Buying the Dip in Retail Trading

Xiao Yin and Dongchen Zou *

May 2026

Abstract

Retail investors have become a major force in U.S. equity markets. Their share of trading volume has tripled over the past decade, to about a third of the total. Using a comprehensive dataset that covers approximately 80 percent of actual U.S. retail trading, we document that retail investors buy the dip: retail flow rises significantly after price declines but is roughly flat after gains. The asymmetry holds across return horizons and is persistent across the sample period. We weigh possible explanations for the pattern and discuss its implications for the broader market.

JEL codes: G11, G12, G14

Keywords: retail trading, buying the dip, retail order flow

*Yin: University College London, xiao.yin@ucl.ac.uk; Zou: Indiana University, dongzou@iu.edu.
We thank seminar participants at the Chicago Booth RP Reunion conference, Deakin University, and Australian National University for helpful comments and suggestions. All errors are our own.

1 Introduction

Retail trading activity in U.S. equity markets has tripled over the past decade. Aggregate retail volume rose from less than 10 percent of total market volume in 2017 to over 30 percent in 2021 (Figure 1).[1] Since the meme-stock episode of early 2021, sustained public and academic attention has turned to retail trading. Yet the prior evidence is divided on a fundamental question: do retail investors lean against returns or chase them? How do retail investors trade in response to market movements, and what does that behavior imply for the rest of the market now that retail participation is substantial?

We answer these questions using a comprehensive dataset of daily retail order flows for individual U.S. equities. The series is constructed from actual retail transactions rather than inferred from microstructure signatures in public tape data. The dataset captures approximately 80 percent of total gross U.S. retail trading activity over the sample period from January 2017 through December 2024.

Using the data, we document a novel and robust stylized fact: retail investors systematically buy the dip. Retail order flow rises markedly following negative returns, but is roughly flat following positive returns. The asymmetric response is robust across return horizons. A piecewise specification that separates the slope of retail flow on positive and negative past returns yields loss-side coefficients roughly an order of magnitude larger in absolute value than their gain-side counterparts at every horizon from one month to twelve months. The pattern is also stable over time. In a tercile sort on the past twelve-month return, average retail flow into the bottom tercile exceeds that into the top tercile in 97 percent of the monthly sample, and the autocorrelation of bottom-tercile flow is 0.7. The behavior persists across the COVID-19 selloff and the meme-stock episode in 2021.

The pattern is specifically a feature of retail flow. The same exercise applied to four non-retail flow measures yields the opposite pattern. A large-trade institutional proxy, the change in 13-F institutional ownership, the change in lendable supply, and

[1] This trend is consistent with industry estimates. J.P. Morgan Asset Management (2025) reports a retail share of approximately 36 percent of U.S. equity market order flow by April 2025.

total initiated order flow all slope upward against past returns on the negative side, suggesting institutions are "selling the dip". Stocks that have fallen receive retail buying and institutional selling. Stocks that have risen receive institutional buying yet only modest retail flow.

The buy-the-dip pattern relates to two strands of prior evidence on retail trading but extends both. The first documents that retail order flow is contrarian against contemporaneous and short-horizon returns and predicts subsequent reversals (Kaniel et al., 2008; Kelley and Tetlock, 2013; Boehmer et al., 2021; Barrot et al., 2016; Luo et al., 2025). Our loss-side response aligns with that body of evidence. The difference is the asymmetric shape, which linear contrarian estimates obscure. The second strand documents that retail investors are concentrated in financially distressed stocks, typically with a jackpot or lottery interpretation in the spirit of Campbell et al. (2008), Conrad et al. (2014), and Barberis and Huang (2008). The conditioning variable in that pattern is a fundamental characteristic. Our conditioning variable is the realized past return, and the response is asymmetric across the sign of that return. We also find that lottery demand alone does not explain the dip-buying pattern.

Two features of buy-the-dip discipline the explanations: (1) the asymmetry across loss and gain states, and (2) the response to signed past returns at multiple horizons. We canvas five candidate explanations against these diagnostics. Attention frameworks predict retail responses to the absolute return rather than to its sign and so cannot generate the asymmetry (Barber and Odean, 2008; Barber et al., 2022). Habitat and lottery preferences describe stable cross-sectional retail preferences rather than return-conditional flow, and the conditional sign on idiosyncratic skewness in our regressions runs opposite the lottery prediction (Kumar, 2009; Dorn and Huberman, 2010; Bali et al., 2011). Extrapolation generates a past-return response at multiple horizons but, under standard symmetric mean-reversion beliefs, also predicts retail selling after gains, which we do not observe. The disposition effect predicts reluctance to sell losers, but retail selling rises rather than slackens after losses, so it does not account for the loss-side response (Shefrin and Statman, 1985;

Odean, 1998). The closest match to both diagnostics is absorption of institutional selling pressure: institutional constraints bind asymmetrically in down states, so the supply pressure that retail absorbs is itself state-contingent (Barrot et al., 2016; Coval and Stafford, 2007; Vayanos and Woolley, 2013). The opposing-slope evidence supports this reading, though it does not separate the channel from a third factor that could move both flows jointly. Each of these candidates may account for part of the pattern, but no single one explains the full asymmetry across horizons. We therefore treat them as suggestive rather than identifying.

We also discuss the implications of buying-the-dip by retail investors in the market. Because retail flow rises after losses, when the non-retail measures we examine move the other way, retail acts as a state-contingent liquidity provider that absorbs supply in down states. The market-quality reading is two-sided. When the price decline reflects transient institutional selling pressure, retail absorption cushions the dislocation and speeds the recovery. When the decline reflects adverse fundamental news, the same absorption slows the incorporation of that news into prices and lowers price informativeness in distressed states. Using flow data alone does not separate the two cases, and a descriptive look at portfolio returns cautions against reading retail dip-buying as profitable liquidity provision.

Related Literature. This paper relates to three strands. The first concerns the measurement of retail trading. Early work infers retail activity from trade size, treating small trades as retail and large trades as institutional (Lee and Radhakrishna, 2000; Hvidkjaer, 2008). Subsequent evidence shows that fixed trade-size cutoffs are fragile, because institutions split large orders and informed trading also appears in small trades (Campbell et al., 2009; Cready et al., 2014). Boehmer et al. (2021) introduce the BJZZ algorithm, which identifies retail trades from sub-penny price improvements granted by wholesalers. Barber et al. (2024) document signing errors in BJZZ and propose a modified measure, BHJOS. Battalio et al. (2023) show that sub-penny-flagged trades have low correlation with known retail imbalances. We contribute a retail flow measure built from actual retail executions and validate it directly against Citadel Securities statistics during the GameStop episode.

The second strand concerns retail investor behavior. Barber and Odean (2008) and Barber et al. (2022) document attention-driven retail trading. Kaniel et al. (2008) and Kelley and Tetlock (2013) document contrarian retail trading at short horizons. Dorn and Huberman (2010), Laarits and Sammon (2025), and Balasubramaniam et al. (2023) document stable cross-sectional retail preferences for particular firm characteristics. Mitton and Vorkink (2007), Kumar (2009), Han and Kumar (2013), and Bali et al. (2011) document retail tilts toward high idiosyncratic volatility and positive skewness. Campbell et al. (2008), Conrad et al. (2014), and Barberis and Huang (2008) document retail demand for financially distressed stocks under a jackpot or lottery interpretation. A related set of papers shows that retail risk-taking is state-contingent on variables other than the stock's own past return. Daniel et al. (2021) show that low interest rates push retail investors toward high-income assets, and Gelman et al. (2026) show that investors who underperform a salient benchmark raise portfolio risk while outperformers do not symmetrically de-risk. These papers establish that retail behavior is conditional and, in the second case, asymmetric, but the conditioning variable is the interest-rate environment or relative performance rather than the realized signed past return. We extend this literature by documenting a return-conditional retail response that is strongly asymmetric across loss and gain states and persistent at horizons up to twelve months, and by showing that the response is not pinned down by attention, by stable cross-sectional preferences, or by symmetric extrapolation.

The third strand concerns retail liquidity provision and price informativeness. Kaniel et al. (2008) and Barrot et al. (2016) study whether retail investors are compensated for providing liquidity to other market participants. A separate literature studies how the composition of demand shapes the informativeness of prices when informed capacity is constrained. Yuan (2005) and Glebkin et al. (2021) show that the equilibrium price function becomes more sensitive to noise in constrained, down states, and Dávila and Parlatore (2025) provide a methodology for measuring price informativeness. We connect to these strands by documenting that retail demand slopes against past returns while institutional and aggregate flows slope with them, and by drawing out what an asymmetric, state-contingent retail demand schedule implies for liquidity provision and for the informational content of prices in distressed states.

2 Data

2.1 Retail Flow

The primary data source is the S&P Retail Flow Database. This dataset provides daily retail buy and sell volume for individual U.S. equities from January 2017 through December 2024. Unlike the TAQ-based approaches described below, the S&P data are constructed from actual retail executions rather than inferred from patterns in market microstructure data. The dataset is accessible through the S&P Global Marketplace.

Coverage is broad. According to S&P Global, the database captures approximately 80 percent of total gross U.S. retail trading activity over the sample period. Figure 1 documents aggregate retail buy and sell volumes over time (Panel A), the retail share of total market volume (Panel B), and aggregate net retail order flow (Panel C). The figure shows that retail trading activity has roughly tripled over the sample. Retail volumes rose from less than 10 percent of market volume in 2017 to over 35 percent during the COVID-19 pandemic and the meme-stock episode of early 2021, receded from the peak, and stabilized at levels above the pre-2020 period.

This featured dataset addresses a tradeoff in the retail-trading literature between coverage and measurement. One strand uses proprietary brokerage records. These data are valuable because they observe actual investor trades, but they often come from one broker, one platform, or a narrow investor group. As a result, the same empirical design can speak to different retail populations: Robinhood users, traditional discount-brokerage customers, or clients of a particular international broker may trade for different reasons and have different market effects (Barber and Odean, 2008; Kumar and Lee, 2006; Barber et al., 2022; Eaton et al., 2022). A second strand uses public TAQ data to infer retail trades. These measures are easier to replicate, but they depend on microstructure signatures that can fail when wholesaler pricing, routing, or market conditions change (Boehmer et al., 2021; Battalio et al., 2023; Barber et al., 2024). The S&P data occupy the middle ground. They are based on actual retail executions, have broad market coverage, and are available to researchers through a commercial data platform. This makes them useful for studying aggregate retail behavior while preserving a common, replicable measurement standard.[2]

We supplement the retail flow data with Trade and Quote (TAQ) data from WRDS to construct benchmark flow measures. We merge in CRSP and Compustat for stock and firm characteristics. We also use securities lending data from S&P Global and 13-F institutional holdings from LSEG Thomson Reuters.

2.2 Variable Definitions and Sample Construction

Our main variable, *Retail Flow*, is monthly retail buy volume minus retail sell volume from S&P Global. For aggregate time-series figures, we report either dollar values in billions or percent of total market dollar volume. For stock-level analyses, we normalize the net flow by total shares outstanding.

We compare Retail Flow to five benchmark flow measures. Two are TAQ-based retail proxies. *BJZZ* is the sub-penny retail flow measure of Boehmer et al. (2021). A trade is classified as retail if it is an off-exchange or internalized execution and the execution price has a sub-penny component, $\mathrm{cf}(p) = 100 \times \mathrm{mod}(\mathrm{p}, 0.01) \notin \{0, 1\}$, that is, the cent fraction of the price is neither zero nor exactly half a cent. Trades that pass this identification filter are signed by the sub-penny pattern of the price. Trades with $\mathrm{cf}(p) \geq 0.6$ are classified as retail buys, trades with $\mathrm{cf}(p) \leq 0.4$ are classified as retail sells, and trades with $\mathrm{cf}(p) \in (0.4, 0.6)$ are excluded as ambiguous. Wholesalers internalize retail buys just below the round-cent ask (cent fraction near one) and retail sells just above the round-cent bid (cent fraction near zero).

Two limitations of the BJZZ measure have been documented in subsequent validation work. The first is identification: Battalio et al. (2023) compare BJZZ-flagged trades against a ground-truth sample of retail brokerage executions and report that BJZZ identifies fewer than one-third of actual retail trades and shows a low correlation with known retail imbalances. The second is signing. The cent-fraction rule aligns

[2]To our knowledge, this is the first paper that examines the S&P Retail Flow dataset.

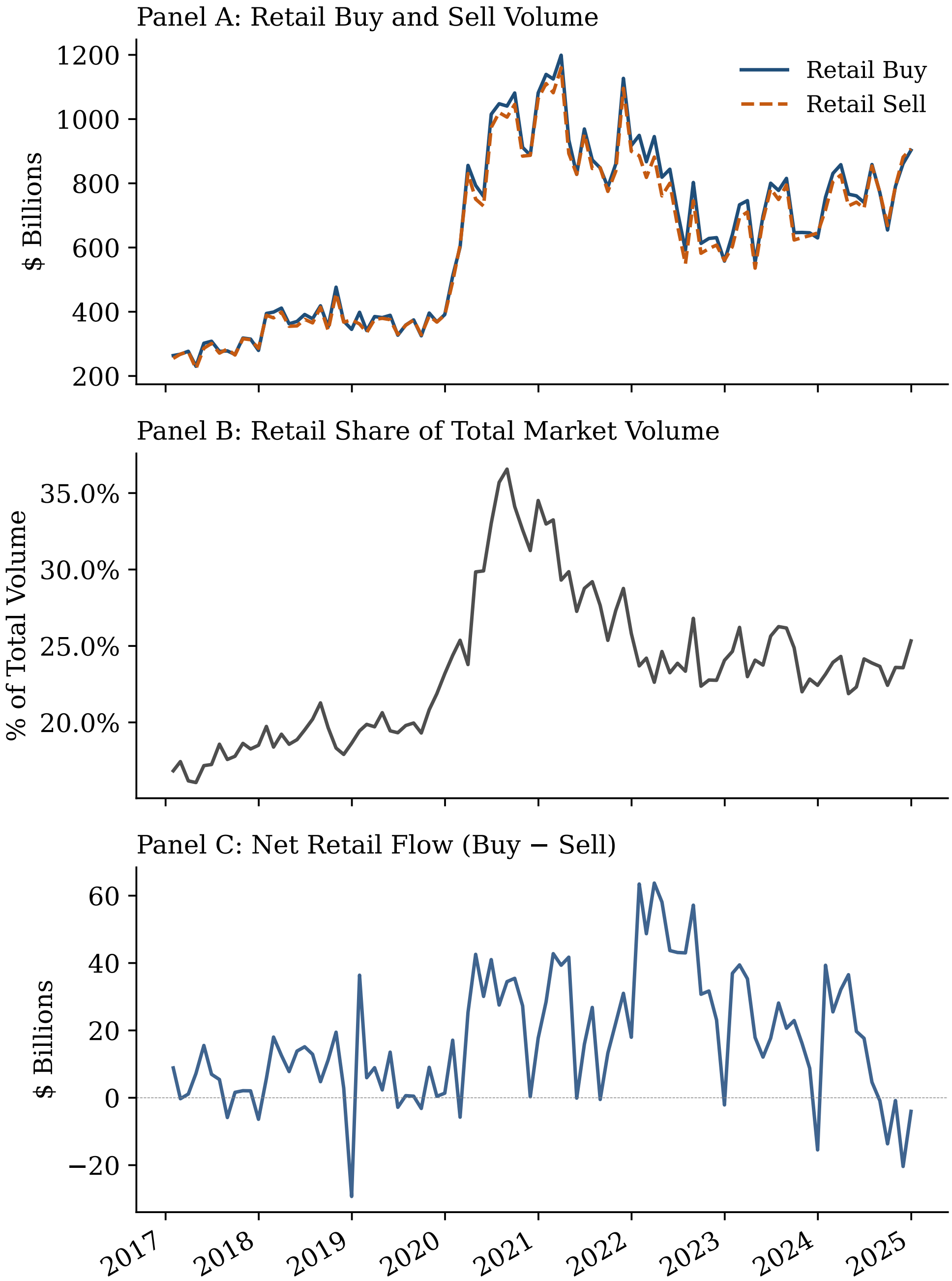

Figure 1. Retail Flows and Volumes

Notes: Panel A: aggregate monthly retail buy and sell volume ($ billions). Panel B: retail share of total market dollar volume (percent). Panel C: aggregate monthly net retail dollar order flow ($ billions), computed as buy volume minus sell volume. The sample covers January 2017 through December 2024. Retail flow is from the S&P Global Retail Flow Database.

with the position of the trade between the bid and the ask only under the one-cent spread regime. When the spread is wider than one cent, the cent fraction of the price no longer reflects whether the trade sits in the buyer half or the seller half of the spread. Barber et al. (2024) highlight this spread-width sensitivity. As a stylized illustration, suppose the bid is \$1.98 and the ask is \$2.00, so the spread is two cents and the midpoint is \$1.99. An execution at \$1.997 has cent fraction 0.7 and sits 0.3 cents below the offer, in the upper half of the spread. An execution at \$1.987 also has cent fraction 0.7 but sits 0.7 cents above the bid, in the lower half of the spread. The BJZZ rule signs both trades identically as retail buys because they share the same cent fraction, even though the second trade is closer to the bid than to the ask and is more naturally classified as a retail sell. Barber et al. (2024) document the resulting signing-error rate in a controlled experiment with roughly 85,000 retail trades of known direction. The BJZZ cent-fraction rule produces signing errors on approximately 28 percent of identified trades, while a position-relative-to-spread rule that uses the bid and the ask reduces signing errors to approximately 5 percent. Inferences about retail trading can be sensitive to these signing errors even when the identification step is held fixed.

BHJOS is the modified retail flow measure of Barber et al. (2024), which retains the BJZZ sub-penny identification filter but replaces the cent-fraction signing rule with a midpoint rule that explicitly uses the spread width. The trade-location statistic is $\text{loc} = (\text{price} - \text{nbb})/\text{spread}$, where nbb is the prevailing national best bid and spread is the prevailing NBBO spread. Trades with $\text{loc} > 0.6$ are classified as retail buys, trades with $\text{loc} < 0.4$ are classified as retail sells, and trades in $[0.4, 0.6]$ are excluded as ambiguous rather than forced into a buy or sell category. Returning to the stylized example above, the trade at \$1.997 has $\text{loc} = 0.85$ and is classified as a retail buy, while the trade at \$1.987 has $\text{loc} = 0.35$ and is reclassified from a BJZZ retail buy to a BHJOS retail sell. BHJOS also excludes trades with notional value above \$100,000 to remove probable institutional executions that occasionally satisfy the sub-penny identification filter.

The remaining measures capture non-retail or aggregate order flow. *Inst50k* is net flow in trades with notional value exceeding \$50,000. Trade-size rules were common in earlier work that inferred trader identity from transactions data (Lee and Radhakrishna, 2000; Hvidkjaer, 2008). Subsequent evidence shows that fixed trade-size cutoffs are fragile because institutions split orders and because small and large trades can both contain institutional information (Campbell et al., 2009; Cready et al., 2014). We therefore treat *Inst50k* as a benchmark, not as a definitive measure of institutional trading. *InstSec* is the monthly change in lendable supply divided by total shares outstanding, constructed from the securities-lending data. This measure serves as a proxy for institutional trades (Barardehi et al., 2024). ΔIO is the quarter-over-quarter change in the fraction of shares outstanding held by 13-F filers, carried forward within the quarter. *Total Flow* is total market net order flow.

We begin with all stocks in the Retail Flow Database that can be merged to CRSP using the PERMNO identifier. We aggregate daily flows to the stock-month level by summing within each calendar month. We then apply three filters. First, we require the month-end stock price to be at least one dollar to exclude penny stocks. Second, we require the absolute value of normalized net retail flow (flow divided by shares outstanding) to be no greater than one. Third, we require positive market equity. This removes observations with invalid or nonpositive price-share information after the CRSP-Compustat merge. After filtering, the matched sample contains 417,519 stock-month observations covering 7,432 distinct stocks and 96 calendar months from January 2017 through December 2024. The average number of stocks per month is 4,349.

2.3 Summary Statistics

Table 1 reports summary statistics for the matched sample. Panel A describes the flow variables. All flow measures are normalized by shares outstanding and expressed in percent. The institutional ownership variable (ΔIO) is the quarter-over-quarter change in institutional ownership, also expressed as percent of shares outstanding.

Retail net flow has a positive mean of 0.23 percent of shares outstanding per month and a median near zero (0.01 percent). The distribution is right-skewed, with a standard deviation of 2.33 percent and a 90th percentile of 0.72 percent. The near-zero median reflects the fact that retail investors are roughly as likely to be net sellers as net buyers in a given stock-month. The positive mean is consistent with a slight secular tendency toward retail accumulation over the sample period.

The TAQ-based retail flow proxies (BJZZ and BHJOS) have smaller absolute means and medians close to zero, which is expected because they measure signed imbalances in a specific sub-penny subset of trades rather than total retail activity. The institutional large-trade proxy ($Inst50k$) has a negative mean of -0.13 percent of shares outstanding, consistent with institutional investors being net sellers on average over this period. The securities lending proxy ($InstSec$) and ΔIO are both near zero on average.

Panel B describes return variables. The average monthly stock return is 1.19 percent, with a standard deviation of 22.75 percent. Return volatility as a stock characteristic is measured using daily returns over a 21-trading-day window. Panel C describes firm characteristics. The average stock in the sample has market equity of $9.4 billion, and the median is $743 million. The average book-to-market ratio is 0.73.

2.4 Validation with Citadel Securities Statistics

We validate the S&P Retail Flow data against public statistics released by Citadel Securities in January 2021. Citadel Securities is one of the largest wholesale market makers, which intermediates a substantial share of U.S. retail order flow via payment for order flow (PFOF) arrangements with retail brokerages. In the aftermath of the GameStop (GME) short squeeze and trading restrictions, Citadel Securities faced heightened congressional and public scrutiny over its role as an intermediary for retail order flow. To provide transparency about its intermediation activity during the episode, Citadel Securities published daily retail buy and sell volumes in shares for GME for January 25–28, 2021. These statistics were publicly available no later than January 29, 2021.

Table 2 reproduces the Citadel Securities statistics. Across the four trading days, retail investors routed 10 to 27 million shares per day through Citadel Securities, which intermediated 28.7 to 34.1 percent of total GME market volume. The episode is informative because the public narrative and the transaction data point in very different directions. Coordinated discussion on the Reddit forum and elsewhere urged retail investors to hold positions through the squeeze. Public commentary at the time treated retail demand as single-sided. The statistics show otherwise.

Table 1. Summary Statistics

Variable	N	Stocks	Mean	Std	p10	p25	p50	p75	p90
Panel A: Flow Variables (normalized × 100, percent of denominator)									
Retail Flow / Shares	417,519	7,432	0.232	2.325	−0.419	−0.126	0.010	0.207	0.715
Retail Flow / Mkt Cap	417,519	7,432	0.268	2.624	−0.413	−0.124	0.011	0.211	0.731
BHJOS Flow / Shares	417,519	7,432	0.045	1.379	−0.140	−0.051	−0.006	0.029	0.135
BJZZ Flow / Shares	417,519	7,432	0.045	1.601	−0.136	−0.043	−0.004	0.026	0.119
Inst50k Flow / Shares	417,519	7,432	−0.129	1.202	−0.561	−0.178	−0.001	0.074	0.296
InstSec Flow / Shares	417,519	7,432	0.149	1.337	−0.649	−0.151	0.023	0.390	1.028
Total Flow / Shares	417,519	7,432	−0.470	3.389	−1.543	−0.611	−0.140	0.161	0.650
ΔIO (qtrly, %)	388,635	7,236	0.004	36.045	−3.851	−1.293	0.102	1.802	5.024
Panel B: Return Variables									
Ret$[t]$	414,239	7,392	0.012	0.228	−0.161	−0.069	0.002	0.070	0.171
Ret$[t-2, t-1]$	406,946	7,342	0.013	0.233	−0.160	−0.068	0.002	0.071	0.172
Ret$[t-3, t-1]$	407,701	7,321	0.024	0.318	−0.225	−0.096	0.005	0.104	0.251
Ret$[t-6, t-1]$	398,213	7,229	0.054	0.507	−0.348	−0.153	0.012	0.176	0.416
Ret$[t-12, t-1]$	379,960	7,014	0.133	0.929	−0.488	−0.222	0.030	0.301	0.695
Idiosyncratic Volatility (21d)	413,237	7,402	0.027	0.033	0.008	0.012	0.020	0.033	0.051
Idiosyncratic Skewness (21d)	413,237	7,402	0.131	0.947	−0.888	−0.364	0.102	0.597	1.205
Panel C: Firm Characteristics									
Market Equity ($M)	417,519	7,432	9,374.9	59,941.8	50.7	175.3	743.1	3,596.2	15,069.6
Log Market Equity	417,519	7,432	6.703	2.189	3.925	5.167	6.611	8.188	9.620
Book-to-Market	371,684	6,498	0.728	1.313	0.103	0.231	0.492	0.875	1.399
Market Beta	320,045	5,687	1.197	0.714	0.417	0.752	1.120	1.548	2.074
Gross Profitability	384,238	6,550	2.088	554.985	0.000	0.043	0.194	0.380	0.604
ROA	391,408	6,677	−0.096	11.432	−0.432	−0.098	0.010	0.052	0.108
Debt-to-Asset	394,409	6,709	0.301	8.975	0.003	0.048	0.200	0.395	0.579
Firm Age (months)	417,519	7,432	280	237	41	85	238	396	617

Notes: The sample covers January 2017 through December 2024 ($N = 417{,}519$ stock-months, 7,432 stocks, 96 months). Filters: month-end price $\geq$ \$1; $|\text{flow/shrout}| \leq 1$; market equity > 0. Panel A: flow variables are monthly net order flow divided by shares outstanding (or by market equity for the Mkt Cap row), multiplied by 100 (percent of denominator). ΔIO is the quarter-over-quarter change in institutional ownership from 13-F filings, expressed as percent of shares outstanding and carried forward within the quarter. Panel B: Ret$[t-k, t-1]$ is the cumulative return from month $t-k$ to $t-1$; idiosyncratic volatility and idiosyncratic skewness are the standard deviation and skewness of the Fama-French three-factor residual returns over a 21-trading-day window. Panel C: market equity is in millions of dollars; firm age is in months since first CRSP appearance.

Net retail flow through Citadel Securities was modestly positive on January 25 when buyers slightly outnumbered sellers, turned decidedly negative on January 26 and 27 as the price continued to climb, and returned to near balance on January 28 when the price retreated. On the squeeze days, retail behavior did not match the buy-and-hold rhetoric: a material share sold into the price increase.

Figure 2 compares our three flow measures—S&P Retail Flow, BJZZ, and BHJOS—against the rescaled Citadel Securities statistics over the broader event window (January 22 through February 5, 2021). Citadel Securities states that it intermediates approximately 39 percent of U.S. retail trading volume, while the S&P Retail Flow data captures approximately 80 percent of gross retail activity. We

Table 2. Citadel Securities: GME Retail Flows, January 25–28, 2021

Date	Price ($)	Retail Buy	Retail Sell	Net (shares)	Mkt. Volume	Mkt. Share (%)	Net Flow (%)
2021-01-25	76.79	26,558,557	24,489,122	+2,069,435	177,874,000	28.7	+1.16
2021-01-26	147.98	24,888,375	26,794,942	−1,906,567	178,587,974	28.9	−1.07
2021-01-27	347.51	12,966,267	13,743,184	−776,917	93,396,666	28.6	−0.83
2021-01-28	193.60	9,972,227	10,078,110	−105,883	58,816,595	34.1	−0.18

Notes: Price is the daily closing price of GME from CRSP. Retail Buy and Retail Sell are the number of retail shares bought and sold through Citadel Securities on each day. Net is the difference (Buy − Sell). Mkt. Volume is total GME market volume in shares. Mkt. Share is the fraction of total GME market volume intermediated by Citadel Securities (percent). Net Flow is net retail shares divided by total market volume (percent). Source: Citadel Securities, published no later than January 29, 2021.

therefore apply a rescaling factor of two ($80/39 \approx 2.05$) to the Citadel series to make it comparable in magnitude to the S&P series. We note that "Mkt. Share (%)" in Table 2 refers to Citadel's share of *total* GME market volume on each day (28.7–34.1 percent), which is a different denominator from the firm-wide retail-trading share used for the rescaling.

The S&P Retail Flow series matches the Citadel Securities statistics closely in both sign and magnitude on January 25–27. Both series show modest net retail buying on January 25, and a shift to net selling on January 26–27. The agreement in direction and scale across three independent trading days provides direct event-level validation that the S&P database captures retail order-flow imbalances in a salient high-volume episode.[3]

The TAQ-based measures tell a very different story. Both BJZZ and BHJOS are inconsistent in sign and magnitude with the reported statistics across the event window. The BJZZ series stays near zero throughout the short squeeze, and does not track either the net buying on January 25 or the net selling on

[3]The one notable discrepancy between the Net Retail Flow series and the Citadel Securities statistics occurs on January 28, 2021. On that day, Robinhood, Webull, and several other retail brokerages placed GME into position-closing-only (PCO) status, meaning retail customers could sell but not open new buy positions. Citadel Securities receives a large share of its retail order flow from Robinhood via PFOF. As a result, Citadel's data for January 28 is disproportionately short of retail buy orders relative to the broader retail universe. The S&P Retail Flow Database aggregates order flow across brokerages, including other brokerage companies that did not restrict buying on that day. Accordingly, the S&P series shows more retail buying on January 28 than the Citadel statistics imply. This discrepancy is a direct consequence of broker-specific restrictions rather than a data quality issue. Outside of this specific episode, the close alignment between the two series provides direct validation that the S&P database accurately measures retail order flow.

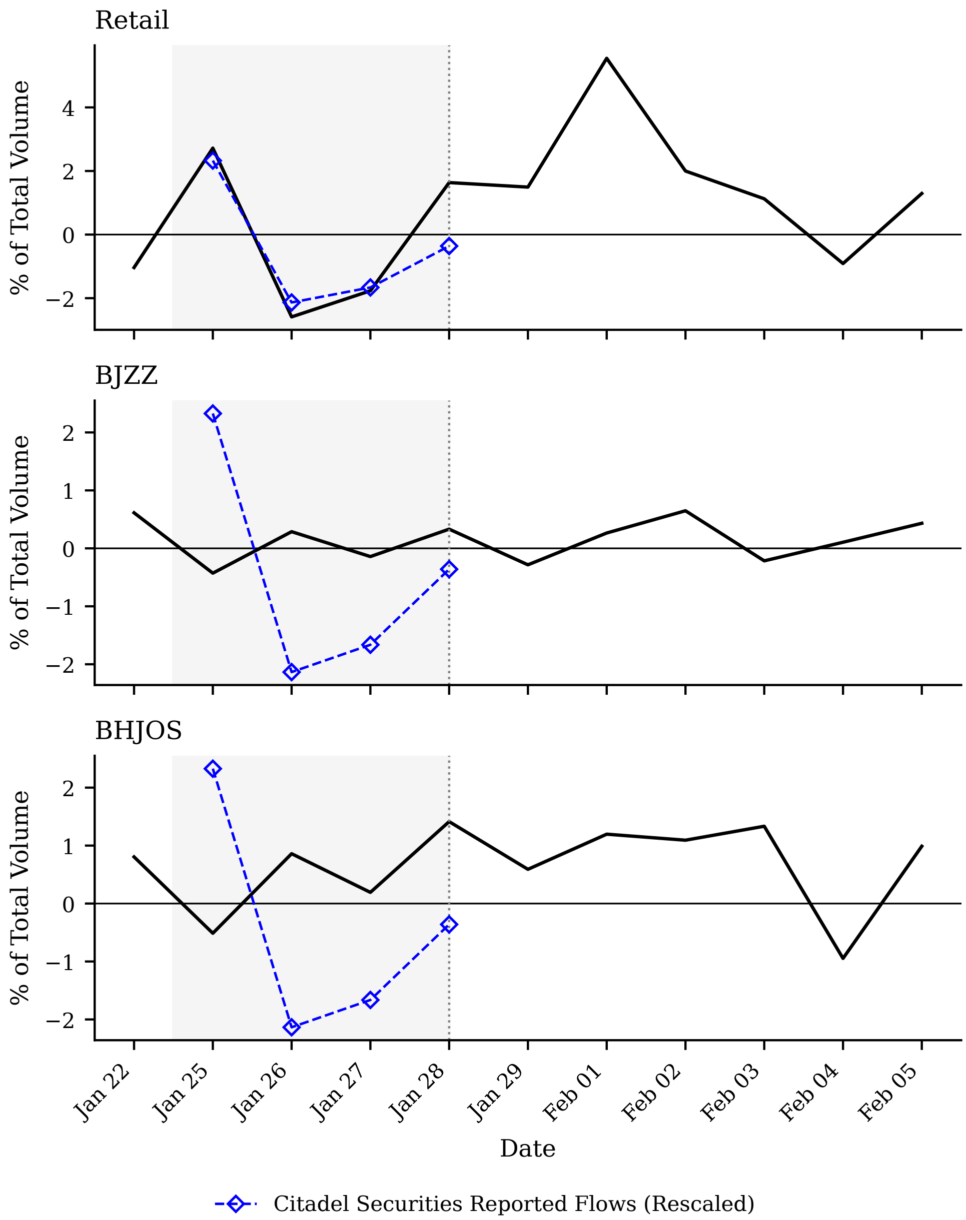

Figure 2. Net Retail Flows vs. Citadel Securities on GameStop

Notes: Each panel plots the daily net retail order flow as a percent of total GME market volume. *S&P Retail Flow*: actual retail flow from the S&P Global Retail Flow Database. *BJZZ*: sub-penny retail flow of Boehmer et al. (2021). *BHJOS*: modified retail flow of Barber et al. (2024). The dashed line with hollow markers shows the Citadel Securities statistics from Table 2, rescaled by a factor of 2 (≈ 80%/39%). The shaded region spans January 25–28, 2021. The vertical dotted line marks January 28, 2021, when several retail brokerages imposed position-closing-only restrictions.

January 26–27 that appear in both the S&P and the Citadel Securities data. The BHJOS series moves more than BJZZ, but its day-to-day signs also disagree with the statistics. After applying the same rescaling factor that aligns the S&P series, BHJOS sits at roughly one-half to one-third the size of the reported net flows. Both measures depend on a sub-penny identification step, which in turn depends on the wholesaler quoting and price-improvement patterns. Widening spreads and altered routing could render the sub-penny signature unreliable, especially during the meme-stock episode. The results are aligned with Battalio et al. (2023), who document that sub-penny-flagged trades show low correlation with known retail imbalances. These failures occur during the episode that attracted the most public attention to retail trading, underscoring why inferred proxies can lead to materially misleading inferences about retail behavior in the cases researchers care most about.

2.5 Comparison with TAQ-Inferred Measures

We compare the S&P Retail Flow series to the five flow measures defined above. Table 3 reports time-series and cross-sectional correlations between S&P Retail Flow and each measure. Time-series correlations aggregate monthly flows across stocks, scale by total market dollar volume, and compute the Pearson correlation of the resulting monthly series. Cross-sectional correlations are the average monthly Pearson correlation across stocks, with all flows normalized by shares outstanding. Panels A through D of Table 3 report results for the full sample, by market capitalization quintile, by momentum quintile, and by sub-period, respectively.

In the full-sample time series (Panel A), retail flow is nearly uncorrelated with BJZZ ($\rho = 0.017$). The sub-penny identification algorithm fails to capture the aggregate timing of retail activity. The BHJOS modification improves this substantially ($\rho = 0.347$), suggesting that the signing refinement recovers information about aggregate retail direction that BJZZ discards. Retail flow is essentially uncorrelated with total market flow in the time series ($\rho = -0.006$), confirming that retail and non-retail order flows are driven by distinct aggregate forces. The negative correlation with ΔIO ($\rho = -0.274$) is consistent with retail investors accumulating shares when institutional investors are reducing their positions at the aggregate level.

Cross-sectional correlations are significantly higher than time-series correlations for both TAQ proxies. In the full sample (Panel A), the average cross-sectional correlation between Retail Flow and BJZZ is 0.478; for BHJOS it is 0.508. These values

Table 3. Correlations of Retail Flow with Other Flow Measures

	Time-Series Correlation						Cross-Sectional Correlation					
	BJZZ	BHJOS	Inst50k	InstSec	ΔIO	Total	BJZZ	BHJOS	Inst50k	InstSec	ΔIO	Total
Panel A: Full Sample												
Full sample	0.017	0.347	−0.092	−0.016	−0.274	−0.006	0.478	0.508	−0.009	0.053	−0.037	−0.092
Panel B: By Market Cap Quintile												
Q1 (Small)	0.272	0.306	−0.336	−0.033	−0.174	−0.164	0.493	0.511	−0.026	0.115	−0.064	−0.115
Q2	0.410	0.456	−0.399	0.056	0.221	−0.448	0.442	0.511	0.028	0.043	−0.058	0.065
Q3	0.216	0.537	−0.449	0.037	−0.036	−0.101	0.346	0.446	−0.015	0.081	−0.028	0.069
Q4	0.062	0.328	0.063	−0.051	−0.336	0.077	0.269	0.446	−0.011	0.048	−0.015	0.076
Q5 (Large)	−0.040	0.302	−0.028	−0.110	−0.110	0.028	0.156	0.455	−0.027	0.053	−0.005	0.053
Panel C: By Momentum Quintile (prior 12-to-1 month return)												
Q1 (Losers)	0.280	0.374	−0.487	0.039	−0.433	−0.335	0.509	0.537	−0.026	0.067	−0.079	−0.133
Q2	0.394	0.483	0.124	−0.249	−0.106	0.421	0.340	0.405	0.014	0.026	−0.049	−0.003
Q3	0.553	0.622	−0.116	−0.139	−0.193	−0.039	0.386	0.426	0.018	0.038	−0.029	0.059
Q4	0.554	0.516	−0.011	−0.144	−0.231	0.109	0.294	0.379	0.019	0.049	−0.031	0.007
Q5 (Winners)	0.102	0.446	−0.667	0.501	−0.237	−0.403	0.351	0.460	−0.016	0.066	−0.014	0.062
Panel D: By Sub-Period												
Pre-COVID (2017–19)	0.597	0.783	0.200	−0.104	−0.089	0.064	0.560	0.626	−0.006	0.026	−0.032	−0.050
COVID/Meme (2020–21)	0.497	0.449	0.375	0.244	−0.231	0.375	0.466	0.475	−0.020	0.070	−0.042	−0.030
Post-meme (2022–24)	−0.075	0.507	−0.803	−0.114	−0.277	−0.074	0.403	0.411	−0.004	0.069	−0.038	−0.176

Notes: This table reports time-series (TS) and average monthly cross-sectional (CS) correlations between S&P Retail net order flow and five alternative flow measures. *BJZZ*: sub-penny retail flow of Boehmer et al. (2021). *BHJOS*: modified retail flow of Barber et al. (2024). *Inst50k*: net flow in trades with notional value above \$50,000, following the trade-size proxy tradition but subject to known classification concerns (Lee and Radhakrishna, 2000; Hvidkjaer, 2008; Campbell et al., 2009; Cready et al., 2014). *InstSec*: monthly change in lendable supply divided by shares outstanding (Barardehi et al., 2024). Δ*IO*: quarter-over-quarter change in 13-F institutional ownership fraction (carried forward within the quarter). *Total*: total market net order flow. TS correlations use monthly flows aggregated across stocks and scaled by total market dollar volume. CS correlations are averaged monthly Pearson correlations with flows normalized by shares outstanding. Panel A uses the full matched sample ($N = 417{,}519$ stock-months, January 2017–December 2024). Panels B and C sort on log market equity and prior 12-to-1 month return, respectively (Q1 = smallest/lowest). Panel D splits by sub-period: Pre-COVID (Jan 2017–Dec 2019), COVID/Meme (Jan 2020–Dec 2021), Post-meme (Jan 2022–Dec 2024).

confirm that the TAQ proxies capture meaningful stock-level variation in retail activity even when they miss aggregate timing. As a benchmark, Barber et al. (2024) report an R^2 of approximately 35 percent between BJZZ-identified flow and actual retail flow in their controlled experiment, implying a correlation of roughly 0.59. Our sample cross-sectional correlation of 0.478 falls somewhat below this level, in part because the S&P database draws from a broader population of retail trades and brokerages than the controlled sample in that study.

When do TAQ-based retail flow proxies work better? The conditional correlations in Panels B–D reveal systematic heterogeneity in proxy quality. Panel B shows that cross-sectional correlations are higher among small stocks ($\rho_{\text{BJZZ}} = 0.493$ for the smallest quintile) than among large stocks ($\rho_{\text{BJZZ}} = 0.156$ for the largest quintile). The BHJOS modification largely closes this size gap, with correlations ranging from 0.511 (Q1) to 0.455 (Q5). The size pattern reflects the fact that sub-penny price improvement is a less reliable marker of retail activity for large, liquid stocks where wholesaler quoting practices differ from those in smaller stocks.

Panel C shows that cross-sectional correlations are highest among past losers. For stocks in the bottom momentum quintile (Q1), the BJZZ correlation is 0.509, compared to 0.351 for past winners (Q5). This pattern has a direct implication for this paper's main analysis. The dip-buying behavior we document is concentrated precisely among past-loser stocks. The higher proxy reliability in that subsample means that evidence in the prior literature based on BJZZ is most credible when studying retail contrarian activity, even if overall alignment has deteriorated in recent periods.

Panel D shows that time-series (cross-sectional) correlations between Retail Flow and BJZZ declined from 0.597 (0.560) in the pre-COVID period from 2017–2019, to 0.497 (0.466) during the COVID and meme-stock period from 2020–2021, and to -0.075 (0.403) in the post-meme period from 2022–2024. BHJOS shows a similar decline in the cross-section. This deterioration is consistent with structural changes in retail broker routing and wholesaler market share following the 2021 meme-stock episode. Researchers applying TAQ-based proxies to post-2021 data face substantially greater measurement noise than calibrations based on earlier samples would suggest.

The correlations also reveal how the relation between retail and institutional flows varies with past returns. The correlation between retail flow and changes in institutional ownership (ΔIO) is negative in the aggregate time series and near zero across stocks. In Panel C, the correlation between Retail Flow and ΔIO is more negative for past losers and near zero for past winners. The asymmetry is suggestive: the negative co-movement between retail and institutional flows is concentrated among past losers, where institutional investors are most likely to demand for immediacy. While we do not interpret this correlation pattern as a mechanism test, we treat it as a descriptive indication that motivates the analyses in subsequent sections, where Section 3.3 documents the opposite-sloping response of retail and institutional flows to past returns and Section 4 returns to its interpretation.

3 Buying the Dip

How do retail investors trade? One body of evidence points to return chasing (Barber et al., 2009a). Another body of evidence points that retail investors are contrarians at (Kaniel et al., 2008). The two readings have coexisted for years.

In this section, we revisit the question with a comprehensive panel of U.S. retail order flow. We find that retail investors are "buying the dip": retail demand is contrarian, but the response is concentrated almost entirely in down states. Retail order flow rises substantially after negative contemporaneous and past returns, whereas positive returns generate little systematic net selling. The asymmetry is persistent across the sample period and holds across various horizons, while non-retail flow measures on average move in the opposite direction.[4]

3.1 Retail Response to Returns

Figure 3 plots bin-averaged retail net flow (as a percent of shares outstanding) against past returns at four horizons. Panel A uses the contemporaneous monthly return $\mathrm{Ret}[t]$. Panels B, C, and D use lagged returns over windows ending at $t-1$: the past month $\mathrm{Ret}[t-2, t-1]$, the past six months $\mathrm{Ret}[t-6, t-1]$, and the past twelve months $\mathrm{Ret}[t-12, t-1]$. The construction follows a standard binscatter approach. Stocks are partitioned into 50 equal-frequency bins by past return within each panel, and the figure plots the mean retail flow within each bin.

The figure delivers two visual findings. First, the relation is strongly negative in the left tail and the slope steepens as returns become more negative. Retail inflows accelerate disproportionately as losses deepen. Second, retail net flow remains close to zero or slightly positive across the upper half of the return distribution. Stocks with positive past returns generate little systematic retail selling. The contrarian response

[4]The "buying the dip" phenomenon has recently been widely observed by market participants and financial media. See, for example, the eToro Retail Investor Beat survey (eToro, 2023), which noted that despite a prolonged bear market, "67% were either ambivalent or positive about 2022 and its impact on their investing mindset...16% saw 2022 as an opportunity to buy the dip, whilst 15% say the bear market actually increased their appetite for investing." J.P. Morgan Asset Management (2025) similarly reports that retail investors continued to purchase U.S. equities despite the market drawdown of early 2025. The financial press frequently echoes these observations; see, e.g., Bloomberg News (2025). Relatedly, Welch (2022) finds that Robinhood investors did not panic-sell during the March 2020 market crash, and were rewarded in the subsequent bull market.

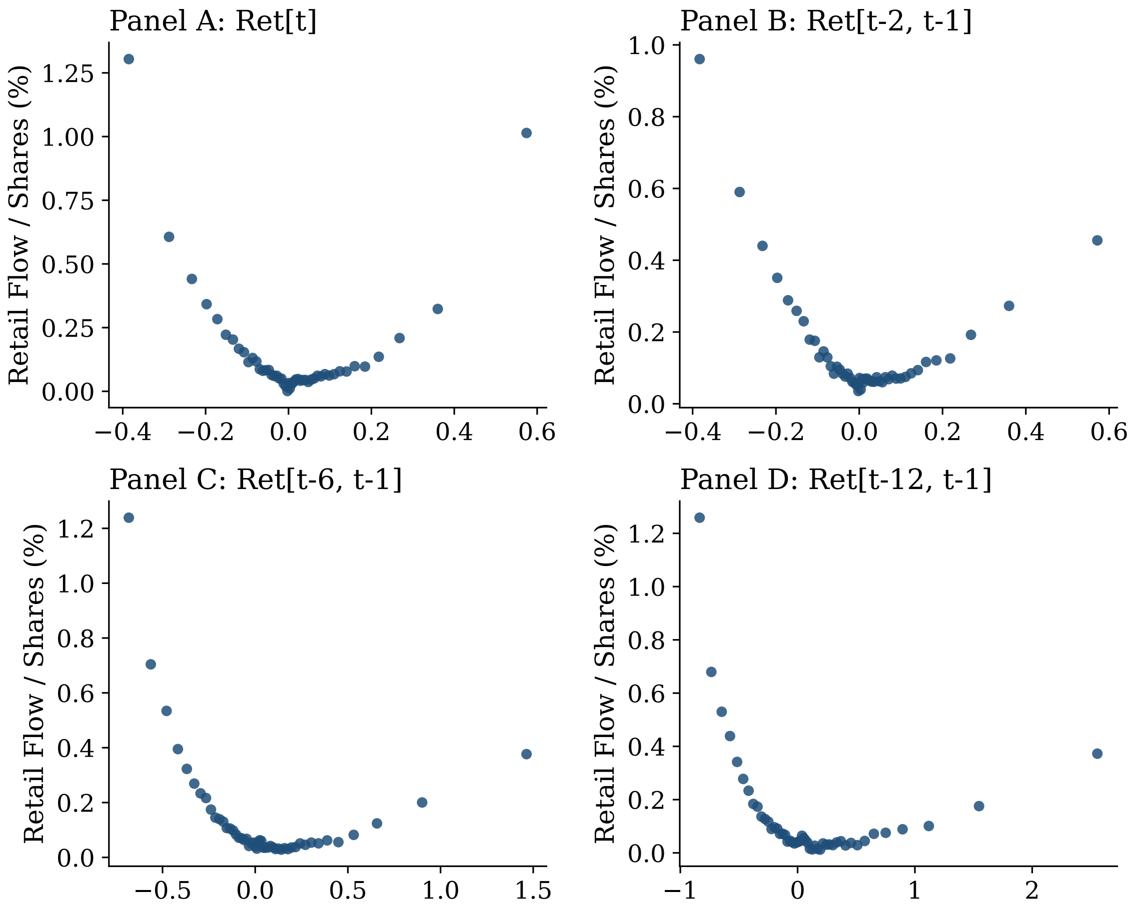

Figure 3. Retail Net Flow Versus Past Returns at Four Horizons

Notes: Each panel is a binscatter with 50 equal-frequency bins. The vertical axis is monthly retail net order flow scaled by shares outstanding (percent). The horizontal axis is the past return measured over the indicated horizon. Panel A uses the contemporaneous return Ret$[t]$; Panels B, C, and D use lagged returns ending at $t-1$. The sample is January 2017 through December 2024. Both the past return and retail flow are winsorized at the 1% and 99% percentiles within each panel.

is therefore concentrated almost entirely in down states. Because this measure is net flow, the downward tilt could in principle reflect a netting artifact rather than directional demand. Section 3.4 decomposes retail buying and selling and shows that it does not.

The pattern holds at various return horizons. Panel A shows that retail demand responds to contemporary price declines. Panels B through D show that the response also tracks lagged returns measured over windows that exclude the current month. The lagged-return panels are important for interpretation. They show that the relation is not driven by co-movement between contemporaneous flows and returns within the same month, given that contemporary returns have low correlations with their past. Retail investors continue to buy stocks that fell over the prior month, prior six months, or prior year, even after the most recent month's return is set aside.

To formalize the visual evidence and to assess robustness to firm characteristics and other flow measures, we estimate panel regressions of retail flow on a piecewise-linear function of past returns. For any return r, write $r^- \equiv \min(r, 0)$ and $r^+ \equiv \max(r, 0)$ for its negative and positive parts. The specification is

$$\begin{aligned}\text{RetailFlow}_{i,t} = a_t + b_i + \beta_0^- \,\text{Ret}[t]^- + \beta_0^+ \,\text{Ret}[t]^+ \\ + \sum_{h \in \{2,3,6,12\}} \left[\beta_h^- \,\text{Ret}[t-h,\, t-1]^- + \beta_h^+ \,\text{Ret}[t-h,\, t-1]^+\right] \\ + \gamma' X_{i,t} + \varepsilon_{i,t},\end{aligned} \tag{1}$$

where $\text{Ret}[t]$ is stock i's contemporaneous monthly return at t, $\text{Ret}[t-h,\, t-1]$ is its cumulative return from month $t-h$ to month $t-1$, a_t and b_i are time and stock fixed effects, and $X_{i,t}$ is a vector of stock-month controls. Splitting each return into its negative and positive parts allows the slope of retail flow on returns to differ between losses and gains: a symmetric linear response would require $\beta_h^- = \beta_h^+$ for every h, while asymmetric dip-buying implies $\beta_h^- < 0$ and $\beta_h^+ \approx 0$.

The control vector $X_{i,t}$ is built up across the columns of Table A.1. Column (1) includes only the contemporaneous return components. Columns (2)–(4) progressively add three sets of controls: return-distribution moments (idiosyncratic volatility, idiosyncratic skewness, and market beta), firm characteristics (log market equity, book-to-market, debt-to-assets, return on assets, firm age, and gross profitability), and three contemporaneous institutional flow proxies discussed in Section 2.5 (the quarterly change in 13-F institutional ownership, the change in securities-lending supply, and the institutional large-trade proxy). Standard errors are two-way clustered by stock and by calendar month.

Figure 4 visualizes the regression coefficients across horizons.[5] Three features

[5] The single-horizon estimates run one regression per horizon and include only that horizon's splits plus the composition controls in $X_{i,t}$. They isolate the loss-side and gain-side slopes at horizon h from the composition channel without conditioning on the other horizons' returns. The multivariate estimates correspond to column (4) of Table A.1 and include all five horizons jointly, so they report the marginal coefficient at horizon h holding all other horizons fixed.

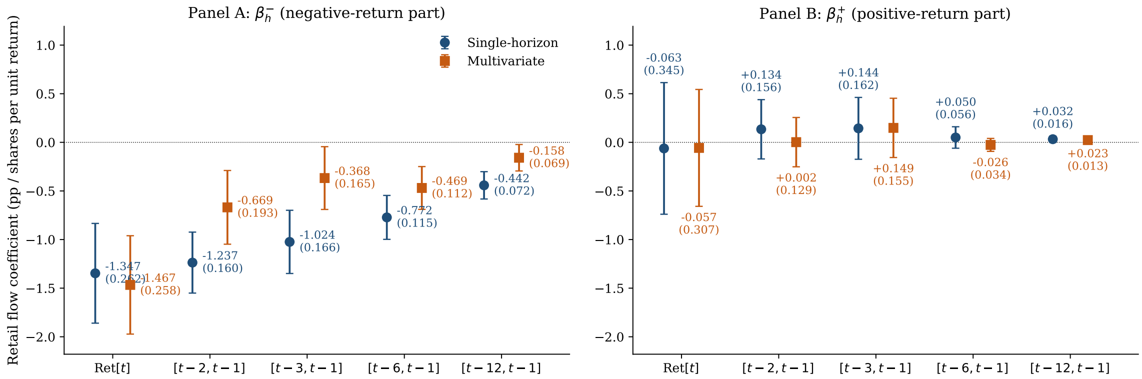

Figure 4. Buying the Dip: Regression Coefficients on Past Returns Across Horizons

Notes: Panel A reports the loss-side coefficients β_h^- and Panel B the gain-side coefficients β_h^+ from estimates of equation (1), for horizons $h \in \{0, 1, 3, 6, 12\}$ months. Navy circles ("Single-horizon") come from a separate panel regression at each horizon that includes only the focal-horizon return splits, the composition controls in $X_{i,t}$, and stock and time fixed effects. Orange squares ("Multivariate") come from column (4) of Table A.1, which includes all five horizons jointly together with $X_{i,t}$ and the same fixed effects. Error bars are 95% confidence intervals based on standard errors clustered by stock and by calendar month. Each marker is annotated with the point estimate and standard error (in parentheses). The sample is January 2017 through December 2024.

stand out. First, the loss-side coefficients β_h^- are negative and tightly estimated at every horizon. The single-horizon estimate at the contemporaneous return is approximately -1.3, indicating that a 1-percentage-point decline in Ret[t] raises monthly retail flow by roughly 1.3 basis points of shares outstanding. Loss-side coefficients decline in absolute value as the horizon lengthens, with the twelve-month coefficient still significantly negative but roughly one-tenth the size of the contemporaneous one. Second, the gain-side coefficients β_h^+ are an order of magnitude smaller in absolute value, and their 95% confidence intervals span or hug zero at every horizon. The asymmetry between β_h^- and β_h^+ is therefore not a feature of the contemporaneous return alone; it holds at every past-return horizon up to twelve months. Third, the single-horizon and multivariate loss-side estimates track each other on the gain side but diverge on the loss side: the multivariate β_h^- is uniformly smaller in absolute value than its single-horizon counterpart. The two specifications agree that the gain-side response is essentially flat, and they together imply that the longer-horizon loss-side response operates partly through the shorter horizons in the multivariate fit.

In terms of economic magnitude, the cross-sectional standard deviation of retail net flow is 2.33 percentage points of shares outstanding per month. The standard deviation of Ret$[t]$ is 22.8 percentage points and that of Ret$[t-12, t-1]$ is 92.9 percentage points (Table 1). Using the column (4) estimates, a one-standard-deviation negative shock to Ret$[t]$ is associated with $1.47 \times 0.228 \approx 0.34$ percentage points of shares outstanding in additional retail buying, or roughly 14% of the cross-sectional retail-flow standard deviation. The gain-side coefficient at the contemporaneous horizon is -0.06 with a standard error of 0.31, statistically indistinguishable from zero. At the past-twelve-month horizon, the marginal monthly effect is smaller, but it accumulates: the coefficient of -0.16 implies that a stock that fell 50 percent over the prior twelve months attracts roughly 0.08 percentage points of shares in additional retail buying every month in which that past-return state persists, or between three and four percent of the retail-flow standard deviation per month.

The positive-side upward tilt visible in Panel A of Figure 3 is therefore not evidence of retail return-chasing on the upside. The bivariate positive-return coefficient at the contemporaneous horizon (1.90 from Table A.1) falls to -0.06 once controls are added, consistent with the gain-side coefficient in Figure 4. The positive-return slope in the binscatter reflects the correlation between the contemporaneous return and other variables that drive retail demand, including idiosyncratic volatility, skewness, and market beta. Once these are controlled for, the positive-side slope disappears, turning marginally negative and statistically indistinguishable from zero. Section 3.2 documents a related feature in the time-series sort and discusses why a contemporaneous-return cut alone is misleading.

Appendix Table A.1 reports the full set of estimates including the control coefficients. The coefficient on idiosyncratic volatility is large and significantly positive across various specifications. Retail investors concentrate buying in stocks with greater idiosyncratic risk, consistent with prior evidence that retail demand is tilted toward speculative and high-volatility stocks (Kumar, 2009; Han and Kumar, 2013). The coefficient on idiosyncratic skewness is modestly *negative*, and the coefficient on market beta is also negative. Conditional on idiosyncratic volatility, retail flow is slightly higher for stocks with less positive return skewness and lower market beta. This nuances the lottery-stock framing in the literature: the volatility tilt is preserved, but the positive-skewness and high-beta tilt that would identify lottery-like demand is absent. We return to this distinction in Section 4. Among the institutional flow proxies, the securities-lending flow enters positively, indicating that retail buying co-moves with lendable supply.

3.2 Buying the Dip over Time

The cross-sectional asymmetry documented in Section 3.1 could in principle be driven by a few episodes in which retail buying was concentrated. The COVID-19 lockdown period and the 2020–2021 meme-stock surge are two natural candidates. To assess whether buying the dip is structural or episodic, we examine the time series of average flows for stocks sorted into return terciles within each calendar month.

Figure 5 reports the results for two sorting variables. Panels A and B sort stocks into terciles by the contemporaneous monthly return $\text{Ret}[t]$ and plot the cross-sectional average of, respectively, retail flow and total initiated flow within each tercile. Panels C and D repeat the exercise sorting by the past twelve-month return $\text{Ret}[t-12, t-1]$. All flows are scaled by shares outstanding and expressed in percent. The shaded regions mark the COVID-19 selloff (February–May 2020) and the meme-stock window (January–March 2021).

Panel A reports the time-series of average retail flow by contemporaneous-return tercile. Both the bottom and the top tercile receive net retail buying. The middle tercile receives the least. Averaged across the 96 months in the sample, monthly retail flow is 0.37% of shares outstanding in the bottom $\text{Ret}[t]$ tercile, 0.05% in the middle tercile, and 0.26% in the top tercile. The bottom-minus-top gap is only 0.10% ($t = 4.09$), and the bottom tercile leads the top tercile in 74% of months. The pattern is best described as a U-shape with a slightly higher left tail.

A simple contrarian-toward-mean interpretation would predict retail flow to fall monotonically across contemporaneous return terciles. The pattern rejects that view. One plausible reason is that the contemporaneous monthly return is jointly determined with within-month order flow and is therefore correlated with several variables that drive retail demand independently of dip-buying. Stocks with highly positive returns in the same month also tend to have high recent volatility, recent

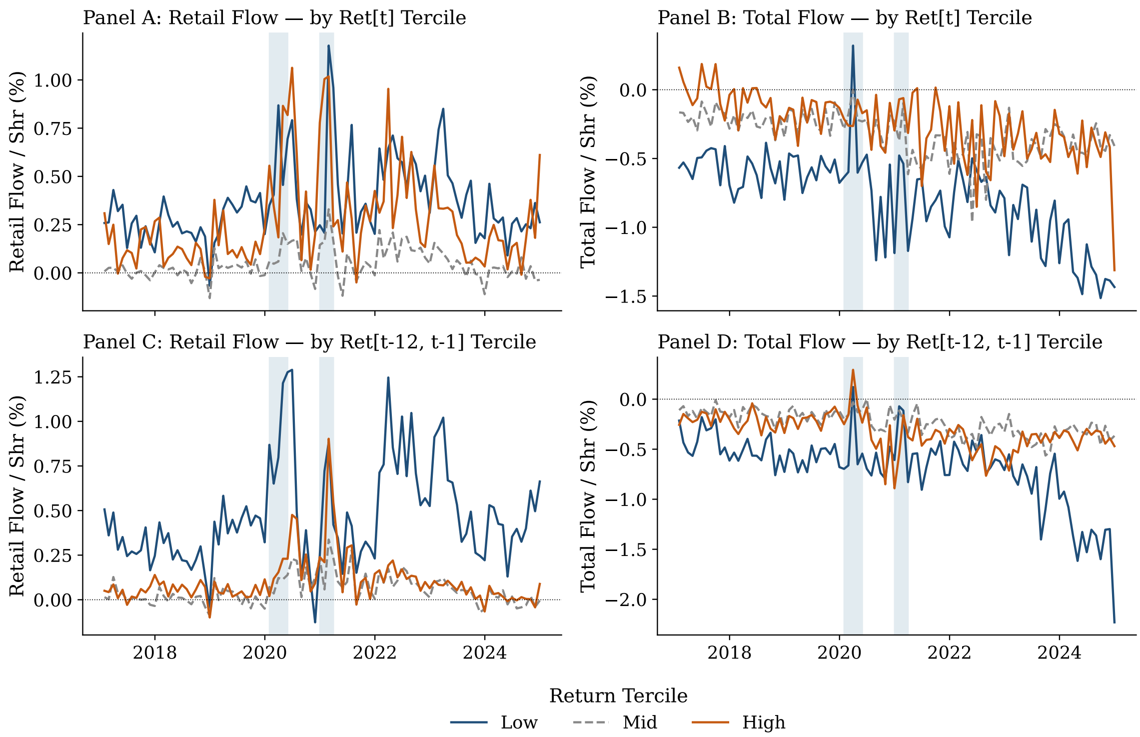

Figure 5. Retail and Total Flows by Return Tercile

Notes: Each panel plots the equal-weighted cross-sectional average of monthly net flow scaled by shares outstanding (percent) for stocks sorted into terciles within each calendar month. Panels A and B sort by contemporaneous return Ret$[t]$. Panels C and D sort by past twelve-month return Ret$[t-12, t-1]$. Panels A and C plot retail flow; Panels B and D plot total initiated flow. Shaded regions mark the COVID-19 selloff (February–May 2020) and the meme-stock window (January–March 2021). The sample is January 2017 through December 2024. Appendix Figure A.1 reports the market-cap weighted version.

news flow, and elevated retail attention. Some of those retail flows show up in the top tercile in Panel A, even though the underlying dip-buying behavior is fundamentally one-sided. The corresponding regression evidence is in Section 3.1: the bivariate positive-return coefficient falls substantially once additional controls are added. The contemporaneous-sort right tail of Panel A reflects this same composition issue. Therefore, it is necessary to also look at the past return sort.

Panel B shows that total initiated order flow is uniformly negative across all three Ret$[t]$ terciles, consistent with average net selling in the sample period. The most negative average flow is concentrated in the bottom tercile -0.78%. The bottom-minus-top gap in total flow is -0.53% ($t = -10.86$). Within a contemporaneous return sort, retail and total flows therefore have opposite signs in the bottom tercile. Retail investors buy net while non-retail investors sell net, even though both the bottom and the top tercile of retail flow are positive on average.

Panels C and D sort by Ret$[t-12, t-1]$, which is predetermined relative to month t and is therefore not jointly determined with month-t flows. The contemporaneous-sort U-shape disappears; instead, we see a consistent L-shape pattern over the sample period. Average monthly retail flow is 0.47% of shares outstanding in the bottom past-return tercile, 0.05% in the middle tercile, and 0.10% in the top tercile. The bottom-minus-top gap is 0.37% ($t = 6.45$), nearly four times larger than under the contemporaneous-return sort. The bottom tercile leads the top tercile in 97% of the monthly sample, that is, in essentially every month. The few exception months cluster in the meme-stock window of early 2021, visible in Panel C of Figure 5. In those months a surge of return-chasing retail inflow into recent winners lifted top-tercile flow above bottom-tercile flow, briefly inverting the usual ordering before it resumed.

The first-order autocorrelation of the bottom-tercile retail flow series is 0.67. This high autocorrelation indicates that strong inflows into recent losers are not isolated to a few unusual months. Heavy retail buying in past-loser stocks persists across months and across distinct market regimes. Panel D shows the total flow is most negative in the bottom past-return tercile (-0.69% per month) and less negative in the top tercile (-0.33% per month). The bottom-minus-top gap is -0.36% ($t = -4.06$), and the bottom tercile lies below the top tercile in only 9.4% of months. The two panels together describe a stable feature of the panel: month after month, the same stocks that retail investors are buying most intensively are the stocks from which non-retail investors are withdrawing.

The shaded regions in Figure 5 mark two of the most disrupted episodes in the sample. The bottom-tercile retail flow series in Panel C lies above the middle and top terciles throughout the sample, including across the COVID-19 selloff and the meme-stock window. The bottom-tercile lead is visible in essentially every month and survives the COVID-19 selloff intact; the meme-stock window is the one brief interruption, after which the ordering resumes. The pattern is not driven by the meme-stock surge in retail participation, although that episode did raise the level of retail flow into all three terciles for several months. The opposing pattern in total flow in Panel D is similarly persistent.

The market-cap weighted version of Figure 5, reported in Appendix Figure A.1, displays the same qualitative pattern. Buying the dip is therefore not concentrated in small-capitalization names; it is also visible when each tercile's average is weighted toward the largest stocks within the tercile. Together, Panels C and D and the persistence statistics establish that buying the dip is a structural feature of retail trading in this sample rather than a phenomenon tied to any single episode or cohort of new retail entrants.

3.3 Other Investor Flows

Sections 3.1 and 3.2 establish that retail flow rises after price declines and that the pattern persists through the full sample. A natural follow-up question is whether dip-buying is specific to retail investors or whether it reflects a broader response shared across initiated order flow. To address this, we apply the binscatter design of Figure 3 to non-retail flow measures using the lagged-return horizon. The lagged horizon mitigates the joint-determination concern raised in Section 3.2. We use the past twelve-month return $\text{Ret}[t-12, t-1]$ for consistency with the time-series sort in Section 3.2.

The four panels show that other investors react to returns differently: institutional proxies and total flow are momentum-following. Panel A plots the institutional large-trade proxy (Inst50k). The slope is positive across the return distribution: large trades are net buyers of past winners and net sellers of past losers. Panel B plots the securities lending supply (InstSec). The lendable supply rises in stocks with strong past returns and falls in past losers, indicating that institutional investors are chasing momentum. Panel C plots total initiated order flow, which exhibits a hump-shaped pattern. The pattern suggests that aggregate market flow tilts away from both past losers and past winners. Panel D plots the quarterly change in 13-F institutional ownership (ΔIO). The slope is positive, indicating that institutional investors sells losers and buy winners.

The retail flow slopes down on past returns; the four non-retail measures slope up. The opposing slopes imply that retail demand and the various institutional and

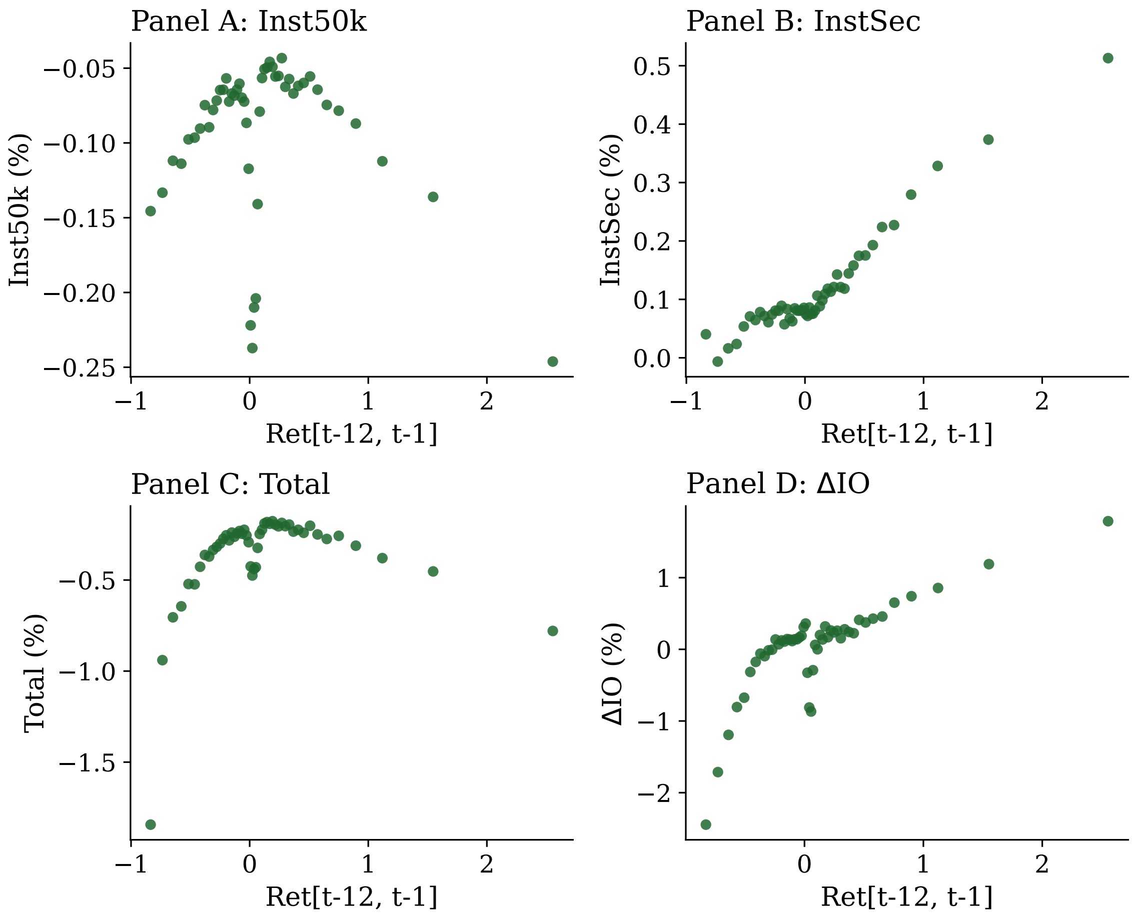

Figure 6. Other Net Flow Measures Versus Past Return

Notes: Each panel is a binscatter with 50 equal-frequency bins. The vertical axis is the indicated monthly net flow scaled by shares outstanding (percent). The horizontal axis is the past twelve-month return Ret$[t-12, t-1]$. Panel A: large-trade institutional proxy (Inst50k, trades above \$50,000 notional). Panel B: institutional securities-lending flow (InstSec). Panel C: total initiated order flow. Panel D: quarterly change in 13-F institutional ownership (ΔIO). Variable definitions follow Section 2.5. Both axes are winsorized at the 1% and 99% percentiles within each panel. The sample is January 2017 through December 2024.

aggregate measures of supply move in opposite directions across the past-return distribution. Stocks that have fallen receive retail buying and institutional selling, while stocks that have risen receive institutional buying and only modest retail flow. The dip-buying response documented earlier in this section is therefore not a generic response of initiated order flow to past returns. It is a specific feature of retail flow.

3.4 Buy versus Sell

A long-standing concern with the retail trading pattern is that it may be a netting artifact (Barber and Odean, 2013). A downward-sloping relation between net flow and past returns need not mean that retail purchases respond contrarily. If gross purchases and sales both move with past returns, net flow can slope down even though each leg moves up. We address this by decomposing retail net flow into its two gross legs. Our data identify retail buying and retail selling separately, and net flow is exactly their difference, RetailFlow = RetailBuy − RetailSell, so the decomposition is exact rather than estimated. Appendix Figure A.2 replicates the event-time exercise of Barber et al. (2009b, Figure 3). Weighting stock-months by retail buy versus sell intensity yields nearly the same pre-trade return path, so on average both trades follow stocks that experienced positive returns. The difference of the two legs tells a different story. The net flow path is hump-shaped: it peaks roughly a year before the trade and then declines steadily into the purchase month, ending below zero. Conditional on net buying, the stock has been losing over the year before retail investors step in. This is the event-time counterpart of the contrarian net behavior.

Separating the return into gain and loss sides, we find that retail purchases also reflect the buying-the-dip phenonmenon. Since the event-time exercise also averages over the sign of the past return, it cannot speak to a second dimension, that retail buying and selling respond asymmetrically to gains and to losses. Appendix Figure A.3 addresses this. It plots bin-averaged retail buying and retail selling against past returns at four horizons, with market turnover on a secondary axis. Two features stand out. First, the response of gross retail trade is not monotone. Retail buying and selling are both elevated when the past return is large in absolute value, of either sign, and are lowest near a zero past return. A single average slope is therefore missed out whether retail investors are contrarian or return-chasing under different return regimes. Second, the buy and sell values are nearly coincident at every horizon. This suggests that both legs are driven by a common component in the return domain. Market turnover is one candidate. From the figure, total turnover traces the same return shape as the two retail legs, so the shared movement of buying and selling is largely a turnover phenomenon.

Appendix Figure A.4 estimates equation (1) separately for retail buying, retail selling, and market turnover, splitting each past return into its negative and positive parts. The regression results confirm that retail purchases exhibit a similar buy-the-dip pattern. Retail purchases are contrarian to negative past returns while chasing positive ones. Moreover, the coefficients are nearly the same for both retail purchases and sales, and they are similar to the coefficients for market turnover at each horizon. This suggests both retail buying and selling are driven by the shared component related to market turnover. A comparison of gross versus net flow coefficients further corroborates the cointegration. Gross coefficients are large but imprecise. At the one-month lagged horizon the loss-side coefficient is -12.54 for buying and -11.30 for selling, with standard errors near 3.9. Once we net out the gross trades, the loss-side coefficient for retail flow is -1.24 with a standard error of 0.16. The standard error falls by roughly a factor of twenty-five when the two legs are differenced, suggesting a large common component that cancels in the net.

These results show that the dip-buying pattern among retail investors is not a netting effect. Gross retail purchases are also contrarian particularly in the negative-return side. Furthermore, gross buying and selling activities are closely aligned, with a common movement related to market turnover. The contrarian tilt in net retail flow is the asymmetric component that remains once the common turnover variation is removed from the gross trades.

3.5 Robustness

The asymmetric dip-buying pattern is not an artifact of how retail flow is scaled or of the linear return measure. Appendix Figure A.5 reproduces the binscatter of Figure 3 using three alternative normalizations of net retail flow, at the contemporaneous and past-twelve-month horizons. The loss-side tilt is preserved under all three: dollar flow scaled by market capitalization traces the same steep left arm as the baseline measure, and the two signed order-imbalance ratios (net flow scaled by retail share volume, and dollar flow scaled by retail dollar volume) slope down monotonically in past returns, turning negative for past winners.

Appendix Figure A.6 re-estimates the coefficient plot of Figure 4 with log returns $\ln(1 + r_h)$ on the right-hand side; the loss/gain sign split is preserved because $\ln(1 + r_h) < 0$ if and only if $r_h < 0$. The asymmetry is unchanged: the loss-side coefficients β_h^- are negative and bounded away from zero at every horizon in both the single-horizon and multivariate specifications, while the gain-side coefficients β_h^+ are economically small with confidence intervals that span zero.

4 Discussion

The buy-the-dip pattern documented in Section 3 is a strongly asymmetric, persistent, and characteristic-level relationship between retail net order flow and past returns. Retail buying intensifies in the left tail of the return distribution. It is roughly flat in the right tail. The shape holds across return horizons from one month to twelve months. It is stable across the 2017–2024 sample. It runs against the slopes of the institutional and aggregate flow measures we observe. This section interprets the pattern. Section 4.1 positions our findings within two strands of prior evidence that speak most directly to dip-buying. Section 4.2 canvasses candidate explanations and asks whether each can generate the two features any account must reproduce, namely the asymmetry across loss and gain states and the response to signed past returns at multiple horizons. Section 4.3 draws out the implications of the pattern for retail liquidity provision and for the informational content of prices in distressed states.

4.1 Reconciling Contrarian Retail Trading and Demand for Distressed Stocks

Contrarian retail trading. Existing work using TAQ-based proxies and proprietary brokerage data has documented that aggregated retail order flow leans against contemporaneous and short-horizon returns and predicts subsequent reversals (Kaniel et al., 2008; Kelley and Tetlock, 2013; Boehmer et al., 2021; Barrot et al., 2016; Luo et al., 2025). The loss-side response in Section 3 is consistent with this body of evidence. Two features push the pattern past what linear contrarian estimates can describe. The first is the asymmetric shape. The piecewise specification in equation (1) shows that the loss-side coefficient is roughly an order of magnitude larger in absolute value than the gain-side coefficient at every horizon from one month to twelve months. The gain-side coefficient is statistically indistinguishable from zero at most horizons once standard controls are included. Linear contrarian specifications implicitly average across these two regimes and mask the kink. The second is the horizon profile. The loss-side coefficient remains negative and tightly estimated all the way out to twelve months, beyond the one-to-three-month windows that have served as the typical estimation horizon in prior work. The asymmetric, multi-horizon shape is the substantive addition that the S&P retail flow data bring to the contrarian-retail literature.

Retail demand for distressed stocks. A separate asset-pricing literature documents that financially distressed stocks earn anomalously low returns and that retail investors are concentrated on the buy side. Distress in this work is a fundamental characteristic, measured from accounting and market data, not a return realization. Campbell et al. (2008) estimate firm-level failure probabilities and show that high-failure-probability stocks deliver low risk-adjusted returns inconsistent with rational compensation for distress risk. Conrad et al. (2014) attribute the anomaly to retail demand for jackpot-like upside in distressed firms, drawing on the stocks-as-lotteries framework of Barberis and Huang (2008). A closely related strand documents the same underlying preference for jackpot-like upside using individual-investor brokerage account data (Mitton and Vorkink, 2007; Kumar, 2009). The stylized fact we document in Section 3 conditions retail demand on contemporaneous and past returns rather than on fundamental distress or jackpot characteristics. The response is asymmetric across loss and gain states in the return at every horizon up to twelve months.

4.2 Candidate Explanations: Asymmetry and Past-Return Reaction

Two features of the dip-buying pattern discipline any candidate explanation. The first is the asymmetry. Retail flow rises markedly after past losses but is statistically indistinguishable from zero after past gains, so an account that generates a symmetric response on either side of the return distribution cannot fit the data. The second is the response to signed past returns at multiple horizons. The trigger is not the absolute return and it is not a stable stock characteristic. It is the signed return measured over windows from one month to twelve months. We canvas the candidate explanations on these two dimensions and treat them as suggestive rather than identifying.

Attention. The first candidate is the attention framework of Barber and Odean (2008), Barber et al. (2022), van der Beck et al. (2026), and Graves (2024). In this reading, retail investors are net buyers of attention-grabbing stocks identified by extreme returns, news coverage, or unusually high volume. The pull from attention is symmetric in the sign of the return, so attention frameworks generate retail flow responses to $|r_{i,t}|$ rather than to $r_{i,t}$. They therefore fail the asymmetry diagnostic in net form. They also fail the past-return diagnostic at the twelve-month horizon, since attention measures typically decay quickly after the triggering event. Attention may contribute to the level of dip-buying in the immediate aftermath of a steep decline, but it cannot by itself generate the asymmetric multi-horizon shape.

Habitat and lottery preferences. A second candidate appeals to stable cross-sectional retail preferences. Laarits and Sammon (2025) document retail concentration in hard-to-value stocks. Balasubramaniam et al. (2023) identify clienteles organized around firm age, share price, and size. Dorn and Huberman (2010) establish a preferred-risk habitat in which investors hold portfolios matching their own risk tolerance. A closely related strand reads retail demand as lottery-seeking: Kumar (2009), Han and Kumar (2013), and Bali et al. (2011) document tilts toward high idiosyncratic volatility, low prices, and high idiosyncratic skewness. Two observations weigh against these as a stand-alone explanation. Habitat and lottery preferences describe stable cross-sectional preferences over stock characteristics, not return-conditional flow, so they do not by themselves generate the asymmetric flow response to signed past returns. Appendix Table A.1 additionally shows that retail flow loads positively on idiosyncratic volatility but *negatively* on idiosyncratic skewness. The volatility tilt is consistent with a preferred-risk habitat. The negative skewness loading is the opposite of the lottery prediction. A habitat for volatility and distress is compatible with the level of dip-buying, but it does not, on its own, pin down the asymmetry across the loss and gain sides of the past-return distribution.

Extrapolation. A third candidate attributes the loss-side response to belief-based mean reversion. If retail investors expect prices to revert toward fundamentals, they should buy stocks with recent losses. Cohen et al. (2002) interpret individual investors as anchoring on past prices and underreacting to cash-flow news. Bastianello and Fontanier (2024) and Bastianello and Fontanier (2025) formalize a partial-equilibrium-thinking framework in which traders learn from prices and produce contrarian flow under normal conditions. Extrapolation satisfies the past-return diagnostic, since beliefs that look back at the realized price path generate flow responses at multiple horizons. The challenge is the asymmetry. A symmetric mean-reversion belief generates retail selling after gains, but our gain-side coefficients are statistically indistinguishable from zero. Rationalizing the asymmetry within an extrapolation framework requires beliefs that switch on only in down states, or asymmetric overreaction to losses and underreaction to gains. Neither is the default in the existing literature.

Disposition and loss realization. A fourth candidate is the disposition effect: investors hold on to losers and sell winners quickly (Shefrin and Statman, 1985; Odean, 1998). This predicts that retail selling should fall after a stock drops, because holders do not want to lock in a loss. The buy and sell decomposition in Section 3.4 shows the opposite. Retail selling rises after losses, not falls. The sell leg in Appendix Figure A.3 is high after large past moves of either sign, and the loss-side sell coefficient in Appendix Figure A.4 is large and close to the loss-side buy coefficient. Retail does not hold losers in the aggregate. The winner side fits the disposition story better. Selling is also high after gains, which lines up with the contrarian households of Grinblatt and Keloharju (2000). Two caveats apply. The disposition effect is about each investor's own purchase price, not a stock's past return over a fixed window, and the two can differ. Our data are aggregate buys and sells with no investor cost basis, so we cannot test it directly. The disposition effect also concerns the sale of existing holdings and says nothing about why retail buys losers, which is the main fact in this paper. The disposition effect is therefore only partially consistent with the evidence. It fits the selling of past winners but is contradicted on the loss side, and it does not explain the asymmetric buying that defines the pattern.

Absorption of institutional selling pressure. The closest match to both diagnostics comes from the institutional side. The opposing-slope evidence in Section 3.3 is the most direct empirical anchor. The institutional large-trade proxy, the institutional securities-lending flow, the change in 13-F institutional ownership, and total initiated order flow all slope positively against past returns when they are negative, while the retail flow slopes in the opposite direction (Section 3.1). On the loss side, retail buys what the rest of the market sells. Coval and Stafford (2007) document that mutual-fund flow-driven sales generate temporary price pressure in commonly held stocks. Vayanos and Woolley (2013) formalize how delegated investment management generates flow-driven trading in equilibrium. This mechanism is also consistent with Barrot et al. (2016), who show that retail traders likely take the other side of these institutional flows and earn a short-horizon return for supplying that liquidity. Under either lens, the asymmetry on the loss side has a direct interpretation. Institutional constraints, redemptions, and risk budgets bind asymmetrically in down states, so the supply pressure that retail absorbs is itself state-contingent. The opposing-slope evidence does not, however, rule out a third factor that drives both retail buying and institutional selling jointly, and it does not identify which institutional friction generates the supply pressure. The reading is therefore the closest match among the candidates we examine, but it is not an identified mechanism.

4.3 Implications for Liquidity Provision and Price Informativeness

The asymmetric, multi-horizon, characteristic-stable shape of retail flow has implications for two aspects of market quality in distressed states: the provision of liquidity and the informativeness of prices. We develop each in turn.

Liquidity provision in distressed states. The opposing-slope evidence in Section 3.3 and the persistence statistics in Section 3.2 together describe a market in which retail flow absorbs the supply that other investors release in distressed states. Retail flow into the bottom past-twelve-month return tercile is positive in 96.9% of months in the sample. The institutional flow proxies slope in the opposite direction over the same return distribution. The interpretation of this absorption depends on the source of the price decline. When the price move reflects transient institutional supply pressure of the kind studied by Coval and Stafford (2007) and Vayanos and Woolley (2013), retail absorption shortens the period of dislocation. When the price move reflects adverse news about fundamentals, the same absorption slows the incorporation of that news into the price. Our flow evidence does not separate these two cases.

A descriptive look at portfolio returns tempers the profitability reading of this liquidity provision. Figure 7 plots cumulative returns for tercile portfolios sorted on the contemporaneous return (Panels A and B) and on the past-twelve-month return (Panels C and D). The past-twelve-month Low tercile, which is the portfolio that receives the strongest retail flow per the persistence statistics in Section 3.2, ends the sample close to zero in equal-weighted form. Past winners outperform past losers over the cumulative window. A short-horizon reversal premium dominated by a longer-horizon momentum drift would generate exactly this picture. The institutional-pressure reading can still hold on horizons shorter than the cumulative window. Our flow data do not have the granularity to map portfolio returns into retail-investor returns, so we treat the figure as a caution against overclaiming profitable liquidity provision rather than as a rejection.

Price informativeness in distressed states. The fundamental-news case has a second formal expression in models of asymmetric information. In the language of standard noisy rational-expectations models, the retail mass that absorbs supply in distressed states is a higher mass of uninformed agents whose demand is the equilibrium response to the realized price rather than a noise residual. Yuan (2005) and Glebkin et al. (2021) show that, in equilibria where informed capacity is constrained in down states, the equilibrium price function becomes more sensitive to noise in the constrained regime. Realized price-change volatility and posterior payoff variance both rise in the distressed state. Our flow evidence is consistent with the demand-side counterpart of that prediction: the same retail demand that supports prices in distressed states also dilutes the informational content of the price in those states. Dávila and Parlatore (2025) provide an empirical methodology for

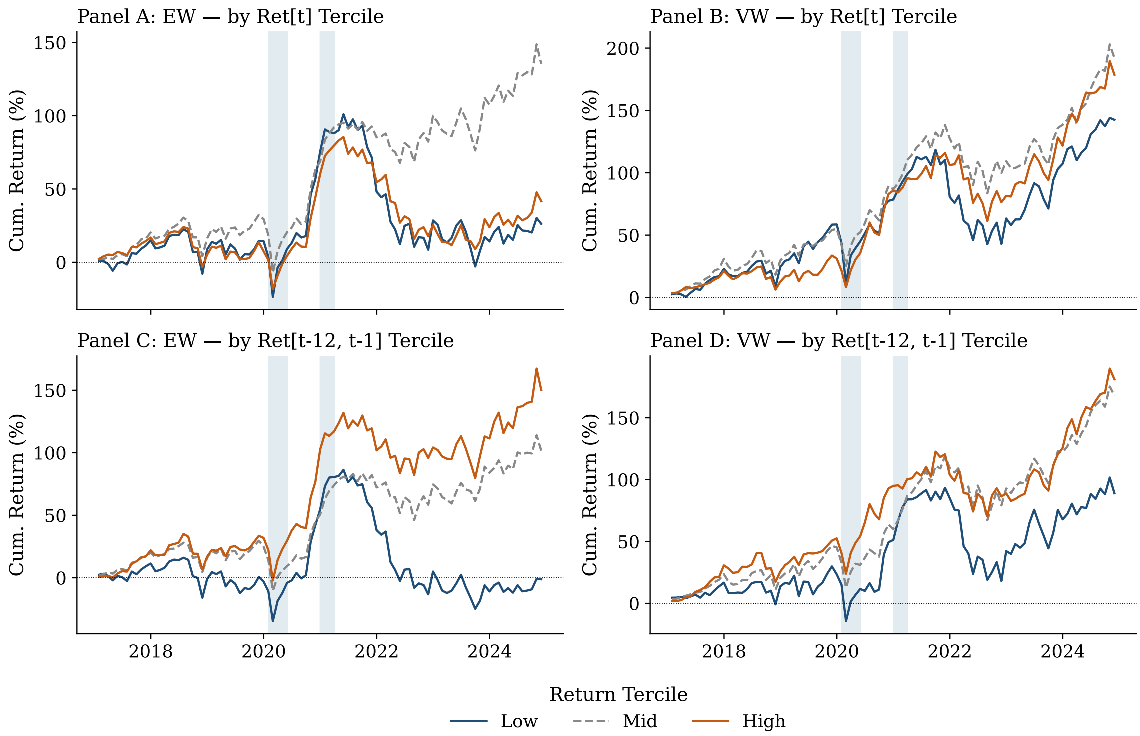

Figure 7. Cumulative returns of tercile portfolios sorted on past returns

Notes: The figure plots cumulative returns from January 2017 through December 2024 for tercile portfolios sorted on contemporaneous return Ret[t] in Panels A (equal-weighted) and B (value-weighted), and on past-twelve-month return Ret[$t-12, t-1$] in Panels C (equal-weighted) and D (value-weighted). Panel C is the past-return sort that aligns with the persistence statistics in Section 3.2. Shaded bands indicate the COVID-19 drawdown of February-March 2020 and the meme-stock rally of January-February 2021. The Low tercile in Panels C and D corresponds to the past-twelve-month loser portfolio that receives the strongest retail flow according to Figure 5 and the persistence statistics.

measuring price informativeness that our flow data alone do not deliver, so we report the implication as a qualitative consequence of the asymmetric schedule rather than as an empirical test.

The market-quality reading of the pattern is therefore two-sided. Retail dip-buying supports prices in distressed states. When the underlying shock is transient institutional supply pressure, this absorption shortens the period of dislocation. When the underlying shock is fundamental, the same absorption slows the incorporation of news into the price and reduces the informational content of the price in those states. The two readings are joint consequences of the same asymmetric demand, and our flow data do not separate them. The pattern is not a market failure in the usual sense, since retail demand is itself the equilibrium response of uninformed agents to a state-contingent price function. The balance between the stabilizing and the informativeness-reducing components depends on parameters outside the scope of our empirical exercise. We leave a fuller welfare characterization to future work that combines our flow data with measures of price informativeness and investor-level market timing.

5 Conclusion

This paper documents a stylized fact about retail trading: retail investors buy the dip. Retail net order flow rises markedly after negative contemporaneous and past returns and is roughly flat after positive returns. The sign-contingent asymmetry holds across horizons and is stable over time.

Several questions remain for future research. First, identifying the institutional channel requires separating pressure absorption from a common factor that could move retail and institutional flows jointly, and pinning down which institutional friction generates the supply. Second, our data measure aggregate retail order flow rather than individual accounts. They speak to retail trading behavior in aggregate but cannot trace holding periods or realized returns. Whether retail investors also exit in time to capture the reversal premium therefore requires more granular data that we leave to future work. Third, a fuller welfare characterization will need to combine flow data with investor-level market timing. We view the asymmetric, state-contingent shape of retail demand as the central object that such work must confront.

References

Vimal Balasubramaniam, John Y. Campbell, Tarun Ramadorai, and Benjamin Ranish. Who owns what? a factor model for direct stock holding. *The Journal of Finance*, 78 (3):1545–1591, 2023. doi: 10.1111/jofi.13220.

Turan G. Bali, Nusret Cakici, and Robert F. Whitelaw. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, 99(2):427–446, 2011. doi: 10.1016/j.jfineco.2010.08.014.

Yashar H. Barardehi, Zhi Da, Peter Dixon, and Junbo L. Wang. You can only lend what you own: Inferring daily institutional trading from lendable equity inventory. Working paper, SSRN, November 2024. Date written: November 15, 2024; posted November 25, 2024; last revised October 30, 2025.

Brad M. Barber and Terrance Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, 21(2):785–818, 2008. doi: 10.1093/rfs/hhm079.

Brad M. Barber and Terrance Odean. The behavior of individual investors. In George M. Constantinides, Milton Harris, and René M. Stulz, editors, *Handbook of the Economics of Finance*, volume 2, chapter 22, pages 1533–1570. Elsevier, 2013. doi: 10.1016/B978-0-44-459406-8.00022-6.

Brad M. Barber, Terrance Odean, and Ning Zhu. Do retail trades move markets? *The Review of Financial Studies*, 22(1):151–186, 2009a. doi: 10.1093/rfs/hhn035.

Brad M. Barber, Terrance Odean, and Ning Zhu. Systematic noise. *Journal of Financial Markets*, 12(4):547–569, 2009b. doi: 10.1016/j.finmar.2009.03.003.

Brad M. Barber, Xing Huang, Terrance Odean, and Christopher Schwarz. Attention-induced trading and returns: Evidence from robinhood users. *The Journal of Finance*, 77(6):3141–3190, 2022. doi: 10.1111/jofi.13183.

Brad M. Barber, Xing Huang, Philippe Jorion, Terrance Odean, and Christopher Schwarz. A (sub)penny for your thoughts: Can analyzing order submissions improve our understanding of retail investor behavior? *The Journal of Finance*, 79(4): 2311–2356, 2024. doi: 10.1111/jofi.13334.

Nicholas Barberis and Ming Huang. Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review*, 98(5): 2066–2100, 2008. doi: 10.1257/aer.98.5.2066.

Jean-Noel Barrot, Ron Kaniel, and David Sraer. Are retail traders compensated for providing liquidity? *Journal of Financial Economics*, 120(1):146–168, 2016. doi: 10.1016/j.jfineco.2016.01.005.

Francesca Bastianello and Paul Fontanier. Partial equilibrium thinking, extrapolation, and bubbles. Working paper, SSRN, 2024.

Francesca Bastianello and Paul Fontanier. Expectations and learning from prices. *The Review of Economic Studies*, 92:1341–1374, 2025. doi: 10.1093/restud/rdae059.

Robert H Battalio, Robert H Jennings, Mehmet Saglam, and Jun Wu. Difficulties in obtaining a representative sample of retail trades from public data sources. *Available at SSRN 4579159*, 2023.

Bloomberg News. YOLO crowd's record dip-buying binge calms a jumpy stock market. Bloomberg News, May 2025. URL https://www.bloomberg.com/news/articles/2025-05-19/yolo-crowd-s-record-dip-buying-binge-calms-a-jumpy-stock-market.

Ekkehart Boehmer, Charles M. Jones, Xiaoyan Zhang, and Xinran Zhang. Tracking retail investor activity. *The Journal of Finance*, 76(5):2249–2300, 2021. doi: 10.1111/jofi.13051.

John Y. Campbell, Jens Hilscher, and Jan Szilagyi. In search of distress risk. *The Journal of Finance*, 63(6):2899–2939, 2008. doi: 10.1111/j.1540-6261.2008.01416.x.

John Y. Campbell, Tarun Ramadorai, and Allie Schwartz. Caught on tape: Institutional trading, stock returns, and earnings announcements. *Journal of Financial Economics*, 92(1):66–91, 2009. doi: 10.1016/j.jfineco.2008.03.006.

Randolph B. Cohen, Paul A. Gompers, and Tuomo Vuolteenaho. Who underreacts to cash-flow news? evidence from trading between individuals and institutions. *Journal of Financial Economics*, 66(2–3):409–462, 2002. doi: 10.1016/S0304-405X(02)00229-5.

Jennifer Conrad, Nishad Kapadia, and Yuhang Xing. Death and jackpot: Why do individual investors hold overpriced stocks? *Journal of Financial Economics*, 113(3):455–475, 2014. doi: 10.1016/j.jfineco.2014.04.001.

Joshua Coval and Erik Stafford. Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, 86(2):479–512, 2007. doi: 10.1016/j.jfineco.2006.09.007.

William Cready, Abdullah Kumas, and Musa Subasi. Are trade size-based inferences about traders reliable? evidence from institutional earnings-related trading. *Journal of Accounting Research*, 52(4):877–909, 2014. doi: 10.1111/1475-679X.12056.

Kent Daniel, Lorenzo Garlappi, and Kairong Xiao. Monetary policy and reaching for income. *The Journal of Finance*, 76(3):1145–1193, 2021. doi: 10.1111/jofi.13004.

Eduardo Dávila and Cecilia Parlatore. Identifying price informativeness. *The Review of Financial Studies*, 2025. doi: 10.1093/rfs/hhaf051.

Daniel Dorn and Gur Huberman. Preferred risk habitat of individual investors. *Journal of Financial Economics*, 97(1):155–173, 2010. doi: 10.1016/j.jfineco.2010.03.013.

Gregory W. Eaton, T. Clifton Green, Brian S. Roseman, and Yanbin Wu. Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics*, 146(2):502–528, 2022. doi: 10.1016/j.jfineco.2022.08.002.

eToro. Retail investor beat. eToro Retail Investor Beat Survey, January 2023. Quarterly retail investor sentiment survey.

Michael Gelman, Liron Reiter-Gavish, and Nikolai Roussanov. FOMO economics: External reference-dependence and risk-taking in household portfolios. Working paper, SSRN, 2026.

Sergei Glebkin, Naveen Gondhi, and John C. Kuong. Funding constraints and informational efficiency. *The Review of Financial Studies*, 34(9):4269–4322, 2021. doi: 10.1093/rfs/hhab001.

Daniel Graves. Attention and the retail alignment puzzle. Working paper, Yale University, 2024.

Mark Grinblatt and Matti Keloharju. The investment behavior and performance of various investor types: A study of finland's unique data set. *Journal of Financial Economics*, 55(1):43–67, 2000. doi: 10.1016/S0304-405X(99)00044-6.

Bing Han and Alok Kumar. Speculative retail trading and asset prices. *Journal of Financial and Quantitative Analysis*, 48(2):377–404, 2013. doi: 10.1017/S0022109013000100.

Soeren Hvidkjaer. Small trades and the cross-section of stock returns. *Review of Financial Studies*, 21(3):1123–1151, 2008. doi: 10.1093/rfs/hhn049.

J.P. Morgan Asset Management. Who is buying U.S. equities? J.P. Morgan Asset Management Market Insights, 2025. URL https://am.jpmorgan.com/us/en/asset-management/liq/insights/market-insights/market-updates/on-the-minds-of-investors/who-is-buying-us-equities/. Accessed March 2026.

Ron Kaniel, Gideon Saar, and Sheridan Titman. Individual investor trading and stock returns. *Journal of Finance*, 63(1):273–310, 2008. doi: 10.1111/j.1540-6261.2008.01316.

Eric K. Kelley and Paul C. Tetlock. How wise are crowds? insights from retail orders and stock returns. *The Journal of Finance*, 68(3):1229–1265, 2013. doi: 10.1111/jofi.12028.

Alok Kumar. Who gambles in the stock market? *The Journal of Finance*, 64(4):1889–1933, 2009. doi: 10.1111/j.1540-6261.2009.01483.x.

Alok Kumar and Charles M. C. Lee. Retail investor sentiment and return comovements. *The Journal of Finance*, 61(5):2451–2486, 2006. doi: 10.1111/j.1540-6261.2006.01063.x.

Toomas Laarits and Marco Sammon. The retail habitat. *Journal of Financial Economics*, 172:104144, 2025. doi: 10.1016/j.jfineco.2025.104144.

Charles M. C. Lee and Balkrishna Radhakrishna. Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets*, 3(2):83–111, 2000. doi: 10.1016/S1386-4181(00)00002-1.

Bryan Luo, Enrichetta Ravina, Marco Sammon, and Luis M. Viceira. Retail investors' contrarian behavior around news, attention, and the momentum effect. Working Paper 34086, National Bureau of Economic Research, 2025.

Todd Mitton and Keith Vorkink. Equilibrium underdiversification and the preference for skewness. *The Review of Financial Studies*, 20(4):1255–1288, 2007. doi: 10.1093/rfs/hhm011.

Terrance Odean. Are investors reluctant to realize their losses? *The Journal of Finance*, 53(5):1775–1798, 1998. doi: 10.1111/0022-1082.00072.

Hersh Shefrin and Meir Statman. The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, 40(3):777–790, 1985. doi: 10.1111/j.1540-6261.1985.tb05002.x.

Philippe van der Beck, Cameron Cohen, and Coralie Jaunin. The equity market implications of the retail investment boom. Working Paper 26-054, Harvard Business School, 2026.

Dimitri Vayanos and Paul Woolley. An institutional theory of momentum and reversal. *The Review of Financial Studies*, 26(5):1087–1145, 2013. doi: 10.1093/rfs/hht014.

Ivo Welch. The wisdom of the robinhood crowd. *The Journal of Finance*, 77(3):1489–1527, 2022. doi: 10.1111/jofi.13128.

Kathy Yuan. Asymmetric price movements and borrowing constraints: A rational expectations equilibrium model of crises, contagion, and confusion. *The Journal of Finance*, 60(1):379–411, 2005. doi: 10.1111/j.1540-6261.2005.00733.x.

A Additional Figures and Tables

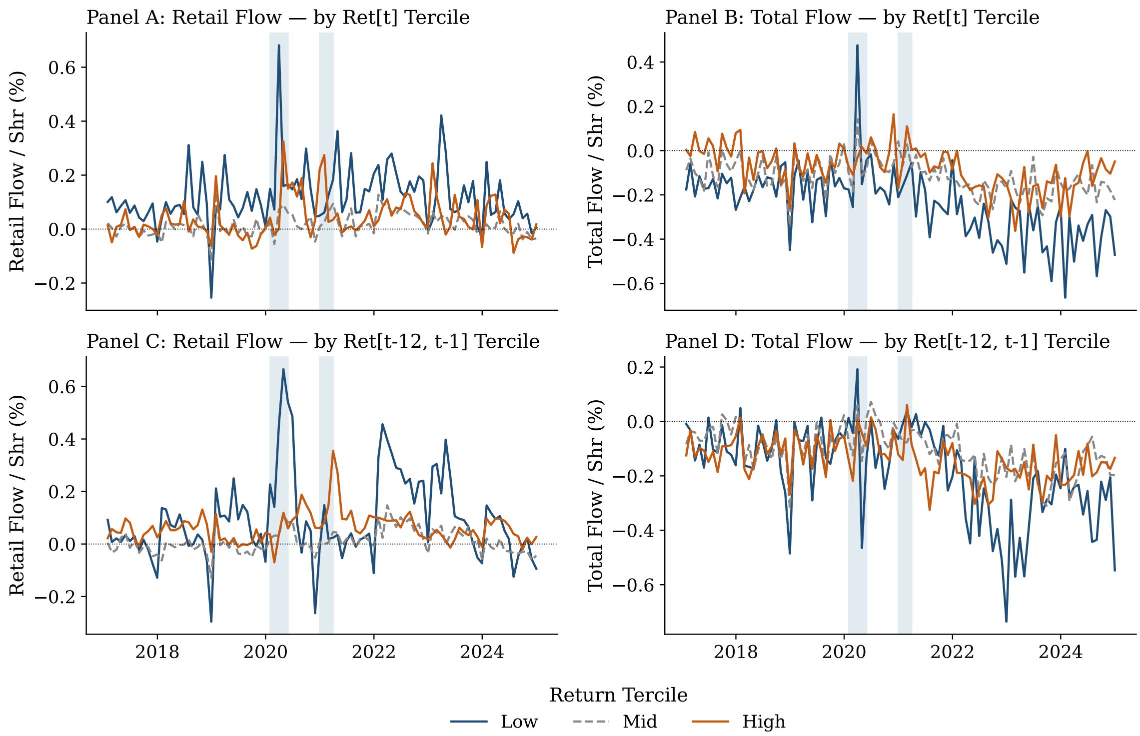

Figure A.1. Retail and Total Flows by Return Tercile, Market-Cap Weighted

Notes: Each panel plots the market-cap-weighted cross-sectional average of monthly net flow scaled by shares outstanding (percent) for stocks sorted into terciles within each calendar month. Weights are stock-level market capitalization at month t. Panels A and B sort by contemporaneous return Ret$[t]$. Panels C and D sort by past twelve-month return Ret$[t-12, t-1]$. Panels A and C plot retail flow; Panels B and D plot total initiated flow. Shaded regions mark the COVID-19 selloff (February–May 2020) and the meme-stock window (January–March 2021). The sample is January 2017 through December 2024. Figure 5 in the main text reports the equal-weighted version.

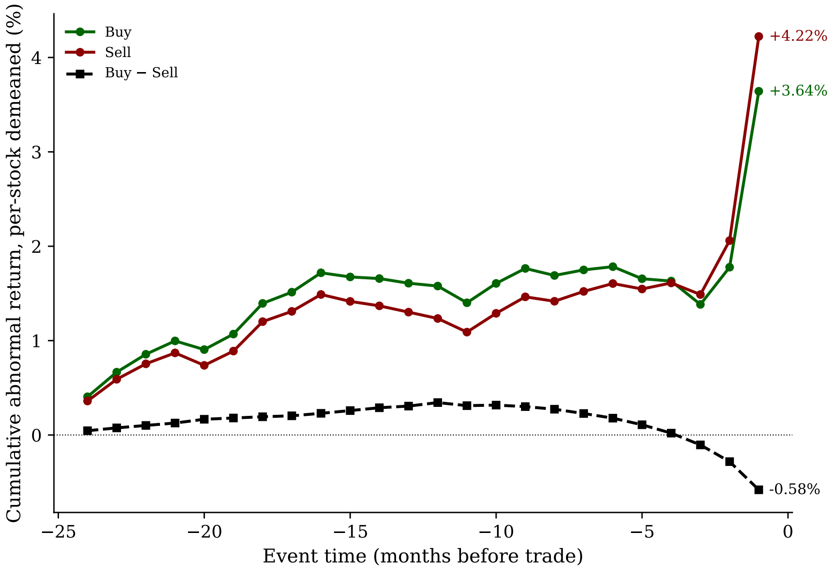

Figure A.2. Pre-Trade Cumulative Market-Adjusted Return: Retail Buys versus Sells

Notes: This figure is the panel analog of Barber et al. (2009b, Figure 3). Each stock-month is an event. The buy path weights each event by retail shares bought scaled by shares outstanding; the sell path weights by retail shares sold scaled by shares outstanding. The plotted quantity is the cumulative market-adjusted return over event months −24 through −1, where the market-adjusted monthly return is the stock return minus the value-weighted panel return that month, winsorized at the 1% and 99% percentiles and then demeaned by stock over the full sample. "Buy − Sell" is the difference of the buy and sell cumulative paths. The sample is January 2017 through December 2024.

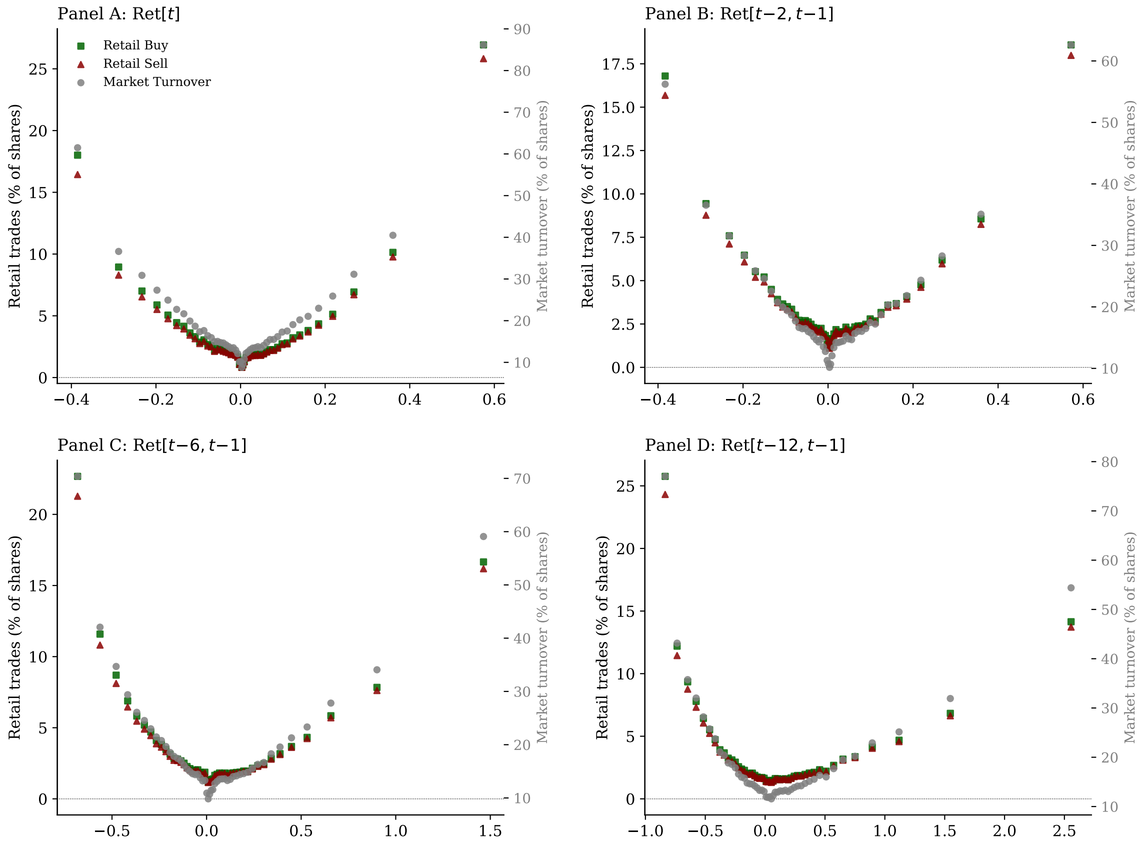

Figure A.3. Retail Buying and Selling Versus Past Returns, with Market Turnover

Notes: Each panel is a binscatter with 50 equal-frequency bins. "Retail Buy" is monthly retail shares bought and "Retail Sell" is monthly retail shares sold, each scaled by shares outstanding (percent) and read on the left axis. "Market Turnover" is total monthly share volume scaled by shares outstanding (percent), read on the right (secondary) axis. Series are pooled across stock-months. The horizontal axis is the past return over the indicated horizon. The past return and each plotted variable are winsorized at the 1% and 99% percentiles within each panel. The sample is January 2017 through December 2024.

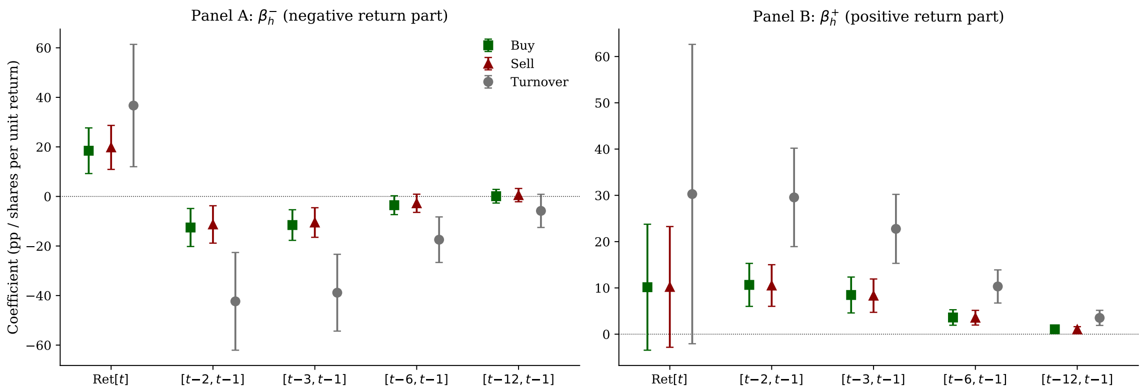

Figure A.4. Loss/Gain-Split Coefficients for Retail Buying, Selling, and Turnover

Notes: Markers report the loss-side coefficients β_h^- (Panel A) and gain-side coefficients β_h^+ (Panel B) from single-horizon panel regressions, splitting each past return into its negative and positive parts. Plotted series are retail buying, retail selling, and market turnover, each scaled by shares outstanding (percent). The net (buying minus selling) coefficients come from the same regressions; they are an order of magnitude smaller and are shown on their own scale in Figure 4. Each regression includes one horizon's negative- and positive-return parts, the composition controls $X_{i,t}$ of equation (1), and stock and time fixed effects. Horizons are $h \in \{0, 1, 3, 6, 12\}$ months. Because net flow equals retail buying minus retail selling and the regressor matrix and sample are identical across the dependent variables at each horizon, the net coefficient equals the buy coefficient minus the sell coefficient by construction. Error bars are 95% confidence intervals from standard errors clustered by stock and by calendar month. Returns are not winsorized. The sample is January 2017 through December 2024.

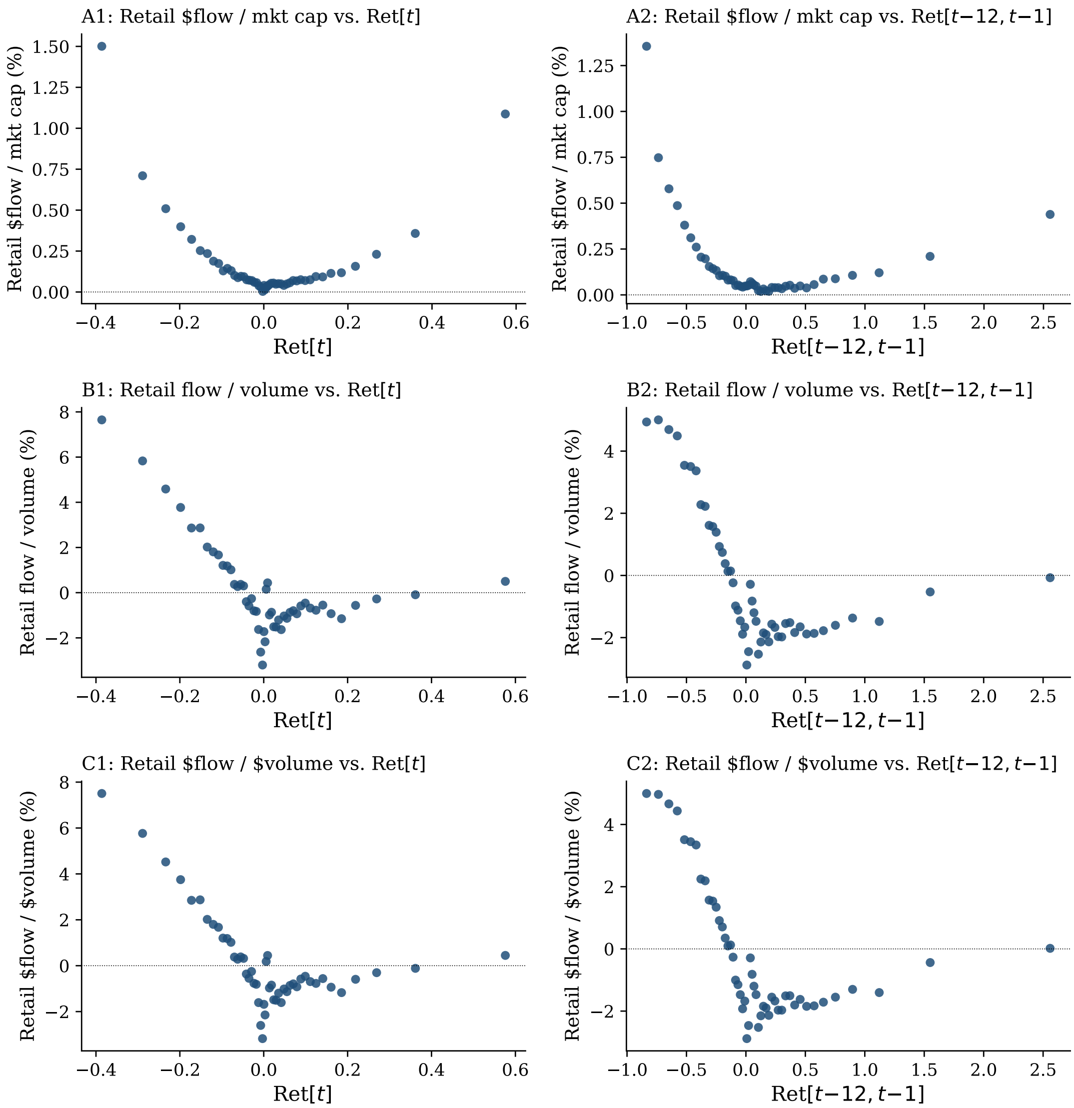

Figure A.5. Buying the Dip Under Alternative Flow Normalizations

Notes: Each panel is a binscatter with 50 equal-frequency bins. The rows use three alternative normalizations of retail net flow: retail dollar flow divided by market capitalization (Row A), retail net share flow divided by retail share volume (Row B), and retail dollar flow divided by retail dollar volume (Row C). The columns use the contemporaneous return Ret$[t]$ and the past twelve-month return Ret$[t-12, t-1]$. Both axes are winsorized at the 1% and 99% percentiles within each panel. The sample is January 2017 through December 2024.

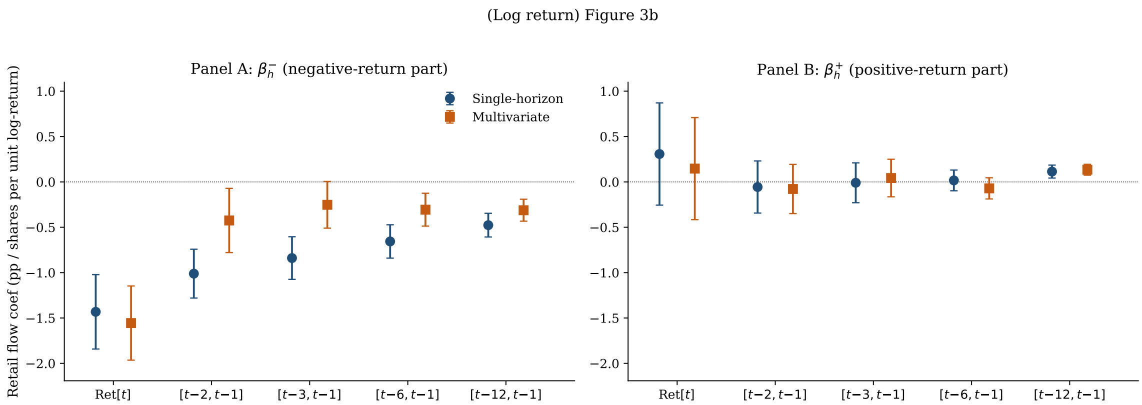

Figure A.6. Buying the Dip With Log Returns

Notes: The figure repeats the loss/gain-split coefficient plot of Figure 4 with each past return r_h replaced by its log transform $\ln(1 + r_h)$, defined where $1 + r_h > 0$. The sign split is preserved because $\ln(1 + r_h) < 0$ if and only if $r_h < 0$. Navy circles are single-horizon estimates and orange squares are the multivariate (all-horizons) estimate, each with the composition controls $X_{i,t}$ of equation (1) and stock and time fixed effects. Error bars are 95% confidence intervals from standard errors clustered by stock and by calendar month. The dependent variable is retail net flow scaled by shares outstanding (percent). The sample is January 2017 through December 2024.

Table A.1. Buying the Dip: Panel Regressions of Retail Flow on Past Returns

	(1)	(2)	(3)	(4)
$\text{Ret}[t]^-$	-3.304***	-1.717***	-1.645***	-1.467***
	(0.258)	(0.268)	(0.272)	(0.258)
$\text{Ret}[t]^+$	1.904***	-0.203	-0.115	-0.057
	(0.298)	(0.273)	(0.299)	(0.307)
$\text{Ret}[t-2, t-1]^-$		-0.716***	-0.733***	-0.669***
		(0.191)	(0.192)	(0.193)
$\text{Ret}[t-2, t-1]^+$		0.015	0.077	0.002
		(0.125)	(0.141)	(0.129)
$\text{Ret}[t-3, t-1]^-$		-0.522***	-0.409**	-0.368**
		(0.167)	(0.172)	(0.165)
$\text{Ret}[t-3, t-1]^+$		0.071	0.088	0.149
		(0.131)	(0.155)	(0.155)
$\text{Ret}[t-6, t-1]^-$		-0.500***	-0.545***	-0.469***
		(0.117)	(0.121)	(0.112)
$\text{Ret}[t-6, t-1]^+$		-0.001	-0.013	-0.026
		(0.029)	(0.033)	(0.034)
$\text{Ret}[t-12, t-1]^-$		-0.175**	-0.171**	-0.158**
		(0.076)	(0.075)	(0.069)
$\text{Ret}[t-12, t-1]^+$		0.023*	0.026*	0.023*
		(0.012)	(0.013)	(0.013)
Idio. Volatility		23.814***	23.985***	22.663***
		(2.211)	(2.184)	(2.335)
Idio. Skewness		-0.035***	-0.041***	-0.039***
		(0.010)	(0.010)	(0.009)
Beta		-0.081***	-0.085***	-0.085***
		(0.022)	(0.022)	(0.021)
Log(ME)			0.025	0.028
			(0.032)	(0.029)
B/M			0.002	0.022
			(0.025)	(0.015)
Debt/Assets			0.170	0.145
			(0.113)	(0.098)
NI/Assets			-0.036	-0.024
			(0.034)	(0.024)
Age			-0.002	-0.004
			(0.002)	(0.004)
Gross Profitability			-0.040	-0.079*
			(0.032)	(0.040)
Δ Inst. Ownership				-0.031
				(0.027)
InstSec Flow / Shr				7.529***
				(1.613)
Inst50k Flow / Shr				0.650
				(5.071)
N	414,239	316,428	298,709	285,742
R^2 (within)	0.03	0.09	0.09	0.09
Stock FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

Notes: Sample is January 2017 through December 2024 ($N \approx$ 414,000 stock-months, 7,392 stocks, 96 months). The dependent variable is monthly retail net order flow as a percentage of shares outstanding (in pp). Past returns are in fractions, so a coefficient of β on $\text{Ret}[\cdot]^-$ means a 100pp move in the negative-return state shifts retail flow by β pp of shares outstanding. For any return r, $r^- \equiv \min(r, 0)$ and $r^+ \equiv \max(r, 0)$ denote its negative and positive parts; $\text{Ret}[\cdot]^-$ and $\text{Ret}[\cdot]^+$ apply this split to the indicated return. All specifications include stock and time fixed effects. Standard errors are clustered by stock and by time and reported in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels.