Retail Trading and Stock Return Predictability

Fabio Moneta
Telfer School of Management
University of Ottawa
55 Laurier E. (DMS 7153)
Ottawa ON K1N 6N5
(613) 562-5800
fmoneta@uottawa.ca

Xiaoyu Sun
Telfer School of Management
University of Ottawa
55 Laurier E. (DMS 6102)
Ottawa ON K1N 6N5
(613) 227-8106
xsun052@uottawa.ca

March 2025

ABSTRACT

Based on the influential work of Boehmer, Jones, Zhang, and Zhang (2021), we estimate the predictability of retail investors for stock returns using a novel retail trading database (VandaTrack) and an extended sample period, from 2014 to 2022. We find that the predictability of retail trading persists for up to five weeks from 2014 to 2016, a period that overlaps with Boehmer et al., but is absent out of sample from 2017 to 2022. The predictability observed in 2014-2016 is concentrated in small and value stocks when buying strongly dominates selling. The predictability for small stocks is seriously weakened because of the improvement in liquidity from 2017 to 2022. When small stocks with low liquidity are excluded and the effect of book-to-market factor on returns is controlled, retail investors lose predictability in 2014-2016. This paper provides evidence that liquidity is an important factor explaining the decline in predictability of retail trading rather than informed trading. A more neutral average sentiment in social media discussions also contributes to the reduced predictability.

Keywords: Retail trading, Predictability, Common Factors, Liquidity

JEL Classification: G12 G17

Retail Trading and Stock Return Predictability

ABSTRACT

Based on the influential work of Boehmer, Jones, Zhang, and Zhang (2021), we estimate the predictability of retail investors for stock returns using a novel retail trading database (VandaTrack) and an extended sample period, from 2014 to 2022. We find that the predictability of retail trading persists for up to five weeks from 2014 to 2016, a period that overlaps with Boehmer et al., but is absent out of sample from 2017 to 2022. The predictability observed in 2014-2016 is concentrated in small and value stocks when buying strongly dominates selling. The predictability for small stocks is seriously weakened because of the improvement in liquidity from 2017 to 2022. When small stocks with low liquidity are excluded and the effect of book-to-market factor on returns is controlled, retail investors lose predictability in 2014-2016. This paper provides evidence that liquidity is an important factor explaining the decline in predictability of retail trading rather than informed trading. A more neutral average sentiment in social media discussions also contributes to the reduced predictability.

Keywords: Retail trading, Predictability, Common Factors, Liquidity

JEL Classification: G12 G17

1. Introduction.

Whether retail trading predicts stock returns is crucial for other market participants to refine their trading strategies and risk management, for behavioral finance researchers to analyze retail investor psychology and activity, for the understanding of market efficiency, and for policymakers to monitor and regulate the stock market. However, earlier literature reaches inconsistent conclusions. Barber and Odean (2000; 2008) find that retail investors are uninformed and inexperienced in stock selection. In contrast, later studies (e.g., Kaniel et al., 2012; Kelly and Tetlock, 2013; Barrot, Kaniel, and Sraer, 2016) suggest that retail investors' trading can predict future stock returns. More recently, Boehmer, Jones, Zhang, and Zhang (BJZZ) (2021) identify trades that execute at share prices with fractional pennies as marketable retail price-improved orders from the Trade and Quote dataset, and find that retail investors can predict future stock returns over the next eight weeks in the 2010-2015 period.

Furthermore, retail trading activities have experienced elevated growth during the last decade, especially during the COVID-19 pandemic when staying at home, federal assistance checks, easier coordination via social media, and commission-free investing platforms boosted retail investors participation in the stock market. The daily net flows of US retail investors rose from \$620 millions in 2014 to \$1.32 billions in 2022, and they poured a record of \$1.51 billions after the earning seasons of Q1 2023.[1] Given the increasing scale of retail trading and the ongoing debate about its informativeness, it is important to reassess the predictability of retail investors over an extended sample period (2014-2022) using a novel database. This paper highlights a

[1] The data is summarized by VandaTrack Research. https://www.visualcapitalist.com/charted-u-s-retail-investor-inflows-2014-2023

declining retail investors' predictive power and provides possible explanations for their predictability in the first part of the sample and of its decline in the most recent sample.

We first examine the predictability of retail equity investors. We use daily flows of retail investors in the US stock market collected by VandaTrack, instead of the BJZZ's methodology.[2] Recent evidence (Barber et al. 2023) raises concerns about the accuracy of BJZZ's algorithm. VandaTrack claims that they improve on the BJZZ methodology by including proprietary adjustments that provide a more accurate reading of retail volume. We use the retail trading activity measure of BJZZ, the retail flow imbalance (*mroibflow*)[3], and estimate its predictive power by regressing future stock returns on retail flow imbalance using Fama-MacBeth (1973) regressions. We find that from 2014 to 2016, retail flow imbalance significantly positively predicts weekly stock returns, with the predictive power lasting up to the next five weeks. These results are consistent with BJZZ, although somewhat weaker given that their predictive power persisted for eight weeks in the 2010-2015 period.[4] However, the predictability of retail trading vanishes in the more recent 2017-2022 period.

We then investigate the reasons for the decline in the predictive power of retail trading by conducting several analyses. First, we explore whether the predictability of retail trading and its evolution stem from the buy side, sell side, or both. We divide retail flow imbalance into quartiles and apply piecewise linear regressions (Sirri and Tufano, 1998) to examine asymmetric

[2] VandaTrack data is widely used by the financial press (e.g., the Wall Street Journal) and industry practitioners. They have been made available to academics only recently. VandaTrack uses a proprietary algorithm to provide daily data on retail buys and sells of US stocks. These data are available from 2014.

[3] $mroibflow = \frac{Buy\ flow - Sell\ flow}{Buy\ flow + Sell\ flow}$, where the *flow* is the total dollar amount bought or sold by retail investors for a given stock.

[4] The difference is not due to the predictability of retail trading in 2016. We check the predictability during the overlapping period of VandaTrack and BJZZ, 2014-2015. The predictability using VandaTrack in 2014-2015 is consistent with that observed in 2014-2016. Due to the data limitation, we cannot determine whether the predictability weakened in the 2014-2016 period compared to the 2010-2013 period. Furthermore, differences in databases may be another reason.

predictability with buy and sell dominance. We show that the persistent predictive power of retail investors during the 2014-2016 period primarily comes from the top 25% of *mroibflow,* suggesting that stocks most heavily bought and least sold by retail investors are expected to have higher future returns. However, the top 25% of *mroibflow* no longer show significant predictive ability during the 2017-2022 period.

Second, to investigate whether the observed predictability of retail investors is due to their exposure to widely used factors[5], we control for the effects of three factors that are well-known by market participants, namely, size, book-to-market, and momentum on future returns, by replacing raw returns with DGTW characteristics-adjusted returns (Danel, Grinblatt, Titman, and Wermers, 1997) and estimate the predictability of retail flow imbalance. We find that, except for the next week look-ahead period, there is no statistically significant predictive power from the top 25% of retail flow imbalance. These results suggest that the observed predictability of retail trading during the 2014-2016 period is primarily attributed to its exposure to size, book-to-market, and momentum factors. There is no significant difference in predictability for DGTW returns in the 2017-2022 period relative to the 2014-2106 period. Next, we examine each of the three factors to determine whether factor tilts in retail trading can explain the decline in predictability. Sorting stocks by size using DGTW characteristic-adjusted portfolios, we find that the positive predictive power of the top 25% *mroibflow* in the 2014-2016 period stems from small stocks. However, for the 2017-2022 period, the top 25% *mroibflow* does not exhibit prediction for small stocks.

[5] Retail investors exhibit varying trading preferences for stocks with different characteristics (size, book-to-market, and momentum), suggesting that their predictive power for stock returns may differ based on these characteristics. The details are shown in the next section.

Accordingly, the observed decline in the predictability of retail flow imbalance can be attributed to the reduced predictive power for small stocks.[6]

Third, we investigate possible explanations for the decline in predictability for small stocks. Small stocks experienced an increase in retail flows, with institutional flows growing even more significantly during the 2017-2022 period. Increased trading activities improved the liquidity of small stocks. The illiquidity level for small stocks (using the measure of Amhuid, 2002) decreases from 304.76 in the 2014-2016 period to 117.28 in the 2017-2022 period.[7] This increased liquidity would reduce the price impact of retail trading and its predive power. Furthermore, we find that the top 25% of *mroibflow* exhibits predictive power for small stocks with low liquidity, but not for small stocks with high liquidity in the 2014-2016 period, suggesting that liquidity is one important factor influencing the predictability of retail trading for small stocks.

We further examine the predictability related to small stocks, which are the driver of the observed reduction in retail trading predictability. In particular, we create a dummy variable that identifies small stocks with low liquidity and interact it with *mroibflow* to isolate the impacts of size and liquidity on predictability. We find that retail flow imbalance does not exhibit significant predictive power in the 2014-2016 period when small stocks with low liquidity are excluded from the sample and the effect of book-to-market factor on returns is controlled for. Portfolios sorted on small stocks with low liquidity and *mroibflow* also show consistent results. A zero-investment strategy that goes long the portfolio of small and illiquid stocks with the top 10% of *mroibflow* and short the portfolio of the stocks with the bottom 10% of *mroibflow* can generate positive and

---

[6] We find retail investors' persistent predictability for value stocks over the next five weeks, but it does not have significant change from 2014 to 2022. Therefore, exposure to value stocks could not explain the decline in predictability. We also examine whether exposure to momentum, investment, or profitability factors explain the retail trading predictability and its evolution. We find no evidence that exposure to momentum, investment, or profitability factors contribute to the predictability or its evolution.

[7] Amhuid (2002) calculates the illiquidity as the ratio of the daily absolute return to the (dollar) trading volume. The statistic is the weekly average and multiplied by 10^6.

significant returns in the next five weeks in the 2014-2016 period. Overall, our results suggest that except for trading in small and illiquid stocks, retail investors cannot significantly predict the future alpha during the 2014-2016 period, as they cannot in the 2017-2022 period. Therefore, these results confirm that the decline in predictability is driven by the increase in stock liquidity.[8]

Finally, we analyze the influence of social media on the predictability of retail trading and its evolution. One important trend in the last decade is the increasing relevance of social media and its impact on retail trading and stock prices as highlighted by the 2021 GameStop episode. Given that the use of social media by retail investors is more prevalent in the post 2016 period than in the 2014-2016 period, it may be considered a contributing factor for the decline in predictability of retail trading. Using tweet publication counts and sentiment analysis on Twitter collected by Bloomberg, we find that higher tweet volumes can stimulate retail trading activities, leading to a larger positive impact on the following week's returns among small stocks with low liquidity in the 2015-2016 period. However, the impact of social media discussion on enhancing the predictability of retail investors in the 2017-2022 period is limited. We also investigate the impact of sentiment on the predictability of retail trading. We further divide tweet sentiment into quartiles and find that during 2015-2016[9], retail trading predictability is most evident when social media sentiment shows clear and directional social media sentiment (i.e., 1st or 4th sentiment quartile). As sentiment becomes more neutral or ambiguous (i.e., 2nd or 3rd quartile), especially with a high tweet volume, retail trading becomes less predictive. This finding suggests that when positive and negative tweets appear in similar large quantities, they provide little help to retail investors in

[8] We also examine whether the decline in predictability is due to the decline in autocorrelation of retail trading. However, the autocorrelation of retail flow imbalance during a month does not change significantly from 2014 to 2022.
[9] Twitter related data provided by Bloomberg starts at 1st January 2015. The predictability of retail trading in the 2015-2016 period is consistent with that in the 2014-2016 period.

predicting future stock returns[10]. Compared to the 2015-2016 period, average sentiment became more neutral in the 2017-2022 period (declining from 0.0138 to 0.0005 in the 2nd quartile and from 0.0705 to 0.0386 in the 3rd quartile), and the retail trading on stocks in the 2nd and 3rd sentiment quartiles does not exhibit predictive power. Furthermore, for stocks receiving the most positive sentiment and high volumes of discussions, retail trading exhibits a negative impact on their future returns in the 2017-2022 period, suggesting that intensive retail buying leads to stock price reversals. The negative impact is consistent with increased liquidity which accelerates price reversals after intensive buying. Overall, the more neutral sentiment observed on social media during the 2017-2022 period may be another contributing factor to the decline in predictability.

Our paper contributes to the literature on retail investors' predictability for stock returns. Earlier results provide inconsistent results using various databases. Barber and Odean (2000; 2008) use data from retail brokerage firms and conclude that retail investors are uninformed. However, later studies (e.g., Kaniel et al., 2012; Kelly and Tetlock, 2013; Barrot et al. 2016) applying account-type data in NYSE, US wholesaler, and Robinhood retail databases reach the opposite conclusion. Barlett et al. (2022) employ fractional share trades in the Consolidated Transaction Reporting System and find that these trades predict future liquidity and volatility. One widely used methodology for identifying retail trades is from BJZZ (2021), who use sub-penny trade prices. However, Barber et al. (2023) doubt the accuracy of BJZZ's algorithm (2021), noting that only 35% of the tested trades are identified as retail. In this paper, we use daily retail flows collected by VandaTrack and find that the predictability documented by BJZZ dissipates in the most recent sample period. In addition, we provide evidence that with the improvement in liquidity due to the

[10] Retail trading may be misled by the mix of positive and negative information, or the inconsistent information may disperse retail buy and sell flows, offsetting the price impacts from both sides on future returns.

increase in total trading volume in recent years, the impact of retail investors on future returns of small stocks has become weaker.

Our paper also relates to the literature on retail trading exposure to factors. Kurmar and Lee (2006) show that retail investors tend to concentrate their holdings on small and value stocks. Betermier et al. (2017) provide possible explanations for household tilting toward value stocks. BJZZ also find a larger coefficient of predictive power for small stocks than for large stocks. We update their findings that in an extended sample period, retail trade maintains similar extents of tilts toward small and value stocks on average and exhibits significant predictability for these stocks during the 2014-2016 period. However, we find that retail investors cannot predict returns for large stocks, and the predictability for small stocks vanishes due to increased liquidity in the 2017-2022 period. Additionally, we show that retail investors have become more contrarian, consistent with the finding of Luo et al. (2023), but their trading fails to predict the returns of losing stocks on the buy side or winning stocks on the sell side. By analyzing the predictability of retail trading and its evolution, controlling for factor exposure over the sample period, this paper shows that the observed predictability in the 2014-2016 period can be largely explained by retail trading exposure on size and book-to-market factors. Our results suggest that, on average, retail investors do not possess superior private information or exhibit significant predictability from exposure to other factors or anomalies.

Moreover, our paper is related to the strand of literature on the role of liquidity in the equity market. Roll et al. (2007) suggest that liquidity enhances the efficiency of the futures-cash pricing system. Chordia et al. (2008) show that liquidity reduces intraday return predictability from order flows, improving market efficiency. Chordia et al. (2014) find that increased liquidity and trading activities have attenuated the economic and statistical significance of anomalies. We extend this literature by providing new evidence from retail order flows, demonstrating that increased liquidity in small stocks is associated with a decline in their predictability.

Finally, our paper is related to the growing literature on social media and retail trading. Whether social media provides useful information to assist retail trading is an ongoing debate, with studies analyzing data from platforms such as StockTwits, Seeking Alpha, Twitter, and WallStreetBets. Farrell et al., (2020) show that shared information on Seeking Alpha helps investors to be better informed using intraday data. Dim (2021) finds that Wallstreetbets (WSB)'s average beliefs predict future abnormal returns and earnings surprises. However, Hu et al., (2023) find that short-sellers' activities are deterred by social media. Lou et al. (2023) show that intense discussions can reduce the production of fundamental related information. Lopez et al. (2023) provide evidence that social media can provide misleading information, making it more difficult for retail investors to select useful and accurate information, thus hindering their ability to forecast future prices. Consistent with the studies of Lou et al. (2023) and Lopez et al. (2023), we do not find evidence that social media enhances the predictability of retail trading. Furthermore, our paper analyzes the evolution of social media discussion features and connects it with the evolution of retail trading predictability. We find that the overall predictability is weakened by more neutral sentiment observed during the 2017-2022 period.

2. Data

We use daily retail flow data (in US dollars) for US stocks provided by VandaTrack from January 3, 2014, to December 31, 2022. The data is widely used by the financial press (e.g., the Wall Street Journal) and industry practitioners. They have only recently become accessible to academics. VandaTrack employs a proprietary algorithm to provide daily data on retail buying and selling activity of US stocks. They claim that they improve upon the BJZZ methodology by incorporating proprietary adjustments that provide a more accurate reading of retail volume.[11] Data on stock characteristics, such as price, return, trading volume, and number of shares outstanding, are retrieved from CRSP daily and monthly stock databases. Fundamental variables are obtained from the COMPUSTAT quarterly and annual databases. Our sample focus on common stocks with share code 10 or 11 listed on the three main US Exchanges. We also exclude stocks with a minimum price below $1 at the previous month-end.

Similar to BJZZ, we measure retail investors' directional trades by computing retail flow imbalance for each stock *i* on each day *t*:[12]

$$mroibflow_{i,t} = \frac{Buy\ flow_{i,t} - Sell\ flow_{i,t}}{Buy\ flow_{i,t} + Sell\ flow_{i,t}}$$

Table 1 Panel A presents the time-series cross-sectional averages for the mean, median, standard deviation, minimum, and maximum calculated from daily data over the period 2014 to 2022. The sample includes an average of 2,441 stocks, with a maximum of 3,005 stocks.[13] Total traded flow is calculated by multiplying the daily trading volume by the closing price, yielding an average of approximately $76.314 million. The average daily retail buy flow is $1.151 million, accounting for 1.51% of total traded flows, while the average daily retail sell flow is slightly lower as $1.068 million, or 1.40% of total flows. The average retail flow imbalance (*mroibflow*) is approximately -0.005 from 2014 to 2022, indicating a slight dominance of selling over buying.

[11] Some academics raised concerns on the BJZZ methodology. For example, Barber et al. (2023) tests the accuracy of BJZZ's algorithm using by placing 85,000 retail trades in six retail brokerage accounts and only 35% of orders are from retail investors.

[12] We also test the measure of BJZZ, retail order imbalance ($mroibvol = \frac{Buy\ vol - Sell\ vol}{Buy\ vol + Sell\ vol}$), where the trading volume is calculated by dividing daily retail flow by daily closing price. The results are consistent.

[13] VandaTrack increases the number of stocks included in the database from 2014 to 2022.

Panel B illustrates changes in retail trading activities across sub-periods. During 2014-2016, the average *mroibflow* was -0.019, suggesting more pronounced net retail selling. In contrast, the average *mroibflow* turned positive (0.003) in the 2017-2019 period, implying net buying. This trend intensified in the 2020-2022 period, where *mroibflow* increased further to 0.013, indicating a substantial rise in net retail buying activity during the post-pandemic period.

To examine the changes of retail trading behavior, we analyse their activities across firms sorted by size, book-to-market ratio, and momentum in sub-periods. Following the methodology of Kumar and Lee (2006), we calculate normalized retail trading activity (NTA I), which captures the relative importance of retail trading compared to overall market activity. Given the increase in institutional flows which may have outpaced retail flows, we also introduce an alternative measure, NTA II. This version replaces the denominator with market capitalization to evaluate the concentration of retail trading. We report the average values for both NTA I and NTA II across different stock portfolios sorted by size, B/M, and momentum characteristics. Stock classification is based on DGTW characteristic-adjusted portfolios. The sample is divided into two phases: 2014-2016 and 2017-2022.

$$NTA\ I_{i,t} = \frac{Retail\ flow\ from\ retail\ investors\ in\ the\ sample}{Total\ flow\ in\ the\ market} \times 10^6$$

$$NTA\ II_{i,t} = \frac{Retail\ flow\ from\ retail\ investors\ in\ the\ sample}{Market\ capitalization} \times 10^6$$

Panel B shows that the average NTA II increases from 0.269 to 0.403, however, NTA I declines from 0.039 to 0.032. The difference implies that despite the rise of retail flows from 2014 to 2022, institutional investor flows have grown at a faster pace, diminishing the relative importance of retail trading. Retail buy flows increase more than sell flows, leading to a positive retail flow imbalance across all stock categories, except for small stocks. In the 2014-2016 period, retail investors show a strong preference for small stocks, with an NTA II of 0.282 compared to 0.216 for large stocks. This preference becomes more pronounced in the 2017-2022 period, where NTA II for small stocks rises sharply to 0.540. However, the relative importance of retail trading in small stocks declines more significantly then in other categories (0.055 vs. 0.044). The relative importance of retail trading in large stocks experience slightly decline (from 0.039 to 0.033). Retail trading maintains tilt toward value stocks relative to growth stocks, with an NTA II of growth and value are 0.384 vs. 0.235 in the 2014-2016 period, and 0.513 vs. 0.398 in the 2017-2022 period. Additionally, retail investors show a stronger tilt toward losing stocks relative to winning stocks (0.373 vs. 0.313 in the 2014-2016 period and 0.643 vs. 0.431 in the 2017-2022 period). The retail flow imbalance turns positive (0.008) in the 2017-2022 period, suggesting an increasingly contrarian strategies (buy "losers"), consistent with the finding of Luo et al. (2023).

[Table 1]

3. Empirical results.

This section describes a decline in the predictability of retail trading from 2014 to 2022. We then analyze factor tilts in retail trading and the predictability for returns across stocks with different factor exposures to explore potential reasons for the decrease in predictability.

3.1. Prediction of retail trading on stock returns

We apply the main methodology of BJZZ to estimate the predictive power of weekly retail flow imbalance for future stock returns. Using retail trading data from VandaTrack, we run Fama-MacBeth regressions (1973) in an extended sample period. In Equation (1), the dependent variable is the weekly return for future k weeks ($k = 1, 2, \ldots n$), calculated as the cumulative firm-level daily stock return from Wednesday to the following Tuesday. The independent variable is the weekly retail flow imbalance, measured as the cumulative daily flow imbalance. Control variables, consistent with BJZZ, include weekly returns and turnover, previous monthly returns and return volatility, and cumulative returns over the prior six months. Considering serial correlation arising from overlapping data, we adjust the time series standard errors using the Newey-West method (1987). Observations from the first quarter of 2020 are excluded to mitigate the impact of market disruptions due to the COVID-19 shock[14].

$$Ret_{w,w+k} = \beta_{0,w} + \beta_{1,w} mroibflow_{i,w} + \beta'_{2,w} controls_{i,w} + \varepsilon_{i,w+k}, k = 1, 2, \dots n. \quad (1)$$

We find significant predictability of retail flow imbalance on future stock returns during the 2014-2016 period, but this predictability becomes insignificant in the subsequent years. Table 2 shows that the coefficient of retail flow imbalance is positive and significant (0.0005) during the 2014-2016 period, suggesting that when retail buying outweighs selling, stock returns tend to increase in the following week, and vice versa. The predictive power persists for up to five weeks. Although this result is somewhat weaker than the finding of BJZZ, which reports that predictability lasts up to eight weeks during 2010-2015, our finding remains consistent in demonstrating that retail trading can positively predict weekly stock returns. Differences in the strength of predictability may be due to differences in the sample periods[15] and/or different databases. However, from 2017 to 2022, retail trading no longer exhibits significant predictive power for future weekly returns. Figure 1 further illustrates this evolution by showing the returns of portfolios sorted by retail flow imbalance over the next seven weeks across different sub-periods. During

[14] Hufner, Strych, and Westerholm (2022) find that retail investors provided liquidity to stock during the COVID-19 shock. However, we do not observe such evidence during normal periods. To assess the general impact of retail trading, we exclude the COVID-19 shock period from our analysis.
[15] The sample period of BJZZ covers the 2010-2015 sample period. We can only start in 2014 because VandaTrack data start in 2014.

2014-2016, there are substantial return gaps between portfolios in the top 10% (strongest buying) and bottom 10% (strongest selling) of *mroibflow* over the next five weeks. In contrast, this return spread narrows significantly in the 2017-2019 period, suggesting a decline in predictive power of retail trading. The declining predictability continues beyond the COVID-19 period[16].

[Table 2]

[Figure 1]

3.2. Sources of predictability

We then investigate whether the predictability of retail flows, and its evolution, originates from the buy side, sell side, or both. We rank stocks into quartiles based on weekly retail flow imbalance, with the bottom 25% representing the highest selling pressure and the top 25% representing the highest buying pressure. To explore potentially asymmetric predictability between buying and selling dominance, we apply piecewise linear regressions (Sirri and Tufano, 1998) which allow us to separately analyze sensitivities of retail flow imbalance to next week's returns across each quartile. We introduce four dummy variables: $D_{1,2,3,4_w}$, which equal one if the stock's retail flow imbalance falls within the respective quartile. We replace the *mroibflow* variable with four interaction terms between the dummy variables and *mroibflow* in Equation (1). The coefficients on these interaction terms capture the conditional predictability of *mroibflow* for next week's returns within each quartile. Control variables and standard errors in the Fama-MacBeth regressions (1973) remain consistent with those specified in Equation (1).

[16] Therefore, we combine the two periods of (2017-2019 and 2020-2022) in our analysis.

Table 3 shows that the persistent predictive power of retail trading over the next five weeks during the 2014-2016 period primarily arises from the top 25% of *mroibflow,* implying that stocks most heavily bought and least sold by retail investors are associated with higher subsequent returns. The bottom 25% of *mroibflow* also exhibits a positive effect, but its coefficient only marginally significant for the following week, suggesting that stocks most heavily sold and least bought by retail investors are associated with lower returns in the next week. In contrast, during the 2017-2022 period, there is no evidence of predictability across any quintiles. Overall, the strong predictability of retail flow imbalance for returns over the next five weeks, which was evident from the most heavily bought stocks in the 2014-2016 period, has disappeared in the most recent period.

[Table 3]

Next, we examine whether the predictive power of retail trading persists after controlling for exposures to size, book-to-market, momentum. We replace raw returns with DGTW characteristic-adjusted returns in Equation (1). Table 4 shows that, during the 2014-2016 period, the top 25% of retail flow imbalance does not exhibit statistically significant predictability beyond the next week. Although the overall coefficients on *mroibflow* remains positive and significant, both the magnitude and significance levels are weaker than those observed in regressions using raw returns. The predictability for the next week may be driven by a demand shock induced by intense retail buying or selling. The findings suggest that the observed predictability is mainly explained by retail trading's tilt toward size, book-to-market, and momentum factors.

[Table 4]

To better understand how common factors explain the predictability of retail trading, we start with the size factor given the aggressive retail trading in small stocks. We classify stocks into three size categories based on quarterly DGTW size portfolio rankings. We then define three dummy variables S_1, S_2, and S_3. Small stocks ($S_1 = 1$) are those ranked in "1" or "2" in DGTW size portfolios, representing the bottom 40% in market capitalization. Medium stocks ($S_2 = 1$) are those in "3", representing the 40%-60%. Large stocks ($S_3 = 1$) are those in "4" or "5", representing the top 40%. We focus on the role of the top 25% of retail flow imbalance, as prior results indicate that predictability in raw returns is concentrated in this quartile. In Equation (2), we construct three interaction terms between the top 25% of *mroibflow* and the size dummy variables: Top 25% *mroibflow**Small, Top 25% *mroibflow**Medium, and Top 25% *mroibflow**Large. We retain the original interactions between *mroibflow* and its quartile dummies (bottom 25%, 25% - 50%, and 50% - 75%) in the Fama-MacBeth (1973) regressions. The dependent variable is the weekly return controlled for book-to-market and momentum factors, calculated as the cumulative daily return minus the cumulative daily return of the stock's DGTW 5x5 B/M-Momentum portfolio. Control variables and Newey-West standard errors are applied as in Equation (1).

$$\begin{aligned} ret_{w,w+k} = \beta_{0,w} + \beta_{1,w} mroibflow * D1_{i,w} * Small \\ + \beta_{2,w} mroibflow * D1_{i,w} * Medium + \beta_{3,w} mroibflow * D1_{i,w} * Large \\ + \beta_{4,w}' controls_{i,w} + \varepsilon_{i,w+k}, k = 1, 2, \dots 5. \end{aligned} \quad (2)$$

Table 5 shows that the predictive power of the top 25% *mroibflow* is primarily observed in small stocks. The coefficient of the top 25% *mroibflow**small interaction is positive and significant (0.0005), indicating that small stocks experiencing strong net retail buying pressure are expected to earn higher returns in the subsequent week. This predictive effect remains significant over a five-week horizon, suggesting a persistent influence of intense retail buying on small stock performance during the 2014-2016 period. We also find that the predictive effect of the bottom 25% *mroibflow* for next week's returns is similarly concentrated in small stocks. However, during the 2017-2022 period, retail flow imbalance does not exhibit predictive power for small stocks. Accordingly, the decline in retail trading predictability can be partially attributed to small stocks.

We then investigate possible reasons behind retail investors' declining predictability for small stocks. As shown in Table 1 Panel B, both retail and institutional investors significantly increased their trading activities in small stocks during the 2017 to 2022 period, enhancing the liquidity of these stocks. Therefore, retail trading would have a diminished impact on the prices and conveyance of information related to prices and returns for small stocks compared to the 2014-2016 period. We apply Amihud's (2002) illiquidity measure, measured weekly for each stock as the average ratio[17] of absolute daily returns to dollar volume, multiplied by 10^6. Following Amihud (2002), we winsorize this measure at the highest and lowest 1% tails of the distribution. Table 5 Panel B displays descriptive statistics for illiquidity across the two sample periods. Small stocks are more illiquid than the overall sample average in both periods (120.976 vs. 46.506 during the 2014-2016 period and 51.868 vs. 20.596 during the 2017-2022 period). However, the levels of illiquidity among small stocks drops by more than half from the 2014-2016 period to the 2017-2022 period (120.976 vs. 51.868). Accordingly, we hypothesize that increased liquidity could be a reason for the decline in retail predictability for small stocks. We split stocks into high- and low-liquidity groups based on their quarterly illiquidity measure. We then construct two liquidity dummy variables and interact them with the top 25% *mroibflow* across size groups (Small,

[17] We use quarterly illiquidity to keep consistence with the size sorting in DGTW characteristic-quarterly-adjusted portfolios. We also check annual and month liquidity measures, and the results are consistent.

Medium, or Large). We also include the remaining three interactions between *mroibflow* and its respective quartile dummies. Panel C shows that predictive power from retail flows in the top 25% is concentrated among small stocks with low liquidity, while no significant predictability is found for small stocks with high liquidity. This finding supports our hypothesis that liquidity is a factor influencing the predictability of retail investors for small stocks in the 2014-2016 period. We conduct a similar test for the bottom 25% of *mroibflow* and again find that the predictability stems only from small stocks with low liquidity. Overall, the declining predictive power of retail trading in recent years can be partially attributed to the increased liquidity of small stocks.

[Table 5]

We apply the same methodology used for the size factor to examine the role of the book-to-market factor in explaining the decline in retail predictability. We classify stocks into three categories based on quarterly DGTW book-to-market portfolio rankings. We then define three dummy variables B_1, B_2, and B_3. Growth stocks ($B_1 = 1$) are those ranked in "1" or "2" in DGTW B/M portfolios, representing the bottom 40% in B/M. Neutral stocks ($B_2 = 1$) are those in "3", representing the 40%-60%. Value stocks ($B_3 = 1$) are those in "4" or "5", representing the top 40%. We create three interaction terms between the top 25% of retail flow imbalance and the B/M-based dummy variables: Top 25% *mroibflow**Growth, Top 25% *mroibflow**Neutral, and Top 25% *mroibflow**Value. We keep the remaining three interactions between *mroibflow* and its respective quartile dummies in the regressions. The dependent variable is the weekly return controlled for size and momentum factors, calculated as the cumulative daily return minus the cumulative daily return of the stock's DGTW 5x5 Size-Momentum portfolio. Table 6 presents that during the 2014-2016 period, the top 25% retail flow imbalance exhibits positive and significant predictive power for value stocks, implying that when retail investors' buying strongly dominates selling, the following week's returns for value stocks are significantly higher. The predictability persists for five weeks, indicating that the book-to-market ratio is another factor contributing to the overall predictive power of retail trading. In the 2017-2022 period, the top 25% retail flow imbalance continues to show positive and significant predictive power for value stocks, persisting for five weeks. The coefficients in the later period are larger than those observed in the 2014-2016 period, probably due to increased buy flows. The sustained predictability for value stock in both sub-periods suggests that the book-to-market ratio does not explain the decline in the predictability of retail trading.

[Table 6]

Finally, we analyze the role of momentum factor. We classify stocks into three categories based on quarterly DGTW momentum portfolio rankings. We then define three dummy variables U_1, U_2, and U_3. Losing stocks ($U_1 = 1$) are those ranked in "1" or "2" in DGTW momentum portfolios, representing the bottom 40% in past 12-month performance. Neutral stocks ($U_2 = 1$) are those in "3", representing the 40%-60%. Winning stocks ($U_3 = 1$) are those in "4" or "5", representing the top 40%. We create three interactions terms between the top 25% of retail flow imbalance and the momentum-based dummy variables: Top 25% *mroibflow**Losing, Top 25% *mroibflow**Neutral, and Top 25% *mroibflow**Winning. The remaining three interactions between *mroibflow* and its respective quartile dummies remain unchanged in the Fama-MacBeth (1973) regressions. The dependent variable is the weekly return controlled for size and momentum, calculated as the cumulative daily return minus the cumulative daily return of the stock's DGTW 5x5 Size-B/M portfolio. Table 7 shows that, beyond the next week's returns, there is no significant predictive power of the top 25% retail flow imbalance across different momentum groups, and the predictive effect does not exhibit clear momentum-based factor exposures. The predictability for next week's return may instead be driven by other reasons. Given that the observed predictability concentrates on small stocks with low liquidity and considering that DGTW-adjusted returns control for the three common factors only at the characteristic portfolio level, we conduct separate regressions for small and illiquid stocks to isolate the size and liquidity effect. The result reveals that the predictive power of the top 25% *mroibflow* on losing, neutral, and winning stocks is evident only in small stocks with low liquidity, suggesting that, once size effect is removed, the top 25% of retail flows imbalance does not predict next week's returns for stocks across different momentum categories. In the 2017-2022 period, we find no evidence of predictability. We further test the predictability using 3-month and 1-month performance as alternative momentum factors, following the portfolio sorting methodology of Daniel, Grinblatt, Titman, and Wermers methodology (1997). The findings are consistent with those from the 12-month momentum factor. Overall, retail investors do not demonstrate predictability for losing or winning stocks, even as they adopt a more contrarian approach.

[Table 7]

So far our finding suggests that the predictability of retail investor activity on future returns is substantially reduced when controlling for size, book-to-market, and momentum factors through DGTW-adjusted returns. Furthermore, we demonstrate that the decline in retail trading predictability can be partially attributed to the diminished influence of retail trading on small stocks, whose liquidity significantly improved over time. To further isolate the effects of size and liquidity on the predictability, we conduct a separate analysis of small stocks with low liquidity, compared to other stocks, during the 2014-2016 period.

We construct two dummy variables: SL and Other. SL equals one if a stock is a small-cap stock with low liquidity, determined by quarterly DGTW size portfolios and the Amihud (2002) illiquidity measure. Other equals one for the remaining stocks. We first interact weekly retail flow imbalance with each of the two stock-type dummy variables (SL and Other). Given that the predictability for value stocks remains stable across the 2014-2022 period, we control for book-to-market effect using DGTW B/M portfolios. The dependent variable is the weekly return controlled for the book-to-market factor, calculated as the cumulative daily return minus the cumulative daily return of the stock's DGTW B/M portfolio. We further decompose retail flow imbalance into quartiles: bottom 25% (strongest selling), 25% - 50%, 50% - 75%, and top 25% (strongest buying). We replace the overall *mroibflow* variable in Equation (3) with quartile-based interactions of *mroibflow*, interacted with the SL and Other dummy variables.

$$ret_{w,w+k} = \beta_{0,w} + \beta_{1,w} mroibflow_{i,w} * SL + \beta_{2,w} mroibflow_{i,w} * Other + \beta'_{3,w} controls_{i,w} + \varepsilon_{i,w+k}, k = 1, 2, \dots n. \qquad (3)$$

Table 8 presents evidence of the predictive power of retail flow imbalance over the next five weeks. Retail flow imbalance shows a positive and statistically significant predictability on future returns of small stocks with low liquidity, whereas its predictability for other stocks is negligible. An analysis of the quartile-based *mroibflow* measures reveals that the persistent predictability mainly stems from the top 25% of *mroibflow*. Although the bottom 25% of *mroibflow* also has a positive effect in both stock groups, the coefficient is larger (0.0016 vs. 0.0003) and statistically more significant (2.39 vs. 1.89) for the small and liquidity stocks. This observed predictability for the next week's returns may be influenced by demand shocks caused by persistent retail selling pressure that push prices downward temporarily. In summary, our results indicate that aside from trading exposure to the three common factors, retail trading does not significantly predict future alpha during the 2014-2016 period, consistent with the findings during the 2017-2022 period. Therefore, the decline in retail trading predictability may not reflect a reduction in the trading ability or information advantage of retail investors. Rather, retail trading exerts a strong and persistent price impact on small stocks with low liquidity than other stocks. During the 2017-2022 period, although retail flows into small stocks increased, institutional flows increased more substantially. Institutional and retail investors both improved the liquidity of small stocks. As a result, the influence of retail trading on small stock prices has weakened, making it more difficult for retail investors to predict future returns in the absence of a superior information.

[Table 8]

We further assess the predictability of retail trading by constructing portfolios during the 2014-2016 period. We sort stocks into deciles based on their weekly retail flow imbalance and form equal-weighted portfolios, where P10 comprises stocks in the top 10% of *mroibflow* (i.e., retail buying strongly dominating selling) and P1 comprises stocks in the bottom 10% of *mroibflow* (i.e., retail selling strongly dominating buying). Small stocks with low liquidity are identified using quarterly DGTW size portfolios and the Amihud (2002) illiquidity measure. Table 9 demonstrates that the P10 portfolio of small and illiquid stocks experience positive and significant returns over the next four weeks, implying that small and illiquid stocks with the retail buy flows strongly dominate sell flows earn higher returns than other stocks with similar retail flow characteristics in the subsequent month. Furthermore, a zero-investment strategy that goes long on the P10 portfolio and short on P1 portfolio yields positive and significant returns among small stocks with low liquidity over the next five weeks, but not among other stocks, supporting our main finding that retail trading predictability during the 2014-2016 period is concentrated in small stocks with low liquidity. The return of this strategy reaches a peak in the first week (0.4491%) and diminishes over the subsequent four weeks. The second zero-investment strategy which is long on the first the zero-investment strategy (P10-P1) applied to small and illiquid stocks and short on the same strategy applied to other stocks also generates positive returns, significant in the first and third week. We also apply excess returns calculated by subtracting the market risk-free rate from raw returns using data from Fama and French website. The results for excess returns are consistent with those using return controlled for the B/M factor, suggesting that the predictability for value stocks does not significantly affect the predictability for small and illiquid stocks and confirming that zero-investment strategies leveraging raw returns can effectively capture the predictive power.

[Table 9]

4. Social media and the predictability of retail trading

We also explore whether social media activity influences the predictability of retail trading. Since 2015, Bloomberg has tracked the daily frequency of stock mentions on Twitter (tweets publication count) and analyzed the average daily sentiment of these tweets (tweet sentiment). Given that Twitter discussions tend to cluster on large-cap stocks, we categorize stocks into deciles based on market capitalization and rank each stock's daily tweet publication count in percentile within each size group. The average daily tweet publication percentile and sentiment score are then averaged at the weekly levels. Table 10 Panel A reports summary statistics of weekly tweet counts and sentiment. The average tweet volume per stock was slightly higher during the 2015-2016 period than in the 2017-2022 period (64.744 vs. 56.828), although the maximum volume was higher in the later period (83,405 vs. 45,562). To capture the level of tweet activity, we define two dummy variables: $SM_{1,w-1}$ equal to one if the previous week's tweet count percentile is above 50% and $SM_{2,w-1}$ equal to one if it is below 50%. These dummies are then interacted with weekly retail flow imbalance to examine how social media intensity influences retail predictability. In the 2015-2016 period, retail flow imbalance exhibits a stronger predictive effect for stocks with higher tweet volumes than those with lower volumes (0.0006 vs. 0.0005). However, there is no significant predictive power in the 2017-2022 period, suggesting the limited effect of social media attention on enhancing predictability. To assess whether social media influences our main findings on small and illiquid stocks, we further intersect the tweet volume dummies with the interaction terms between retail flow imbalance and small stocks with low liquidity. Results show that when small stocks with low liquidity are highly discussed on Twitter, retail investors exert a greater predictive impact on next week's returns compared to other small stocks with low liquidity (0.0006 vs. 0.0005). Moreover, retail trading displays weak predictive power on the next week's return in other stocks with high social media discussion (0.0002). This is probably due to persistent retail flows stimulated by intense discussions. In contrast, during the 2017-2022 period, we do not observe significant impact of social media discussions on predictability.

Similarly, we examine whether sentiment expressed on Twitter influences the predictive power of retail trading. The average sentiment score was lower in the 2017-2022 period than in the 2015-2016 period (0.02 vs. 0.041), with a higher percentage of stock exhibiting negative sentiment in the later period (36.908% vs.29.446%). The increase negative sentiment may appeal to retail investors employing contrarian trading strategies, contributing to the positive retail flow imbalance toward losing stocks during the 2017-2022 period shown in Table 1 Panel B. We introduce two dummy variables to capture sentiment polarity: $SM_{1,w-1}$ equal to one if the previous week's average sentiment is positive, and $SM_{2,w-1}$ equal to one if the sentiment is negative. In Panel C, during the 2015-2016 period, overall *mroibflow* exhibits a significantly positive effect on next week's returns when sentiment is positive (0.0007) and this predictability is mainly driven by the stocks in the top 25% of *mroibflow*, suggesting that positive sentiment amplifies buying pressure. This result is consistent with Hu et al. (2021), which report that positive tone predicts higher returns using WallStreetBets data. Examining retail trading responses to Twitter sentiment, we find that positive sentiment in the previous week encourages retail buying, while negative sentiment drives retail selling. However, when isolating small and illiquid stocks, the predictability under positive or negative sentiment remains similar (0.0007 vs. 0.0006), indicating that low liquidity in small stocks remains a dominant factor in explaining predictability, despite negative sentiment's mitigating effect on buying. In the 2017-2022 period, we find no significant difference in the predictive power of retail trading based on sentiment conditions. Overall, while increased tweet volume and positive sentiment may stimulate retail trading activity and amplify impacts on next week stock returns, social media's impact on enhancing weekly predictability remains limited.

The results in Panel C do not eliminate the possibility that social media attributes to the observed decline in retail trading predictability. Prior research provides evidence that social media activity can be detrimental to market participants and market efficiency. Lou et al. (2023) document that intense discussions on platforms, such as StockTwits, Seeking Alpha, and WallStreetBets can reduce the production of fundamental information. Lopez Avila et al., (2023) find that social media may disseminate misleading information, making it more difficult for retail investors to extract valuable insights and thus impeding their ability to forecast future prices. However, these earlier studies do not examine the evolution of social media activity and its possible influence on the evolution of retail trading predictability. Given the decreased in average tweet sentiment observed in Panel A during the 2017-2022 period (from 0.041 to 0.027), we categorize weekly tweet sentiments into quartiles: bottom 25% (most negative), 25% - 50% (negative), 50% - 75% (positive), and top 25% (most positive) to investigate which sentiment level contribute to this decline. We also categorize tweet volume into two groups based on tweet publication percentile rankings: above 50% (high volume) and below 50% (low volume). Panel B presents summary statistics of tweet volumes and sentiment for the periods 2015-2016 and 2017-2022. Average sentiments in the 2^{nd} and 3^{rd} sentiment quartiles became more neutral during 2017-2022, suggesting a greater presence of negative sentiment within these previously moderate categories. Stocks with the most negative or positive sentiment and high tweet volume also display a decrease or increase in average sentiment (-0.017 and 0.030). The coexistence of positive and negative tweets, especially under conditions of high tweet volumes, may confuse retail investors and increase the likelihood of inaccurate forecasts. Furthermore, if retail investors are limited in attention or exhibit confirmation bias, as suggested by Cookson et al. (2023), they may selectively process information that aligns with their prior beliefs, either buying on positive sentiment or selling on negative sentiment. Consequently, the dominant trading direction (i.e., buying or selling) may drive short-term price movements, increasing price volatility and impairing the predictability of retail investors. We therefore hypothesize that more neutral sentiment, especially in the presence of intense social media discussions, may attribute to the observed decline in retail trading predictability. We construct eight interaction terms by combining the sentiment quartiles with the tweet volume groups. Each interaction term is then multiplied by *mroibflow* to capture the sensitivity of retail flows on the next week's returns in the range of sentiment levels and social media intensity.

Table 10 Panel D illustrates that during the 2015-2016 period, retail trading has a significantly positive impact on the next week's returns for stocks with the most positive sentiment on Twitter, especially when accompanied by high tweet volumes (0.0017). Retail trading also has a significantly positive impact on stocks with the 3rd sentiment quartile but low tweet volumes (0.0007). However, when tweet volume is high within the same sentiment quartile, this effect becomes insignificant. A large volume of mixed sentiment tweets may simultaneously stimulate retail trading activity on both buy and sell sides. Retail demand shocks may push prices temporarily in either direction depending on which side dominates. The resulting increase in trading volatility leads to larger standard errors, weakening the statistical significance of the retail flow coefficients. We do not observe significant retail trading predictability in the 2nd sentiment quartile. There is a weak negative impact when sentiment is in the most negative quartile and tweet volumes are high (0.0003). These findings highlight that retail trading predictability is most evident when social media sentiment shows clear and directional social media sentiment. As sentiment becomes more neutral or ambiguous, especially with a high tweet volume, retail trading becomes less predictive. In the 2017-2022 period, sentiment in the 2^{nd} and 3^{rd} quartile becomes more neutral (see Panel B) and the predictability of retail trading can be weakened by this trend. Furthermore, in the 2017-2022 period, retail trading under the most positive sentiment combined with high tweet volumes shows a significantly negative impact on subsequent weekly returns (-0.0009). Intensified retail buying may prompt faster prices reversals, probably due to increased liquidity in recent years, which facilitates quicker absorption of demand shocks and limits persistent price impacts.

[Table 10]

5. Other results.

This section examines other widely used factors that may influence the predictability of retail trading. We further decompose the predictability to explore its underlying explanations.

5.1. Investment and profitability factors

We also test whether retail investors exhibit predictability concerning the investment and profitability factors in Fama and French's (2015) five-factor model. Our findings reveal predictive power for the subsequent week's returns (as shown in Table A1 and A2), but no distinct factor exposure, similar to the momentum factor results. However, when isolating small and low-liquidity stocks, we observe that the predictability is primarily driven by this subsample.

5.2. Autocorrelation of retail trading

We also investigate the autocorrelation of retail trading in the 2014-2016 and the 2017-2022 periods to assess whether predictability in the first period arises from demand shocks caused by persistent retail flows. A high autocorrelation of retail order imbalance would suggest that retail buying or selling pressure persists across several weeks, influencing future stock prices. High correlations of retail trading and its lags may be linked to herding behavior. As retail investors increasingly access broader information channels, we might expect a reduction in both herding and trading correlations. In Equation (4), the dependent variable is the retail flow imbalance in the next week and the independent variables include retail flow imbalance this week and in the three previous weeks. The regression employs a Fama-MacBeth (1973) regression with Newey-West standard errors for 20 lags. Table 9 shows that the impacts of past *mroibflow* on the next week's *mroibflow* remains consistent across both the 2014-2016 and 2017-2022 periods, indicating that the persistence of retail flows has not change significantly over time and therefore, does not account for the observed decline in the predictability.

6. Conclusion

This paper examines the decline in the predictability of retail trading on stock returns during the 2014-2022 period. Retail investors can positively predict stock returns over the subsequent five weeks from 2014 to 2016, but this predictive power declines during 2017 to 2022. Using DGTW characteristic-adjusted portfolios sorted on size, book-to-market, and momentum factors, this paper reveals that the decline is particularly pronounced among small stocks with low liquidity. In the 2014-2016 period, retail investors' extensive buying activity on certain small stocks with low liquidity creates a larger and more persistent price impact. However, as liquidity improved during 2017-2022 and the relative influence of retail trading diminished, retail investors exerted a weaker impact on the prices of small stocks. Without access to as much information as institutional investors, it is increasingly difficult for retail investors to predict future returns. Retail investors also demonstrate persistent predictive power for value stocks throughout the 2014-2022 period. Beyond trading exposure to the size, value, and momentum factors, retail investors were unable to significantly predict future returns either in the 2014-2016 or 2017-2022 period. Thus, the observed decline in predictability is not primarily attributable to a reduction in retail investors' trading ability or access to information.

Other factors, such as momentum, book-to-market, investment, and profitability do not yield significant results in explaining the decline in retail trading predictability. While social media discussions and a positive sentiment bias can stimulate increased retail trading and amplify its impact on next week's returns, their effect on enhancing predictability remains limited. The shift toward neutral sentiment appears to contribute to the decline in predictability. When a high volume of both positive and negative information is presented simultaneously, retail investors may selectively process information that aligns with their prior views or fail to absorb all relevant information. As retail investors increasingly adopt diverse strategies, their trading becomes less likely to exert a consistent and discernible impact on future returns, except in cases where sentiment is overwhelmingly positive. Furthermore, increased liquidity facilitates quicker price adjustments, further diminishing predictability. Generally, beyond the size factor and social media, other factors may also explain the decline in predictability. For example, the rise in institutional trading activity in small stocks during the 2017-2022 period has reduced the relative influence of retail trading. As retail investors become more active and their trading clusters around news events (e.g., Hirshleifer and Sheng, 2022; Farrell et al., 2020), institutional investors may exploit price bubbles created by retail trading, either by riding the trend or trading against it, to capitalize on profit opportunities.

References

Amihud, Yakov. "Illiquidity and stock returns: cross-section and time-series effects." *Journal of financial markets* 5, no. 1 (2002): 31-56.

Barber, Brad M., and Terrance Odean. "Boys will be boys: Gender, overconfidence, and common stock investment." *The quarterly journal of economics* 116, no. 1 (2001): 261-292.

Barber, Brad M., and Terrance Odean. "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors." *The review of financial studies* 21, no. 2 (2008): 785-818.

Barber, Brad M., Shengle Lin, and Terrance Odean. "Resolving a paradox: Retail trades positively predict returns but are not profitable." *Journal of Financial and Quantitative Analysis* (2021): 1-67.

Barber, Brad M., Xing Huang, Philippe Jorion, Terrance Odean, and Christopher Schwarz. "A (sub) penny for your thoughts: Tracking retail investor activity in TAQ." *The Journal of Finance* (2023).

Bartlett, Robert P., Justin McCrary, and Maureen O'Hara. "Tiny trades, big questions: Fractional shares." *Journal of Financial Economics* 157 (2024): 103836.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang, and Xinran Zhang. "Tracking retail investor activity." *The Journal of Finance* 76, no. 5 (2021): 2249-2305.

Coval, Joshua D., David A. Hirshleifer, and Tyler Shumway. "Can individual investors beat the market?." (2005): 02-45.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam. "Liquidity and market efficiency." *Journal of financial Economics* 87, no. 2 (2008): 249-268.

Chordia, Tarun, Avanidhar Subrahmanyam, and Qing Tong. "Have capital market anomalies attenuated in the recent era of high liquidity and trading activity?." *Journal of Accounting and Economics* 58, no. 1 (2014): 41-58.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers. "Measuring mutual fund performance with characteristic-based benchmarks." *The Journal of finance* 52, no. 3 (1997): 1035-1058.

Fama, Eugene F., and Kenneth R. French. "Common risk factors in the returns on stocks and bonds." *Journal of financial economics* 33, no. 1 (1993): 3-56.

Fama, Eugene F., and Kenneth R. French. "A five-factor asset pricing model." *Journal of financial economics* 116, no. 1 (2015): 1-22.

Fama, Eugene F., and James D. MacBeth. "Risk, return, and equilibrium: Empirical tests." *Journal of political economy* 81, no. 3 (1973): 607-636.

Farrell, Michael, T Clifton Green, Russell Jame, and Stanimir Markov, 2022, The democratization of investment research and the informativeness of retail investor trading, *Journal of Financial Economics* 145, 616–641

Gompers, Paul A., and Andrew Metrick. "Institutional investors and equity prices." *The quarterly journal of Economics* 116, no. 1 (2001): 229-259.

Hu, Danqi, Charles M. Jones, Siyang Li, Valerie Zhang, and Xiaoyan Zhang. "The rise of reddit: How social media affects retail investors and short-sellers' roles in price discovery." *Available at SSRN 3807655* (2021).

Hirshleifer, David, and Jinfei Sheng. "Macro news and micro news: complements or substitutes?." *Journal of Financial Economics* 145, no. 3 (2022): 1006-1024.

Hüfner, Felix, Jan-Oliver Strych, and P. Joakim Westerholm. "The Impact of Retail Investors on Stock Liquidity and Crash Risk." *Available at SSRN 3973202* (2022).

Kaniel, Ron, Gideon Saar, and Sheridan Titman. "Individual investor trading and stock returns." *The Journal of finance* 63, no. 1 (2008): 273-310.

Kelley, Eric K., and Paul C. Tetlock. "How wise are crowds? Insights from retail orders and stock returns." *The Journal of Finance* 68, no. 3 (2013): 1229-1265.

Korniotis, George M., and Alok Kumar. "Do older investors make better investment decisions?." *The review of economics and statistics* 93, no. 1 (2011): 244-265.

Kang, Namho, Xiaoxia Lou, Gideon Ozik, Ronnie Sadka, and Siyi Shen. "Innocuous Noise? Social Media and Asset Prices." *Social Media and Asset Prices (April 21, 2022)* (2022).

Kumar, Alok, and Charles MC Lee. "Retail investor sentiment and return comovements." *The Journal of Finance* 61, no. 5 (2006): 2451-2486.

Liu, Hongqi, Lin Peng, and Yi Tang. "Retail attention, institutional attention." *Journal of Financial and Quantitative Analysis* 58, no. 3 (2023): 1005-1038.

Luo, C., E. Ravina, M. Sammon, and L. M. Viceira. "Retail investors' contrarian behavior around news." *Attention, and the Momentum Effect* (2022).

Lopez Avila, Edna, Charles Martineau, and Jordi Mondria. "When Crowds Aren't Wise: Biased Social Networks and its Price Impact." *Available at SSRN 4439793* (2023).

Roll, Richard, Eduardo Schwartz, and Avanidhar Subrahmanyam. "Liquidity and the law of one price: The case of the futures‐cash basis." *The Journal of Finance* 62, no. 5 (2007): 2201-2234.

Sirri, Erik R., and Peter Tufano. "Costly search and mutual fund flows." *The journal of finance* 53, no. 5 (1998): 1589-1622.

Table 1. Summary statistics of retail trading activity and stock characteristics

Table 1 Panel A displays time-series cross-sectional averages of mean, median, standard deviation, minimum, and maximum of retail flows and stock characteristics in 2014-2022. The sample covers 2,266 trading days. Total flow traded is calculated as the total trading volume multiplied by the close price for each day. Turnover is defined as the ratio of total trading volume to total shares outstanding. The daily DGTW return is stock's raw return minus the return of its value-weighted DGTW portfolio, where portfolios are sorted quarterly. Stocks with minimum prices below $1 at the previous month-end are excluded from the sample. Panel B displays two normalized measures of retail trading activities: NTA I and NTA II, as well as retail flow imbalance across different stock categories for two sub-periods: 2014-2016 and 2017-2022. NTA I and NTA II are computed using the following equations. Stock categories defined based on the DGTW 125 portfolios.

$$NTA\ I_{i,t} = \frac{Retail\ flow\ from\ retail\ investors\ in\ the\ sample}{Total\ flow\ in\ the\ market} \times 10^6$$

$$NTA\ II_{i,t} = \frac{Retail\ flow\ from\ retail\ investors\ in\ the\ sample}{Market\ capitalization} \times 10^6$$

Panel A. Summary statistics of retail flows and stock characteristics in 2014-2022

Variable	Mean	Median	Std Dev	Minimum	Maximum
Mroibflow	-0.005	0.009	0.025	-0.114	0.076
Buy Flow (Millions of $)	1.151	1.112	0.371	0.317	2.890
Sell Flow (Millions of $)	1.068	1.037	0.341	0.224	2.651
Total Flow Traded (Millions of $)	76.314	73.294	17.134	25.640	185.089
Turnover (%)	2.105	2.038	0.440	0.550	4.955
Return (%)	0.051	0.083	1.326	-12.267	9.901
DGTW Return (%)	0.041	0.066	1.403	-12.823	10.185
Price ($)	48.071	47.866	4.745	31.148	60.627
Market Capitalization (Millions of $)	9694.410	9460.973	1398.090	7169.440	13053.290
Number of stocks	2440.980	2387	393.195	1887	3005

Panel B. Retail trading activities and factor exposures

		2014-2016			2017-2022		
	Stock category	NTA I	NTA II	*Mroibflow*	NTA I	NTA II	*Mroibflow*
Overall		0.039	0.269	-0.019	0.032	0.403	0.003
Size	Small	0.055	0.282	-0.029	0.044	0.540	-0.002
	Big	0.039	0.216	0.006	0.033	0.229	0.020
B/M	Growth	0.043	0.384	-0.009	0.035	0.513	0.010
	Value	0.035	0.235	-0.026	0.031	0.398	0.001
Momentum	Losing	0.032	0.373	-0.022	0.030	0.643	0.008
	Winning	0.047	0.313	-0.014	0.034	0.431	0.003

Table 2. Prediction of retail trading

Table 2 displays the predictability of retail flow imbalance on weekly returns over the next seven weeks during the periods 2014-2016 and 2017-2022. The independent variable is the weekly retail flow imbalance, calculated as the cumulative daily imbalances. The dependent variable is the weekly stock return, measured by the cumulative daily returns. Control variables include weekly returns and turnover, previous monthly returns and return volatility, and cumulative returns over the prior six months. Monthly return volatility is computed as the standard deviation of daily returns. Turnover is defined as total trading volume divided by total shares outstanding. All weekly variables are measured from Wednesday to the following Tuesday. Observations during the Covid-19 shock period (Q1 2020) are excluded from the sample. We estimate Fama-MacBeth (1973) regressions and report the time-series average of the regression coefficients. T-statistics are computed using Newey-West standard errors. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

Panel A. Prediction of retail flow imbalance on next-week returns

	2014-2016		2017-2022	
Dependent variable:	Ret (w, w+1)	T-stats	Ret (w, w+1)	T-stats
mroibflow (i, w)	0.0005***	(6.60)	0.0001	(0.85)
Weekly return (i, w)	-0.0124	(-1.48)	-0.0216	(-3.81)
Turnover (i, w)	-0.0003	(-3.35)	-0.0108	(-2.12)
Monthly return (i, m-1)	-0.0025	(-0.82)	-0.0046	(-1.21)
Monthly return (i, m-7, m-2)	-0.0042	(-3.80)	-0.0021	(-1.42)
Monthly volatility (i, m-1)	0.0209	(0.53)	0.0260	(0.62)
Intercept	0.0020	(1.52)	0.0017	(1.32)
Observations	272,766		582,355	
Adjusted R-squared	0.0426		0.0262	

Panel B. Prediction of retail flow imbalance on weekly returns in longer horizon

	2014-2016					
	Ret (w, w+2)	Ret (w, w+3)	Ret (w, w+4)	Ret (w, w+5)	Ret (w, w+6)	Ret (w, w+7)
mroibflow (i, w)	0.0003***	0.0004***	0.0002***	0.0003***	0.0000	0.0001
	(-3.26)	(-3.85)	(-3.05)	(-2.60)	(-0.32)	(-1.01)
	2017-2022					
mroibflow (i, w)	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	(-0.71)	(-0.77)	(-0.73)	(-0.76)	(-1.14)	(-1.20)

Figure 1.

Figure 1 illustrates the portfolio returns sorted by weekly retail flow imbalance over the next seven weeks during three distinct periods: 2014-2016, 2017-2019, and 2020-2022. Each week, stocks are divided into deciles based on their weekly retail flow imbalance (*mroibflow*). We focus on the top decile (blue lines), representing stocks with the highest 10% of *mroibflow* and the bottom decile (orange lines), representing stocks with the lowest 10%. We report the equal-weighted weekly returns of these portfolios from Week 1 (W1) to week 7 (W7). All weekly variables are measured from Wednesday to the following Tuesday. All returns are expressed in percentages. Observations during the Covid-19 shock period (Q1 2020) are excluded from the sample.

Table 3. Decomposition of retail trading and its prediction on weekly returns

Table 3 displays the predictability of retail flow imbalance quartiles on stock returns over the next five weeks during the periods 2014-2016 and 2017-2022. Stocks are divided into quartiles based on their weekly retail flow imbalance: bottom 25% (strongest selling), 25% - 50%, 50% - 75%, and top 25% (strongest buying). The independent variables are interaction terms between the retail flow imbalance and quartile dummy variables. The dependent variable is the weekly return, computed as the cumulative daily returns from Wednesday to the following Tuesday. Control variables include weekly returns and turnover, previous monthly returns and return volatility, and cumulative returns over the prior six months. Monthly return volatility is computed as the standard deviation of daily returns. Turnover is defined as total trading volume divided by total shares outstanding. Observations during the Covid-19 shock period (Q1 2020) are excluded from the sample. We estimate Fama-MacBeth (1973) regressions and report the time-series average of the regression coefficients. T-statistics are computed using Newey-West standard errors. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

Panel A. Prediction of retail flow imbalance quartiles on future stock returns during 2014-2016					
	Dependent variable: Ret (w, w+k)				
	w+1	w+2	w+3	w+4	w+5
mroibflow (i, w, B25%)	0.0011*	0.0006	0.0004	0.0000	0.0000
	(1.94)	(1.11)	(0.55)	(0.05)	(-0.07)
mroibflow (i, w, B25%-B50%)	0.0012	0.0011	0.0013	0.0004	0.0006
	(1.30)	(1.84)	(1.19)	(0.89)	(1.09)
mroibflow (i, w, B50%-B75%)	-0.0006	-0.0006	0.0003	0.0000	-0.0014
	(-0.44)	(-0.71)	(0.30)	(0.01)	(-1.34)
mroibflow (i, w, T25%)	0.0004***	0.0002*	0.0004***	0.0003**	0.0003***
	(4.64)	(1.86)	(3.23)	(2.36)	(2.64)
mroibflow (i, w)	0.0005***	0.0003***	0.0004***	0.0003***	0.0003***
	(6.60)	(3.26)	(3.85)	(3.05)	(2.60)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	272,766	270,459	268,920	267,348	265,093
Adjusted R-squared	0.0475	0.041	0.0384	0.038	0.0376

Panel B. Prediction of retail flow imbalance quartiles on future stock returns during 2017-2022					
	Dependent variable: Ret (w, w+k)				
	w+1	w+2	w+3	w+4	w+5
mroibflow (i, w, B25%)	0.0000	0.0001	0.0000	-0.0001	0.0000
	(0.03)	(0.20)	(0.01)	(-0.08)	(-0.01)
mroibflow (i, w, B25%-B50%)	-0.0012	-0.0013	0.0013	0.0014	0.0011
	(-0.52)	(-0.72)	(0.57)	(0.76)	(0.85)
mroibflow (i, w, B50%-B75%)	-0.0004	0.0000	0.0003	-0.0006	-0.0016
	(-0.26)	(0.02)	(0.22)	(-1.46)	(-2.02)
mroibflow (i, w, T25%)	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.65)	(0.61)	(0.70)	(0.58)	(0.70)

mroibflow (i, w)	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.85)	(0.71)	(0.77)	(0.73)	(0.76)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	585,498	582,300	579,318	576,274	573,275
Adjusted R-squared	0.0544	0.0499	0.0444	0.0455	0.044

Table 4. Prediction of retail trading and DGTW returns

Table 4 displays the predictability of retail flow imbalance quartiles on stock returns over the next five weeks during the periods 2014-2016 and 2017-2022. Stocks are divided into quartiles based on their weekly retail flow imbalance: bottom 25% (strongest selling), 25% - 50%, 50% - 75%, and top 25% (strongest buying). The independent variables are interaction terms between the retail flow imbalance and quartile dummy variables. The dependent variable is the weekly DGTW return, calculated as the cumulative daily return minus the cumulative daily return of the stock's value-weighted DGTW portfolio, where portfolios are sorted quarterly. Control variables include weekly returns and turnover, previous monthly returns and return volatility, and cumulative returns over the prior six months. Monthly return volatility is computed as the standard deviation of daily returns. Turnover is defined as total trading volume divided by total shares outstanding. All weekly variables are measured from Wednesday to the following Tuesday. Observations during the Covid-19 shock period (Q1 2020) are excluded from the sample. We estimate Fama-MacBeth (1973) regressions and report the time-series average of the regression coefficients. T-statistics are computed using Newey-West standard errors. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

Panel A. Prediction of retail flow imbalance quartiles on future stock DGTW returns during 2014-2016					
	Dependent variable: DGTW Ret (w, w+k)				
	w+1	w+2	w+3	w+4	w+5
mroibflow (i, w, B25%)	0.0014***	0.0006	0.0005	0.0002	0.0001
	(2.49)	(1.21)	(1.27)	(0.52)	(0.59)
mroibflow (i, w, B25%-B50%)	0.0010	0.0008	0.0001	0.0015	0.0008
	(1.07)	(0.85)	(0.07)	(1.79)	(1.21)
mroibflow (i, w, B50%-B75%)	0.0011	-0.0004	0.0022	-0.0001	-0.0011
	(0.70)	(-0.68)	(0.81)	(-0.06)	(-1.01)
mroibflow (i, w, T25%)	0.0004***	0.0001	0.0001	0.0001	0.0001
	(5.14)	(0.57)	(1.44)	(0.98)	(1.33)
mroibflow (i, w)	0.0004***	0.0001	0.0002*	0.0001	0.0002
	(6.33)	(1.50)	(1.69)	(1.48)	(1.30)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	274,609	272,686	270,769	270,210	268,073
Adjusted R-squared	0.0253	0.0292	0.0205	0.0258	0.0256

Panel B. Prediction of retail flow imbalance quartiles on future stock DGTW returns during 2017-2022					
	Dependent variable: DGTW Ret (w, w+k)				
	w+1	w+2	w+3	w+4	w+5
mroibflow (i, w, B25%)	0.0010	0.0003	0.0000	0.0001	0.0004
	(1.40)	(0.43)	(0.03)	(0.14)	(0.49)
mroibflow (i, w, B25%-B50%)	-0.002	-0.001	0.0016	0.0010	0.0023
	(-1.12)	(-0.58)	(0.68)	(0.44)	(1.10)
mroibflow (i, w, B50%-B75%)	0.0000	-0.0007	0.0000	-0.0009	-0.0015
	(-0.04)	(-0.47)	(0.01)	(-0.97)	(-1.98)
mroibflow (i, w, T25%)	0.0000	0.0001	0.0001	0.0002	0.0001

	(-0.19)	(0.59)	(1.25)	(1.36)	(0.57)
Cum_ Mroibflow (i, w)	0.0000	0.0001	0.0002	0.0001	0.0001
	(0.15)	(0.80)	(1.46)	(1.48)	(1.45)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	585,409	582,360	579,318	576,274	573,235
Adjusted R-squared	0.0262	0.0299	0.029	0.0394	0.0318

Table 5. Prediction of retail trading, size, and liquidity

Table 5 Panel A displays the predictability of the top 25% of retail flow imbalance on small, medium, and large stocks over the next five weeks during the periods 2014-2016 and 2017-2022. We classify stocks into three size categories based on quarterly DGTW size portfolio rankings. We then define three dummy variables S_1, S_2, and S_3. Small stocks ($S_1 = 1$) are those ranked in "1" or "2" in DGTW size portfolios, representing the bottom 40% in market capitalization. Medium stocks ($S_2 = 1$) are those in "3", representing the 40%-60%. Large stocks ($S_3 = 1$) are those in "4" or "5", representing the top 40%. We then construct three interaction terms between the top 25% of retail flow imbalance and the size dummy variables: Top 25% *mroibflow**Small, Top 25% *mroibflow**Medium, and Top 25% *mroibflow**Large. The remaining three interactions between *mroibflow* and its respective quartile dummies remain unchanged in the Fama-MacBeth (1973) regressions. The dependent variable is the weekly return controlled for book-to-market and momentum factors, calculated as the cumulative daily return minus the cumulative daily return of the stock's DGTW B/M-Momentum portfolio. Control variables include weekly returns and turnover, previous monthly returns and return volatility, and cumulative returns over the prior six months. Monthly return volatility is computed as the standard deviation of daily returns. Turnover is defined as total trading volume divided by total shares outstanding. All weekly variables are measured from Wednesday to the following Tuesday. Panel B reports time-series cross-sectional averages of illiquidity for small stocks and for all stocks in the sample during the periods 2014-2016 and 2017-2022. Following Amihud (2002), weekly illiquidity is computed as the average of the daily ratio of absolute return to (dollar) trading volume, multiplied by 10^6. The illiquidity measure is winsorized at the top and bottom 1% of the distribution. Panel C displays the predictability of the top 25% of retail flow imbalance on small, medium, and large stocks, conditional on liquidity levels, for the 2014-2016 period. Stocks are divided into high- and low-liquidity groups based on their quarterly illiquidity levels, and corresponding dummy variables are constructed. We generate six interaction terms by combining the Top 25% *mroibflow**Size category (Small, Medium, Large) with the liquidity dummy variables (High, Low). Control variables and the remaining three interactions about *mroibflow* are the same as those in Panel A. Observations during the Covid-19 shock period (Q1 2020) are excluded from the sample. T-statistics are computed using Newey-West standard errors. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

Panel A. Prediction of the top 25% of retail flow imbalance on stock returns conditional on firm size					
	Dependent variable: Ret-B/M-MOM (w, w+k)				
	2014-2016				
	w+1	w+2	w+3	w+4	w+5
T25% *mroibflow* *Small	0.0005***	0.0002*	0.0004***	0.0002***	0.0002***
	(4.65)	(1.95)	(3.34)	(2.55)	(2.72)
T25% *mroibflow* *Medium	0.0001	-0.0005	0.0000	-0.0004	-0.0004
	(0.37)	(-1.90)	(0.04)	(-1.44)	(-1.16)
T25% *mroibflow* *Large	-0.0004	-0.0002	-0.0003	-0.0005	-0.0003
	(-0.73)	(-0.34)	(-0.67)	(-1.50)	(-0.70)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	274,598	273,203	272,150	270,465	269,736
Adjusted R-squared	0.0355	0.0397	0.0365	0.0366	0.0383
	2017-2022				
	w+1	w+2	w+3	w+4	w+5
T25% *mroibflow* *Small	0.0000	0.0001	0.0009	0.0001	0.0001
	(-0.25)	(0.43)	(0.92)	(0.69)	(0.68)
T25% *mroibflow* *Medium	0.0002	0.0002	-0.0017	-0.0001	0.0001
	(0.49)	(0.57)	(-1.63)	(-0.33)	(0.34)

T25% *mroibflow* *Large	0.0000	0.0004	-0.0002	-0.0001	0.0003
	(0.01)	(0.63)	(-0.07)	(-0.12)	(0.42)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	581,090	578,041	574,998	571,959	568,921
Adjusted R-squared	0.0387	0.0427	0.0417	0.0415	0.0408

Panel B. Summary statistics of illiquidity						
	2014-2016					
	Number of periods	Mean	Median	Std Dev	Minimum	Maximum
Small stock	156	120.9759	115.8655	60.5158	34.9240	371.9393
Total	307	46.5064	40.5910	22.5925	14.0249	139.559
	2017-2022					
Small stock	156	51.8680	46.4767	36.0235	6.7032	179.3907
Total	307	20.5962	18.3502	14.2853	2.6594	71.9191

Panel C. Prediction of the top 25% of retail flow imbalance on stock returns conditional on size and liquidity during 2014-2016					
	Dependent variable: Ret-B/M-MOM (w, w+k)				
	w+1	w+2	w+3	w+4	w+5
T25% *mroibflow* *Small *High	-0.0001	0.0001	-0.0007	-0.0006	-0.0001
	(-0.06)	(0.53)	(-1.18)	(-1.59)	(-0.18)
T25% *mroibflow* *Small *Low	0.0004***	0.0002**	0.0004***	0.0003**	0.0003**
	(4.47)	(1.99)	(3.29)	(2.26)	(2.54)
T25% *mroibflow* *Medium *High	-0.0002	-0.0011	-0.0001	-0.0004	0.0000
	(-0.35)	(-2.21)	(-0.26)	(-1.37)	(-0.09)
T25% *mroibflow* *Medium *Low	0.0003	-0.0005	0.0000	0.0001	-0.0004
	(0.93)	(-0.94)	(0.04)	(0.11)	(-1.08)
T25% *mroibflow* flow *Large *High	-0.0005	-0.0004	-0.0004	-0.0003	-0.0005
	(-0.90)	(-1.02)	(-0.84)	(-0.85)	(-1.39)
T25% *mroibflow* *Large *Low	-0.0003	-0.0003	-0.0005	0.0000	0.0006
	(-1.23)	(-0.57)	(-1.04)	(-0.03)	(0.83)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	272,568	271,615	271,585	266,622	262,019
Adjusted R-squared	0.0349	0.0416	0.0378	0.0385	0.0386

Table 6. Prediction of retail trading and the book-to-market ratio

Table 6 displays the predictability of the top 25% of retail flow imbalance on growth, neutral, and value stocks over the next five weeks during the periods 2014-2016 and 2017-2022. We classify stocks into three categories based on quarterly DGTW book-to-market portfolio rankings. We then define three dummy variables B_1, B_2, and B_3. Growth stocks ($B_1 = 1$) are those ranked in "1" or "2" in DGTW B/M portfolios, representing the bottom 40% in B/M. Neutral stocks ($B_2 = 1$) are those in "3", representing the 40%-60%. Value stocks ($B_3 = 1$) are those in "4" or "5", representing the top 40%. We create three interaction terms between the top 25% of retail flow imbalance and the B/M-based dummy variables: Top 25% *mroibflow**Growth, Top 25% *mroibflow**Neutral, and Top 25% *mroibflow**Value. The remaining three interactions between *mroibflow* and its respective quartile dummies remain unchanged in the Fama-MacBeth (1973) regressions. The dependent variable is the weekly return controlled for size and momentum factors, calculated as the cumulative daily return minus the cumulative daily return of the stock's DGTW Size-Momentum portfolio. Control variables include weekly returns and turnover, previous monthly returns and return volatility, and cumulative returns over the prior six months. Monthly return volatility is computed as the standard deviation of daily returns. Turnover is defined as total trading volume divided by total shares outstanding. All weekly variables are measured from Wednesday to the following Tuesday. Observations during the Covid-19 shock period (Q1 2020) are excluded from the sample. T-statistics are computed using Newey-West standard errors. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

Prediction of the top 25% of retail flow imbalance on stock returns conditional on the book-to-market ratio					
	Dependent variables: Ret-MOM-Size (w, w+k)				
	2014-2016				
	w+1	w+2	w+3	w+4	w+5
mroibflow *Growth	0.0001	0.0000	0.0001	0.0001	0.0000
	(0.95)	(-0.01)	(0.81)	(0.35)	(-0.12)
mroibflow *Neutral	0.0004**	0.0002	0.0003	0.0003	0.0001
	(2.20)	(1.26)	(1.65)	(1.41)	(1.39)
mroibflow *Value	0.0006***	0.0003*	0.0004**	0.0002*	0.0002*
	(4.27)	(1.89)	(1.99)	(1.93)	(1.71)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	274,598	274,277	273,758	272,199	269,032
Adjusted R-squared	0.0312	0.0354	0.0333	0.0323	0.0329
	2017-2022				
	w+1	w+2	w+3	w+4	w+5
mroibflow *Growth	0.0002	-0.0002	-0.0001	0.0001	-0.0001
	(1.23)	(-0.61)	(-0.30)	(0.27)	(-0.72)
mroibflow *Neutral	-0.0001	-0.0001	0.0000	0.0001	-0.0001
	(-0.65)	(-0.97)	(-0.16)	(0.54)	(-0.72)
mroibflow *Value	0.0004**	0.0006***	0.0006**	0.0005**	0.0003*
	(2.28)	(2.68)	(1.98)	(2.26)	(1.75)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	665,734	662,645	659,723	656,568	653,727
Adjusted R-squared	0.0392	0.0477	0.0342	0.0352	0.0346

Table 7. Prediction of retail trading and stock momentum

Table 7 displays the predictability of the top 25% of retail flow imbalance on losing, neutral, and winning stocks over the next five weeks during the periods 2014-2016 and 2017-2022. We classify stocks into three categories based on quarterly DGTW momentum portfolio rankings. We then define three dummy variables U_1, U_2, and U_3. Losing stocks ($U_1 = 1$) are those ranked in "1" or "2" in DGTW momentum portfolios, representing the bottom 40% in past 12-month performance. Neutral stocks ($U_2 = 1$) are those in "3", representing the 40%-60%. Winning stocks ($U_3 = 1$) are those in "4" or "5", representing the top 40%. We create three interactions terms between the top 25% of retail flow imbalance and the momentum-based dummy variables: Top 25% *mroibflow**Losing, Top 25% *mroibflow**Neutral, and Top 25% *mroibflow**Winning. The remaining three interactions between *mroibflow* and its respective quartile dummies remain unchanged in the Fama-MacBeth (1973) regressions. The dependent variable is the weekly return controlled for size and momentum, calculated as the cumulative daily return minus the cumulative daily return of the stock's DGTW Size-B/M portfolio. Control variables include weekly returns and turnover, previous monthly returns and return volatility, and cumulative returns over the prior six months. Monthly return volatility is computed as the standard deviation of daily returns. Turnover is defined as total trading volume divided by total shares outstanding. All weekly variables are measured from Wednesday to the following Tuesday. Observations during the Covid-19 shock period (Q1 2020) are excluded from the sample. T-statistics are computed using Newey-West standard errors. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

Prediction of the top 25% of retail flow imbalance on stock returns conditional on past 12-month performance					
	Dependent variables: Ret-B/M-Size (w, w+k)				
	2014-2016				
	w+1	w+2	w+3	w+4	w+5
mroibflow *Losing	0.0004**	0.0001	0.0005	0.0000	0.0002
	(2.42)	(0.44)	(1.31)	(0.20)	(1.06)
mroibflow *Neutral	0.0004**	0.0001	0.0000	0.0002	0.0002
	(2.05)	(0.65)	(0.23)	(1.02)	(1.20)
mroibflow *Winning	0.0003**	0.0001	0.0001	0.0000	0.0001
	(2.39)	(0.93)	(0.96)	(-0.17)	(0.61)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	266,721	264,721	263,581	262,943	260,979
Adjusted R-squared	0.0341	0.0377	0.0357	0.0349	0.0357
	2017-2022				
	W+1	W+2	W+3	W+4	W+5
mroibflow *Losing	0.0002	0.0001	0.0004	0.0002	0.0003
	(0.75)	(0.49)	(1.01)	(1.09)	(1.28)
mroibflow *Neutral	-0.0003	0.0001	-0.0001	0.0002	0.0001
	(-1.48)	(0.29)	(-0.53)	(0.64)	(0.52)
mroibflow *Winning	0.0000	-0.0001	-0.0001	-0.0001	0.0001
	(0.31)	(-0.82)	(-0.49)	(-0.65)	(0.30)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	665,738	662,736	659,744	656,758	653,077
Adjusted R-squared	0.0352	0.0375	0.0385	0.0385	0.0389

Table 8. Predictability of retail trading on small stocks with low liquidity

Table 8 displays the predictability of retail flow imbalance on stock returns conditional on stock types over the next five weeks during the 2014-2016 period. We construct two dummy variables: SL and Other. SL equals one if a stock is a small-cap stock with low liquidity, determined by quarterly DGTW size portfolios and the Amihud (2002) illiquidity measure. Other equals one for the remaining stocks. We first interact weekly retail flow imbalance with each of the two stock-type dummy variables (SL and Other). The dependent variable is the weekly return controlled for the book-to-market factor, calculated as the cumulative daily return minus the cumulative daily return of the stock's DGTW B/M portfolio. We further decompose retail flow imbalance into quartiles: bottom 25% (strongest selling), 25% - 50%, 50% - 75%, and top 25% (strongest buying). The independent variables are eight interaction terms, constructed by combining quartile-based interactions of *mroibflow* with the SL and Other dummy variables. Control variables include weekly returns and turnover, previous monthly returns and return volatility, and cumulative returns over the prior six months. Monthly return volatility is computed as the standard deviation of daily returns. Turnover is defined as total trading volume divided by total shares outstanding. All weekly variables are measured from Wednesday to the following Tuesday. We estimate Fama-MacBeth (1973) regressions and report the time-series average of the regression coefficients. T-statistics are computed using Newey-West standard errors. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

Prediction of retail flow imbalance on stock returns conditional on size and liquidity during 2014-2016					
	Dependent variable: Ret-B/M (w, w+k)				
Independent variables	w+1	w+2	w+3	w+4	w+5
mroibflow (i, w, B25%) *SL	0.0016**	0.0006	0.0000	0.0003	-0.0007
	(2.39)	(0.89)	(0.03)	(0.36)	(-1.22)
mroibflow (i, w, B25%) *Others	0.0003*	0.0011	0.0004	0.0005	0.0012
	(1.89)	(1.34)	(0.76)	(1.27)	(1.29)
mroibflow (i, w, B25%-B50%) *SL	-0.0003	0.0001	0.0004	-0.0004	0.0002
	(-0.17)	(0.11)	(0.23)	(-0.23)	(0.12)
mroibflow (i, w, B25%-B50%) *Others	0.0003	0.0007	0.0016	0.0003	0.001
	(1.48)	(1.84)	(1.11)	(1.88)	(1.29)
mroibflow (i, w, B50%-B75%) *SL	0.0003	0.0018	0.0023	0.0018	0.0017
	(0.99)	(0.80)	(1.39)	(0.35)	(0.95)
mroibflow (i, w, B50%-B75%) *Others	-0.0006	-0.001	-0.0021	-0.0022	-0.0024
	(-0.34)	(-0.92)	(-1.56)	(-1.65)	(-1.59)
mroibflow (i, w, T25%) *SL	0.0004***	0.0002***	0.0004***	0.0002**	0.0002**
	(4.50)	(2.66)	(3.12)	(2.45)	(2.27)
mroibflow (i, w, T25%) *Others	0.0000	-0.0001	-0.0001	-0.0004	-0.0001
	(0.24)	(-0.59)	(-0.46)	(-1.70)	(-0.20)
mroibflow (i, w) *SL	0.0005***	0.0002***	0.0004***	0.0002***	0.0002**
	(5.95)	(3.12)	(3.43)	(2.74)	(2.12)
mroibflow (i, w) *Others	0.0002	0.0001	0.0000	0.0000	0.0001
	(1.56)	(0.49)	(0.27)	(0.00)	(0.29)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	273,101	272,963	272,758	267,761	262,568
Adjusted R-squared	0.0392	0.0460	0.0420	0.0417	0.0429

Table 9. Retail flow imbalance and future weekly stock returns.

Table 9 displays the time-series cross-sectional averages of portfolio returns over the next five weeks sorted by retail flow imbalance during the period of 2014-2016. We divide stocks into two groups: small stocks with low liquidity (SL) and other stocks using quarterly DGTW size portfolios and Amihuid (2002) illiquidity measure. Each week, both groups of stocks are split into deciles based on weekly retail flow imbalance. Portfolio P10 comprises stocks experiencing the highest retail buying while the lowest selling. Portfolio P1 comprises stocks with the highest retail selling while the lowest buying. The P10-P1 spread represents a zero-investment strategy. The return difference is defined as the return from going long on the zero-investment strategy among SL stocks and short on the zero-investment strategy among other stocks. Ret-B/M is defined as the weekly return controlled for book-to-market factor, calculated as the stock return minus the return of its DGTW B/M portfolio. Excess return is computed as the stock returns minus the risk-free rate retrieved from French's data library. Weekly returns are calculated as the cumulative daily returns from Wednesday to the following Tuesday. Observations during the Covid-19 shock period (Q1 2020) are excluded from the sample. All returns are expressed in percentages. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

Future weekly stock returns sorted on retail flow imbalance during 2014-2016								
	Small stock with low liquidity			Other stocks			Return	Number
Weekly returns	P10	P1	P10-P1	P10	P1	P10-P1	difference	of periods
Ret-B/M (w, w+1)	0.3115***	-0.1376	0.4491***	0.0751	-0.0480	0.1231	0.3260***	156
T-stats.	(2.89)	(-1.21)	(5.79)	(0.76)	(-0.05)	(1.56)	(3.53)	156
Excess return (w, w+1)	0.3158***	-0.1211	0.4369***	0.0930	-0.0224	0.1154	0.3215***	156
T-stats.	(2.68)	(-1.08)	(5.64)	(0.93)	(-0.23)	(1.61)	(3.50)	156
Ret-B/M (w, w+2)	0.1955*	0.0298	0.1657**	0.1129	0.0232	0.0897	0.0760	156
T-stats.	(1.85)	(0.39)	(2.21)	(1.08)	(0.24)	(1.02)	(0.68)	156
Excess return (w, w+2)	0.1851*	0.0374	0.1477**	0.1272	0.0253	0.1019	0.0458	156
T-stats.	(1.74)	(0.33)	(2.00)	(1.29)	(0.25)	(1.16)	(0.43)	156
Ret-B/M (w, w+3)	0.2514**	0.0084	0.2431***	0.0346	0.0165	0.0181	0.2250**	156
T-stats.	(2.33)	(0.08)	(3.39)	(0.38)	(0.18)	(0.25)	(2.29)	156
Excess return (w, w+3)	0.2636**	0.0160	0.2476***	0.0527	0.0161	0.0367	0.2109**	156
T-stats.	(2.47)	(0.16)	(3.21)	(0.57)	(0.17)	(0.50)	(2.18)	156
Ret-B/M (w, w+4)	0.1791*	0.0400	0.1391*	-0.0224	-0.0441	0.0217	0.1174	156
T-stats.	(1.73)	(0.45)	(1.88)	(-0.27)	(-0.46)	(0.31)	(1.28)	156
Excess return (w, w+4)	0.1871*	0.0443	0.1428**	-0.0137	-0.0435	0.0298	0.113	156
T-stats.	(1.89)	(0.53)	(1.98)	(-0.16)	(-0.44)	(0.41)	(1.24)	156
Ret-B/M (w, w+5)	0.1262	-0.0324	0.1586**	0.0790	0.0120	0.0670	0.0916	156
T-stats.	(1.17)	(-0.39)	(2.13)	(0.96)	(0.12)	(0.87)	(1.03)	156
Excess return (w, w+5)	0.1121	-0.0443	0.1564**	0.1089	0.0130	0.0960	0.0604	156
T-stats.	(0.98)	(-0.41)	(2.01)	(1.28)	(0.13)	(1.22)	(0.68)	156

Table 10. Prediction of retail trading and social media

Table 10 displays the summary statistics of social media variables derived from Twitter and the predictability of retail trading conditional on tweet volume and sentiment during the periods 2015-2016 and 2017-2022. Panel A reports the summary statistics of weekly tweet publication counts and average sentiment. Weekly variable is computed as the means of daily variables from Wednesday to the following Tuesday. A positive (negative) sentiment indicates a predominance of positive (negative) tone in tweets. Panel B reports the summary statistics of weekly tweet counts and average sentiment across tweet volume and sentiment groups. We divide weekly tweet sentiment into quartiles: the bottom 25% (most negative), 25% - 50% (negative), 50% - 75% (positive), and the top 25% (most positive). We first divide stocks into deciles based on sizes. Within each size decile, we further classify stocks into two groups based on the percentiles of their tweet publication counts. Stocks with tweet counts above the median (>50%) are classified as high tweet volume, while those below the median (<50%) are classified as low tweet volume. We also report the differences of average tweet volume and sentiment between the 2015-2016 and 2017-2022 periods. Panel C displays the predictability of retail flow imbalance on next- week returns across stock types and social media conditions. In the first two columns, the independent variable are interaction terms by combining *mroibflow* and the two stock-type dummies (SL and Other) with the two tweet volume groups (SM>50% or SM<50%). In the last two columns, the independent variable are interaction terms by combining *mroibflow* and two stock-type dummies with the two tweet sentiment groups (SM>0 or SM<0). The dependent variable is the weekly return controlled for book-to-market factor, calculated as the cumulative daily return minus the cumulative daily return of its DGTW B/M portfolio. Control variables include weekly returns and turnover, previous monthly returns and return volatility, and cumulative returns over the prior six months. Monthly return volatility is computed as the standard deviation of daily returns. Turnover is defined as total trading volume divided by total shares outstanding. All weekly variables are measured from Wednesday to the following Tuesday. T-statistics are computed using Newey-West standard errors. Panel D displays the predictability of retail flow imbalance on next-week returns across tweet volume and sentiment levels. The independent variables are eight interaction terms formed by combining *mroibflow* with sentiment quartiles and the two tweet volume groups. Control variables are the same as those in Panel C. Observations during the Covid-19 shock period (Q1 2020) are excluded from the sample. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively.

Panel A. Summary statistics of tweets publication counts and sentiments								
				Tweets publication counts				
Period	Mean	Median	Std Dev	Minimum	Maximum	Obs.		
2015-2016	64.7447	16.0000	418.6792	0.0000	45,562.8000	178,400		
2017-2022	56.8284	14.8000	387.1354	0.0000	83,405.2000	727,472		
				Tweets sentiment				
Period	Mean	Median	Std Dev	Minimum	Maximum	% of positive obs.	% of negative obs.	Obs.
2015-2016	0.0409	0.0126	0.1361	-1.0000	1.0000	70.5542%	29.4458%	176,680
2017-2022	0.0267	0.0005	0.1376	-1.0000	1.0000	63.0923%	36.9077%	722,091

Panel B. Weekly changes of Tweet counts and sentiments						
	2015-2016		2017-2022			
	Sentiment	Tweets counts	Sentiment	Tweets counts	Sentiment difference	Tweets count difference
Most negative &Top 50% of Tweets counts	-0.1072	137.501	-0.0902	125.22	-0.0170	12.2810

Most negative &Bottom 50% of Tweets counts	-0.1083	18.3625	-0.1168	15.3488	0.0085	3.0137
2nd quartile &Top 50% of Tweets counts	0.0124	210.919	-0.0002	209.3129	0.0126	1.6061
2nd quartile &Bottom 50% of Tweets counts	0.0158	20.2350	-0.0005	19.6842	0.0163	0.5508
3rd quartile &Top 50% of Tweets counts	0.0694	88.4755	0.0351	91.8363	0.0343	-3.3608
3rd quartile &Bottom 50% of Tweets counts	0.0705	18.7746	0.0386	18.3887	0.0319	0.3859
Most positive &Top 50% of Tweets counts	0.2306	69.0938	0.2005	63.8457	0.0301	5.2481
Most positive &Bottom 50% of Tweets counts	0.2219	20.665	0.2196	15.1801	0.0023	5.4849

Panel C. Prediction of retail flow imbalance on next week returns conditional on social media				
	Tweet volume		Tweet sentiment	
	2015-2016	2017-2022	2015-2016	2017-2022
	Dependent variable: Ret-B/M (w, w+1)			
mroibflow *SL*SM>50% (>0)	0.0006***	0.0005	0.0007***	0.0003
	(2.74)	(1.60)	(4.51)	(1.36)
mroibflow *SL*SM<50% (<0)	0.0005**	0.0004	0.0006**	-0.0001
	(2.19)	(1.12)	(2.07)	(-0.18)
mroibflow *Other*SM>50% (>0)	0.0002*	0.0001	0.0000	-0.0005
	(1.97)	(0.62)	(0.17)	(-0.88)
mroibflow *Other*SM<50% (<0)	0.0001	0.0000	0.0001	0.0001
	(1.23)	(-0.26)	(1.33)	(0.12)
mroibflow (i, w) *SM>50% (>0)	0.0006***	0.0004	0.0007***	0.0002
	(3.04)	(1.52)	(4.22)	(1.00)
mroibflow (i, w) *SM<50% (<0)	0.0005***	0.0000	0.0002	-0.0004
	(4.53)	(-0.43)	(1.20)	(-0.88)
Controls	Yes	Yes	Yes	Yes
Observations	153,821	442,312	153,195	434,787
Adjusted R-squared	0.0419	0.0365	0.0499	0.0338

Panel D. Prediction of retail flow imbalance on next week returns conditional on social media		
	2015-2016	2017-2022
	Ret (i, w+1)	Ret (i, w+1)
mroibflow *Most Negative*High	0.0003*	-0.0004
	(1.84)	(-0.70)
mroibflow *Most Negative*Low	-0.0002	-0.0002

	(-0.74)	(-0.68)
mroibflow *2nd quartile*High	0.0001	0.0002
	(0.42)	(0.53)
mroibflow *2nd quartile*Low	0.0002	-0.0002
	(0.66)	(-0.44)
mroibflow *3rd quartile*High	0.0007	0.0009
	(1.16)	(1.14)
mroibflow *3rd quartile*Low	0.0007***	0.0002
	(2.62)	(0.47)
mroibflow *Most Positive*High	0.0017***	-0.0009*
	(4.27)	(-1.96)
mroibflow *Most Positive*Low	0.0007***	0.0002
	(2.58)	(1.00)
Controls	Yes	Yes
Observations	129,430	468,126
Adjusted R-squared	0.0499	0.0488