.22	0.0034	4.01	0.0690	5.46	0.0068	7.37
Controls	Yes		Yes		Yes		Yes	
Day FE	Yes		Yes		Yes		Yes	
Interquartile	0.0929		0.3897		0.0929		0.3897	
Interquartile Dep.var Diff	0.0094		0.0013		0.0064		0.0026	
Adj.R2	0.789		0.794		0.728		0.730	

Panel B. Retail trading predicts effective spread in the long run

Dep.var	k^{th} week Effspr		k^{th} week Effspr		k^{th} week Intvol		k^{th} week Intvol	
Retail	Actvol		Oibvol		Actvol		Oibvol	
k	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat
1	0.0810	10.71	0.0004	0.59	0.0682	7.73	0.0062	11.01
2	0.0758	8.93	0.0002	0.36	0.0443	5.15	0.0009	1.42
3	0.0715	10.07	0.0000	0.06	0.0492	5.60	0.0003	0.57
4	0.0770	10.22	-0.0004	-0.66	0.0486	5.50	0.0008	1.29
5	0.0788	10.62	-0.0006	-0.86	0.0482	5.38	0.0006	1.03
6	0.0716	9.47	0.0001	0.11	0.0474	4.97	0.0003	0.43
7	0.0911	11.49	0.0005	0.82	0.0540	5.97	-0.0004	-0.63
8	0.0751	9.72	-0.0008	-1.20	0.0543	6.26	0.0005	0.87
9	0.0713	9.29	-0.0003	-0.52	0.0507	5.55	0.0008	1.30
10	0.0726	10.03	0.0000	0.02	0.0474	5.09	-0.0004	-0.71
11	0.0579	8.41	-0.0004	-0.62	0.0455	5.01	-0.0002	-0.28
12	0.0612	8.41	0.0003	0.52	0.0517	6.25	-0.0005	-0.93

Table IV. Retail Trading Predicts High Frequency and Short Selling

This table reports estimation results on whether retail trading activity can predict high frequency trading and short selling from one day to 12 weeks. Our sample period is January 2020 to March 2022, and our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. We estimate panel regressions, as specified in equation (4). The dependent variables *Y* are next day high frequency trading and short selling in Panel A, and next *k*th week high frequency trading and short selling scaled by number of days involved in Panel B. High frequency trading proxy is cancel to trade ratio *HFTCancel*, while short selling proxy is days to cover ratio, *SDTCR*. The independent variables are retail trading activity, *Actvol*, and retail order imbalance, *Oibvol*. Controls include previous day return, previous week return, previous month return, size, book to market ratio, previous month turnover and volatility of daily returns. We include second-order lagged dependent variable to control for its persistency and also include day fixed effect. The standard errors are double clustered at day and stock level.

Panel A. Retail trading predicts next day cancel to trade and shorting flow

Regression	I		II		III		IV	
Dep.var	HFTCancel		HFTCancel		SDTCR		SDTCR	
Retail	Actvol		Oibvol		Actvol		Oibvol	
	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat
Retail(-1)	-0.2300	-17.12	-0.0027	-2.62	-1.1141	-12.28	-0.0505	-6.06
Controls	Yes		Yes		Yes		Yes	
Date FE	Yes		Yes		Yes		Yes	
Interquartile	0.0929		0.3897		0.0929		0.3897	
Interquartile Dep.var Diff	-0.0214		-0.0011		-0.1035		-0.0197	
Adj.R2	0.563		0.566		0.707		0.714	

Panel B. Retail trading predicts cancel to trade and shorting flow in the long run

Dep.var	k^{th} week HFTCancel		k^{th} week HFTCancel		k^{th} week SDTCR		k^{th} week SDTCR	
Retail	Actvol		Oibvol		Actvol		Oibvol	
k	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat
1	-0.1002	-9.89	-0.0052	-7.15	-0.8233	-11.31	-0.0520	-7.82
2	-0.1101	-10.23	0.0009	1.37	-0.7948	-10.74	-0.0031	-0.49
3	-0.0885	-8.36	0.0010	1.34	-0.7713	-10.68	0.0069	1.06
4	-0.0770	-7.24	0.0019	2.65	-0.7927	-11.25	0.0013	0.21
5	-0.0763	-7.18	0.0011	1.70	-0.7476	-10.27	-0.0074	-1.06
6	-0.0747	-7.12	0.0006	0.87	-0.6602	-8.89	-0.0090	-1.31
7	-0.0751	-7.01	0.0005	0.67	-0.7038	-10.25	0.0052	0.81
8	-0.0597	-5.59	-0.0006	-0.86	-0.6605	-9.42	-0.0132	-2.18
9	-0.0723	-6.61	0.0011	1.69	-0.5827	-8.14	-0.0152	-2.43
10	-0.0698	-6.76	-0.0004	-0.52	-0.6478	-8.75	0.0002	0.03
11	-0.0774	-8.02	0.0013	1.83	-0.6534	-8.91	-0.0053	-0.91
12	-0.0756	-7.90	0.0006	0.80	-0.6029	-8.05	-0.0012	-0.18

Table V. Retail Attention and Retail Trading

This table examines retail investors' attention and their trading. Our sample period is January 2020 to March 2022, and our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. We focus on the abnormal Google search volume index (*ASVI*), from Da, Engelberg, and Gao (2011), which is defined as the log *SVI* during the current week minus the log median *SVI* during the previous eight weeks. Panel A examine whether *ASVI* could predict future retail investors trades. Panel B decompose retail investor trading into attention components (*Att*) and orthogonal component (*AttOrth*), and examine the predictive power of attention-driven retail flows and the orthogonal components. Panel C examine whether retail order imbalances could predict earnings announcement *CAR* over different horizon.

Panel A. Attention predict retail trading

Regression	I		II		III		IV	
Dep.var	Oibvol		Oibvol		Actvol		Actvol	
	Coef.	*t*-stat	Coef.	*t*-stat	Coef.	*t*-stat	Coef.	*t*-stat
ASVI(-1)			0.0091	10.44			0.0012	4.50
Retail(-1)	0.0364	25.96	0.0371	26.17	0.4862	117.33	0.5306	126.11
Ret(-1)	0.0135	1.32	0.0187	1.88	-0.0190	-5.36	-0.0220	-6.64
Ret(w-1)	-0.4186	-18.74	-0.4309	-18.02	-0.0252	-2.26	-0.0250	-2.38
Ret(m-1)	-0.4709	-10.21	-0.4757	-10.50	-0.1523	-6.99	-0.1413	-6.91
Lsize	0.0046	14.82	0.0046	14.70	-0.0102	-60.14	-0.0091	-59.14
Lbm	0.0020	6.01	0.0019	5.34	-0.0033	-28.33	-0.0029	-25.10
Lturnover	0.0161	4.38	0.0118	3.12	0.0350	10.32	0.0326	11.05
Lmvol	0.1391	8.11	0.1320	7.71	0.2153	22.78	0.2041	21.90
Intercept	-0.0561	-19.47	-0.0557	-19.50	0.1034	66.00	0.0940	64.96
Adj.R2	0.0032		0.0035		0.4494		0.5000	

Panel B. The predictive power of attention-driven retail flows and the orthogonal component

Regression	I	II	III	IV	V
Dep.var	Ret	Effspd	IntVol	HFTcancel	SDTCR
Retail	Oibvol	Actvol	Actvol	Actvol	Actvol
Att(-1)	0.3337	-0.1022	0.1665	-0.5511	-3.0180
t-Stat	1.46	-2.90	2.24	-6.70	-5.52
AttOrth(-1)	0.0009	0.1077	0.1047	-0.2304	-1.1794
t-Stat	9.32	10.79	7.61	-16.47	-12.48
Controls	Yes	Yes	Yes	Yes	Yes
Adj.R2	0.084	0.801	0.743	0.573	0.725

Panel C. Use *Oibvol* to predict earnings news

Dep.var	CAR[0,1]		CAR[2,60]		CAR[2,250]	
	Coef.	*t*-stat	Coef.	*t*-stat	Coef.	*t*-stat
Retail(-1)	0.0029	2.02	0.0384	2.90	0.0144	1.89
Control	Yes		Yes		Yes	
Adj.R2	0.0095		0.0655		0.0514	

Table VI. Retail Trading over Different Subperiods

This table reports retail investor trading to predict future stock returns, liquidity, volatility, high frequency trading, and short seller trading in high VIX period and three distinct periods. Our sample period is January 2020 to March 2022, and our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least \$1. We split our sample into different subperiods in two ways. Panel A follow Barrot et al. (2016) and split the period into high and low VIX days using CBOE VIX Index. The *DVIX* dummy takes one when the day's VIX index is higher than the sample 75th percentile. Panel B select three distinct subperiods contained in our sample, the Covid strike period, the Gamestop period and the War period. The *DCovid* takes 1 for March 2020. The *DGME* takes 1 when the time ranges from January 13th 2021 to February 12th 2021. The *DUKR* takes 1 for days from February 21th 2022 to our sample end. For regression methodology, we adopt Fama-Macbeth method to estimate the coefficients of retail order imbalances predicting next day return as in equation (3), and report the coefficients interacted with these subperiod dummies. We use panel regressions to estimate the coefficients of retail trades predicting next day liquidity, volatility, high frequency trading and short selling as in equation (4), and report the coefficients interacted with these subperiod dummies.

Panel A. Retail trading in high/low VIX period

Regression	I	II	III	IV	V
Dep.var	Ret	Effspd	IntVol	HFTcancel	SDTCR
Retail	Oibvol	Actvol	Actvol	Actvol	Actvol
Intercept (Retail)	0.0008	0.0576	0.0419	-0.2599	-1.1741
t-Stat	7.92	5.82	3.17	-18.35	-10.95
DVIX	0.0007	0.1860	0.1129	0.1227	0.2426
t-Stat	3.65	6.50	4.06	3.94	1.30
Adj.R2	0.021	0.789	0.728	0.563	0.707

Panel B. Retail trading in three different periods

Regression	I	II	III	IV	V
Dep.var	Ret	Effspd	IntVol	HFTcancel	SDTCR
Retail measures	Oibvol	Actvol	Actvol	Actvol	Actvol
Intercept (Retail)	0.0009	0.0804	0.0491	-0.2475	-1.1327
t-Stat	9.34	7.93	3.80	-18.20	-11.48
DCovid	0.0024	0.6622	0.2397	0.3034	0.5050
t-Stat	5.33	5.64	2.24	2.61	2.04
DGME	0.0007	-0.1021	0.1161	-0.0446	-0.8038
t-Stat	1.63	-3.11	1.89	-0.69	-2.64
DUKR	0.0000	0.0266	0.1274	0.1354	0.4627
t-Stat	0.08	1.11	2.69	3.17	1.31
Adj.R2	0.046	0.789	0.728	0.563	0.707

Table VII. Retail Trading for Different Subgroups

This table reports retail investor trading to predict future stock returns, liquidity, and volatility, interacted with different stock characteristics, and Robinhood 50 restricted list. Our sample period is January 2020 to March 2022, and our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. Panel A examines subgroups results by stock characteristics. We separate sort all firms into three groups based on previous month-end market cap and turnover and estimate the regressions for each subgroup. Panel B examines the Robinhood 50 stocks in the restricted list as the interacted group. If a stock *i* belongs to the list, then the dummy $RH50_i$ takes 1, and 0 otherwise. For regression methodology, we adopt Fama-Macbeth method to estimate the coefficients of retail order imbalances predicting next day return as in equation (3), and report the coefficients interacted with these dummies. We use panel regressions to estimate the coefficients of retail trading predicting next day liquidity, volatility, high frequency trading and short selling as in equation (4), and report the coefficients interacted with these dummies. *p<.1; **p<.05; ***p<.01.

Panel A. Retail trading in different stock characteristics

Regression	I	II	III	IV	V
Dep.var	Ret	Effspd	IntVol	HFTcancel	SDTCR
Retail measures	Oibvol	Actvol	Actvol	Actvol	Actvol
SmallSize	0.0013***	0.1223***	0.0609***	-0.2092***	-1.1005***
MediumSize	0.0006***	-0.1497***	0.0094***	-0.4116***	-0.0828***
LargeSize	0.0007***	0.1077***	0.1375***	-0.3117***	-1.6200***
LowTurnover	0.0006***	0.2608***	0.0601***	-0.0119***	-0.8456***
MediumTurnover	0.0010***	0.0481***	0.0348***	-0.3665***	-0.9690***
HighTurnover	0.0025***	0.0212***	0.0835***	-0.3287***	-1.3601***

Panel B. Retail trading in Robinhood restricted 50 stocks

Regression	I	II	III	IV	V
Dep.var	Ret	Effspd	IntVol	HFTcancel	SDTCR
Retail	Oibvol	Actvol	Actvol	Actvol	Actvol
Non Robinhood 50 stocks	0.0010***	0.1020***	0.0671***	-0.2294***	-1.1238***
Robinhood 50 stocks	0.0169***	0.0850***	0.1834***	-0.2641***	-0.5311***

Table VIII. Retail Trading over Different Intraday Horizons

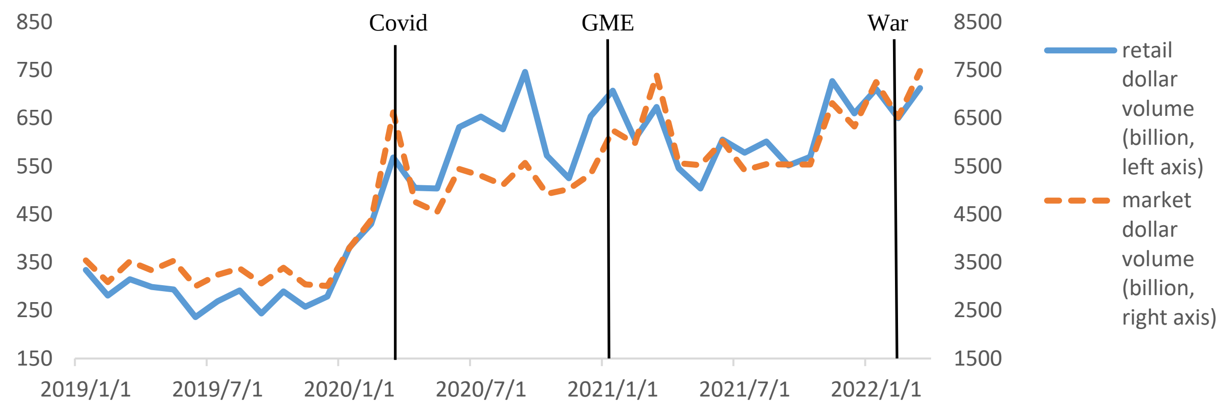
Our sample period is January 2020 to March 2022. We use bid-ask average prices to calculate returns for each interval. Every 30 minutes is an intraday interval, starting from 09:30 to 16:00, that provides 13 intervals per day. We estimate Fama-MacBeth regressions, similar to equation (3), and panel regressions similar to equation (4), the independent variables are retail trading from each 30 minutes. The row stands for retail trading in each interval, and the column stands for returns in different intervals.

Regression	I	II	III	IV	V	VI
Retail measures	Oibvol	Oibvol	Actvol	Actvol	Actvol	Actvol
Dep.var	Next 30 min ret	Next day ret	Next 30 min effspd	Next day effspd	Next 30 min intvol	Next day intvol
09:30-10:00	0.000120	0.000103	0.0962	0.0204	0.1941	0.0291
t-Stat	6.76	1.69	11.26	4.28	13.92	4.73
10:00-10:30	0.000119	0.000096	0.0246	0.0333	0.1106	0.0249
t-Stat	7.17	1.53	5.61	8.83	17.02	5.58
10:30-11:00	0.000110	0.000060	0.0192	0.0268	0.0766	0.0234
t-Stat	7.23	1.06	4.45	6.95	13.18	5.21
11:00-11:30	0.000110	0.000177	0.0071	0.0287	0.0698	0.0259
t-Stat	9.51	3.20	1.94	7.67	14.02	5.86
11:30-12:00	0.000079	0.000161	0.0052	0.0233	0.0680	0.0225
t-Stat	7.65	3.02	1.54	6.83	14.10	5.22
12:00-12:30	0.000080	0.000154	0.003	0.0218	0.0629	0.0267
t-Stat	6.89	2.99	0.94	6.51	14.29	6.39
12:30-13:00	0.000089	0.000197	0.0048	0.0136	0.0452	0.0256
t-Stat	7.66	3.13	1.48	4.03	9.63	5.75
13:00-13:30	0.000090	0.000200	0.0006	0.0206	0.0463	0.0243
t-Stat	8.58	3.66	0.16	5.95	10.79	5.55
13:30-14:00	0.000086	0.000273	0.0115	0.0199	0.0501	0.0213
t-Stat	8.22	4.76	2.98	6.32	9.92	4.92
14:00-14:30	0.000064	0.000176	0.0023	0.0204	0.0473	0.0266
t-Stat	5.60	3.10	0.67	5.44	9.72	5.80
14:30-15:00	0.000071	0.000221	0.0175	0.024	0.0501	0.0284
t-Stat	6.53	3.72	4.68	6.92	10.13	6.13
15:00-15:30	0.000019	0.000325	0.0197	0.0244	0.0487	0.0373
t-Stat	1.36	4.80	3.40	5.98	6.70	7.31
15:30-16:00		0.000656		0.0379		0.0518
t-Stat		9.96		5.76		6.25

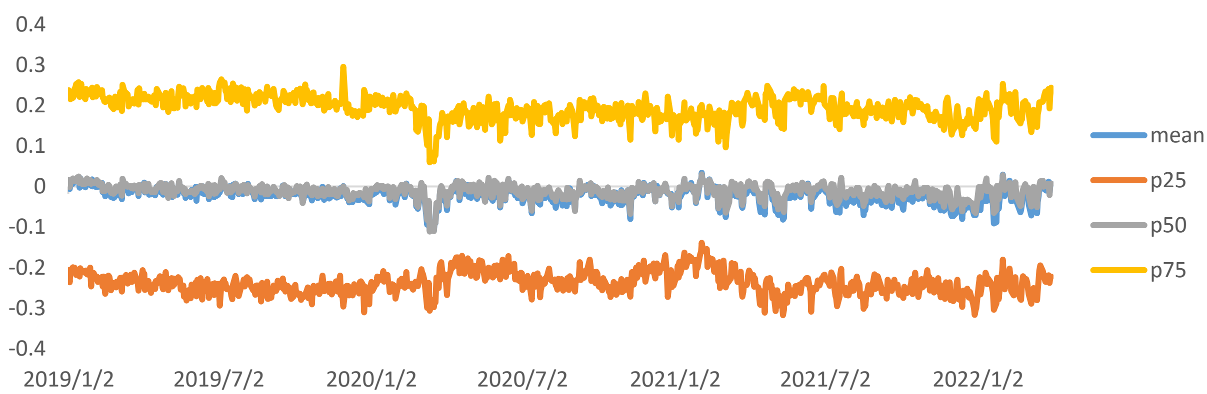
Figure I. Marketable Retail Investors Trading Flow

These figures plot the time-series statistics of the marketable retail investor trading activity from January 2019 to March 2022, identified using algorithm from Boehmer et al. (2021). Our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. Panel A presents the aggregate retail investor and market dollar volume across all stocks for each month. Panel B presents the daily cross-sectional distribution of marketable retail order imbalance, *Oibvol*, as defined in equation (1). Panel C presents the daily cross-sectional distribution of marketable retail trading activity, *Actvol*, as defined in equation (2).

Panel A. Aggregate retail investor and market dollar volume for each month



Panel B. Cross-sectional distribution of *Oibvol* for each day



Panel C. Cross-sectional distribution of *Actvol* for each day

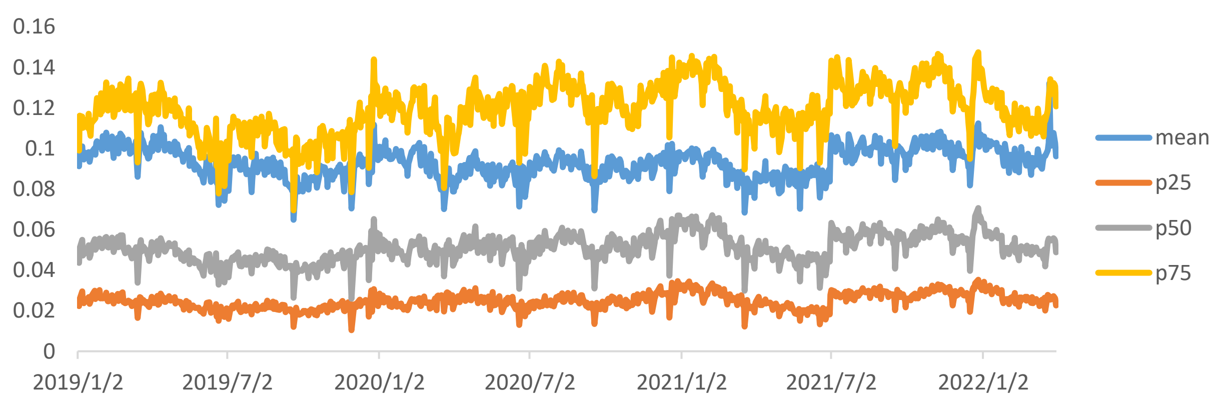
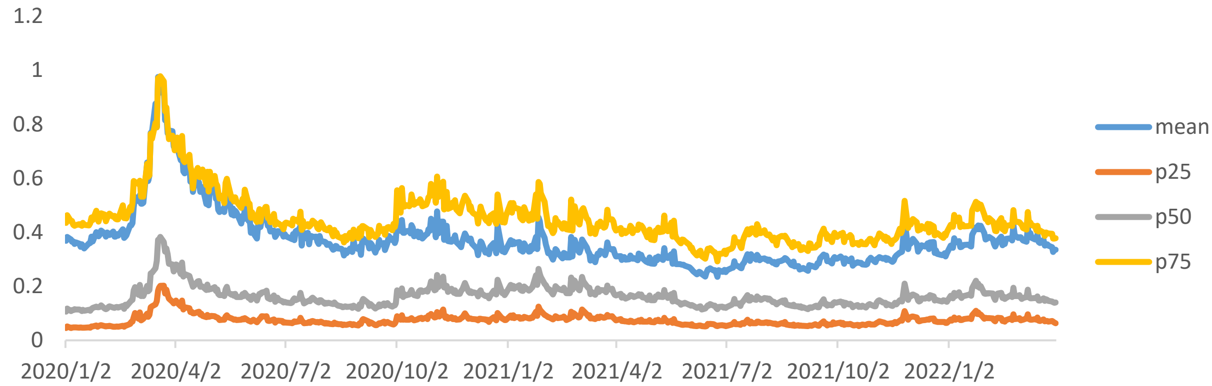


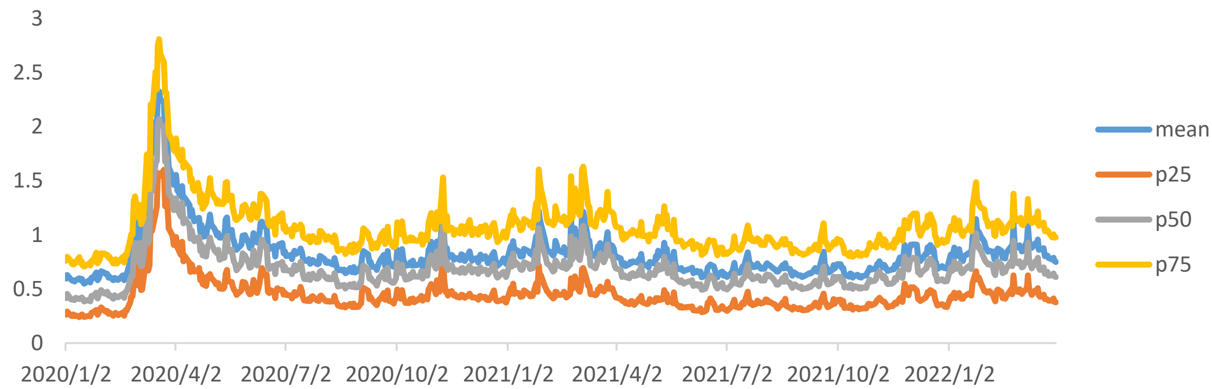
Figure II. Market Liquidity, Volatility, High Frequency Trading and Short Selling

These figures plot the time-series statistics of the market liquidity, volatility, high frequency trading and short selling from January 2020 to March 2022. Our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. Panel A presents the effective spread (%) proxy for liquidity. Panel B presents the intraday volatility constructed by Andersen et al. (2001). Panel C presents the cancel to trade ratio proxy for high frequency trading. Panel D presents the days to cover ratio proxy for short selling.

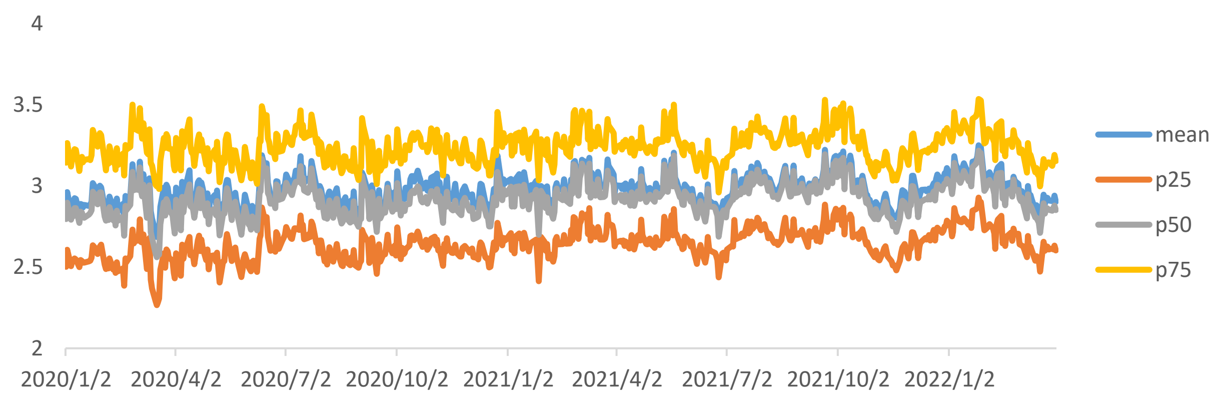
Panel A. Cross-sectional distribution of effective spread (%), *Effspr*



Panel B. Cross-sectional distribution of intraday volatility, *Intvol*



Panel C. Cross-sectional distribution of high frequency trading cancel to trade, *HFTCancel*



Panel D. Cross-sectional distribution of short selling days to cover ratio, *SDTCR*

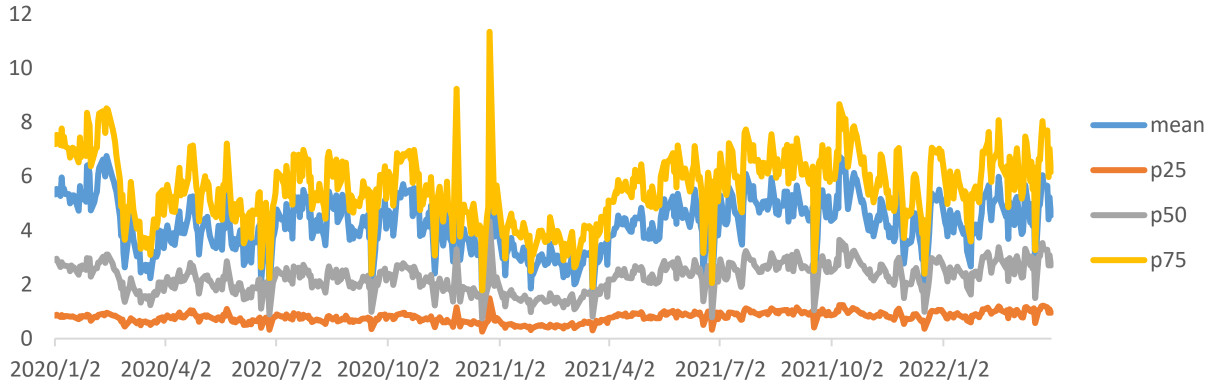
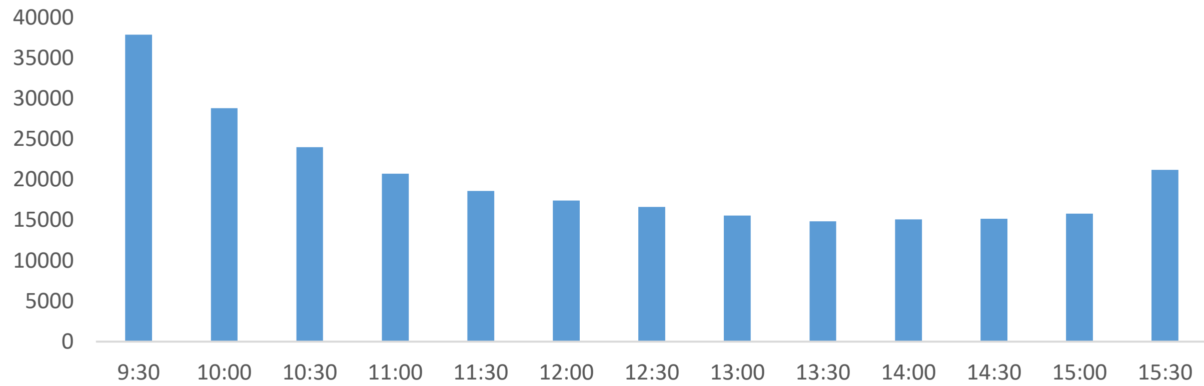


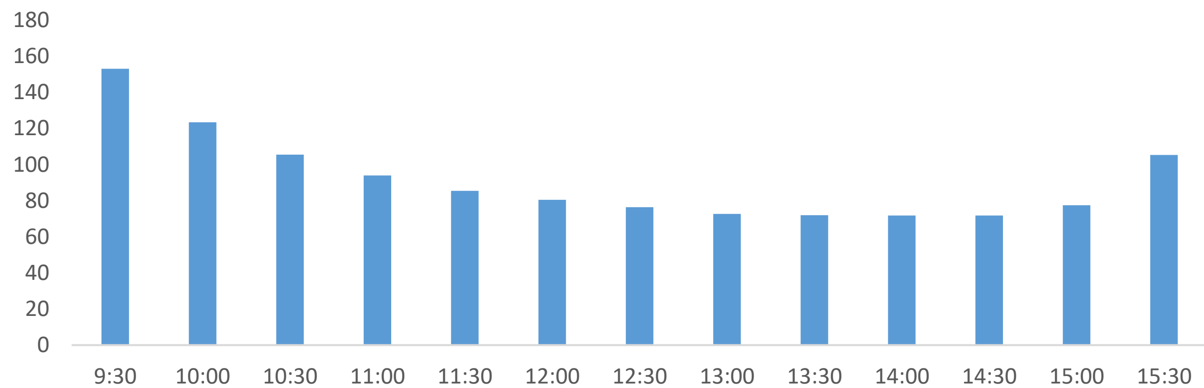
Figure III. Retail Intraday Trading

These figures report summary statistics for the intraday marketable retail investor trading volume in Panel A (number of trades in Panel B), identified using algorithm from Boehmer et al. (2021). Our sample period starts from January 2020 to March 2022 and our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. On each panel, we divide one trading day into 13 intervals, starting from 09:30am to 16:00pm. For each interval, we report the time-series average of cross sectional mean of marketable retail investor trading volume (trades).

Panel A. Trading volume



Panel B. Number of trades



Appendix Table A.1 Retail Attention and Stock Returns in the Long Run

This table examines retail investors' attention and the stock return in the long run. Our sample period is January 2020 to March 2022, and our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. We focus on the abnormal Google search volume index (*ASVI*), from Da, Engelberg, and Gao (2011), which is defined as the log *SVI* during the current week minus the log median *SVI* during the previous eight weeks. We decompose retail order imbalances into attention components (*Att*) and orthogonal component (*AttOrth*), and examine the retail order imbalances return predictive power is driven by attention-driven retail flows or the orthogonal components for the next 4 to 52 weeks.

Dep.var	Future Cumulative *w* weeks return			
Retail measures	Att		AttOrth	
	Coef.	*t*-Stat	Coef.	*t*-Stat
w=4	0.01386	0.60	0.00017	5.70
w=8	0.01097	0.57	0.00010	4.47
w=12	0.01646	1.31	0.00010	5.37
w=16	0.01693	1.14	0.00009	4.95
w=20	0.02186	1.33	0.00007	4.34
w=24	0.01493	1.15	0.00008	4.63
w=28	0.01450	1.10	0.00007	4.47
w=32	0.00777	0.84	0.00008	4.72
w=36	0.00345	0.38	0.00008	4.51
w=40	0.00525	0.70	0.00008	4.30
w=44	0.00615	0.70	0.00009	4.37
w=48	0.00562	0.54	0.00008	4.34
w=52	0.00533	0.45	0.00008	4.32

Appendix Table A.2 Retail Trading in High Retail Activity Stocks

This table reports retail investor trading predicting future stock returns, liquidity, volatility, high frequency trading and short selling, interacted with high retail activity subgroups. Our sample period is January 2020 to March 2022, and our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. If a stock i's *Actvol* ranks on the top quintile in the previous month of day t, then the high retail activity dummy $RAct_{i,t}$ takes 1, and 0 otherwise. For regression methodology, we adopt Fama-Macbeth to estimate the coefficients of retail order imbalances predicting next day return as in equation (3), and report the coefficients interacted with *RAct* dummy. We use panel regressions to estimate the coefficients of retail trades predicting next day liquidity, volatility, high frequency trading and short selling as in equation (4), and report the coefficients interacted with *RAct* dummy.

Regression	I	II	III	IV	V
Dep.var	Ret	Effspd	IntVol	HFTcancel	SDTCR
Retail measures	Oibvol	Actvol	Actvol	Actvol	Actvol
Retail(-1)	0.0007	0.0958	0.0199	-0.2078	-0.7656
t-Stat	7.85	8.51	1.59	-14.75	-7.60
Retail(-1)*RAct(m-1)	0.0011	0.0031	0.0610	-0.0245	-0.5372
t-Stat	4.18	0.35	5.83	-2.42	-6.85
Adj.R2	0.0800	0.790	0.730	0.564	0.709

Appendix Table A.3 Retail Trading over Weekly Horizon

This table reports the retail activity predict future stock returns, liquidity, volatility, high frequency trading and short selling over weekly horizon. Our sample period is January 2020 to March 2022, and our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. Panel A present weekly retail order imbalances predict next week measures. Panel B present weekly retail order imbalances predict cumulative *n*-week ahead stock returns, scaled by involved weeks.

Panel A. Weekly retail trading predicts next-week return, market quality and other market participant activity measures

Regression	I		II		III		IV		V	
Dep.var	Ret		Effspr		Intvol		HFTCancel		SSDTCR	
Retail	Oibvol		Actvol		Actvol		Actvol		Actvol	
	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat
Retail(-1)	0.0008	4.68	0.1410	9.80	0.0864	5.90	-0.2555	-9.53	-1.5357	-8.67
Controls	Yes		Yes		Yes		Yes		Yes	
Adj.R2	0.072		0.886		0.858		0.681		0.809	

Panel B. Weekly retail trading predicts returns over the long run

Retail measures	Oibvol Coef.	*t*-Stat
w=1	0.00080	4.68
w=2	0.00059	4.94
w=3	0.00050	5.00
w=4	0.00051	4.90
w=5	0.00047	5.21
w=6	0.00038	4.76
w=7	0.00030	3.82
w=8	0.00030	4.06
w=9	0.00030	4.39
w=10	0.00030	4.72
w=11	0.00030	5.01
w=12	0.00031	5.39