When Price Discovery and Market Quality Are Most Needed:

The Role of Retail Investors Around Pandemic

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

First version: March 2022
This version: January 2025

ABSTRACT

We identify a broad swath of marketable retail investor orders in the U.S. stock market over 2018 to 2023. The average marketable retail daily trading volumes rapidly rise from $17 billion between 2018-2020 to $32 billion between 2020-2021, and stay high for the next two years. We study the trading patterns of retail investors in three aspects. First, we examine policy and technology factors that might lead to more retail trading, and find that government's relief checks, the Fed's monetary policies, retail investors' rising attention towards trading apps and social media all contribute to the increase of retail trading, especially the latter two. Second, the retail order flows positively predict cross-sectional returns over daily and weekly horizons, with stronger predictive power during and after the pandemic than beforehand. We investigate the predictive information embedded in retail flows, and find they are associated with future news sentiment. Third, for market quality measures, higher retail trading is associated with wider future effective spreads and higher future volatility measures throughout our sample period. Interestingly, these relations are mostly weaker during the pandemic, suggesting that retail investors likely demand less liquidity and generate less uncertainties during this special period. Future participations of high-frequency traders and short-sellers are negatively associated with retail trading.

Keywords: pandemic, retail investors, market quality, high frequency trading, short selling.

JEL codes: G11, G12, G14, G23.

---

*Charles M. Jones was at Columbia Business School, Lin Tan is at the PBC School of Finance at Tsinghua University, Xiaoyan Zhang is at the PBC School of Finance at Tsinghua University, and Xinran Zhang is at Central University of Finance and Economics. Xiaoyan Zhang acknowledges the financial support from China NSF Project 72350710220 and Beijing NSF Project IS2317. Xinran Zhang acknowledges the financial support from China NSF Project 72303268. We acknowledge financial support from Plato. 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.

When Price Discovery and Market Quality Are Most Needed:

The Role of Retail Investors Around Pandemic

ABSTRACT

We identify a broad swath of marketable retail investor orders in the U.S. stock market over 2018 to 2023. The average marketable retail daily trading volumes rapidly rise from $17 billion between 2018-2020 to $32 billion between 2020-2021, and stay high for the next two years. We study the trading patterns of retail investors in three aspects. First, we examine policy and technology factors that might lead to more retail trading, and find that government's relief checks, the Fed's monetary policies, retail investors' rising attention towards trading apps and social media all contribute to the increase of retail trading, especially the latter two. Second, the retail order flows positively predict cross-sectional returns over daily and weekly horizons, with stronger predictive power during and after the pandemic than beforehand. We investigate the predictive information embedded in retail flows, and find they are associated with future news sentiment. Third, for market quality measures, higher retail trading is associated with wider future effective spreads and higher future volatility measures throughout our sample period. Interestingly, these relations are mostly weaker during the pandemic, suggesting that retail investors likely demand less liquidity and generate less uncertainties during this special period. Future participations of high-frequency traders and short-sellers are negatively associated with retail trading.

Keywords: pandemic, retail investors, market quality, high frequency trading, short selling.

JEL codes: G11, G12, G14, G23.

1. Introduction

The outbreak of the COVID-19 pandemic in the spring of 2020 witnesses two significant changes in the U.S. stock market. First, large negative shocks brought by the pandemic significantly reduce market returns by 13.3% in March, while market uncertainty measured by VIX shot up from 18% to 83%, and market liquidity quickly worsens, with effective spreads quadrupling at the same time. Second, retail investors' involvement in the U.S. stock market significantly increases after the pandemic starts. Here we identify the marketable retail order flow using the algorithm in Barber et al. (2024, BHJOS hereafter), and Table I Panel A shows that the daily aggregate marketable retail trading volumes rapidly increase from $17.26 billion for the pre-pandemic period (January 2018 to February 2020) to $32.25 billion for the pandemic period (March 2020 to April 2021), and remain high at $31.86 billion for the post-pandemic period (May 2021 to December 2023). Lessons from previous crisis periods show that the smooth functioning of the capital market during volatile times is extremely important. Do large inflows of retail investors help to improve price discovery, provide liquidity and calm the market during these special years?

Previous literature on retail investors provides many mixed results supporting different hypotheses on retail investors' role in price discovery and market quality.[1] We believe the pandemic period and the increased participation of retail investors provide an ideal setting for understanding the dynamics of retail investors and overall market dynamics. Specifically, we answer the following research questions: why retail trading increases significantly during the pandemic and whether they remain in the market after the pandemic; how the mix of retail

[1] In Appendix A, we provide a summary of related papers and make comparisons of their predictions and findings.

investors, including both experienced and novice investors on Robinhood,[2] contributes to price discovery and market quality before, during, and after the pandemic.

To understand the reasons for the increase of retail trading after the outbreak of the pandemic, we investigate two policy triggers and two technology triggers. Greenwood et al. (2023) find that fiscal policy of the government, such as three rounds of relief checks entitled the "Economic Impact Payments", lead to more retail trading. Cox et al. (2020) suggest positive monetary policy of the Federal Reserve, such as rate cuts of FOMC meetings can be another reason. We also consider two wide-spread technologies over this sample period, trading applications and social medias. Barber et al. (2021) show that the new generation of trading platforms offers commission-free trading, is simple and engaging, and might encourage retail investors to participate. Pedersen (2022) discusses the 2021 Gamestop trading frenzy, and suggests that social media has significant influences on trading behaviors of retail investors.

Empirically, we find that the relief checks and rate cuts from the Fed are both significantly associated with higher retail activities. The aggregate retail trading volume increase by 2.05% on the arrival of relief checks days, and 1.01% on the announcement of rate cuts of FOMC days. More interestingly, retail investors' growing attention in popular trading apps and social media are even more significantly associated with higher retail trading. A 100% increase of Google search on trading apps is associated with a 6.19% increase of aggregate retail trading volume, while a 100% increase of social media attention is associated with a 6.04% increase of aggregate retail trading volume. Since social media data is also available at the firm-level, we examine the relation between firm-level retail volume and social media attention, and find the two are also significantly and

[2] According to Welch (2022), Robinhood investors triple their trading and account for 21% of the retail trading volume during the second quarter of 2020. Figure A1 in Bryzgalova et al. (2023) also shows that Robinhood takes around 20% of all brokerages' payment for order flow from 2020 to 2021.

positively related. That is, retail investors prefer to trade stocks with higher social media attention.

Before we examine retail investors' role in price discovery and market quality, here we briefly summarize the large volumes of existing work on retail investors to introduce economic intuitions, which we separate into three broad groups. We refer to Group I as the "pure noise retail investors hypothesis". Classical papers, such as Black (1986), assume that pure noise traders have no information regarding firm value. They can neither predict future price movements nor contribute to price discovery, but provides liquidity and improves market quality through trading, as in Glosten and Milgrom (1985) and Kyle (1985). Group II contains works consistent with the "informed retail investors hypothesis". Kelley and Tetlock (2013) find evidence that marketable retail investors are informed about future firm fundamentals and their trading shows return predictability. If retail investors are informed, their trades demand liquidity from other market participants, which would increase transaction costs, and reduce market quality. Group III is referred to as "uninformed retail investors hypothesis". Studies in this group assume that the retail investors are uninformed about future firm fundamentals. However, the activities from these retail investors, through different channels, such as the attention channel from Da et al. (2011), and social media channel as in Pedersen (2022), still predict future prices movements. Notice that the three groups of hypotheses are actually not completely exclusive. In fact, group I and III have similar assumptions, and group II and III have overlapping empirical predictions. More discussions on the literature are summarized in Appendix A.

In terms of price discovery, or retail order flows' predictive power for future returns, Boehmer, Jones, Zhang, and Zhang (2021, BJZZ hereafter) examine the U.S. equity market between 2010 and 2015, and find that the retail order imbalances significantly predict future stock returns. However, over the more recent years, 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 significantly changes. With the heightened uncertainty and crash risks brought by the pandemic, it is unclear whether retail order flows can still predict returns in the cross section. It is also possible that the retail traders might have more time and resources to acquire and process information than usual, due to stay-at-home mandate and the rise of social medias, which may potentially help them to have stronger predictive power for future returns.

Our empirical results show that, for the pre-pandemic period, the retail order imbalances positively predict future one day ahead return. An interquartile increase in retail order imbalance is associated with an increase in daily return of 3.81 bps. For the pandemic and the post-pandemic period, the interquartile return difference becomes 5.91 bps and 6.69 bps, respectively. Comparing these numbers, we draw two inferences. First, retail order flows are always predictive of future returns across all three periods, which is consistent with previous literature. Second, it is intriguing to find that the predictive power of retail order flows actually increases from pre-pandemic to pandemic and post-pandemic. We investigate the predictive information embedded in retail flows, and find retail trading significantly predicts next day news sentiment. It is possible that the retail investors become more experienced and informed during and after pandemic about news sentiment.

In terms of market quality measures, we first focus on liquidity, and volatility measures. Take liquidity measures as an example, before pandemic, we find retail trading activities are positively associated with future effective spreads, and an interquartile increase in retail trading is associated with an increase of 1.42 bps for effective spreads. During and after pandemic, we find retail trading activities are still positively associated with future effective spreads, but the interquartile difference decreases to 0.27 bps (1.42 bps – 1.15 bps), respectively. Similar results are observed also for volatility measures. All these findings provide two implications: First, higher retail trading is generally associated with higher illiquidity, and higher volatility, which is consistent with the informed retail hypothesis. Second, the above associations are mostly weaker during pandemic, suggesting that some of the retail investors might behave like noise traders during these periods, and demand less liquidity and generate less volatility.

To further look into this possibility, we follow Greenwood et al. (2023) and use the arrival of relief checks as positive exogenous shocks to retail trading and further identify retail trading's potential impacts on market quality. This identification exercise shows that during pandemic, the exogenous shock to retail trading is associated with significantly lower effective spreads. That is, retail order flows help to provide some liquidity during pandemic. Results on volatility are insignificant.

Finally, we investigate how professional investors trade in the existence of increased retail order flows. The impacts the pandemic brings to professional investors probably significantly differ from those to retail investors. Presumably, the environment of low liquidity and high volatility during pandemic leads 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). HFT are generally believed to trade on arbitrage opportunities and are sensitive to short-term changes in prices and liquidity, while SS are assumed to be informed pessimistic investors who collect and process information regarding future price movements. We find that before pandemic the general activity levels of retail investors are associated with significantly lower activity levels of both HFT and SS, and the relations are particularly stronger during the pandemic. It is possible that increased trading by retail investors makes it harder for

HFT and SS to trade profitably, or the lower market quality associated with heightened retail trading activity makes it less attractive for HFT and SS to trade. The negative relation between retail volumes and HFT becomes significantly weaker after pandemic, while the negative relation between retail volumes and SS are still negative.

Our paper contributes to the emerging literature on retail trading activities during the pandemic, and the debate on how retail trading contributes to price discovery and market quality. Ozik et al. (2021), Eaton et al. (2022) and Baig et al. (2023) document the increase of retail trading activity at the beginning of the COVID-19. However, the patterns for retail trading dynamics during and after the pandemic remain unclear. Moreover, many studies provide mixed theoretical predictions and empirical evidence for price discovery and market quality. As for price discovery, while Kaniel et al. (2008), Kelley and Tetlock (2013), Barrot et al. (2016) and Boehmer et al. (2021) find retail order imbalances positively predict stock returns, some papers find retail trading lose money (Barber and Odean (2000), Barber et al. (2009)). Recent studies focus on Robinhood investors during the pandemic, but the results are still mixed: Welch (2022) show that the aggregated crowd consensus portfolio has good timing and good alpha, while Barber et al. (2022) find that Robinhood investors engage in more attention-induced trading, which results in negative returns. As for market quality, Barrot et al. (2016) and Ozik et al. (2021) find retail investors provide liquidity especially in volatile market, Eaton et al. (2022) show decreased Robinhood participation is associated with higher market liquidity, and Baig et al. (2023) find a heightened negative impact of retail trading on the stability of financial markets during the pandemic.

Compared to the existing literature, our study makes three important new contributions. First, we extensively investigate alternative hypotheses on driving forces of the increases in retail trading, and find that the relief checks, the Fed's rate cuts, the utilization of zero-commission trading apps, and the growing influence of social media are all significant contributors. Second, our six-year sample allows us to rigorously examine the differences in retail trading's predictive power for future returns before, during and after the pandemic. Interestingly, we find the predictive power of retail order flows is even stronger with the pass of the Pandemic. We provide supportive evidence that retail order flows are associated with future news sentiment. Third, we examine how the market quality measures evolve around retail trading over pre-, during and post-pandemic periods. Generally, higher retail trading is associated with higher illiquidity and higher volatility, but the association is weaker during the pandemic. To establish causality, we adopt distribution of relief checks as exogenous shocks, and find that during pandemic, the positive shocks to retail trading is associated with significantly lower effective spread, suggesting that retail order flows help to provide some liquidity during the pandemic.

2. The Rise of Retail Investors Around the Pandemic

2.1 Data

Our sample starts on January 1 of 2018, and ends on December 31 of 2023, a total of 1509 trading days. To separate our sample into clear pre- and post- pandemic periods, we rely on the U.S. weekly COVID-19 new cases of deaths reported to the World Health Organization (WHO). Figure I Panel A shows that the COVID-19 new cases experience a rapid surge starting in March 2020, followed by a decline to a relatively stable low level after April 2022, when the head of the WHO announces that the reported COVID-19 cases and deaths decline significantly.[3] Therefore, we separate our sample into pre-pandemic (January 2018 to February 2020), pandemic (March 2020 to April 2022), and post-pandemic (May 2022 to December 2023) periods.

[3] https://news.un.org/en/story/2023/05/1136367

We use the TAQ dataset to identify the marketable retail order buys and sells. Boehmer et al. (2021, BJZZ thereafter) point out that retail trades often receive price improvements in fractions of a penny and are routed to the FINRA trade reporting facility (TRF), and develop an algorithm to identify retail buy and sell orders using sub-penny digits for all marketable orders. Given that there are more wider-than-1-cent orders over recent periods, a recent study by Barber et al. (2024, BHJOS hereafter) provides a modified algorithm to identify the retail buy and sell direction using the quoted spread midpoints. We use the BHJOS algorithm for our main results. Notice that the above algorithms only identify marketable retail orders, which tend to be more aggressive, but not limit order submitted by retail investors, which tend to be less aggressive.

Following existing literature, we merge the retail trading data with stock returns and accounting data from CRSP and Compustat. We adopt the following standard filters: common stocks with share code of 10 or 11 (which excludes mainly ETFs, ADRs, and REITs); listed on the NYSE, NYSE MKT (formerly the Amex), or NASDAQ; and stock price higher than $1 on the previous trading day. For each day, we have an average of around 3,300 firms included in the sample. Overall, we have 5.1 million stock-day observations.

2.2 Increases in Retail Trading

To understand the aggregate magnitude and trend of marketable retail order flows in our sample period, we first calculate market-level aggregated activity of retail investors ($RVolume_t$) by summing up both retail buy dollar volumes ($RetailBuyVol_{i,t}$) and retail sell dollar volumes ($RetailSellVol_{i,t}$), identified by BHJOS algorithm for stock *i* on day *t* across all stocks, and dividing by the total dollar volumes ($TotalVol_{i,t}$) over all stocks:

$$RVolume_t = \frac{\sum_{i=1}^{N} RetailBuyVol_{i,t} + \sum_{i=1}^{N} RetailSellVol_{i,t}}{\sum_{i=1}^{N} TotalVol_{i,t}}. \quad (1)$$

Here N represents total number of stocks.

Table I Panel A presents the aggregate daily retail trading pre-, during and post-pandemic. The daily aggregate retail trading is on average $17.26 billion before the pandemic, almost doubles to $32.25 billion during pandemic and stays high at $31.86 billion after the pandemic. As for the daily trading volume as percentage of total volume, $RVolume_t$, retail trading gradually increases from 9.89% before the pandemic, to 11.31% during the pandemic, and remains high at 12.28% after the pandemic. Overall, these results show that retail trading increases during the pandemic, and remains high after the economy reopens. We present the time-series plot in Figure I Panel B. Other than the general increasing trend, we also observe spikes in retail trading during the episode of GameStop in January 2021, and expectation of Ukraine War in late 2021. Panel C presents the percentage of retail trading volume as a percentage of market total trading volume. Retail trading accounts for around 10% of total market trading volume during the year 2018, which increases to about 13% during the year 2023.

2.3 What drives the increases in aggregate retail trading?

It is important to understand the reasons behind the rise of retail investors trading during and after the pandemic. Here we list four possible hypotheses from policy and technology perspectives. According to Greenwood et al. (2023), the arrivals of relief checks provide retail investors with additional funds, some of which might be directly invested in the stock market. There are three rounds of relief checks in the U.S. during the pandemic: April 13, 2020, December 30, 2020 and March 12, 2021. Therefore, we define event-days dummy for relief checks, $DRelief_t$, as an indicator variable that takes the value of one in the 3-day window, [0, +2], upon and following each of the three first actual payment dates. It takes the value of one for a total of 9 days of the sample: April 13, 2020, and the subsequent 2 trading days; December 30, 2020, and the subsequent 2 trading days; March 12, 2021, and the subsequent 2 trading days.

Meanwhile, Cox et al. (2020) point out that the decline in interest rates, coupled with the Federal Reserve's commitments to maintaining a low target rate for a longer period, bolsters market confidence and motivates retail trading. For Federal Reserve's decisions to cut interest rates and reaffirming indications to zero lower bound on interest rates during the pandemic period, we obtain the scheduled and unscheduled Federal Open Market Committee (FOMC) announcement dates from the Federal Reserve Board website.[4] Our sample period contains 49 FOMC days. After the outbreak of COVID-19 pandemic, the Fed rapidly lowers the target range for the federal funds rate to bolster the economy. The Fed holds two unscheduled meetings on March 3 and March 15 of 2020 and cuts the target range for the federal funds rate by a total of 1.5 percentage points, dropping it to near zero. Moreover, the FOMC states that it intends to maintain the zero lower bound on interest rates for a good while and leaves its key interest rate unchanged for the next 2 years. Appendix B presents these positive FOMC announcement dates during the pandemic. We define the event-days dummy for positive FOMC announcements, $DFOMC_t$, as an indicator variable that takes the value of one on the days when the FOMC decides to cut interest rates or reaffirms to maintain the zero lower bound on interest rates, which includes 17 FOMC announcement dates.

For technology, Barber et al. (2022) point out that the introduction of retail trading platforms, particularly those with zero trading commissions, lowers the barriers to entry for retail investors and facilitates retail participation. For the rising popularity of trading apps, we consider four retail platforms: Robinhood, TD Ameritrade, Charles Schwab and E-Trade. We use the Google Trends of retail platforms' names to proxy for their popularity, and obtain the updated Google Trends of their names. Google Trends provides search results by monthly frequency with

[4] https://www.federalreserve.gov.

a normalized structure, where the values only range from 0 to 100. Variable $TrdApp_t$ is the average of Google Trends measures of four retail brokers' names, divided by 100 to range between 0 and 1. We plot the time-series of Google Trends in Figure II. The attention towards retail brokerages significantly increases during the pandemic, and are still relatively higher after pandemic than before pandemic. One obvious spike is in January 2021, potentially linked to the Gamestop episode.

Finally, as in Pedersen (2022), the exponentially growth of social media provides channels for retail investors to gather information and interact with others, especially during the pandemic period when people have more time to engage online. To construct the social media attention, we adopt two datasets. The first dataset is from Cookson et al. (2024) with the sample period from 2018 to 2021, compiled from three major platforms: Twitter, StockTwits, and Seeking Alpha. It provides the firm-day level common social attention across different platforms. To be comparable with other measures, we first scale the magnitude of the firm-level attention indices to have a minimum value of zero and a maximum value of one, and obtain firm-level proxy for social media attention, $SocMed_{it}$. Then we compute market aggregate social attention proxy, $SocMed_t$, by taking the average of the firm-level measures. The second dataset is compiled from the Reddit forum "WallStreetBets" with the sample period from 2020 to 2023, as in Hu et al. (2024). It provides number of firm-day level posts on the "WallStreetBets". We use the daily total number of firm-level posts as a proxy for social media attention. Again, we scale the raw data to have a minimum value of zero and a maximum value of one to be comparable with other measures. Variable $SocMed_t$ and $SocMed_{it}$ are defined similarly to the Cookson et al. (2024)'s dataset.

To examine how aggregate retail trading activities are related to the four proposed explanations, we estimate time-series regressions of the following form:

$$RVolume_{t+1} = a + b_1 DRelief_{\mathrm{t}} + b_2 DFOMC_t + b_3 TrdApp_t + b_4 SocMed_t + u_t. \quad (2)$$

Here, the dependent variable represents the daily market-level retail trading activities as defined in equation (1). We use Newey-West standard errors with optimal lag numbers to account for the serial dependence in retail trading.[5] If any of the candidate explanations is important, we expect a significant coefficient.

We report the estimation results in Table II Panel A. In regressions I and II, the coefficients on event-days dummies for relief checks and positive FOMC announcements, *DRelief* and *DFOMC*, are 0.0205 and 0.0101, with *t*-statistics of 2.92 and 2.68, respectively. The positive and significant coefficients reflect that the retail trading activities increase 2.05% on relief checks days and 1.01% on positive FOMC announcement days. The adjusted R^2 is 0.005 for the relief checks days and 0.002 for the FOMC announcement days, suggesting that these events only explain a small part of the increasing of retail trading. In regression III, the coefficient for *TrdApp* is 0.0619, with significant *t*-statistics of 8.38, indicating that a 100% increase in search of trading apps is associated with a 6.19% increase of retail trading activities. More importantly, the adjusted R^2 is 0.324, suggesting that the trading apps explain a substantial part of retail trading activities.

In regression IV (Cookson sample) and V (Reddit sample), the coefficients for *SocMed* are 0.0604 and 0.0683, both are statistically significant, indicating that a 100% increase in social media attention is associated with an increase of around 6% for retail trading activities. The adjusted R^2 is 0.186 for Cookson sample, and 0.051 for Reddit sample, suggesting that social media is also an important driver of retail trading activities.

[5] The number of lags equals to seven in our whole sample period from January 2018 to December 2023, as well as the sample period from January 2018 to December 2021, following Newey and West (1987), where we use integer $[T^{1/4}$+1.5] to calculate the optimal bandwidth parameter with number of days in the two samples T equals 1509 and 1008 respectively. Our results are robust if we use other algorithms to compute the optimal lag.

In regression VI and VII, when we consider all factors simultaneously, the coefficients on trading apps and social media attention remain significant, while the coefficients on relief checks and positive FOMC announcements become insignificant, suggesting that the trading apps and social media are the main driver of the rise of retail trading. The adjusted R^2 are quite high at 0.516 for Cookson attention sample and 0.087 for Reddit sample, suggesting that these factors explain a significant part of the time-variation of aggregate retail trading.

Ozik et al. (2019) document that retail trading increases at the outbreak of pandemic because retail investors have ample free time and access to trading platforms during that period. Barber et al. (2021) find that the innovations by Robinhood brokerage attract concentrated retail trading. Martineau and Zoican (2023) document a sharp increase in retail trading at the beginning of 2020. In comparison with these papers, our study has three new findings. First, we examine the rising retail trading not only at the outbreak of pandemic, but also throughout the pandemic and till a few years after, and thus provide a comprehensive perspective. Second, we combine all factors from existing studies, and jointly examine their explanatory power. We find that all four factors lead to long term increase in retail trading during and after the pandemic, especially the attention to trading apps and social media. Third, we provide results at firm level and show retail trading is significantly associated with higher social media activities. These new results clearly broaden our understanding of retail trading and contributing factors.

Needless to say, the pandemic itself plays an important role, too, because it is the reason for the fiscal and monetary policy interventions, and it provides an ideal setting for the quick and wide spread of trading apps and social media influences. Cong et al. (2024) document significant economic benefits of digitization in increasing SMEs' resilience against the COVID-19 pandemic, and finds the digitization of firms does not revert after the pandemic. Our finding is consistent with

Cong et al. (2024) in the sense that the pandemic accelerates the application of technology, which does not revert back after the reopening, and lead to long-term or permanent changes in trading.[6]

2.4 What drives the increases in firm level retail trading?

We are also interested in understanding which factors explain retail trading at firm level. Notice that relief checks, rate cuts, and search volumes for trading apps are market aggregates rather than firm level observations. So here we focus on whether higher firm-level social media attention, $SocMed_{it}$, is associated with more retail trading. Still, we have two firm-level social media attention measures from Cookson et al. (2024) and Hu et al. (2024). We compute firm-level activity of retail investors ($RVolume_{i,t}$) by summing up both retail buys and sells volumes for each stock on day *t,* and dividing by the stock's total buys and sells from all investors:

$$RVolume_{i,t} = \frac{RetailBuyVol_{i,t} + RetailSellVol_{i,t}}{TotalVol_{i,t}} \tag{3}$$

Table I Panel B presents firm level summary statistics on our sample firms. We compute the mean, and standard deviation of the pooled stock-day sample. Before pandemic, the average of firm level daily retail dollar volume is \$5.255 million, and it increases to \$9.3954 million during pandemic, and remains high after pandemic. These patterns are consistent with the aggregate retail trading volumes. But, as percentage of total trading volumes across all investors, the firm level $RVolume_{it}$ has a mean of 0.0970 before pandemic, and it decreases to 0.0891 and 0.0898 during and after pandemic. This might appear inconsistent with the aggregate numbers presented in Panel A. We examine the raw data and find that retail volume is particularly high for some groups during

[6] Similar patterns are also observed in China. We are fortunate to obtain the proprietary trading data from one major stock exchange in China, which includes retail trading data from January 2018 to May 2020, during which January 2018-December 2019 is pre-pandemic, and January 2020-May 2020 is pandemic. We find that increase of retail investors trading is closely related to trading apps (proxied by Baidu search of trading apps) and social media (proxied by Guba comments). The empirical results are presented in Appendix Table I.

and after pandemic, but not for all firms, and that's why the means across all firms don't display upward trends.[7]

To better understand which factors can predict retail trading, we estimate a Fama-MacBeth (1973) regression to predict next day retail trading. That is, for each day *t*, we estimate the following cross-sectional specification:

$$RVolume_{i,t+1} = b_{0t} + b_{1t} RSocMed_{i,t} + b_{2t}{}' Controls_{i,t-1} + u_{2,t}. \quad (4)$$

We include both social media information and firm level characteristics as independent variables. Since we now focus on the cross-sectional dimension, we adopt a rank transformation for the two version of firm-level social media attention to minimize impact of outliers (such as Gamestop), as in Cao and Narayanamoorthy (2012) and Livnat and Mendenhall (2006). For each day *t*, we first rank the social median attention variable cross-sectionally into 100 groups, from the lowest to the highest. Then we use the rank variable divided by 100 as a new social media attention variable, $RSocMed_{i,t}$, for firm *i* on day *t*. The regression coefficient on this rank variable thus can be comparable for different versions of social media attentions we examine. We follow existing studies to select important firm characteristics as control variables. 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.

After we obtain the time-series of parameter estimates $\{b_{1t}, b_{2t}'\}$ from the cross-sectional regressions, we conduct inferences on the mean and standard errors of these parameter estimates

[7] We report the cross-sectional distribution of marketable retail activities in Appendix Figure I Panel A. The time-series of the mean, 25th percentile, the median, and the 75th percentile of *RVolume* is relatively stable over time.

over pre-, during and post-pandemic periods, while the standard errors are adjusted using Newey-West (1987) approach for daily regressions.[8] If social media attention or firm characteristics predict future retail activity for any period, we expect a significant coefficient of b_1 or b_2, the time-series average of b_{1t} or b_{2t} over that particular period.

The estimation results are reported in Table II Panel B. For the Cookson sample, the coefficients for *RSocMed* are 0.0230 and 0.0296 for before and during pandemic, both with significant *t*-statistics, indicating that firm-level retail trading activities are positively related to social media attention. An interquartile difference in social media attention is associated with an increase of 1.13% and 1.45% in next-day retail trading activities, respectively. When we use the second measure of Reddit social media attention, the patterns are similar but the magnitudes are smaller. For other control variables, we find that retail volumes are higher for firms with lower previous day returns, smaller firms, firms with lower BM ratios, and firms with higher turnovers and volatilities. The adjusted R^2 are mostly above 60%, suggesting that social media, together with firm level characteristics explain the majority cross-sectional variations in retail trading volumes.

Overall, in this section, we document that there is a significant rise of retail investors trading during pandemic, which stays high for the next few years. At market level, we find that the distribution of relief checks, positive monetary policy from the Fed, the attention zero-commission trading apps, and the growing influence of social media, especially the latter two, all lead to long term increase in retail trading during and after the pandemic. Firm level evidence also shows that social media attention is an important contributor to retail trading.

[8] The number of lags equals six in each of our pre-pandemic, pandemic and post-pandemic periods. Following Newey and West (1987), we use integer [$T^{1/4}$+1.5] to calculate the optimal bandwidth parameter, with number of days in our three subsamples, *T*, equals 543, 547 and 419 respectively. Our results are robust if we use other algorithms to compute the optimal lag.

3. Retail Trading and Return Predictability

3.1 Data and Methodology

To examine whether order flows predict future returns, we compute signed retail order imbalances measure as follows:

$$ROib_{i,t} = \frac{RetailBuyVol_{i,t} - RetailSellVol_{i,t}}{RetailBuyVol_{i,t} + RetailSellVol_{i,t}}. \tag{5}$$

The order imbalance measure reflects net buy and sell directions for retail investors, and is mainly used for predicting directional changes in future stock prices. From Table I Panel B, the retail order imbalance measure, *ROib*, has a mean of -0.0421 and a standard deviation of 0.4176 before pandemic. The mean of retail order imbalance increases to -0.0224 during pandemic, suggesting that retail investors are less likely to sell during pandemic. 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.[9]

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.[10]

To establish the predictive relation between retail trading and future returns, we mainly use current retail trades to predict next period movements in prices, and other variables. To be more specific, for predicting returns, we estimate a Fama-MacBeth (1973) regression, similar to the one

[9] We report the cross-sectional distribution of marketable retail order imbalances in Appendix Figure I Panel B. The time-series of the mean, 25^{th} percentile, the median, and the 75^{th} percentile of *ROib* is quite stable over time, except that there is a clear dip in March 2020 due to selling pressures for large market downturns.

[10] We also consider using number of trades rather than share volumes in equation (5). Results are similar to those using share volumes and are available on request.

in Kelley and Tetlock (2013). That is, for each day *t*, we estimate the following cross-sectional specification:

$$Ret_{i,t+1} = b_{0t} + b_{1t}ROib_{i,t} + b'_{2t}Controls_{i,t-1} + u_{i,t}. \tag{6}$$

Here the dependent variable $Ret_{i,t+1}$ refers to returns of stock *i* on day t+1, and the independent variable, $ROib_{i,t}$, refers to retail order imbalance measure on day *t*. We obtain the time-series of parameter estimates $\{b_{1t}, b'_{2t}\}$ from the cross-sectional regressions, and conduct inferences on the mean and standard errors of these parameter estimates over pre-, during and post-pandemic periods, while the standard errors are adjusted using Newey-West (1987) approach for daily regressions. If retail trading can predict future returns in the correct direction for any particular period, we expect a significant and positive coefficient of b_1, the time-series average of b_{1t} over particular periods.

3.2 Return Prediction Results

We first use retail order imbalance to predict next day return, as in equation (5), during pre-pandemic, pandemic and post-pandemic sample periods, and report the estimation results in Table III Panel A. The coefficients on *ROib* for pre-pandemic, pandemic and post-pandemic periods are 0.0009, 0.0017, and 0.0018, respectively, with corresponding *t*-statistics being 13.22, 10.92 and 15.29. The positive and significant coefficients suggest 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, during both normal and pandemic periods. For economic magnitude, an interquartile increase in retail order imbalance is associated with an increase in daily returns of 0.0381% (0.0381% * 250 days = 9.5% annualized). The magnitude increases to 0.0591% (0.0591% * 250 days = 14.8% annualized) for pandemic period, and 0.0669% (0.0669% * 250 = 16.7% annualized) for post-pandemic period. The differences between pre-pandemic and pandemic periods and between pre-pandemic and post-pandemic are both statistically significant at 1%.

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. Therefore, we extend equation (6) to longer horizons. That is, we use retail order imbalance measures to predict *k*-week ahead returns, $Ret_{i,w+k}$, with *k*=1 to 12. Instead of using a cumulative return over *n* weeks, we focus on a weekly return over a one-week period to observe the decay speed. To be specific, for *k*=1, $Ret_{i,w+1}$ is the weekly return from day *t+1* to day *t*+5; for *k*=12, $Ret_{i,w+12}$ is the weekly return from day *t*+56 to day *t*+60. If the marketable retail order imbalances have only short-lived predictive power for future returns, the coefficient b_1 would decrease to zero quickly. Alternatively, if the marketable retail order imbalances have longer predictive power, the coefficient b_1 should remain positive for a long period.

We report the results in Panel B of Table III. For the pre-pandemic period, when the estimation window is extended from 1 to 12 weeks, the interquartile weekly return differences on *ROib* gradually decrease from 0.0893% to 0.0261%. The coefficients on *ROib* are statistically significant for the first 4 weeks. For the pandemic period and post-pandemic period, we find the interquartile weekly return differences are relatively larger than those for pre-pandemic period, and they are significant for the next 5-6 weeks.

The general finding of retail order flow positively and increasingly predicting next day and next few weeks of returns does not support the "pure noise retail investor hypothesis", but is consistent with the other two hypotheses. Given that the predictive power of the retail order imbalances persists for multiple weeks, they potentially capture relatively longer-term information.

More interestingly, retail order flow's predictive power for future returns seems to be stronger during and after pandemic.

3.3 Why Retail's Predictive Power Increases for Pandemic and Post-Pandemic Periods?

These findings might be surprising to readers who have been paying attention to Robinhood investors, and their lack of trading experiences. There are two possibilities. First, even though Robinhood investors attract substantial attention from media and regulators, according to Welch (2022) and Bryzgalova et al. (2023), they still only account for around 20% of the total retail order flows at the outbreak of the pandemic, and probably cannot overturn the empirical pattern that the more general retail flow has return predictability. 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 or news sentiment. Second, the unexperienced retail investors at the beginning of pandemic might become more experienced during pandemic and can process information better which leads to the stronger predictive power.

From previous literature, future news contains comprehensive information, such as Jeon et al. (2022) and Dang et al. (2015). Ravenpack news sentiment include all news related to firms and already include earnings announcement events. We obtain data on all public news events over year 2018-2023 from RavenPack Equity Module database. Three filters are applied to the data. First, to include the most relevant news events related to the stock, we require the relevance score to be 100, which means that the stock is prominent in the news story. Second, to filter out news related to public price and return data, which doesn't contain much new information, we restrict the subject or theme of events to be "business", and exclude three groups of events "stock-prices", "order-imbalances", and "technical-analysis", which are usually press releases summarizing stock's recent price movements and past returns. Third, given that our aim is novel information and not stale news, we require the Event Similarity Days (*SIM*) to be more than 90 days, meaning that the news is a novel and has no proceeding similar reporting in the previous 90 days. In total, we have over 10 million intraday firm news releases. The event sentiment score (ESS) is a stock-event sentiment score between -1 and +1, computed by Ravenpack using its proprietary algorithm. This score is negative for negative news, 0 for neutral news, and positive for positive news.

We follow Engelberg et al. (2012) and use the following specification to examine whether retail investors trading can predict arrivals of future news:

$$FutureNews_{i,t+1} = b_{0t} + b_{1t}ROib_{it} + b_{2t}'Controls_{it} + \epsilon_{i,t+1}. \qquad (8)$$

The dependent variable, $FutureNews_{i,t+1}$, is the average ESS of the news release for firm *i* between market closes on day *t* and day *t*+1. If retail investors can predict the forthcoming news, we expect the coefficient b_1 (average of time-series of b_{1t}) to be positive. We estimate equation (8) using Fama-MacBeth regressions and adjust standard errors using Newey-West (1987) with six lags.

The estimation results are presented in Table IV. Before pandemic, the coefficient on *ROib* is 0.0010 for the pre-pandemic period, with a significant *t*-statistic of 5.00. An interquartile increase in retail order imbalance is associated with an increase in news sentiment of 0.0421%. The regression coefficients are 0.0016 for pandemic period, and 0.0014 for post-pandemic period, both highly significant. The interquartile differences are 0.0557% and 0.0522%, for these two periods, both significantly higher than the pandemic magnitude.

A direct reading of these results suggests that retail investors in general become better in predicting news sentiment. Notice that the predictive pattern doesn't necessarily mean that retail investors possess some private information or have advantages in information processing. It is possible that pandemic allows these retail investors to have more time to process information, or become more experienced. These effects seem to persist for the 2 years after pandemic. Interestingly, Farrell et al. (2022) show that social media actually help retail investors share information and help retail investors become better informed, which might be another possible explanation.

BJZZ (2021) find that retail order imbalances positively predict future returns over 2010-2015. Also, Eaton, Green, Roseman, and Wu (2022) find that BJZZ imbalances predict returns over the January 2019-August 2020 period. Our findings are consistent with these earlier findings. More importantly, we have a novel finding that the predictive power of retail flows is actually stronger during and after pandemic. To provide economic mechanism, we relate retail order flows to future news sentiment, and find significant and positive relations.

4. Retail Trading and Market Quality Around Pandemic

4.1 Data and Empirical Methodology

How the activity of retail investors associates with future firm-level liquidity and volatility? Recall the hypotheses in the introduction. The "pure noise" hypothesis predicts that retail investors provide liquidity to the market and reduce volatility. The "informed" retail investor hypothesis suggests that informed retail flows probably decrease market liquidity and increase volatility. This is consistent with inventory risk models, such as Ho and Stoll (1981) and Hendershott and Menkveld (2014), where market maker bears non-diversifiable inventory risk from retail trading, and retail trading lowers market liquidity. The "uninformed" retail investor hypothesis, on the other hand, is silent on this issue. For our setting, if retail investors provide liquidity to the market, then higher retail activity would be associated with lower effective spreads in the future. If retail trading stabilizes the market, we expect that higher retail activity is associated with lower volatility for the future.

Previous studies use many alternative liquidity measures, such as percentage effective spread, quoted spread, price impact and realized spreads. Given that trades can happen within the quoted bid and ask prices, the percentage effective spread is more precise, so we choose it as our main liquidity measure (denoted as effective spread hereafter).[11] 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 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, and $M_{i,k}$ 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).

Our data on volatility is also obtained from WRDS Intraday Indicators. The trade-based second-by-second intraday volatility of stock *i* on day *t* is calculated as: $IVol_{i,t} = \frac{\sum_{s=1}^{T}(Ret_{i,s}-\overline{Ret}_{i,s})^2}{T-1}$, where $Ret_{i,s} = Ln\frac{P_{k',s}}{P_{k,s-1}}$. Here *s* refers to second, *T* is the number of seconds that stock *i* trades on day *t* and P_k is the price of trade k. The annualized intraday volatility measure for stock *i* on day *t* is defined as $IntVol_{i,t} = \sqrt{250 * IVol_{i,t}}$, assuming there are 250 trading days in one year.

Table I Panel B presents the summary statistics of our market quality measures. The effective spreads, *EffSprd*, has a median of 0.1385 and a standard deviation of 0.7307 before

[11] Results using alternative measures are similar and available on request.

pandemic. The median of *Effsprd* increases during the pandemic, suggesting that market liquidity decreases in this period. For volatility, the median of *IntVol* is 0.1467 and it increases during the pandemic, indicating that market volatility increases in this period. [12]

Petersen (2009) suggests 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 and volatility, panel regression is preferred. The auto-correlations of daily firm level liquidity and volatility measures are 0.86 and 0.85. Given the strong persistence pattern, we adopt the panel regression,

$$Y_{i,t+1} = b_0 + b_1 RVolume_{i,t} + b_2 RVolume_{i,t} * Pandemic_t + b_3 RVolume_{i,t} * PostPandemic_t + b_4{}' Controls_{i,t} + \gamma_t + u_{i,t+1}. \tag{9}$$

Here the dependent variable, $Y_{i,t+1}$, represents the liquidity and volatility measures for stock *i* at time *t*+1, and the independent variable, $RVolume_{i,t}$, represents retail trading activities at time *t*. Dummy variable $Pandemic_t$ equals one for the days between March 2020 and April 2022 and zero otherwise. Dummy variable $PostPandemic_t$ equals one for days between May 2022 and December 2023. 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. Following previous literature, the standard errors are double clustered at day and firm level. If retail activities are related to future market quality measures differently during and after the pandemic, we expect coefficients b_2 and b_3 to be significant.

[12] To illustrate the cross-sectional pattern over time, we also plot their cross-sectional distributions in Appendix Figure I Panel C and D, where we present the cross-sectional means, p25 (the 25th percentile), p50 (the 50th percentile), and p75 (the 75th percentile) for each day or each month. In Panel C, the *EffSpd* is expressed in % and it spikes drastically from around 0.5% to nearly 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. We present time-series of volatility in Panel D, and observe a sharp increase of *IntVol* during March and April of 2020. Volatilities are relatively higher around the GME episode, Ukraine War and the collapse of Silicon Valley Bank.

4.2 Market Quality Results

We report the estimation results in Table V. In Panel A, when we use *RVolume* and its interaction terms with the *Pandemic* and *PostPandemic* dummy variables to predict next day effective spread, the coefficient on *RVolume* is 0.1889 for the pre-pandemic period, with a significant *t*-statistic of 19.74. An interquartile increase in retail volume is associated with an increase of 1.42 bps in effective spread, which is sizeable, given that the median effective spread is 16 bps for our sample. The positive coefficient indicates that, more activities from retail investors are associated with higher effective spread, or lower liquidity on the next day. This is consistent with the earlier finding that retail flows have predictive power for future returns, and is likely demanding liquidity.

More interestingly, the relation between retail activity and effective spread stays positive and significant, but is significantly weaker during pandemic and after pandemic. The coefficient on *RVolume***Pandemic* is -0.1531 (*t*-stat = -11.02), and the coefficient on *RVolume***PostPandemic* is -0.0252 (*t*-stat = -1.94). An interquartile increase for *RVolume* is associated with an increase of daily effective spread of 0.27 bps (0.0142-0.0115 = 0.27 bps) during pandemic and 1.23 bps (0.0142-0.0019 = 1.23 bps) post-pandemic, both significantly lower than the 1.42 bps for pre-pandemic period.

When we use our retail trading measures to predict next-day stock intraday volatility in Panel B, similar pattern emerges. The coefficient on *RVolume* is 0.1238, statistically significant for the pre-pandemic period. The coefficient on *RVolume***Pandemic* is -0.1019 (*t*-stat = -8.60), and the coefficient on *RVolume***PostPandemic* is 0.0194 (*t*-stat = 1.70). An interquartile increase for *RVolume* is associated with an increase of intraday volatility of 0.93%, 0.16% (0.0093-0.0077=0.16%) and 1.08% (0.0093+0.0015=1.08%) for these three periods, with the during-period significantly lower than the pre-pandemic and post-pandemic period. That is, in general, higher retail trading is associated with higher volatility, but this relation becomes significantly weaker during and post-pandemic.

To investigate whether the patterns persist in the long run, we replace the dependent variables in equation (9) from next-day market quality measures to market quality measures for the next 12 weeks. We report the difference in these measures associated with an interquartile increase of the retail volume measure in Panel B of Table V. Take liquidity as example, for the pre-pandemic period, the interquartile difference in effective spread decreases from 0.0117 to 0.0104, while the *t*-statistics remain highly significant for all weeks. During pandemic and post-pandemic, the negative changes in interquartile differences, ranging between -0.0068 and -0.0078, -0.0017 and -0.0030, meaning that the interquartile differences are much smaller than those in pre-pandemic period.[13]

Combining these findings together, retail investors' activities are significantly and positively associated with future illiquidity, and volatilities, over short and long horizons. These findings are generally consistent with the predictions from "informed retail investors hypothesis", but not that of "pure noise retail investors hypothesis". More intriguingly, the above relations are significantly weaker during and after the pandemic for liquidity and volatility. That is, retail trading activities in general are associated with worse market quality, but if we compare pre-pandemic with the other two periods, the "worse" part is less severe during and post the pandemic. It is possible that before pandemic, the mixture of retail investors is dominated by relatively

[13] In addition to examine the impact of retail trading on market quality over weeks 1 to 12, we also examine a similar pattern in week -12 to -1. Take week +1 and -1 as example, coefficient of b1 is 0.1058 and the coefficient of b1 0.1069, and the coefficients difference is -0.0011. These results suggest that that positive and predictive relation between past retail trading and future market quality is likely not driven by endogeneity. Section 4.3 provides an identification exercise using relief checks to alleviate the endogeneity concern. We thank our referee for this suggestion.

experienced investors, and their trades demands liquidity and increase volatility, and thus retail trading is associated with lower liquidity and higher volatility. As documented in previous sections, a large quantity of retail investors, experienced or unexperienced, participates in the stock market during the pandemic and they stay after pandemic. Some of these retail investors, potentially unexperienced ones, might actually provide liquidity, and cushion volatility to some extent.

4.3 Causal Analysis

To further look into this possibility, we introduce exogenous shocks to retail trading and further identify retail trading's potential impacts on market quality. Meanwhile, there is a concern that many events happen during our period, and it is possible that our results on market quality measures are driven by omitted variables. A formal identification strategy would help to address this issue. Previous literature provides two types of exogenous shocks: Eaton et al. (2022) use brokerage platform outages (outages at Robinhood vs. outages at traditional retail brokerages) to examine the impact of retail investors on market quality. They find that for stocks with high retail interest, the negative shocks to Robinhood investor participation are associated with increased market liquidity and lower return volatility, whereas the opposite relations hold following outages at traditional retail brokerages. A recent study by Greenwood et al. (2023) uses the relief checks to examine retail trading and future stock returns, but they do not examine the relation of retail trading and market quality. No studies that we are aware of examine the relation between retail trading and market quality measures, and our study provides new insights to the literature.

To be specific, we estimate the following panel specification:

$$Y_{i,t+1} = b_0 + b_1 RVolume_{i,t} + b_2 * DRelief_t * RVolume_{i,t} + b_3{}' Controls_{it} + \gamma_t + u_{i,t+1}. \quad (10)$$

Here the dependent variables include next day liquidity, and volatility measures. Variable *DRelief*, as an indicator variable that takes a value of one in the 3-day window [0, +2] upon and following one of the three first actual payment dates. We include day fixed effect. The standard errors are double clustered at day and stock level. If the exogenous increase of retail trading during the pandemic is associated with better market quality, i.e. higher liquidity and lower market volatility, we expect coefficient b_2 to be significant negative.

We estimate equation (9) and report the estimation results in Table VI. For predicting next day *EffSpr*, and *IntVol*, the coefficients of *RVolume* are 0.1419, and 0.1008, both positive and statistically significant. More importantly, the coefficient b_2 on the interaction term of (*RVolume*DRelief*) are -0.1665 (*t*-Stat = -2.63) for effective spreads, suggesting that retail investors trading decreases the effective spreads on relief checks days. Economically speaking, an interquartile increase of retail trading indicates a decrease of 1.25 bps on relief checks days relative to other days. The negative and significant coefficient suggests that the positive exogenous shock from relief checks, which potentially increases retail trading, is associated with smaller effective spreads, meaning that increased retail trading might provide liquidity. For intraday volatility, the coefficient is negative, suggesting that increased retail trading on these days might also lower volatility, but the coefficient is statistically insignificant. These findings suggest that retail investors in general demand liquidity, but they likely demand less liquidity to certain degrees during the pandemic.

Our results are different from Eaton et al. (2022), in the sense that we use new exogenous shocks of relief checks to identify retail trading's connection with future market quality measures, and find different results from Eaton et al. (2022). While Eaton et al. (2022) find the Robinhood investors demand liquidity using outage as exogenous shocks, we find the increased retail trading leads to better liquidity on the relief check distribution days, suggesting that some retail investors might provide liquidity to the market. This different finding provides new insights to the question and complements results from previous literature.

5. Retail Trading and Other Investors

There are many participants in the stock market. Given the rise of retail investors during the pandemic, what are the dynamics between retail investor and 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 towards the price informativeness and market quality conditions. There are competing hypotheses regarding high frequency traders and short-sellers. On the one hand, as high frequency traders are believed to trade on arbitrage opportunities (Hendershott et al. (2011)), and short sellers are assumed to be informed investors (Boehmer et al. 2008), they might increase their trading when "pure noise" retail trading contains profitable arbitrage opportunities (Barber et al. (2022)). On the other hand, if retail flows have return predictability and their trades are associated with low liquidity and high volatility, then it would be hard for high frequency traders and short-sellers to trade profitably when retail investors are active.

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 the 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.

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.[14]

From Table I, the mean of the cancel to trade ratio *HFTCancel* is 2.9478, 2.9858 and 3.0083 for the pre-, during, and post-pandemic periods, respectively, indicating there are slight increases in high frequency trading over our sample period. For short selling measure, the mean is 6.2293, 4.2143 and 5.6294 for the pre-, during, and post-pandemic periods, respectively, showing that short-selling dips during pandemic period and bounce back after pandemic. For *SDTCR*, the median is 3.0459 with a standard deviation of 8.5609 in the pre-pandemic period.[15]

We estimate equation (9) with the dependent variables being *HFTCancel* for high frequency trading, and *SDTCR* for short-selling. To save space, we directly report in Table VII the changes in the *HFTCancel* and *SDTCR* measures when there is an interquartile increase in the retail trading. From the left half panel, there is an interquartile increase in the retail trading, the next-day *HFTCancel* significantly decreases by 0.0095, 0.0147 (-0.0095-0.0052=-0.0147) and 0.0044 (-0.0095+0.0051=-0.0044) for the pre-pandemic, pandemic and post-pandemic periods. Considering that the standard deviation for *HFTCancel* is 0.51 for this variable, then the above magnitude is economically meaningful. The first two interaction coefficients are statistically significant. The negative interquartile difference on the interaction coefficient implies that high

---

[14] Results using alternative measures, such as short supply and short flow, are similar and available on request.

[15] To illustrate the cross-sectional pattern over time, we also plot their cross-sectional distributions in Appendix Figure I, Panels F and G, where we present the cross-sectional means, p25 (the 25th percentile), p50 (the 50th percentile), and p75 (the 75th percentile) for each day. For the high frequency cancel-to-order ratio in Panel F, the cross-sectional distributions are relative stable over time and there is no obvious time trend. In Panel G, the shorting days-to-cover ratio spikes drastically adjacent to Christmas Day and Black Friday. It's possible that there is a decline in trading volume adjacent to these holidays, and days-to-cover ratio increase accordingly.

retail activities are associated with lower activities from high frequency traders on the next day, and the relation is the strongest during the pandemic. Similar patterns persist for the next 12 weeks. As mentioned earlier, the lower market quality associated with heightened retail trading activity likely makes it less profitable for HFTs to trade, especially during the special period of pandemic, which explains the large decrease in HFT. After pandemic, the negative relation between HFT and retail trading becomes weaker. It is possible that either some of the retail trades provide more arbitrage opportunities for HFT, or market quality improves (relative to pandemic) and more HFT returns to the market.

When we use *RVolume* to predict next day short selling in the right half panel, an interquartile increase in the retail trading is associated with significant decreases in short-selling, with the magnitude being -0.0112, -0.1047 (-0.0112-0.0935=-0.1047) and -0.0368 (-0.0112-0.0256=-0.0368) for the pre-pandemic, pandemic and post-pandemic periods. Similar patterns persist for the next 12 weeks. These negative coefficients suggest when retail investors trading more on a given day, there is less shorting of that stock the next day. This relation is the strongest during pandemic. Similar to the HFT, the lower market quality associated with heightened retail trading activity likely makes it less profitable for short-sellers to trade, especially during the special period of pandemic, which explains the large decrease in short-selling. But there is one difference, the Gamestop episode, during which famous short-sellers are squeezed when retail activities are extremely high. So after pandemic, the negative relation between SS and retail trading remain significant, even though that market quality generally improves (relative to pandemic), short-sellers still reduce their activity when there is more retail trading.

6. Conclusions

We study how retail investors are related to price discovery and market quality before, during and after the pandemic. First, we identify the marketable retail investors' orders in the U.S. market between January 2018 and December 2023. We clearly observe that marketable retail trading volumes almost doubles from a daily volume of $17 billion before pandemic to $32 billion during pandemic, and remain high at $32 billion after pandemic. We identify policy and technology factors that might lead to the rise of retail trading, and find that government's relief checks, the Fed's monetary policies, retail investors' rising attention towards trading apps and social media all contribute to the increase of retail trading, especially the latter two.

For return prediction and price discovery, retail order flows always positively and significantly predict future returns for all three periods, which is consistent with findings in previous literature. More interestingly, we observe that the predictive power of retail flows is actually stronger during and after pandemic. To find out why, we connect retail trading with firm news sentiment. Empirical results show significant and positive connection between retail trades and next-day news sentiment. It is possible that the retail investors become more experienced and informed during and after pandemic about news sentiment.

We also carefully investigate how the market quality measures evolve around retail trading over pre-, during and post-covid periods. Our results show that higher retail trading is associated with wider future effective spreads, higher volatilities, and lower participation of other investors in general. It is more intriguing to find that these relations are mostly much weaker during and after the pandemic period. To establish causality, we adopt distribution of relief checks as exogenous and positive shocks to retail trading, and find these shocks to retail trading are associated with smaller effective spreads, and lower market volatility. This finding suggests retail investors likely provide liquidity and help to cushion uncertainties to certain degree during the pandemic. Previous literature provides three broad types of hypotheses for retail investors trading: pure noise, informed and uninformed retail investors. It is possible that the retail population is a mix of all three types of investors.

Given retail investors trading's impact on price discovery, market quality, and other market participants during the Pandemic, regulators may need to carefully consider modifying policies on retail investors protection and their impact on market quality.[16] Still, our study clearly leaves many interesting questions unsolved. For example, is retail investors' trading a new form of systemic risk? We leave these interesting and important questions to future research.

[16] 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

Ang A, Hodrick RJ, Xing Y, Zhang X (2006) The cross-section of volatility and expected returns. *The Journal of Finance* 61(1):259–299.

Baig AS, Blau BM, Butt HA, Yasin A (2023). Reprint of: Do retail traders destabilize financial markets? An investigation surrounding the COVID-19 pandemic. *Journal of Banking & Finance* 147:106744.

Barber BM, Huang X, Jorion P, Odean T, Schwarz C (2024) A (sub) penny for your thoughts: Tracking retail investor activity in TAQ. *The Journal of Finance*, *79*(4), 2403-2427.

Barber BM, Huang X, Odean T, Schwarz C (2022) Attention-induced trading and returns: Evidence from Robinhood users. *The Journal of Finance* 77(6):3141–3190.

Barber BM, Lee YT, Liu YJ, Odean T (2009) Just how much do individual investors lose by trading? *Review of Financial Studies* 22(2):609–632.

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

Barrot JN, Kaniel R, Sraer D (2016) Are retail traders compensated for providing liquidity? *Journal of Financial Economics* 120(1):146–168.

Black F (1986) Noise. *The Journal of Finance* 41(3):528–543.

Blume ME, Stambaugh RF (1983) Biases in computed returns. *Journal of Financial Economics* 12(3):387–404.

Boehmer E, Jones CM, Zhang X (2008) Which shorts are informed? *The Journal of Finance* 63(2):491–527.

Boehmer E, Jones CM, Zhang XY, Zhang XR (2021) Tracking retail investor activity. *The Journal of Finance* 76(5):2249–2305.

Bryzgalova S, Pavlova A, Sikorskaya T (2023) Retail trading in options and the rise of the big three wholesalers. *The Journal of Finance*, *78*(6), 3465-3514.

Cong LW, Yang X, Zhang X (2024) Small and medium enterprises amidst the pandemic and reopening: Digital edge and transformation. *Management Science* 70(7):4564–4582.

Cookson JA, Lu R, Mullins W, Niessner M (2024) The social signal. *Journal of Financial Economics*, *158*, 103870.

Cox J, Greenwald D, Ludvigson S (2020) What explains the COVID-19 stock market? NBER Working Paper.

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

Dang TL, Moshirian F, Zhang B (2015) Commonality in news around the world. *Journal of Financial Economics* 116(1):82–110.

Eaton GW, Green TC, Roseman BS, Wu Y (2022) Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics* 146(2):502–528.

Engelberg JE (2012) How are shorts informed? Short sellers, news, and information processing. *Journal of Financial Economics* 105(2):260–278.

Fama EF, MacBeth JD (1973) Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81(3):607–636.

Farrell M, Green TC, Jame R, Markov S (2022) The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics* 145(2):616–641.

Glosten LR, Milgrom PR (1985) Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14(1):71–100.

Greenwood R, Laarits T, Wurgler J (2023) Stock market stimulus. *The Review of Financial Studies*, *36*(10), 4082-4112.

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

Hendershott T, Jones CM, Menkveld AJ (2011) Does algorithmic trading improve liquidity. *The Journal of Finance 66*(1), 1-33.

Hendershott T, Menkveld AJ (2014) Price pressures. *Journal of Financial Economics* 114(3):405–423.

Ho T, Stoll HR (1981) Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics* 9(1):47–73.

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

Hong H, Li W, Ni SX, Scheinkman JA, Yan P (2016) Days to cover and stock returns. NBER Working Paper.

Hu D, Jones CM, Zhang V, Zhang X (2024) The rise of Reddit: How social media affects belief formation and price discovery. Working Paper.

Jeon Y, McCurdy TH, Zhao X (2022) News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *Journal of Financial Economics* 145(2):1–17.

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

Kelley EK, Tetlock PC (2013) How wise are crowds? Insights from retail orders and stock returns. *The Journal of Finance* 68(3):1229–1265.

Kyle AS (1985) Continuous auctions and insider trading. *Econometrica* 53(6):1315.

Lee CMC, Ready MJ (1991) Inferring trade direction from intraday data. *The Journal of Finance* 46(2):733–746.

Newey WK, West KD (1987) Hypothesis testing with efficient method of moments estimation. *International Economic Review* 28(3):777–787.

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

Pedersen LH (2022) Game on: Social networks and markets. *Journal of Financial Economics* 146(3):1097–1119.

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

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

Weller BM (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 periods of Pre-Pandemic, Pandemic and Post-Pandemic are January 2018 to February 2020, March 2020 to April 2022 and May 2022 to December 2023, respectively. Our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. Panel A reports the total retail trading volume in billion and as percentage of total stock market volume. Panel B reports firm level summary statistics. For retail trading, we report retail trading dollar volume in million and as percentage of total stock volume (*RVolume*), and retail order imbalance measure (*ROib*). The daily stock return (*Ret*) is computed using bid-ask average prices and reported in percentage. For firm level news, we include Ravenpack news sentiment (*NewsSent*). For liquidity, volatility and price efficiency, we report effective spread (*EffSpr*) in percentage, annualized intraday volatility (*IntVol*). For other investors, we report high frequency trading measure cancel to trades (*HFTCancel*), and short selling days to cover ratio (*SDTCR*). For stock characteristics, we report market capitalization (*Size*) in billions, book to market ratio (*Lbm*), monthly stock turnover (*Lturnover*) and monthly return volatility (*Lmvol*).

Panel A. Aggregate daily retail trading

Variables	Pre-pandemic period	Pandemic period	Post-pandemic period
Total retail dollar volume ($ billion)	17.26	32.25	31.86
Total retail dollar volume as percentage of total volume	9.89%	11.31%	12.28%

Panel B. Firm level summary statistics

	Pre-pandemic period		Pandemic period		Post-pandemic period	
Variables	Mean	Std	Mean	Std	Mean	Std
Retail Dollar Volume ($ million)	5.2550	46.6442	9.3954	104.0190	9.0552	137.4080
RVolume	0.0970	0.1184	0.0891	0.0870	0.0898	0.0929
ROib	-0.0421	0.4176	-0.0224	0.3612	-0.0456	0.3806
Ret	0.0002	0.0387	0.0012	0.0524	0.0006	0.0510
NewsSent	0.0203	0.1154	0.0276	0.1333	0.0389	0.1510
EffSprd(%)	0.4448	0.7307	0.3767	0.5602	0.4313	0.6629
IntVol	0.3956	0.5950	0.3686	0.4983	0.4153	0.6048
HFTCancel	2.9478	0.5749	2.9858	0.5094	3.0083	0.5052
SDTCR	6.2293	8.5609	4.2143	5.5077	5.6294	7.2600
Size ($Bil.)	8.60	39.62	11.34	68.15	11.56	72.89
Lbm	0.64	1.85	0.66	1.87	0.75	1.83
Lturnover	0.27	10.33	0.41	2.18	0.31	2.32
Lmvol	0.03	0.03	0.04	0.04	0.03	0.05

Table II. Triggers for the Rise of Retail Trading Activity Around the Pandemic

This table reports results on how retail trading activities are related to three rounds of fiscal relief checks, the positive indications from monetary FOMC meetings, and the rising attention towards trading apps of retail brokerages and social media. The sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. Our sample periods of Pre-Pandemic, Pandemic and Post-Pandemic are January 2018 to February 2020, March 2020 to April 2022 and May 2022 to December 2023, respectively. Panel A estimates the time-series regressions, as specified in equation (2). The dependent variable is the daily market level retail activity measure, *RVolume*, as defined in equation (1). The key independent variables are the event-day dummy for relief checks, *DRelief*, in Regression I; the event-day dummy for positive indications from FOMC meetings, *DFOMC*, in Regression II; and the average of monthly updated Google Trends of four retail brokerages (Robinhood, TD Ameritrade, Charles Schwab and E-Trade), *TrdApp*, which is normalized to have a minimum value of zero and a maximum value of one, in regression III. Regression IV and V report how retail trading relate to market level social media attention, *SocMed*. Here we adopt two measures of social media attention, one is from Cookson et al. (2024) with sample period from 2018 to 2021, another is from Hu et al. (2024) with sample period from 2020 to 2023. We normalize the firm-level attention in Cookson et al. (2024) to have a minimum value of zero and a maximum value of one and then aggregate to the market level. For the social media attention from Hu et al. (2024), we use the daily aggregation of firm-level posts on "WallStreetBets", a Reddit forum, and normalized to have a minimum value of zero and a maximum value of one. Regression VI and VII include all the independent variables mentioned above. Panel B reports the estimation results for determinants of firm-level retail trading activity before, during, and after the pandemic using Fama-MacBeth regressions, as specified in (4). The dependent variable is the daily firm level retail activity measure, *RVolume*, as defined in equation (3). For regression I and II, we use the firm-level attention from Cookson et al. (2024) with sample period from 2018 to 2021; for regression III and IV, we use data from Hu et al. (2024) and measure the firm-level attention by the number of firm-level posts. To make the two datasets more comparable in the cross sectional, we adopt a rank transformation for the two version of firm-level social media attention. For each day t, we first rank the social media attention variable cross-sectionally into 100 groups, from the lowest to the highest, and then use the rank variable divided by 100 as a new social media attention variable, *RSocMed*. We use Newey-West standard errors with six lags to account for the serial dependence in retail trading. To understand the economic magnitude and compare between different periods, we multiply the interquartile range for social media with the coefficient of *RSocMed*, and generate the interquartile retail activity, *RVolume*, differences.

Panel A. Predict aggregate retail trading activity

	I		II		III		IV		V		VI		VII	
Dep. Var	RVolume(*t*+1)		RVolume(*t*+1)		RVolume(*t*+1)		RVolume(*t*+1)		RVolume(*t*+1)		RVolume(*t*+1)		RVolume(*t*+1)	
Period	2018-2023		2018-2023		2018-2023		Cookson 2018-2021		Reddit 2020-2023		Cookson 2018-2021		Reddit 2020-2023	
	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat
DRelief	0.0205	2.92									0.0005	0.04	0.0019	0.22
DFOMC			0.0101	2.68							0.0011	0.27	-0.0036	-1.06
TrdApp					0.0619	8.38					0.0616	7.83	0.0196	2.49
SocMed							0.0604	7.24	0.0683	4.76	0.0342	5.99	0.0361	2.71
Adj.R2	0.005		0.002		0.324		0.186		0.051		0.516		0.087	

Panel B. Predict firm-level retail trading activity before, during, and after the Pandemic

Dep. Var	RVolume (*t+1*)		RVolume (*t+1*)		RVolume (*t+1*)		RVolume (*t+1*)	
Period	Cookson Pre-Pandemic		Cookson Pandemic		Reddit Pandemic		Reddit Post-Pandemic	
	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat
RSocMed(*t*)	0.0230	34.01	0.0296	50.41	0.0138	25.01	0.0111	13.46
RVolume (*t*)	0.6458	131.23	0.5938	122.83	0.7653	157.86	0.7948	121.54
Ret(*t*)	-0.0339	-10.03	-0.0430	-17.18	-0.0309	-10.71	-0.0323	-5.74
Ret(*t*-1,*t*-5)	-0.0082	-1.36	-0.0154	-2.79	0.0117	1.77	0.0213	1.95
Ret(*t*-6,*t*-25)	-0.0033	-0.25	-0.0323	-3.04	-0.0238	-1.65	-0.0221	-0.83
Lnsize	-0.0085	-49.67	-0.0072	-56.95	-0.0024	-18.40	-0.0001	-1.08
Lbm	-0.0008	-4.47	-0.0002	-1.16	0.0001	0.25	0.0021	5.96
Lturnover	0.0004	2.51	0.0003	5.10	0.0014	7.42	0.0028	6.45
Lmvol	0.0200	3.18	0.0070	1.06	0.0764	8.61	0.2072	12.42
Interquartile next-day Rvolume diff	1.13%***		1.45%***		0.69%***		0.63%***	
Adj.R2	0.687		0.659		0.766		0.779	

Table III. Retail Order Imbalances Predict Future Stock Returns Before, During and After the Pandemic

This table reports estimation results on marketable retail investor order imbalance predicting future stock returns before/during/after the pandemic. Our sample periods of Pre-Pandemic, Pandemic and Post-Pandemic are January 2018 to February 2020, March 2020 to April 2022 and May 2022 to December 2023, respectively. 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 (6). The dependent variable is the next day returns computed using the end-of-day bid-ask average price in Panel A and next k-th week returns in Panel B. The independent variable is the scaled marketable retail order imbalance measure, *ROib*, as defined in equation (5). Control variables include previous day return, *Ret*(t), previous week return, *Ret*(t-1,t-5), previous month return, *Ret*(t-6,t-25), log market capitalization, *Lsize*, book to market ratio, *Lbm*, monthly stock turnover, *Lturnover*, and monthly return volatility *Lmvol*. The standard deviations are adjusted using Newey-West (1987) with six lags. To understand the economic magnitude and compare between different periods, we multiply the interquartile range for retail order imbalances with the coefficient of *ROib*, and generate the interquartile return differences.

Panel A. Retail order imbalances predict next day returns

Dep.var	Ret(t+1)		Ret(t+1)		Ret(t+1)	
Period	Pre-pandemic		Pandemic		Post-Pandemic	
	Coef.	t-Stat	Coef.	t-Stat	Coef.	t-Stat
ROib(t)	0.0009	13.22	0.0017	10.92	0.0018	15.29
Ret(t)	-0.0147	-3.37	-0.0141	-2.68	0.0020	0.52
Ret(t-1,t-5)	-0.0048	-0.67	-0.0188	-1.88	-0.0276	-3.64
Ret(t-6,t-25)	-0.0052	-0.45	0.0076	0.53	-0.0013	-0.10
Lsize	0.0000	0.32	-0.0002	-2.60	0.0000	0.10
Lbm	0.0000	-0.40	0.0002	2.64	0.0002	2.38
Lturnover	-0.0003	-2.33	0.0000	-0.13	-0.0002	-3.74
Lmvol	0.0061	1.10	-0.0106	-1.15	-0.0032	-0.37
Intercept	0.0000	-0.10	0.0010	1.56	0.0003	0.57
Interquartile next-day return diff	0.0381%***		0.0591%***		0.0669%***	
Adj.R2	0.033		0.059		0.038	

Panel B. Retail order imbalances predict next 12 weeks returns, interquartile return differences for different horizons

Dep.var	Future Returns	Future Returns	Future Returns
Period	Pre-Pandemic	Pandemic	Post-Pandemic
Week 1	0.0893%***	0.0897%***	0.1110%***
Week 2	0.0141%*	0.0317%***	0.0244%**
Week 3	0.0189%**	0.0203%*	0.0218%**
Week 4	0.0182%**	0.0266%**	0.0254%**
Week 5	0.0057%	0.0287%***	0.0264%**
Week 6	0.0044%	0.0230%***	0.0098%
Week 7	0.0030%	-0.0049%	0.0028%
Week 8	0.0026%	0.0079%	0.0034%
Week 9	0.0196%**	0.0000%	0.0076%
Week 10	0.0139%	0.0119%	0.0260%**
Week 11	0.0147%	0.0086%	0.0107%
Week 12	0.0261%**	0.0213%*	0.0195%*

Table IV. Retail Order Imbalances Predict News Sentiment Before, During, and After the Pandemic

This table reports estimation results on marketable retail investor order imbalance predicting future Ravenpack news sentiment before/during/after the pandemic. Our sample periods of Pre-Pandemic, Pandemic and Post-Pandemic are January 2018 to February 2020, March 2020 to April 2022 and May 2022 to December 2023, respectively. 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 (8). The dependent variables are next day Ravenpack news sentiment released for firm *i* between market closes on day *t* and day *t*+1. The independent variables are retail order imbalances, *ROib*. Control variables include previous day return, *Ret*(*t*), previous week return, *Ret*(*t*-1,*t*-5), previous month return, *Ret*(*t*-6,*t*-25), log market capitalization, *Lsize*, book to market ratio, *Lbm*, monthly stock turnover, *Lturnover*, and monthly return volatility *Lmvol*. The standard deviations are adjusted using Newey-West (1987) with six lags. To understand the economic magnitude and compare between different periods, we multiply the interquartile range for retail order imbalances with the coefficient of *ROib*, and generate the interquartile news sentiment differences.

Dep.var	NewsSent(*t*+1)		NewsSent(*t*+1)		NewsSent(*t*+1)	
Period	Pre-Pandemic		Pandemic		Post-Pandemic	
	Coef.	*t*-Stat	Coef.	*t*-Stat	Coef.	*t*-Stat
ROib(*t*)	0.0010	5.00	0.0016	7.76	0.0014	5.18
NewsSent(*t*)	0.0866	15.44	0.0897	33.30	0.1639	47.06
Ret(*t*)	0.0206	6.13	0.0286	8.24	0.0155	3.03
Ret(*t*-1,*t*-5)	0.0498	5.94	0.0810	8.73	0.0374	3.88
Ret(*t*-6,*t*-25)	0.0512	4.14	0.0925	5.90	0.0711	4.92
Lsize	0.0042	35.81	0.0055	35.16	0.0073	68.16
Lbm	-0.0003	-6.11	-0.0005	-8.95	-0.0010	-13.68
Lturnover	0.0007	3.63	0.0002	2.86	0.0004	4.47
Lmvol	0.0193	2.93	0.0115	1.50	0.0083	1.04
Interquartile next-day sentiment diff	0.0421%***		0.0557%***		0.0522%***	
Adj.R2	0.019		0.021		0.047	

Table V. Retail Trading Activity Predict Future Liquidity and Volatility Before, During and After the Pandemic

This table reports estimation results on retail trading activity predicting future liquidity and volatility before/during/after the pandemic. Our sample periods of Pre-Pandemic, Pandemic and Post-Pandemic are January 2018 to February 2020, March 2020 to April 2022 and May 2022 to December 2023, respectively. 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 (9). The dependent variables are next day liquidity and volatility in Panel A, next 12 weeks liquidity, volatility in Panel B. Liquidity proxy is the effective spread, *EffSpr(%)*, and volatility proxy is the intraday volatility *IntVol*. The independent variable is daily retail trading activity, *RVolume. Pandemic* equals one for the days between March 2020 and April 2022 and zero otherwise. *PostPandemic* equals one for May 2022 and after. Controls include lagged dependent variable, and other controls are same as before. We include day fixed effect. The standard errors are double clustered at day and stock level. To understand the economic magnitude and compare between different periods, we multiply the interquartile range for retail trading activities with the coefficient of *RVolume*, and generate the interquartile liquidity and volatility differences.

Panel A. Retail trading activity predict next-day liquidity and volatility before, during and after the Pandemic

Dep.var	EffSpr(*t*+1)		IntVol(*t*+1)	
	Coef.	*t*-Stat	Coef.	*t*-Stat
RVolume(*t*)	0.1889	19.74	0.1238	15.07
RVolume(*t*)*Pandemic	-0.1531	-11.02	-0.1019	-8.60
RVolume(*t*)*PostPandemic	-0.0252	-1.94	0.0194	1.70
Ret(*t*)	-0.0301	-5.20	-0.0237	-4.37
Ret(*t*-1,*t*-5)	-0.1310	-11.44	-0.0985	-10.18
Ret(*t*-6,*t*-25)	-0.0110	-1.56	-0.0139	-2.20
Dep.var(*t*)	0.4804	177.41	0.4708	204.34
Dep.var(*t-1*)	0.3704	154.89	0.3682	167.75
Lsize	-0.0245	-34.22	-0.0239	-35.75
Lbm	0.0020	3.47	0.0014	2.85
Lturnover	-0.0017	-2.15	-0.0013	-1.84
Lmvol	-0.1107	-3.95	-0.0869	-3.93
Intercept	0.0489	27.97	0.0529	33.17
Day FE	Yes		Yes	
Interquartile next-day Dep.var diff	0.0142***		0.0093***	
Change of interquartile next-day Dep.var diff in Pandemic	-0.0115***		-0.0077***	
Change of interquartile next-day Dep.var diff in PostPandemic	-0.0019*		0.0015*	
Adj.R2	0.792		0.783	

Panel B. Retail trading activity predict future 12 weeks liquidity and volatility before, during and after the Pandemic, interquartile differences for different horizons

Dep.var	Future EffSpr			Future IntVol		
	RVolume(*t*)	Retail(t)*Pandemic	Retail(t)*PostPandemic	Retail(t)	Retail(t)*Pandemic	Retail(t)*PostCovid
Week1	0.0117***	-0.0068***	-0.0017	0.0093***	-0.0040***	0.0005
Week2	0.0095***	-0.0074***	-0.0025**	0.0077***	-0.0056***	-0.0008
Week3	0.0107***	-0.0101***	-0.0038***	0.0084***	-0.0073***	-0.0019**
Week4	0.0114***	-0.0107***	-0.0048***	0.0091***	-0.0083***	-0.0023**
Week5	0.0109***	-0.0102***	-0.0029**	0.0091***	-0.0081***	-0.0009
Week6	0.0111***	-0.0100***	-0.0028**	0.0086***	-0.0073***	-0.0007
Week7	0.0107***	-0.0084***	-0.0032***	0.0080***	-0.0062***	-0.0007
Week8	0.0106***	-0.0083***	-0.0032***	0.0083***	-0.0060***	-0.0010
Week9	0.0103***	-0.0080***	-0.0024**	0.0084***	-0.0060***	-0.0011
Week10	0.0100***	-0.0076***	-0.0030**	0.0083***	-0.0059***	-0.0009
Week11	0.0106***	-0.0084***	-0.0027**	0.0082***	-0.0060***	-0.0006
Week12	0.0104***	-0.0078***	-0.0030**	0.0086***	-0.0062***	

Table VI. Causal Analysis on Retail Trading Activity and Market Quality on Relief Checks Days

This table reports causal analysis on retail trading activities predicting market quality, using three rounds of stimulus checks as exogenous shocks. Our sample period is January 2018 to December 2023, 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 panel regressions, as specified in equation (10). The dependent variables are next day liquidity, next day volatility, and next month price efficiency. The independent variable is daily retail trading activity, *RVolume.* Following Greenwood et al. (2023), *DRelief* is a dummy variable capturing the following 9 days: April 13, 2020, and the subsequent 2 trading days; December 30, 2020, and the subsequent 2 trading days; March 12, 2021, and the subsequent 2 trading days. Controls include lagged dependent variable, and others are same as before. We include day fixed effect. The standard errors are double clustered at day and stock level.

Dep.var	EffSpr(*t*+1)		IntVol(*t*+1)	
	Coef.	*t*-Stat	Coef.	*t*-Stat
RVolume(*t*)	0.1419	20.73	0.1008	16.18
RVolume(*t*)*DRelief	-0.1665	-2.63	-0.1090	-1.23
Day FE	Yes		Yes	
Double Cluster	Yes		Yes	
Change of interquartile next-day diff on relief checks days	-0.0125***		-0.0082	
Adj.R2	0.792		0.783	

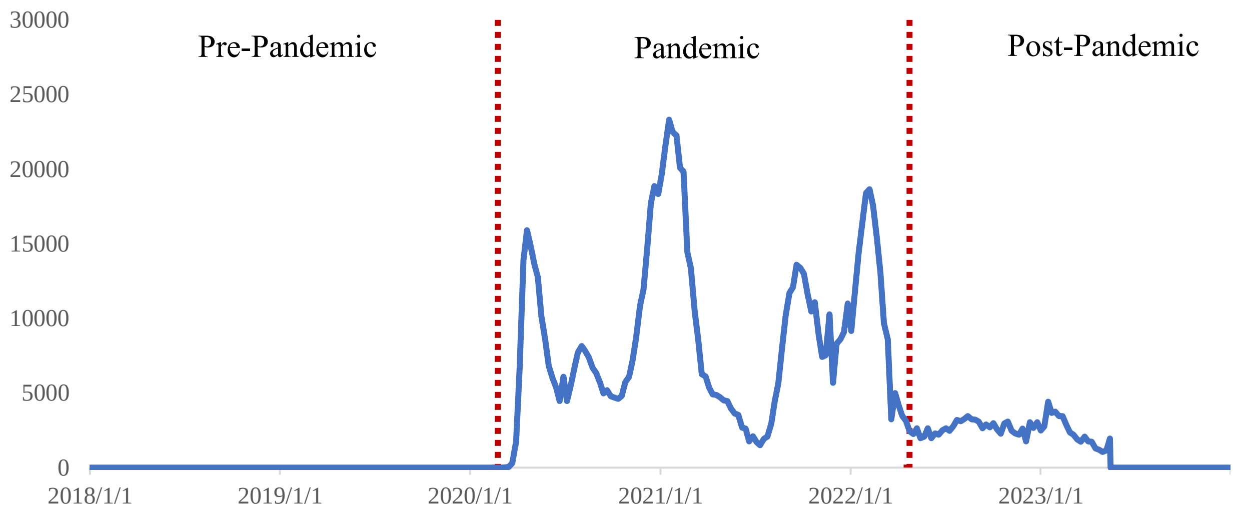
Table VII. Retail Trading Activity Predict Future High Frequency Trading and Short Selling Before, During, After the Pandemic
This table reports estimation results on retail trading activities predicting future high frequency trading and short selling from one day to 12 weeks before/during/after the pandemic. Our sample periods of Pre-Pandemic, Pandemic and Post-Pandemic are January 2018 to February 2020, March 2020 to April 2022 and May 2022 to December 2023, respectively. 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 (9). The dependent variables are the next day and next *k*-th week high frequency trading and short selling scaled by number of days involved. High frequency trading proxy is the cancel to trade ratio *HFTCancel*, while short selling proxy is the days to cover ratio, *SDTCR*. The independent variable is total retail activity measure, *RVolume*. *Pandemic* equals one for the days between March 2020 and April 2022 and zero otherwise. *PostPandemic* equals one for May 2022 and after. Controls include lagged dependent variable, and others are same as before. We also include day fixed effect. The standard errors are double clustered at day and stock level. To understand the economic magnitude and compare between different periods, we multiply the interquartile range for retail trading activities with the coefficient of *RVolume*, and generate the interquartile dependent variable differences. To save space, we only report the interquartile differences.

Dep.var	Future HFTCancel			Future SDTCR		
	RVolume(*t*)	Rvolume(*t*) *Pandemic	Rvolume(*t*) *PostPandemic	RVolume(*t*)	Rvolume(*t*) *Pandemic	Rvolume(*t*) *PostPandemic
1 day	-0.0095***	-0.0052***	0.0051***	-0.0112	-0.0935***	-0.0256*
Week1	-0.0089***	0.0000	0.0076***	-0.0267***	-0.0588***	-0.0243**
Week2	-0.0113***	0.0009	0.0071***	-0.0245***	-0.0410***	-0.0103
Week3	-0.0097***	-0.0014	0.0056***	-0.0245***	-0.0460***	-0.0143
Week4	-0.0092***	-0.0014	0.0052***	-0.0255***	-0.0443***	-0.0184*
Week5	-0.0096***	-0.0013	0.0060***	-0.0272***	-0.0414***	-0.0139
Week6	-0.0095***	-0.0005	0.0062***	-0.0248***	-0.0387***	-0.0084
Week7	-0.0091***	-0.0002	0.0066***	-0.0251***	-0.0382***	-0.0082
Week8	-0.0091***	0.0003	0.0076***	-0.0189***	-0.0400***	-0.0107
Week9	-0.0090***	-0.0004	0.0076***	-0.0241***	-0.0346***	-0.0103
Week10	-0.0087***	-0.0015	0.0073***	-0.0236***	-0.0359***	-0.0089
Week11	-0.0091***	-0.0007	0.0066***	-0.0272***	-0.0369***	-0.0120
Week12	-0.0092***	-0.0006	0.0073***	-0.0262***	-0.0364***	-0.0140

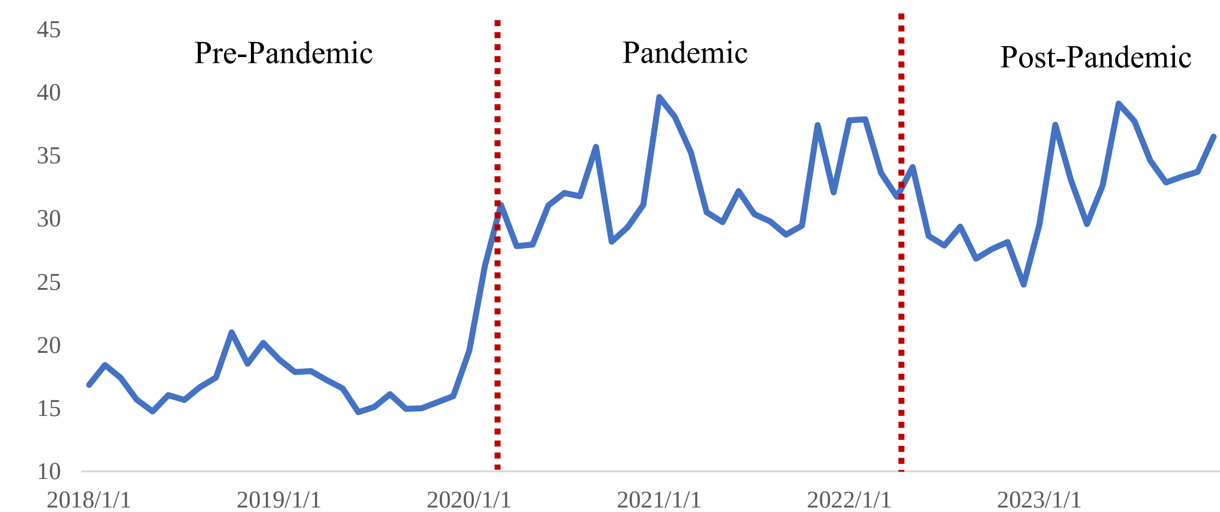
Figure I. Marketable Retail Investors Trading Flows Around Pandemic

This figure plots marketable retail investors trading flows around Pandemic. Our sample periods of Pre-Pandemic, Pandemic and Post-Pandemic are January 2018 to February 2020, March 2020 to April 2022 and May 2022 to December 2023, respectively. Our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1. Panel A plots the U.S. weekly COVID-19 deaths reported to WHO from January 2018 to December 2023. Panel B presents the daily aggregate retail trading volume. Panel C presents the daily percentage of retail trading volume, *RVolume*, as specified in equation (1). To facilitate clear presentation, we present the monthly average of daily retail trading activity.

Panel A. number of deaths due to pandemic



Panel B. Retail dollar volume ($billion)



Panel C. Retail dollar volume as percentage of total trading volume, *RVolume*

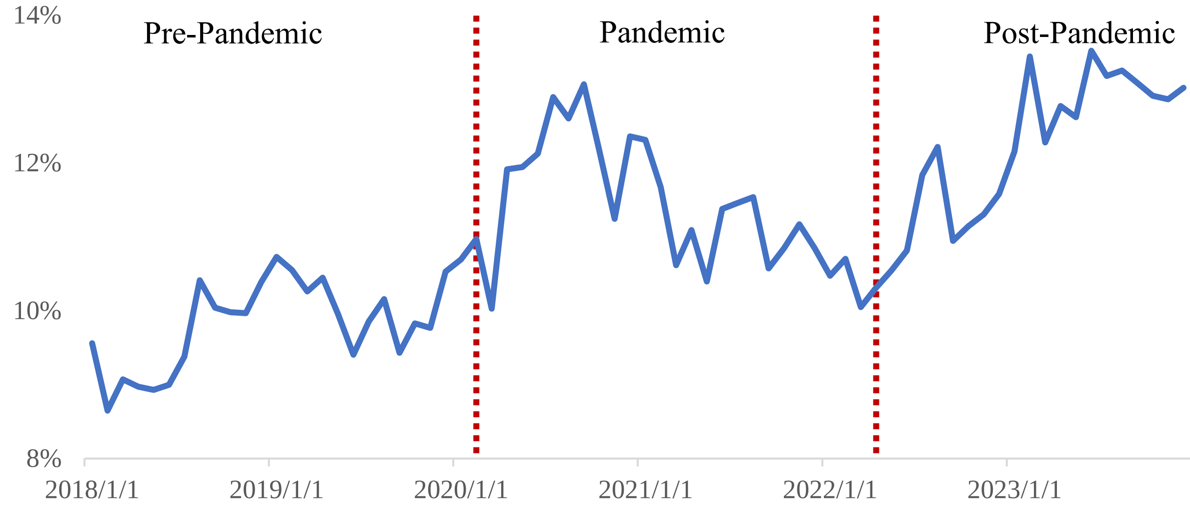
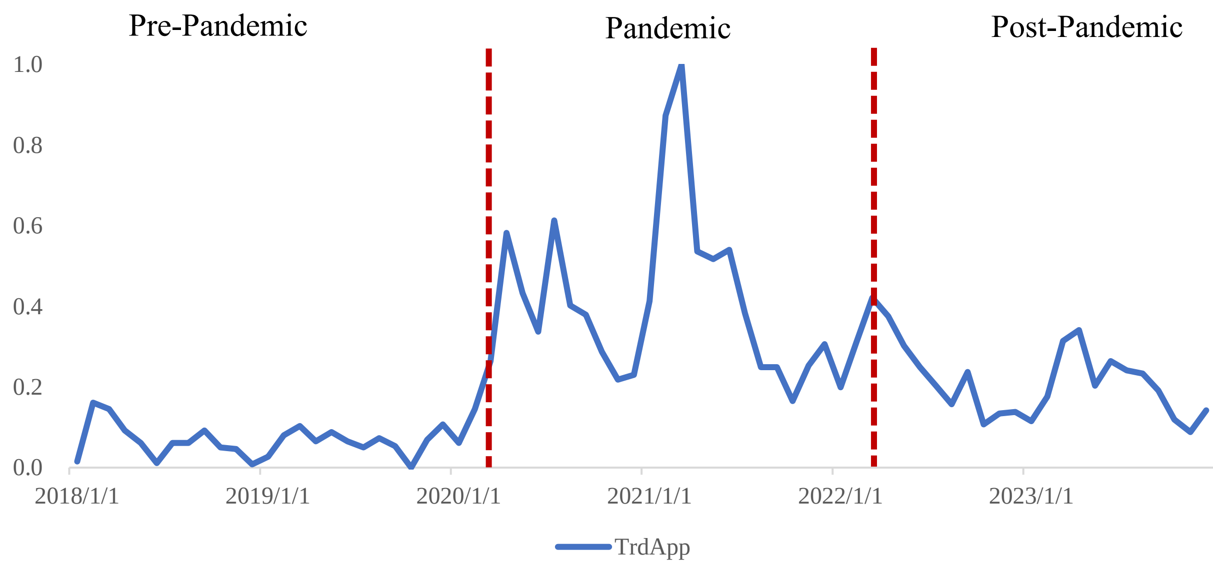


Figure II. Google Search of Trading Applications

This figure plots the time-series patterns of the average Google Trends of four retail brokerages (Robinhood, TD Ameritrade, Charles Schwab, E-Trade) from January 2018 to December 2023. Google Trends provides monthly search frequency with a relative (normalized) structure, where the values range from 0 to 100. We further normalize the measure to have a minimum value of 0 and a maximum value of 1.



Appendix A. Previous Studies on Types of Retail Investors

Hypotheses	Previous studies
Group I: Pure noise retail investors (1) Retail investors can't predict return (2) Retail investors provide liquidity	**Glosten and Milgrom (1985 JFE)** argue that informed traders lead to a positive bid-ask spread, and uninformed investors provide liquidity. **Kyle (1985 Econometrica**) model the insider, noise traders, and market makers, and argue that informed insider make profits, while noise traders provide liquidity. **Black (1986 JF)** argue that noise traders have no information of firm value and provide liquidity to make trading in financial markets possible. **Barber and Odean (2000 JF)** use individual investors trading from discount brokerage over 1991 to 1996 and find their trading performance are poor. **Peress and Schmidt (2020 JF)** exploit episodes of sensational news that distract noise traders and find trading activity, stock liquidity and volatility decrease on distraction days.
Group II: Informed retail investors (1) Retail investors could predict short term and long-term return, and firm fundamental information (2) Retail investors demand liquidity and increase volatility	**Kelley and Tetlock (2013 JF)** use retail orders from off-exchange brokers from 2003 to 2007 and find both market and limit retail order imbalance positively predict future stock returns. Further evidence suggests retail market orders are informed, while retail limit orders provide liquidity. **Boehmer, Jones, Zhang, and Zhang (2021 JF)** provide an algorithm to identify marketable retail purchases and sales from TAQ over 2010 to 2015, and find retail order imbalance positive predict future stock return and may contain firm-level information.
Group III: Uninformed retail investors (1) Retail activities predict short term return, but reverse in long run, and can't predict firm fundamental information (2) Mixed predictions for liquidity and volatility	**Grossman and Miller (1988 JF)** model the market liquidity as the demand and supply of immediacy, and market makers need to bear the inventory risks when facing momentum return. **Kaniel, Saar, and Titman (2008 JF)** use all retail orders from NYSE CAUD file between 2000 and 2003 and find positive excess returns following their intense net buying. **Da, Engelberg and Gao (2011, JF)** find an increase in retail attention proxied by Google Search Volume Index (SVI) predicts higher stock prices in the next 2 weeks and an eventual price reversal within the year.

	Barrot, Kaniel, and Sraer (2016 JFE) use the retail orders from a leading European online broker between 2002 and 2010 and find individual investors provide liquidity to the stock market, but they are not compensated. **Ozik, Sadka, and Shen (2021 JFQA)** use Robinhood investors holding data from May 2018 to August 2020 and find retail trading attenuated the rise in illiquidity by roughly 40% during the Covid pandemic lockdown. **Welch (2022 JF)** use Robinhood investors holding data from May 2018 to August 2020, and find the aggregated Robinhood portfolio had both good timing and good alpha. **Barber, Odean, Huang, and Schwarz (2022 JF)** use Robinhood investors holding data from May 2018 to August 2020 and find Robinhood investors engage in more attention-induced trading than other retail investors. Intense buying by Robinhood users forecasts negative returns. **Eaton, Green, Roseman, and Wu (2022 JFE)** find that Robinhood outages are associated with increased market liquidity, and lower return volatility, whereas the opposite relations hold following outages at traditional retail brokerages. **Hendershott, Menkveld, Praz, and Seasholes (2022 RFS)** model inattentive investors arriving stochastically to trade, and add to return prediction with new insights based on the inattention friction. **Pedersen (2022 JFE)** model four types of investors trade assets over time: naive investors who learn via a social network, "fanatics" possibly spreading fake news, and rational short- and long-term investors. The naïve investors are influenced by fanatic and rational views. **Hu, Jones, Li, Zhang and Zhang (2024)** show activities on Reddit can significantly predict future stock price movements and volatility.

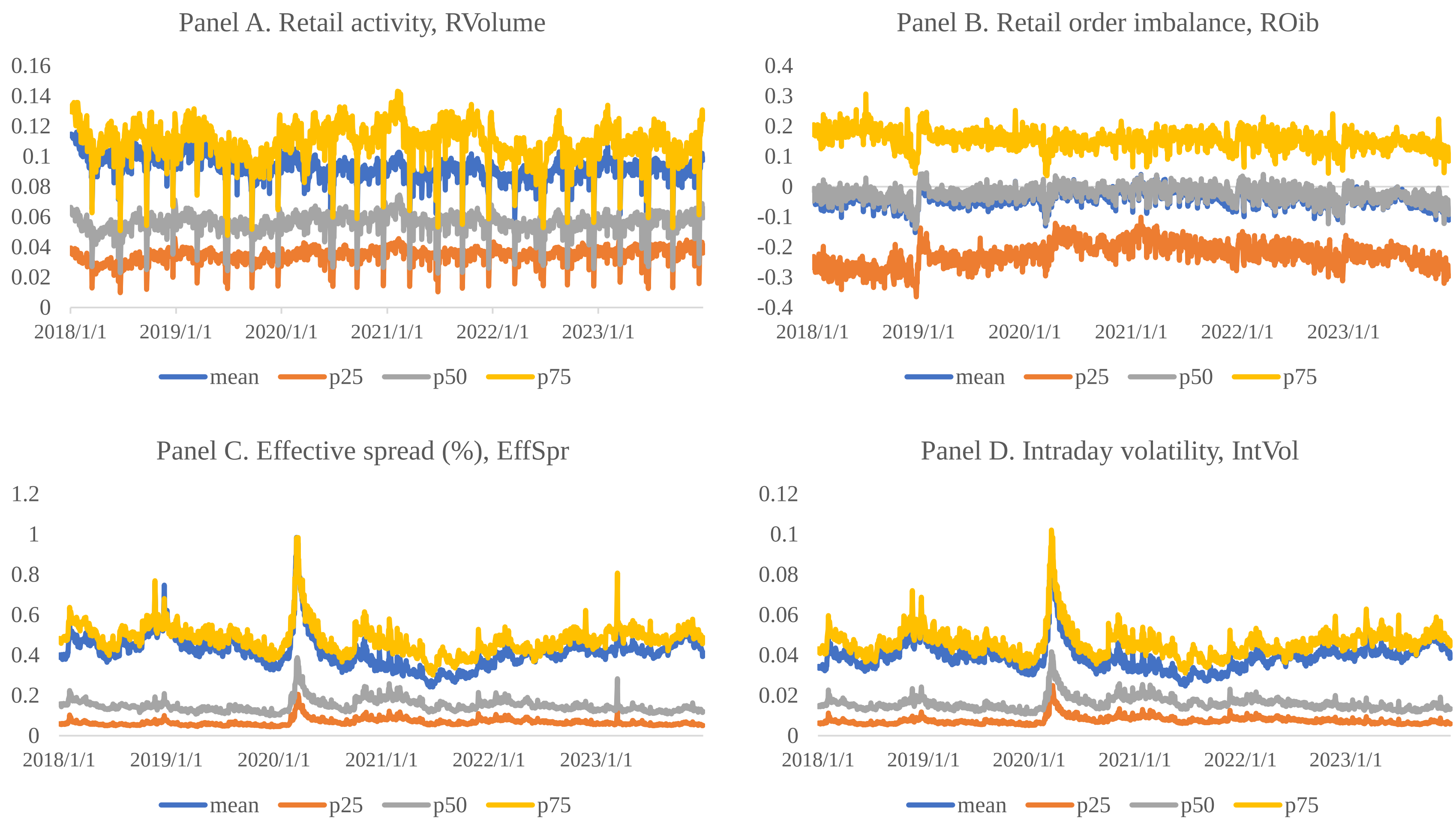
Appendix B. FOMC Announcements Around Pandemic

This table reports the timing schedule of FOMC announcements, from January 2018 to December 2023. Our sample periods of Pre-Pandemic, Pandemic and Post-Pandemic are January 2018 to February 2020, March 2020 to April 2022 and May 2022 to December 2023, respectively. The FOMC holds eight regularly scheduled meetings during the year. Adjustment with a '*' indicates unscheduled FOMC date; Adjustment with a '+' ('-') are FOMC dates when the Committee decides to raise (lower) the target range for the federal funds rate; Adjustment without '+' and '-' are FOMC dates when the Committee decides to maintain the target range for the federal funds rate.

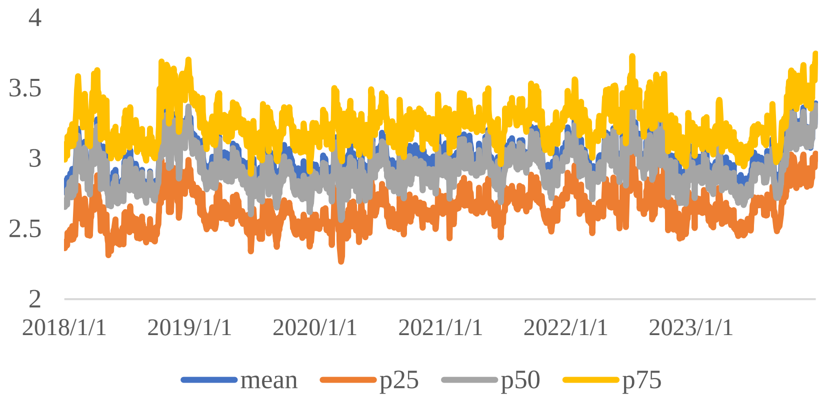
Pre-Pandemic FOMC Date	Adjustment	Pandemic FOMC Date	Adjustment	Post-Pandemic FOMC Date	Adjustment
01/31/18		03/03/20	- *	05/04/22	+
03/21/18	+	03/15/20	- *	06/15/22	+
05/02/18		04/29/20		07/27/22	+
06/13/18	+	06/10/20		09/21/22	+
08/01/18		07/19/20		11/02/22	+
09/26/18	+	09/16/20		12/14/22	+
11/08/18		11/05/20		02/01/23	+
12/19/18	+	12/16/20		03/22/23	+
01/30/19		01/27/21		05/03/23	+
03/20/19		03/17/21		06/14/23	
05/01/19		04/28/21		07/26/23	+
06/19/19		06/16/21		09/20/22	
07/31/19	-	07/28/21		11/01/23	
09/19/19	-	09/22/21		12/13/23	
10/30/19	-	11/03/21			
12/11/19		12/15/21			
01/29/20		01/26/22			
		03/16/22	+		

Appendix Figure I. Cross-sectional Distribution of Key Variables Over Time

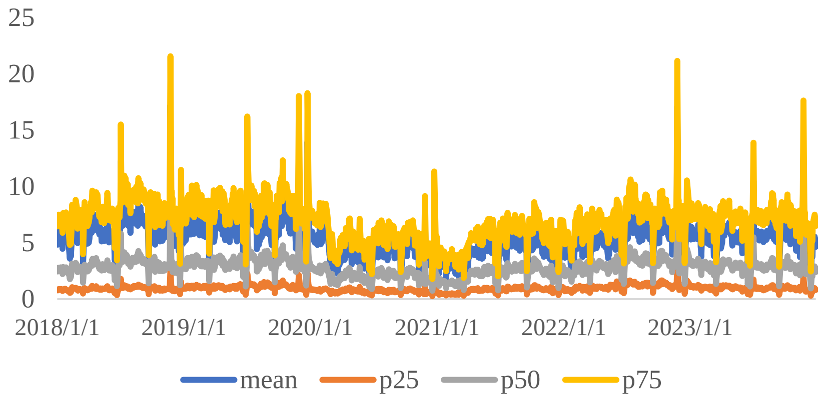
These figures plot the time-series statistics of retail activity, retail order imbalance, effective spread, intraday volatility, cancel-to-trade ratio of high frequency trading, days-to-cover-ratio of short selling, from January 2018 to December 2023. Our sample firms are all common stocks listed on U.S. stock exchanges with a share price of at least $1.



Panel E. High frequency trading, Cancel to Trade



Panel F. Short selling, SDTCR



Appendix Table I. Triggers for the Rise of Retail Trading Activity in the Chinese Stock Market

This table reports results on how retail trading activity are related to social media attention and usage of trading apps. We obtain the proprietary trading data from one major stock exchange in China, which includes retail trading data from January 2018 to May 2020. Since the outbreak of Covid-19 in China starts from January 2020, we define January 2018–December 2019 as pre-pandemic period, and January 2020–May 2020 as pandemic period. Panel A estimates the time-series regressions, as specified in equation (2), and Panel estimate the Fama-MacBeth regression, as specified in equation (4). The dependent variables are the daily market level retail activity measure in Panel A, and firm level retail activity measure in Panel B. The key independent variables are trading apps, *TrdApp*, proxied by Baidu search of retail trading apps (TongHuaShun, Eastmoney Securities, CITIC Securities, and Huatai Securities), and social media, *SocMed*, proxied by the Guba posts.

Panel A. Predict aggregate retail trading activity in Chinese stock market

Dep.Var	RVolume(t+1)		RVolume(t+1)		RVolume(t+1)	
	Coef.	t-Stat	Coef.	t-Stat	Coef.	t-Stat
SocMed	0.0395	2.57			-0.0261	-1.64
TrdApp			0.1187	5.90	0.1336	6.27
Adj.R2	0.031		0.253		0.262	

Panel B. Predict firm-level retail trading activity in Chinese stock market, before and during the Pandemic

Dep. Var	RVolume(*t*+1)		RVolume(*t*+1)	
	Pre-Pandemic		Pandemic	
	Coef.	*t*-Stat	Coef.	*t*-Stat
RSocMed(*t*)	0.0284	39.90	0.0261	15.02
RVolume(*t*)	0.8417	268.25	0.8774	179.90
Ret(*t*)	0.0756	5.26	0.0033	0.14
Ret(*t*-1,*t*-5)	-0.0112	-2.52	-0.0244	-2.94
Ret(*t*-6,*t*-25)	-0.0070	-2.81	-0.0060	-1.81
Lnsize	-0.0131	-39.30	-0.0106	-23.45
Lbm	-0.0002	-3.00	0.0003	2.17
Lturnover	0.1024	11.73	0.0060	0.52
Lmvol	-0.1445	-5.45	-0.1773	-7.93
Interquartile next-day Rvolume diff	1.42%		1.31%	
Adj.R2	0.828		0.867	