

Buying the Dip in Retail Trading

Xiao Yin and Dongchen Zou *

July 2026

Abstract

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 persists throughout the sample period. A structural model with retail contrarian trading and a short-sale constraint reproduces the asymmetry. The model predicts that greater retail participation initially reduces volatility through dip-buying but eventually raises it by displacing institutional risk-bearing capacity.

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 Rob Battalio, Xavier Gabaix, Nick Roussanov, John Shim, and seminar participants at the Chicago Booth RP Reunion conference, Wabash River Finance 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. A one-standard-deviation decline in the current (prior) month's return is worth about 0.34 (0.16) percentage points of shares outstanding. This pattern holds across firm size, value, and market beta groups. It also holds in periods of both high and low institutional flows, suggesting that retail dip-buying is not solely an accommodation of institutional trading.

The pattern of retail buying the dip is 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

¹This 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.

bottom-tercile flow is 0.7. The behavior persists across the COVID-19 selloff and the meme-stock episode in 2021. Moreover, retail investors in aggregate also buy the dip of the market: a 10 percent decline in the S&P 500 is followed by about 8 basis points of aggregate market value in extra retail net buying in the following month, roughly \$35 billion, and nearly twice the lagged buying that follows a rise.

The dip-buying pattern is a specific feature of retail flow. We run the same exercise on four non-retail flow proxies: a large-trade institutional proxy, the change in 13F institutional ownership, the change in lendable supply, and total initiated order flow. On the loss side, all these measures slope upward: as past returns fall, their flow falls with them. In other words, non-retail investors are “selling the dip”.

To examine the market implications of buying the dip, we develop a structural model disciplined by two features of retail trading: the asymmetry across loss and gain states, and the response to both current and past returns. Retail investors expect reversals, so lower current or past returns raise their perceived gains from buying. A short-sale constraint translates these symmetric contrarian beliefs into asymmetric demand: investors increase purchases after losses but choose zero holdings when expected gains are insufficient to justify buying. Institutions absorb the remaining supply according to their valuations of future dividends and their capacity to bear risk. We estimate the model by simulated method of moments, matching the current- and past-return dip-buying slopes and additional moments of flows and price informativeness.

With the model disciplined to the dip-buying facts, we ask how this behavior affects price dynamics. The natural conjecture is that buying the dip stabilizes prices. Retail investors buy at the time prices are falling, absorbing selling pressure and reducing the price adjustment required for markets to clear. This liquidity provision should cushion declines and dampen subsequent reversals, echoing the “stabilizing speculation” argument of [Friedman \(1953\)](#). The model supports this conjecture for buying the current dip. Because retail demand responds to the contemporaneous price, it absorbs pressure as it arises. In the calibrated model, increasing retail investors’ expectation sensitivity to current returns significantly reduces the initial

decline following positive supply shocks.

Meanwhile, buying the past dip has a different effect. A previous price decline encourages purchases by retail traders in the next period. These purchases strengthen the rebound. This delayed response can generate further reversals, amplifying return volatility following exogenous return changes. In the calibrated model, increasing retail investors' expectation sensitivity to past returns raises the rebound following initial price decline largely compared with the benchmark case. Thus, contrarian demand can cushion the initial adjustment or amplify subsequent fluctuations, depending on which returns investors respond to.

A larger retail share combines these demand effects with a reduction in institutional risk-bearing capacity. Greater retail capacity strengthens dip buying, while lower institutional capacity makes prices more sensitive to the supply institutions must absorb. Their interaction produces a shallow U-shaped relation between retail share and unconditional return volatility. The market implications of retail participation therefore depend jointly on the timing of contrarian demand and the institutional capacity it replaces.

Related Literature. This paper relates to three strands of literature. 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 verify it against publicly released retail trading statistics during the GameStop episode.

The second strand concerns retail investor behavior. Prior work documents that retail trading is attention-driven (Barber and Odean, 2008; Barber et al., 2022) and contrarian at short horizons (Kaniel et al., 2008; Kelley and Tetlock, 2013). A large literature documents stable cross-sectional tilts toward high idiosyncratic volatility and positively skewed payoffs (Mitton and Vorkink, 2007; Kumar, 2009; Han and Kumar, 2013; Bali et al., 2011) and toward financially distressed, lottery-like stocks (Campbell et al., 2008; Conrad et al., 2014; Barberis and Huang, 2008). Recent studies find that retail risk-taking is state-contingent, but on variables other than the stock's own past return: low rates push investors toward high-income assets (Daniel et al., 2021), and underperformance against a salient benchmark prompts asymmetric risk-taking (Gelman et al., 2026). Our dip-buying pattern differs in three respects. First, retail flow responds asymmetrically to past returns, rising steeply after losses but staying flat after gains, whereas a linear specification averages the two sides and labels retail simply as contrarian. Second, the pattern survives controls for lottery demand and firm distress, pointing to a distinct mechanism. Third, it persists at horizons up to twelve months, beyond the short horizons at which contrarian retail trading has been documented.

The third strand concerns retail liquidity provision and market outcomes. 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. In addition, retail traders could destabilize the market if they chase the trend as in De Long et al. (1990). Our structural model connects to these studies and identifies retail traders as a source of demand during market downturn. The model highlights the distinct role of buying the current dip versus buying the past dip. In particular, While buying the current dip cushions downward price pressure, buying the past dip amplifies reversal after short-lived supply shocks.

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 comprehensive. 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

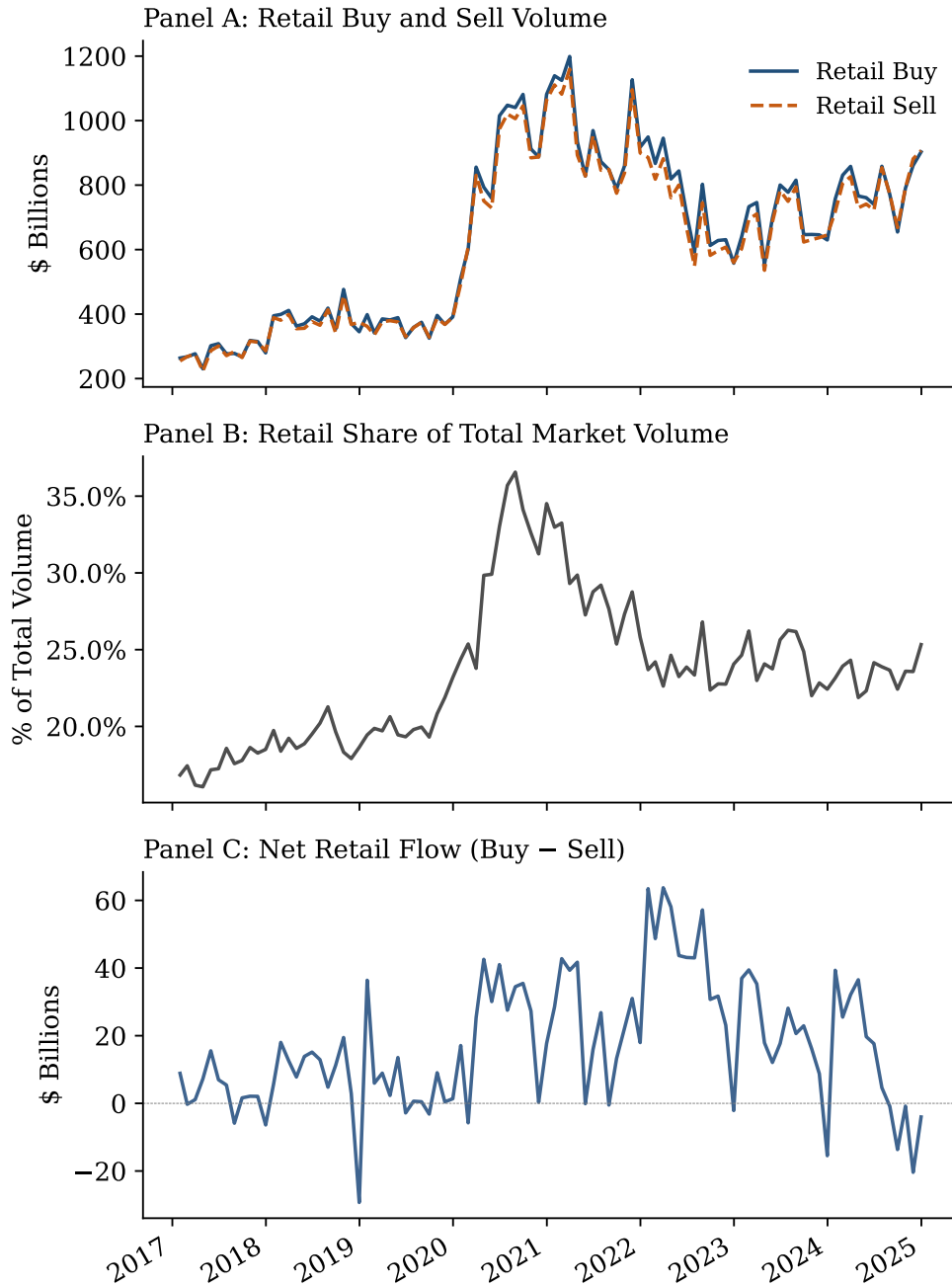


Figure 1. Retail Flows and Volumes. Panel A: aggregate monthly retail buy and sell volume (\$ billions). Panel B: retail share of total market dollar volume (percent). Panel C: aggregate monthly net retail dollar order flow (\$ billions), computed as buy volume minus sell volume. The sample covers January 2017 through December 2024. Retail flow is from the S&P Global Retail Flow Database.

retail behavior while preserving a common, replicable measurement standard.²

We supplement the retail flow data with Trade and Quote (TAQ) data from WRDS

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

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 also construct five benchmark flow measures for retail and non-retail flows. 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, $cf(p) = 100 \times \text{mod}(p, 0.01) \notin \{0, 1\}$, 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 $cf(p) \geq 0.6$ are classified as retail buys, trades with $cf(p) \leq 0.4$ are classified as retail sells, and trades with $cf(p) \in (0.4, 0.6)$ 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). The *BJZZ* rule has two documented limitations. It identifies fewer than one-third of actual retail trades ([Battalio et al., 2023](#)), and its cent-fraction signing is reliable only when the spread is one cent; with a wider spread the same cent fraction can sit near the bid or near the ask ([Barber et al., 2024](#)).³ *BHJOS* is the modification proposed by [Barber et al. \(2024\)](#): it keeps the sub-penny identification but signs each trade by its position within the quoted spread, buys in the upper 40 percent, sells in the lower 40 percent, the middle excluded, and it drops trades above \$100,000 in notional value.

Appendix [B](#) validates the S&P measure against the daily retail flow statistics that Citadel Securities published for GameStop in January 2021, and compares it with both proxies. The S&P series matches the Citadel statistics in sign and in magnitude,

³In Appendix [B](#), we demonstrate the *BJZZ* signing problem with an example.

while the TAQ-based proxies miss both. The proxies align best with actual retail flow among small stocks and past losers, with alignment deteriorating after 2021.

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 proxy rather than 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 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.

Appendix Table B.2 reports the time-series and cross-sectional correlations between S&P Retail Flow and each of these flow measures. Retail flow is nearly uncorrelated with BJZZ in the aggregate time series, correlated about 0.5 with both TAQ-based retail proxies across stocks, and negatively correlated with ΔIO in the time series.

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

Table 1. Summary Statistics

This table reports summary statistics for the sample, which covers January 2017 through December 2024 ($N = 417,519$ stock-months, 7,432 stocks, 96 months). Panel A reports flow variables, each measured as monthly net order flow divided by shares outstanding (or by market equity for the Mkt Cap row) and 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 reports return variables, where $Ret[t - k, t - 1]$ is the cumulative return from month $t - k$ to $t - 1$, and 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 reports firm characteristics, with market equity in millions of dollars and firm age in months since first CRSP appearance.

Variable	N	Stocks	Mean	Std	p10	p25	p50	p75	p90
<i>Panel A: Flow Variables (normalized $\times 100$, percent of denominator)</i>									
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
<i>Panel B: Return Variables</i>									
$Ret[t]$	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
<i>Panel C: Firm Characteristics</i>									
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

is persistent across the sample period and holds across various horizons, while non-retail flow measures on average move in the opposite direction.⁴

⁴The “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.

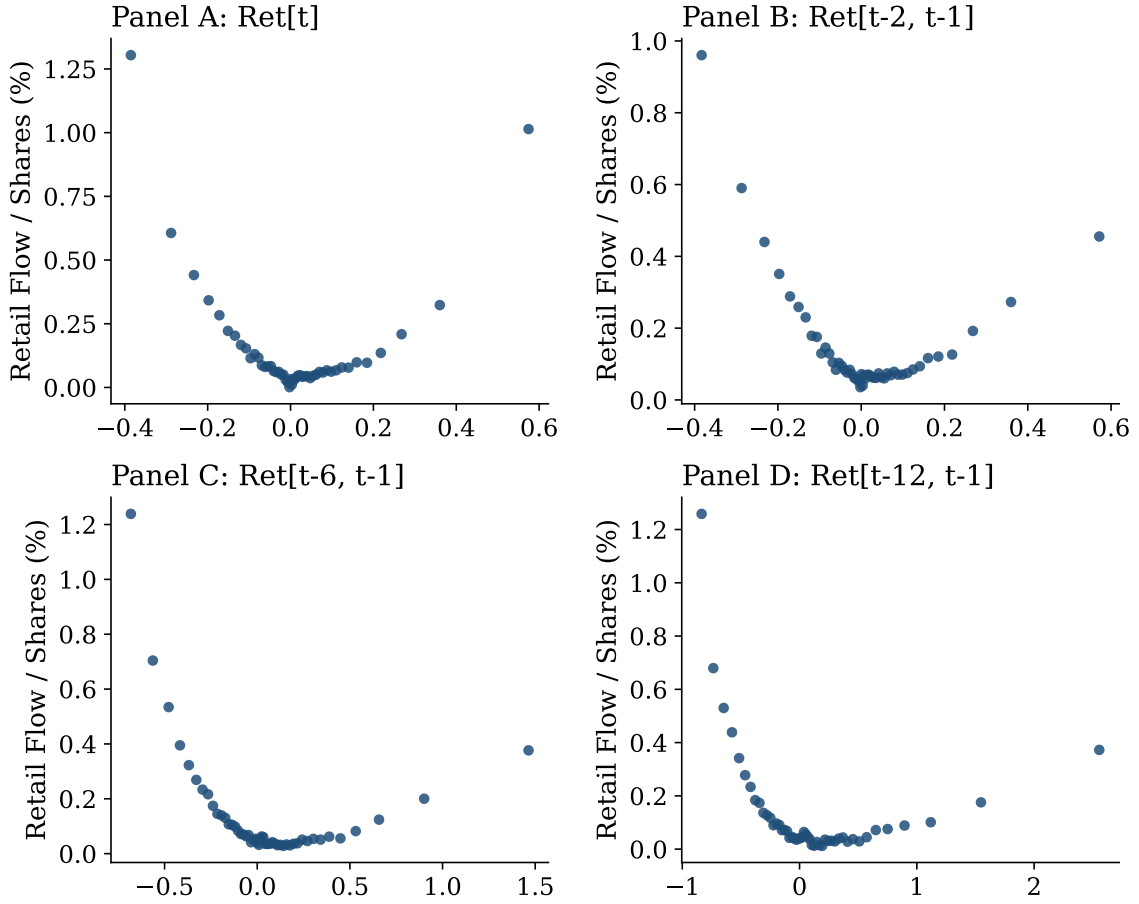


Figure 2. Buying the Dip: Retail Net Flow Versus Returns Across Horizons. Each panel is a binscatter with 50 equal-frequency bins. The vertical axis is monthly retail net order flow scaled by shares outstanding (percent). The horizontal axis is the past return measured over the indicated horizon. Panel A uses the contemporaneous return $\text{Ret}[t]$; 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.

3.1 Retail Response to Returns

Figure 2 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 $\text{Ret}[t]$. Panels B, C, and D use lagged returns over windows ending at $t - 1$: the past month $\text{Ret}[t - 2, t - 1]$, the past six months $\text{Ret}[t - 6, t - 1]$, and the past twelve months $\text{Ret}[t - 12, t - 1]$. 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 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 3.6 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

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

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 2.2 (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.

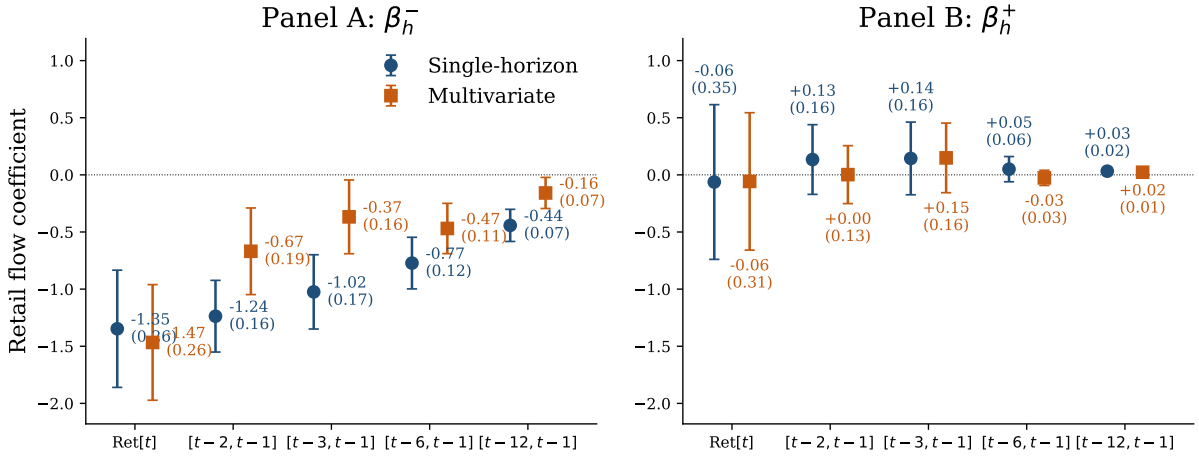


Figure 3. Regression Coefficients of Retail Flows on Returns Across Horizons. Panel A reports the loss-side coefficients β_h^- and Panel B the gain-side coefficients β_h^+ from estimates of equation (1), for horizons $h \in \{0, 1, 3, 6, 12\}$ months. Navy circles (“Single-horizon”) come from a separate 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.

Figure 3 visualizes the regression coefficients across horizons.⁵ Three features 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 approx-

⁵The 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.

imately -1.3 , indicating that a 1-percentage-point decline in $\text{Ret}[t]$ 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 $\text{Ret}[t]$ is 22.8 percentage points and that of $\text{Ret}[t - 12, t - 1]$ is 92.9 percentage points (Table 1). Using the column (4) estimates, a one-standard-deviation negative shock to $\text{Ret}[t]$ 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 2 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 3.

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. 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 4 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

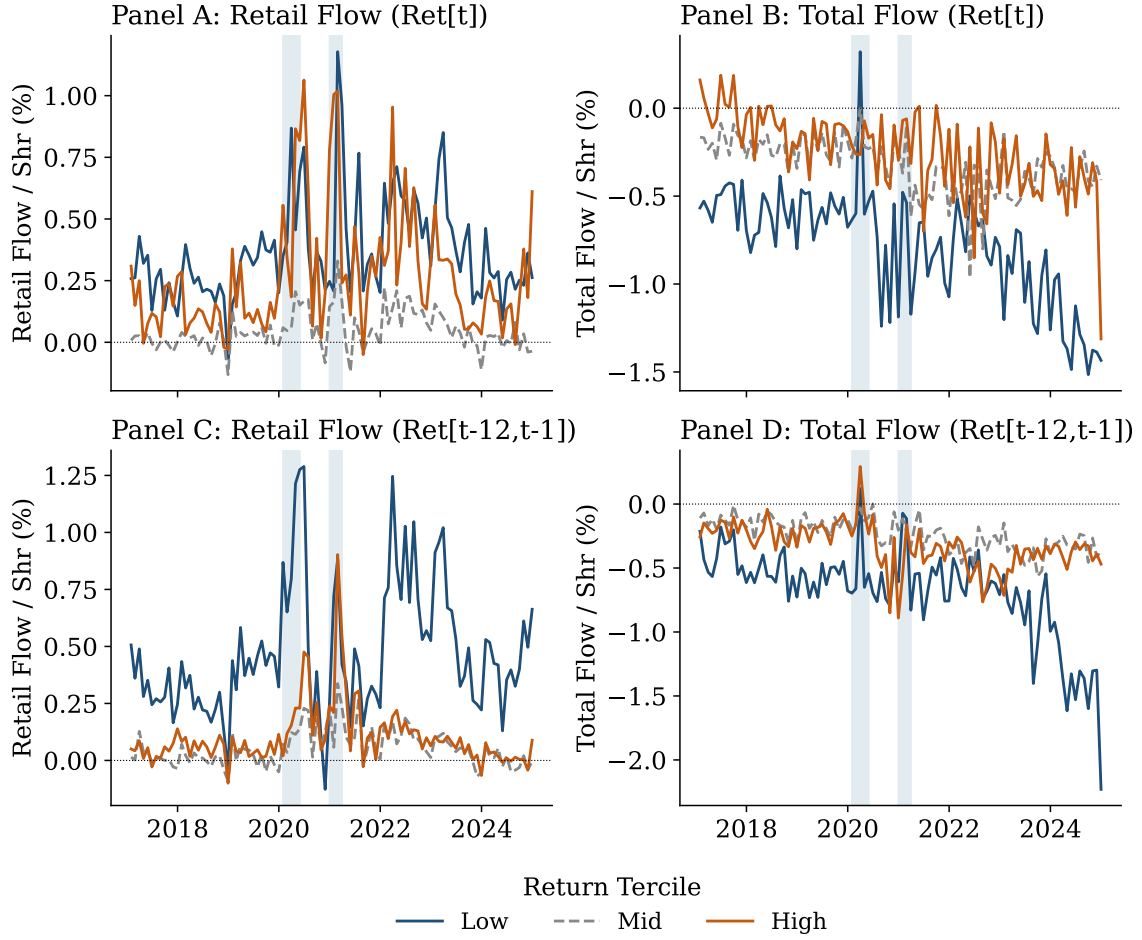


Figure 4. Retail and Total Flows by Return Tercile. Each panel plots the equal-weighted cross-sectional average of monthly net flow scaled by shares outstanding (percent) for stocks sorted into terciles within each calendar month. Panels A and B sort by contemporaneous return $\text{Ret}[t]$. Panels C and D sort by past twelve-month return $\text{Ret}[t - 12, t - 1]$. 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.

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 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 $\text{Ret}[t]$ 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 $\text{Ret}[t - 12, t - 1]$, 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 4. 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 4 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 4, 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 Aggregate Flow and the Market Return

Do retail buy the market as a whole after it falls? We sum retail net dollar flow and total order flow across all stocks in each month, scale both by the aggregate market

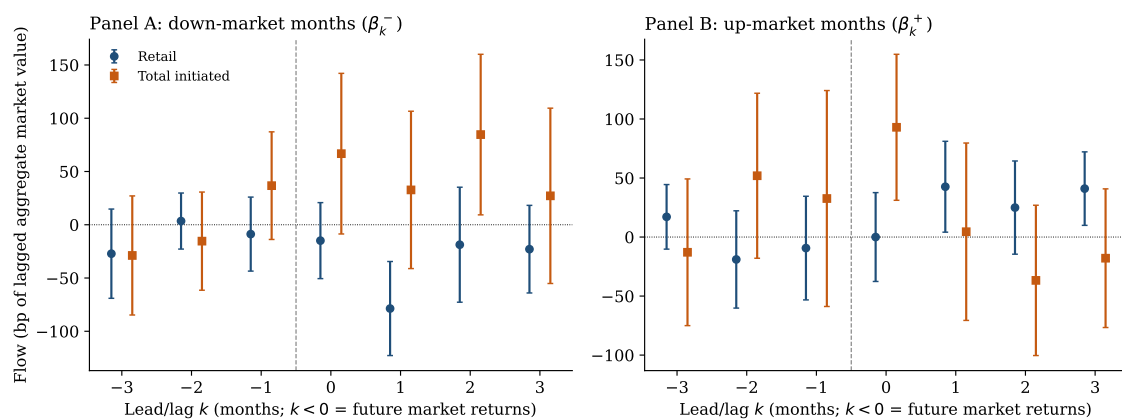


Figure 5. Aggregate Retail and Total Flow Around Market Returns. Coefficients from regressions of aggregate monthly retail net dollar flow (circles) and aggregate total order flow (squares), each in basis points of the previous month's aggregate market value, on the S&P 500 monthly total return split at zero, with three leads ($k < 0$) and three lags ($k > 0$) of single-month returns. Panel A reports the coefficients on the negative part of the return at lead or lag k ; Panel B those on the positive part. A negative coefficient in Panel A means more net buying after a decline. Retail flow is net retail buying; total flow is buy-initiated minus sell-initiated dollar volume across all trades (Section 2.2). Bars are 95 percent intervals from Newey-West standard errors with three lags. The sample is January 2017 through December 2024.

value at the end of the previous month, and regress each on the S&P 500 monthly total return split at zero, with three leads and three lags of single-month returns.⁶ Retail are net buyers of stocks in most months of the sample (Figure 1, Panel C); the question here is when that buying arrives relative to market moves.

Figure 5 shows that retail investors in aggregate buy the dip of the market. In a month in which the market falls, retail flow does not move in that month. Retail buy in the following month. The response is economically significant: a 10 percent market decline is followed by about 8 basis points of extra retail net buying in the next month, roughly \$35 billion. In a month in which the market rises, retail follow the rise with buying spread over the next three months. Retail flow does not respond to future market returns.

Total flow moves on a different clock and in different directions. Total flow moves with the market and in the same direction as the return on both sides. It is net selling in the month of a decline and net buying in the month of a rise. Retail flows move with a one-month lag, against the return after a decline while with it after a rise. The

⁶Flows are in basis points of lagged aggregate market value. Standard errors are Newey-West with three lags. The sample is January 2017 through December 2024.

retail response is also asymmetric in amount: the buying one month after a decline is nearly twice the lagged buying after a rise. These results reaffirm the asymmetric and delayed contrarian response of retail investors in the aggregate time series.

3.4 Other Investor Flows

Sections 3.1 to 3.3 establish that retail flow rises after price declines, that the pattern persists through the full sample, and that it holds for the market as a whole. 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 2 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 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.

The patterns of non-retail flows can invite a mechanical reading: institutions

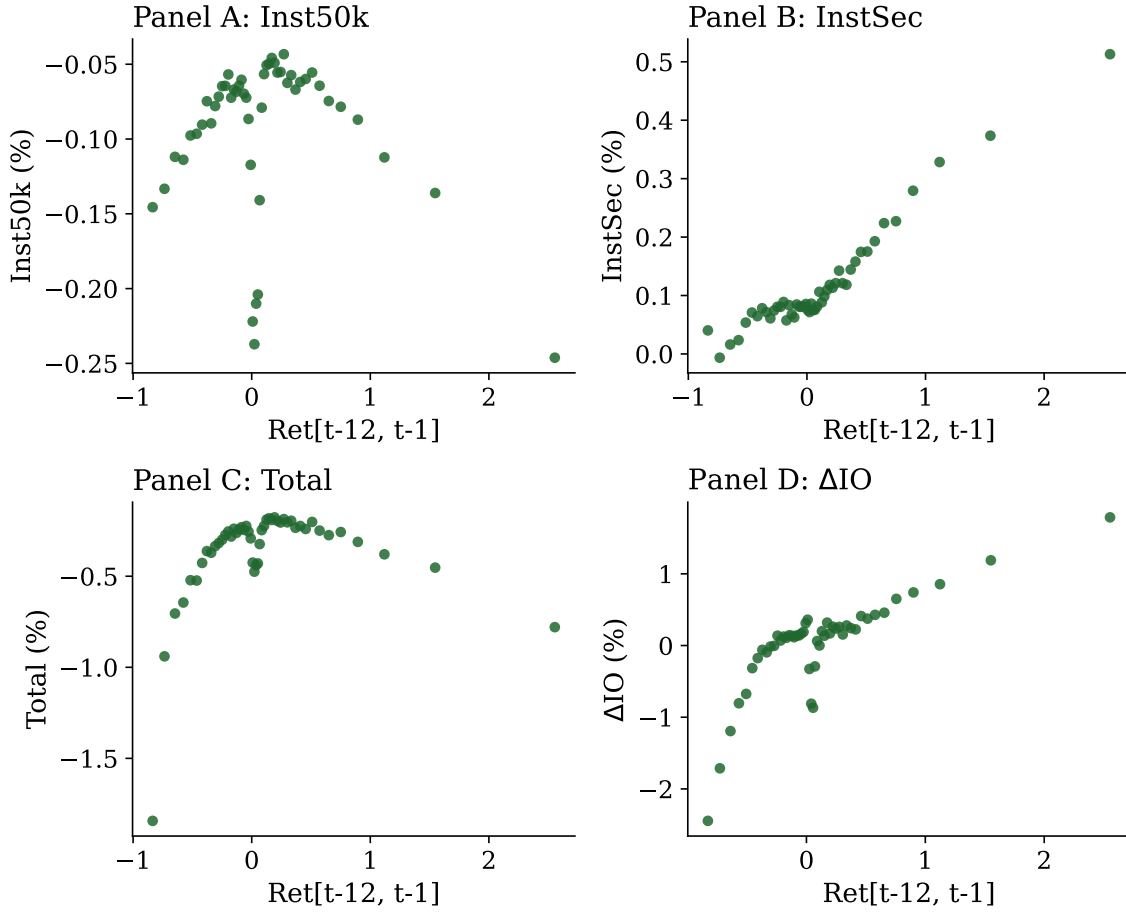


Figure 6. Other Net Flow Measures Versus Past Return. Each panel is a binscatter with 50 equal-frequency bins. The vertical axis is the indicated monthly net flow scaled by shares outstanding (percent). The horizontal axis is the past twelve-month return $\text{Ret}[t - 12, t - 1]$. Panel A: large-trade 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 2.2. Both axes are winsorized at the 1% and 99% percentiles within each panel. The sample is January 2017 through December 2024.

sell losers for reasons of their own, retail take a fixed share of whatever institutions sell, and the retail is a counterpart that mirrors institutional selling. Three results, reported in Appendix C, suggest that retail dip-buying is significantly deviated from passive absorption of institutional selling. First, the retail demand on returns is the same in stocks that institutions are buying and in stocks they are selling in the same month. Splitting stock-months at the within-month median of each institutional flow proxy, the loss-side slope is about as steep in the two groups, and adding the proxies as controls in Table A.1 moves it by about a tenth. Second, using mutual funds' flow-induced trading pressure in Lou (2012) as indicators for institutional trades,

prices and institutional ownership move monotonically, but leaves retail flow with a flat and asymmetric pattern, in reminiscence of buying the dip in returns: retail investors purchase when funds are forced to sell, but they do not sell into forced fund buying. Third, part of what retail accumulate in the persistent buying actually comes from firms rather than from institutions: retail absorb about six percent of newly issued shares in the month of issuance, whether or not the price fell. Retail can be counterparties to institutional sellers, but their demand actively responds to the price, whichever side of the market institutions are on. The model of Section 4 builds on this assignment of roles.

3.5 Heterogeneity Across Stocks

Does buying the dip hold within different groups of stocks, and is the effect stronger for some groups? We split stocks each month at the within-month median of three characteristics measured at $t - 1$: size, book-to-market, and idiosyncratic volatility. Figure 7 applies the binscatter design of Figure 2 within each group at the contemporaneous and past twelve-month horizons. Appendix Figure A.7 estimates the single-horizon specification of equation (1) within each group at the horizons of Figure 3.

Buying the dip is present in every group. In Figure 7, retail flow rises as losses deepen in all six groups at both horizons, and the loss-side coefficient is negative and significant in every group at every horizon. Retail's dip-buying response is stronger among small, volatile, and high book-to-market stocks, consistently across return horizons. Large stocks nonetheless respond at every horizon, which restates the value-weighted evidence of Section 3.2 that buying the dip is not a micro-cap phenomenon.

At the contemporaneous horizon the positive-side upward tilt of Panel A of Figure 2 reappears only in the small and volatile groups, while large and low-volatility stocks show a flat gain side. In the regressions, the gain-side response coefficients are muted in most groups. At the past twelve-month horizon the gain side is flat in all six groups: stocks with positive past returns receive little net retail flow whether they are small or large, value or growth, volatile or not. The results suggest that the tilt is

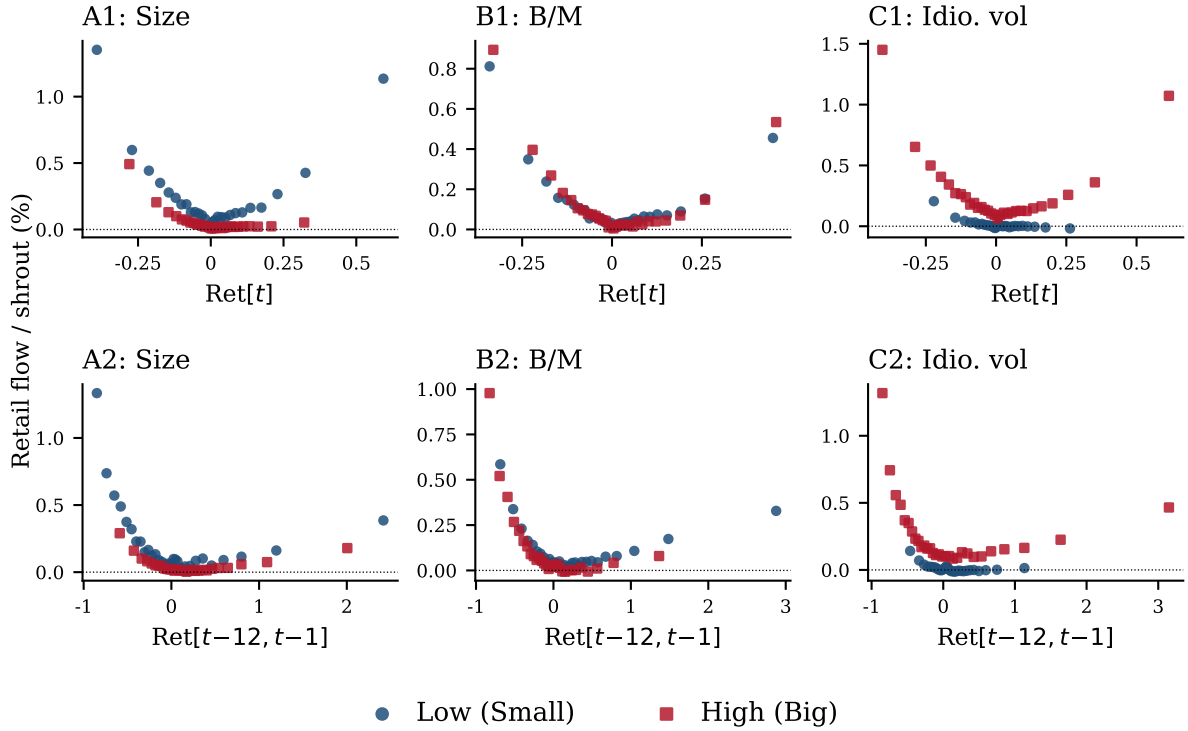


Figure 7. Buying the Dip by Stock Characteristic. Each panel is a binscatter plot grouped by stock characteristics. Columns split stocks at the within-month median of market equity (Panels A), book-to-market (Panels B), and idiosyncratic volatility (Panels C) respectively, each measured at $t - 1$. The vertical axis is monthly retail net order flow scaled by shares outstanding (percent). The horizontal axis is the contemporaneous return $\text{Ret}[t]$ (top panels) or the past twelve-month return $\text{Ret}[t - 12, t - 1]$ (bottom panels). Both axes are winsorized at the 1% and 99% percentiles within each group and panel. The sample is January 2017 through December 2024.

the composition effect, where net buying of volatile small stocks in months when they happen to rise, rather than being a response to the gain. In the regressions, gain-side coefficients are smaller relative to their loss-side counterparts in every group, and no group significantly buys winners. The exception is high book-to-market stocks, which are consistently net sold after past gains at the one- to six-month horizons. Since book-to-market at $t - 1$ embeds the past price decline, this group is largely made up of stocks that have already fallen; retail buy further declines in these stocks and sell their rebounds.

3.6 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, $\text{RetailFlow} = \text{RetailBuy} - \text{RetailSell}$, 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 phenomenon. Appendix Figure A.3 plots bin-averaged retail buying and selling against past returns at four horizons, with market turnover on a secondary axis. The gross response is not monotone: buying and selling are both elevated when the past return is large in absolute value, of either sign, and lowest near zero. Thus, a single average slope obscures the asymmetry. Moreover, the buy and sell legs are nearly coincident at every horizon, and market turnover traces the same shape, indicating a common turnover component in both legs. Appendix Figure A.4 estimates equation (1) separately for retail buying and selling, and market turnover, splitting each past return into its negative and positive parts. Gross retail purchases are contrarian to losses while chasing gains, and the buying, selling, and turnover coefficients are similar at each horizon. The common component is clearest in the gross-versus-net comparison: 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, while netting the two legs leaves a loss-side coefficient of -1.24 with a standard error of 0.16. The roughly twenty-five-fold collapse in the standard error confirms that a large common component 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.

3.7 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 2 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 3 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.

3.8 Relation to Prior Empirical Evidence

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 we document 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 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](#)). Our stylized fact conditions retail demand on contemporaneous and past returns, which remain robust after controlling for fundamental distress or jackpot characteristics.

3.9 Alternative explanations

The model delivers the two features of dip-buying, the asymmetry across loss and gain states and the response to signed past returns at multiple horizons, from a single set of primitives. We now consider whether alternative behavioral mechanisms in the prior literature reproduce both. Some of the mechanisms fail because they generate

symmetric response in the sign of the return. The attention or salience framework of Barber and Odean (2008), Barber et al. (2022), van der Beck et al. (2026), and Graves (2024) predicts retail buying of attention-grabbing stocks identified by extreme returns, news, or volume, so it generates a response to $|r_{i,t}|$ rather than to $r_{i,t}$, and the attention measures decay too quickly to match the longer horizon. Mean-reverting beliefs, in which retail expects prices to revert toward fundamentals (Cohen et al., 2002; Bastianello and Fontanier, 2024, 2025), matches retail response at the loss side, but predicts retail selling after gains that is absent in the data. Extrapolative beliefs (Barberis et al., 2015) can generate retail responses to past returns but fail to create contrarian trading. Our model delivers the asymmetry from a complementary route, a short-sale constraint layered on a mean-reversion belief that is itself symmetric in the supply-pressure state.

The other candidates fail because they are not return-conditional or predict the wrong sign. Habitat and lottery preferences describe stable cross-sectional tilts toward hard-to-value, high-volatility, or high-skewness stocks (Laarits and Sammon, 2025; Balasubramaniam et al., 2023; Dorn and Huberman, 2010; Kumar, 2009; Han and Kumar, 2013; Bali et al., 2011), not flow conditioned on the signed past return. Appendix Table A.1 further shows retail flow loading negatively on idiosyncratic skewness, the opposite of the lottery prediction. The disposition effect (Shefrin and Statman, 1985; Odean, 1998) predicts that retail selling slackens after losses, yet the buy-sell decomposition in Section 3.6 shows the loss-side gross sell response rising rather than falling, close in magnitude to the loss-side buy response (Appendix Figures A.3 and A.4). Each mechanism may contribute to the level of dip-buying in particular states or stocks, but none reproduces the asymmetric, multi-horizon shape that the single-mechanism model delivers.

4 A Model of Buying the Dip

Section 3 documents that retail investors buy more following negative current and past returns, while their purchases respond much less to positive returns. In this section, we develop a model to study how this behavior affects price dynamics.

4.1 Setup

The economy has discrete time and an infinite horizon. A single risky asset trades each period and pays dividend v_t . Its ex-dividend price is p_t , and return is defined as

$$r_t = p_t - p_{t-1}.$$

Dividends contain persistent and transitory components:

$$\begin{aligned} v_t &= \bar{v} + \theta_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2), \\ \theta_t &= \rho_\theta \theta_{t-1} + \sigma_\theta z_{\theta,t}, & z_{\theta,t} &\sim N(0, 1), \quad |\rho_\theta| < 1. \end{aligned}$$

The stationary component of the asset supply absorbed by active investors follows

$$\lambda_t = \bar{\lambda} + \rho_\lambda (\lambda_{t-1} - \bar{\lambda}) + \sigma_\lambda z_{\lambda,t}, \quad z_{\lambda,t} \sim N(0, 1), \quad |\rho_\lambda| < 1.$$

All innovations are mutually independent and independent over time. A positive supply innovation increases the quantity investors must absorb, while persistent dividend news changes expected future cash flows.

There is a continuum of investors of mass one. Retail investors have mass $\alpha \in (0, 1)$ and institutional investors have mass $1 - \alpha$. The return-dependent component of a retail investor's expected gain is

$$\mu_t^R = \chi - b_c r_t - b_p r_{t-1}, \quad b_c, b_p \geq 0. \quad (2)$$

The evidence on whether investors form extrapolative or contrarian return expectations is mixed. [Greenwood and Shleifer \(2014\)](#) document that surveyed return expectations increase with past market returns. Other studies including [Dominitz and Manski \(2011\)](#), [Laudenbach et al. \(2026\)](#), and [Nagel and Xu \(2023\)](#) document contrarian return expectations among professional and retail investors. We adopt this parsimonious forecasting rule to capture current and past dip buying without imposing a particular mechanism of belief formation, which our trading data do not separately identify.

The forecasting rule (2) nevertheless admits an interpretation based on the law of small numbers: investors expect recent declines to be offset by subsequent gains, even when the underlying sequence is approximately independent. [Rabin \(2002\)](#) and [Rabin and Vayanos \(2010\)](#) formalize this tendency to expect reversals in random

sequences.⁷ A lower current return raises the expected gain through b_c , while a lower previous return does so through b_p . Note that μ_t^R is a function of current price. Thus, an implicit assumption of (2) is that retail traders do not consider their own price impact. Equation (2) allows us to vary these responses separately, holding the remaining parameters fixed, and quantify their respective effects on equilibrium price dynamics.

The belief rule governs new purchases, while the persistence of net retail demand also depends on when existing positions are liquidated. Each period, an incumbent retail investor exits with probability $\delta \in (0, 1]$, independently across investors and over time, and independently of dividend and supply innovations. Exiting investors liquidate their positions and are replaced by an equal mass of entrants. Entrants choose their active positions once and retain them until exit; surviving incumbents do not rebalance. The holding duration T therefore satisfies

$$\Pr(T = k) = \delta(1 - \delta)^{k-1}, \quad k \geq 1, \quad E[T] = \frac{1}{\delta}.$$

A position purchased at date t first faces liquidation at $t + 1$. The constant exit probability provides a parsimonious representation of variation in holding durations and gradual liquidation. This matters because the empirical object is net retail buying: purchases responding to past losses can be offset by sales of previously acquired positions. Allowing positions to survive beyond one period separates dip buying from the unwinding of existing holdings and helps the model jointly match contemporaneous and lagged flow responses.

Institutions observe the persistent dividend state θ_t and use a constant valuation discount factor $\beta \in (0, 1)$. Prices and discounted cash flows are expressed in date- t money units. Within a period, dividend and supply innovations are realized first. Exiting retail investors then liquidate, and a new cohort enters. Entrants and institutions submit price-contingent demands, and the price adjusts to clear the market. Purchases and liquidations occur at the same price.

⁷In Rabin (2002), an observer treats independent signals as draws without replacement from a finite urn whose composition reflects the population distribution. A recent excess of low outcomes therefore raises the perceived probability of high outcomes in subsequent draws.

4.2 Optimality conditions

Let $\eta_R > 0$ be individual retail risk-bearing capacity. Given the expected gain in equation (2), an entrant solves

$$\max_{q \geq 0} \left\{ \mu_t^R q - \frac{q^2}{2\eta_R} \right\}.$$

The quadratic term represents the cost of bearing risk. Under a subjective normal-payoff interpretation with fixed perceived variance σ_R^2 , individual capacity is $\eta_R = (\gamma_R \sigma_R^2)^{-1}$. We maintain this cost separately from the belief distortion; random exit and the retail constraint need not produce normally distributed equilibrium payoffs.

The necessary and sufficient optimality conditions are

$$q_t^R \geq 0, \quad \frac{q_t^R}{\eta_R} - \mu_t^R \geq 0, \quad q_t^R \left[\frac{q_t^R}{\eta_R} - \mu_t^R \right] = 0.$$

The optimal position and aggregate purchases by the entering cohort of mass $\alpha\delta$ are

$$q_t^R = \eta_R [\chi - b_c r_t - b_p r_{t-1}]_+, \\ G_t = \alpha\delta q_t^R,$$

where $[z]_+ = \max\{z, 0\}$. An entrant purchases a positive quantity when the expected gain is positive and chooses zero holdings otherwise. Lower current or past returns raise the incentive to buy. Retail investors are price takers: an individual treats the candidate price as given, while the aggregate price effect of their purchases is included in market clearing.

Institutions evaluate a position using the discounted stream of all future dividends,

$$\mathcal{D}_t = \sum_{j=1}^{\infty} \beta^j v_{t+j}, \quad V_t = E_t[\mathcal{D}_t], \quad s_D^2 = \text{Var}_t(\mathcal{D}_t). \quad (3)$$

Their date- t portfolio objective is

$$\max_{q \in \mathbb{R}} \left\{ E_t[q(\mathcal{D}_t - p_t)] - \frac{\gamma_I}{2} \text{Var}_t[q\mathcal{D}_t] \right\} = \max_{q \in \mathbb{R}} \left\{ (V_t - p_t)q - \frac{\gamma_I s_D^2}{2} q^2 \right\}. \quad (4)$$

Under Gaussian dividend innovations, $\mathcal{D}_t$ is conditionally normal and this criterion also has a CARA certainty-equivalent interpretation. Appendix B shows that

$$V_t = \frac{\beta \bar{v}}{1 - \beta} + a_V \theta_t, \quad a_V = \frac{\beta \rho_\theta}{1 - \beta \rho_\theta}. \quad (5)$$

The conditional variance s_D^2 is constant. Defining $\eta_I = (\gamma_I s_D^2)^{-1}$, aggregate institutional demand is

$$I_t = \tau_I(V_t - p_t), \quad \tau_I = (1 - \alpha)\eta_I. \quad (6)$$

Institutions update their positions each period under this valuation criterion. A lower τ_I means that a larger discount is required to induce institutions to hold a given quantity. Institutions remain informed about θ_t ; capacity governs their willingness to bear risk.⁸

4.3 Equilibrium

Let H_{t-1} denote active retail holdings after trading at $t - 1$. Independent exit shocks imply liquidation sales $D_t = \delta H_{t-1}$ and carryover holdings $C_t = (1 - \delta)H_{t-1}$. Active holdings and active net retail flow after entry are

$$H_t = C_t + G_t, \quad F_t = G_t - D_t = H_t - H_{t-1}. \quad (7)$$

New purchases are nonnegative, but net flow can be negative when liquidation sales exceed entry purchases. In particular, $F_t \geq -\delta H_{t-1}$.

In addition to active trading, the retail sector accumulates a passive position B_t through regular contributions. This position is separate from the holdings subject to liquidity-driven exit. Let

$$B_t = B_{t-1} + \bar{F}, \quad \bar{F} \geq 0. \quad (8)$$

Total retail holdings are $H_t + B_t$, and observed net retail purchases are

$$Q_t^R = \Delta(H_t + B_t) = F_t + \bar{F}. \quad (9)$$

We allow the level of outside asset supply to contain the same deterministic component, $\Lambda_t = \lambda_t + B_t$. Total market clearing is $(H_t + B_t) + I_t = \lambda_t + B_t$. Cancelling the common baseline component gives

$$H_t + I_t = \lambda_t, \quad p_t = V_t + \frac{C_t + G_t - \lambda_t}{\tau_I}. \quad (10)$$

Since purchases depend on the current return, price and entrant demand must be

⁸The institutional objective is a period-by-period portfolio criterion over discounted dividends. It abstracts from the unrestricted dynamic trading problem with endogenous resale values and intertemporal hedging. The dividend valuation and its conditional variance are derived in Appendix B.

determined jointly. Define the no-entry price and attractiveness of entry at that price as

$$p_t^0 = V_t + \frac{C_t - \lambda_t}{\tau_I}, \quad A_t = b_c(p_{t-1} - p_t^0) - b_p r_{t-1} - \chi. \quad (11)$$

The exact equilibrium is

$$G_t = \frac{\tau_E \tau_I}{\tau_I + \tau_E b_c} [A_t]_+, \quad p_t = p_t^0 + \frac{G_t}{\tau_I}. \quad (12)$$

The price is unique because excess demand is continuous and strictly decreasing. Entry purchases raise the clearing price and weaken the current-return signal that induced entrants to buy. Both this feedback and the participation boundary are retained in the solution.

Away from the entry boundary, local supply impact is

$$\frac{\partial p_t}{\partial \lambda_t} = \begin{cases} -1/(\tau_I + \tau_E b_c), & A_t > 0, \\ -1/\tau_I, & A_t < 0. \end{cases} \quad (13)$$

Current dip buying adds depth when entrants are active. Past dip buying changes the participation decision and the purchases of later cohorts. Liquidation carries the effects of inherited positions into subsequent periods. Increasing α raises entrant capacity and lowers institutional capacity. The first effect can cushion declines; the second increases price sensitivity, especially when entry is inactive. The two capacity effects therefore need not have the same implications for price sensitivity.

The difference between the two contrarian channels is also visible in the return equation. When entry is active,

$$(\tau_I + \tau_E b_c) r_t = \tau_I \Delta V_t - \Delta \lambda_t - \delta H_{t-1} - \tau_E b_p r_{t-1} - \tau_E \chi. \quad (14)$$

Current dip buying increases the denominator relating contemporaneous pressure to returns. Past dip buying instead introduces a delayed response to the previous return. A decline can therefore induce buying after the initial pressure has begun to dissipate. The subsequent withdrawal of this demand, together with liquidation, can generate further reversals. These local mechanisms do not imply that unconditional volatility is monotone in either belief coefficient.

4.4 Calibration and SMM results

We calibrate the model in two steps. Panel A of Table 2 reports the fixed parameters. We set the monthly discount factor to 0.995, dividend-state persistence to 0.96, and supply persistence to 0.50 motivated by Coval and Stafford (2007), Lou (2012), and Vayanos and Woolley (2013). Dividend news is therefore persistent, while supply pressure dissipates more quickly. The baseline retail share is 0.250, corresponding to the average in our sample. We fix individual retail risk aversion at 5.2 following the estimate from Gorodnichenko and Yin (2026).

Panel B reports the reference moments. We then estimate seven parameters jointly by SMM: institutional risk-bearing capacity τ_I , the two contrarian belief coefficients, b_p and b_c , and average return expectation χ , the retail exit probability δ , the supply innovation SD σ_λ , and the persistent dividend innovation SD σ_θ . We target the four current- and past-return dip-buying slopes, the total-to-retail flow SD ratio, the raw retail-flow SD, and the price informativeness measure as in Dávila and Parlato (2025), which measures the incremental explanatory power of future dividend growth for current price growth. At each parameter evaluation, we choose the transitory dividend variance to match dividend autocorrelation of 0.820. Given dividend-state persistence, this autocorrelation determines the relative importance of persistent and transitory dividend variation; the informativeness target supplies an additional restriction on their scale. After solving the model, we re-scale returns and retail flows to match the empirical measure. In particular, the retail-flow SD target of 0.014 anchors the quantity scale. Returns are rescaled after solving the model to have an SD of 0.183.

Panels C and D of Table 2 report the SMM fit and joint estimates. All three starting values converge to the same solution and match the seven targets to numerical precision in the estimation sample. The estimated persistent dividend innovation SD is 0.050. The exit probability implies an average holding duration of 2.946 periods. Aggregate institutional capacity is 0.058, corresponding to individual capacity of 0.077. The two contrarian belief coefficients are 0.180 and 0.139. Therefore, retail investors are more contrarian to current returns than past returns.

Table 2. Calibration and SMM Results

<i>Panel A. Fixed parameters</i>			<i>Panel B. Reference moments</i>		
Parameter	Value		Moment	Value	
Discount factor, β	0.995		Dividend autocorrelation	0.820	
Dividend persistence, ρ_θ	0.960		Return SD normalization	0.183	
Supply persistence, ρ_λ	0.500		Retail-flow SD target	0.014	
Retail share, α_0	0.250		Mean retail flow/SD	0.150	
<i>Panel C. SMM moments</i>			<i>Panel D. Parameter estimates</i>		
Moment	Target	Model	Parameter	Estimate	SE
Current loss-side slope	−0.057	−0.057	τ_I	0.058	0.002
Current gain-side slope	0.000	0.000	χ	-0.235	0.013
Past loss-side slope	−0.031	−0.031	b_c	0.180	0.008
Past gain-side slope	0.000	0.000	b_p	0.139	0.007
Total/retail flow SD ratio	1.500	1.500	δ	0.339	0.016
Raw retail-flow SD	0.014	0.014	σ_λ	0.018	0.000
Price informativeness	0.032	0.032	σ_θ	0.050	0.000

Notes: The model period is one month. Panel A gives the parameter values set before running SMM by matching the moments in Panel B. Panel C reports simulated moments, using 64 paths with 4,096 retained periods after a burn-in of 4,096 periods. Panel D reports estimates of the parameters.

4.5 Buying the dip

Figure 8 plots simulated net retail flow against current and past returns. Each curve connects conditional bin means. The model reproduces the asymmetric response used in calibration: retail purchases increase as returns become more negative, while the gain-side relation is much flatter. The current-return loss-side slope is −0.057 and the past-return slope is −0.031 in the estimation sample, matching the empirical estimates perfectly.

The asymmetry follows from the entry decision. A sufficiently negative signal raises the perceived gain above the hurdle and induces purchases. When the signal does not support entry, the new cohort chooses zero rather than a short position. The smoothness of the plotted conditional means reflects variation in inherited holdings and the other return signal. Individual demand remains piecewise linear, and the participation boundary need not coincide with a zero current or past return.

Positive average retail accumulation is consistent with net selling in some states. The calibrated passive inflow shifts both conditional-flow curves upward without

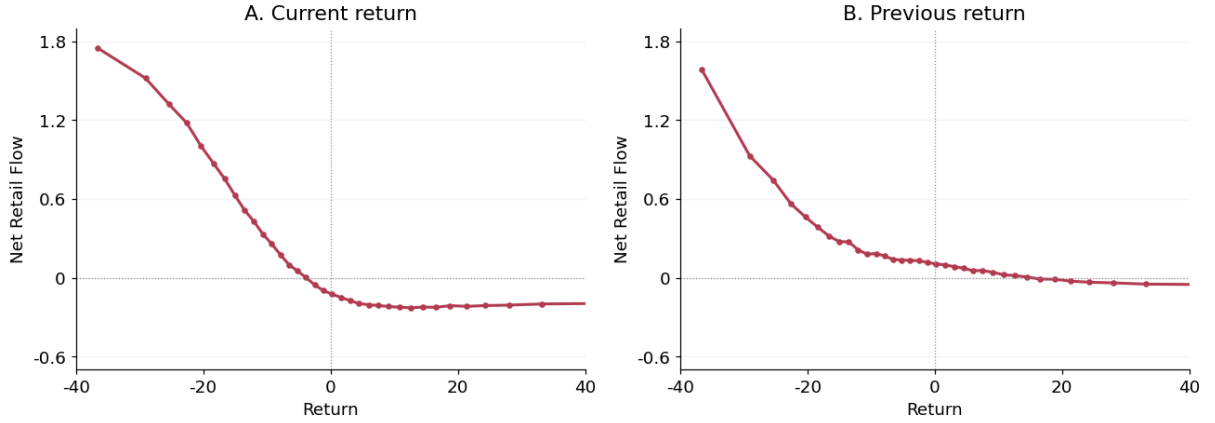


Figure 8. Current and Past Dip Buying For both panels, the y-axis is the simulated net retail flow. The x-axis of Panel A and Panel B are respectively the current return and previous period’s return. Each curve connects 35 equal-count binned averages pooled across 128 simulated paths, each containing 8,192 periods after a 4,096-period burn-in. The outer 0.1% of observations in each return tail are omitted from the displayed bins.

changing their slopes. It does not make retail flow positive at every date: liquidation sales can exceed the combined purchases of entrants and passive investors. The model separates this average accumulation from the return-dependent trading mechanism.

4.6 Dip buying and return dynamics

We now perform counterfactuals to understand the impacts of buying the dip. Our first exercise is to study return dynamics. Note that contrarian expectation formation is captured by two parameters, b_c and b_p , respectively capturing being contrarian to current and past return. To study the impact of b_c and b_p on return, we vary these belief coefficients separately, holding the remaining parameters and the retail share fixed. We then study the impulse-response functions (irf) of returns after a one-time supply shock. All impulse responses compare a shocked economy with an otherwise identical unshocked economy under the same belief specification and common innovations to retain each specification’s pre-shock distribution.

Figure 9 reports the responses to supply innovations that generate baseline impact returns of -30% and 30% . A positive supply innovation produces the decline; a negative innovation produces the increase. The two raw shock sizes are chosen separately at the baseline and then held fixed across belief specifications. The comparison therefore holds the underlying shock fixed within each panel.

From Panel A, stronger current dip buying cushions the initial decline. Doubling b_c reduces the impact loss from 30% to about 20% as retail traders absorb more of the initial selling pressure. This also reduces the next-period rebound from around 18% to 12% given less initial deviation from fundamental values. Meanwhile, from Panel B, after a positive price impact, the effect is more modest: the short-sale constraint, which creates the asymmetry of contrarian trading, limits the extent to which retail can accommodate buying pressure by reducing positions.

In comparison, buying the past dip has an opposite impact on return. Stronger past dip buying changes the timing of price adjustment. First, doubling b_p does not affect the initial price movement. In the next period, after a previous dip, the same decline is followed by a stronger reversal when b_p is larger given stronger demand from retail (Panel C). The subsequent reversal in period three is also stronger due to larger deviation from fundamental values after the extra retail demand in period two. Therefore, doubling b_p creates a more oscillating return path after positive supply shocks. Similarly, from Panel D, returns also get more fluctuated when b_p is doubled and after negative supply shock. However, due to the asymmetric dip-buying behaviors, initial price increase is not followed by more selling in the next period. Therefore, doubling b_p does not affect return dynamics in the first two periods. While in the period after the mean-reverting supply shock, return becomes negative. So retail demand gets stronger in the next period. For this period onwards, the dynamics gets similar to that after positive supply shocks.

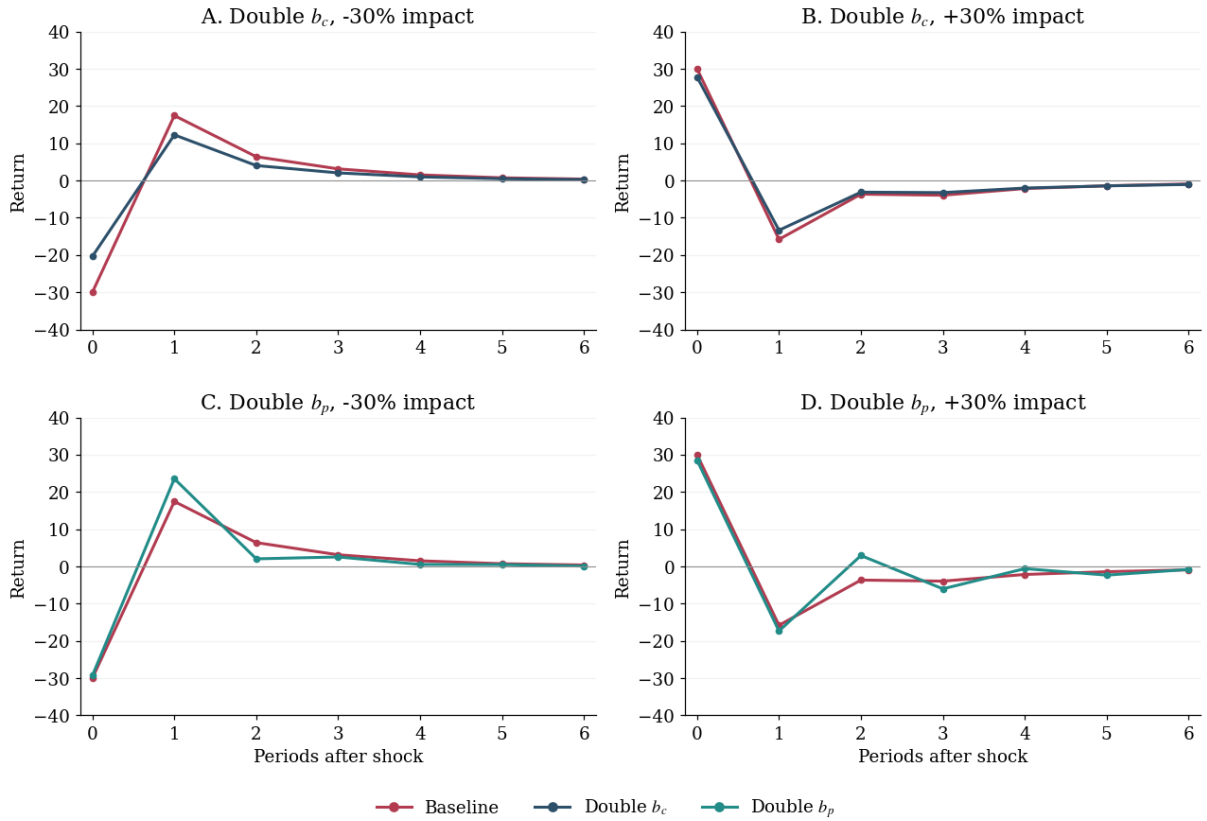


Figure 9. Buying the Dip and Return Dynamics This figure shows how stronger current and past dip buying affect return responses to supply shocks. Panels A and B double the current-return belief coefficient b_c , while Panels C and D double the past-return belief coefficient b_p . Shock sizes are chosen to generate baseline impact returns of -30% in the left panels and +30% in the right panels, and are held fixed across specifications. Each curve averages the return difference between shocked and otherwise identical unshocked simulations over 8,192 paired paths, following a 4,096-period burn-in.

These responses also suggest opposite implications on return volatility for buying the current dip or buying the past dip. Stronger current dip buying absorbs selling pressure as it arises, reducing both the initial price decline and the subsequent reversal. It therefore dampens return fluctuations. Stronger past dip buying, however, responds to a decline after the initial adjustment has occurred. The resulting purchases can push prices up when they are already recovering, amplifying the rebound. As successive cohorts respond to preceding returns, this delayed response can also generate further reversals. Thus, a faster recovery in prices need not imply lower return volatility: stronger rebounds and subsequent oscillations can increase fluctuations over time. Although both forms of dip buying reflect contrarian beliefs, their timing determines whether retail demand cushions price movements or

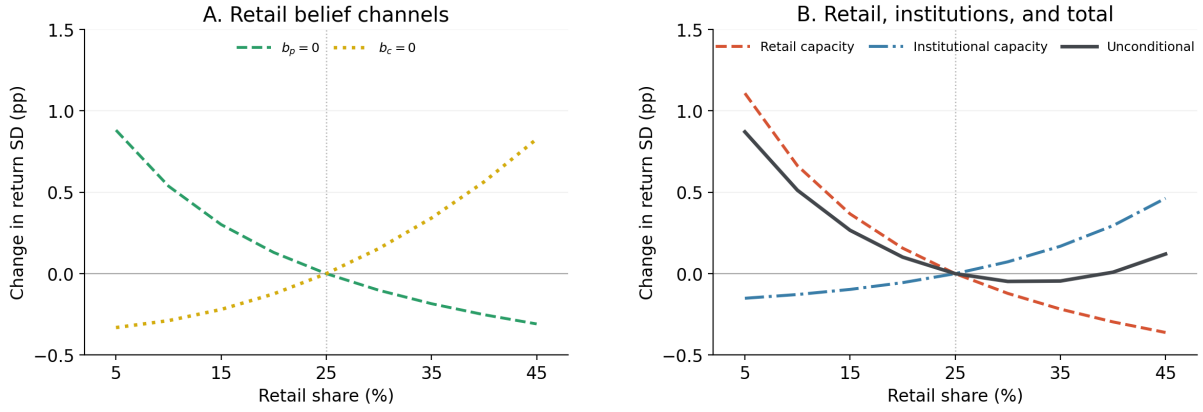


Figure 10. Retail Capacity, Contrarian Beliefs, and Return Volatility This figure plots changes in return volatility against retail share. Panel A compares the retail-capacity effect with both belief channels, with $b_c = 0$, and with $b_p = 0$. Panel B compares the retail-capacity effect, the institutional-capacity effect, and the full-model response, retaining both belief channels. The retail-capacity curves vary τ_E while holding τ_I fixed; the institutional-capacity curve varies τ_I while holding τ_E fixed. The full-model curve varies both capacities. All changes are measured in percentage points relative to each specification's 25% retail-share economy. Simulations use 128 paths with 8,192 retained periods after a burn-in of 4,096 periods.

amplifies subsequent adjustment.

4.7 Retail share and return volatility

Motivated by the findings in Figure 9, we now examine how retail participation affects market volatility. A larger retail share changes both sides of market clearing: it increases entrant capacity, $\tau_E = \delta\alpha\eta_R$, while reducing institutional capacity, $\tau_I = (1 - \alpha)\eta_I$. Figure 10 separates these effects and examines how they depend on retail beliefs. Individual risk-bearing capacities, belief coefficients, and the dividend and supply processes remain fixed, except where a belief channel is explicitly removed.

Panel A shows that the effect of greater retail capacity depends on the timing of dip buying. With only current dip buying ($b_p = 0$), greater retail capacity lowers volatility: entrants absorb contemporaneous selling pressure, cushioning the initial decline and reducing its subsequent reversal. With only past dip buying ($b_c = 0$), greater retail capacity instead raises simulated volatility. Purchases respond to an earlier decline and can strengthen the rebound, generating further price fluctuations through delayed demand.

Panel B compares the retail and institutional capacity effects while retaining both belief channels. Each counterfactual varies one capacity while holding the other at

its value in the estimated economy, $\alpha_0 = 0.25$. From the red dashed line, even if b_c and b_p have opposite effects on volatility, in net, the effects through buying the current dip dominates, and increasing retail capacity lowers volatility. Reducing institutional capacity has the opposite effect: a given imbalance requires a larger price adjustment. Institutions remain equally informed about the dividend state; the effect operates through their willingness to bear risk. Increasing retail share from 25% to 45% lowers return SD by 0.36 percentage points in the retail-capacity counterfactual and raises it by 0.46 percentage points in the institutional-capacity counterfactual. These effects need not add to the full-model response because the equilibrium effect of either capacity depends on the other.

Together, the two capacity changes produce the shallow U-shaped full-model response in Panel B. Although the figure reports changes relative to the 25% benchmark, the corresponding return SD falls from 19.2% at a retail share of 5% to 18.3% at 25%, reaches approximately 18.2% around a retail share of one-third, and then rises to 18.4% at 45%. Thus, greater retail participation initially reduces volatility, but the loss of institutional risk-bearing capacity eventually outweighs the stabilizing effect of additional retail capacity.

4.8 Trading profits

In the end, we study investor profitability implied by our model. Let $Q_t^R = H_t$ and $Q_t^I = I_t$ denote aggregate retail and institutional holdings. Retail includes all outstanding purchase cohorts, including non-rebalancing investors, and excludes permanent holdings. The benchmark-adjusted profit on positions carried into period t is

$$\Pi_t^g = Q_{t-1}^g x_t, \quad x_t = p_t - p_{t-1} + v_t - \kappa p_{t-1}, \quad \kappa = \beta^{-1} - 1, \quad (15)$$

for $g \in \{R, I\}$. The adjustment charges longs and credits shorts at a common rate. This valuation-based benchmark is a reporting convention, not a separately estimated financing rate. We scale average profit by the value of gross positions carried into each period:

$$Y_g = 100 \frac{\sum_{n=1}^N \sum_{t=1}^T \Pi_{n,t}^g}{\sum_{n=1}^N \sum_{t=1}^T |Q_{n,t-1}^g p_{n,t-1}|}, \quad (16)$$

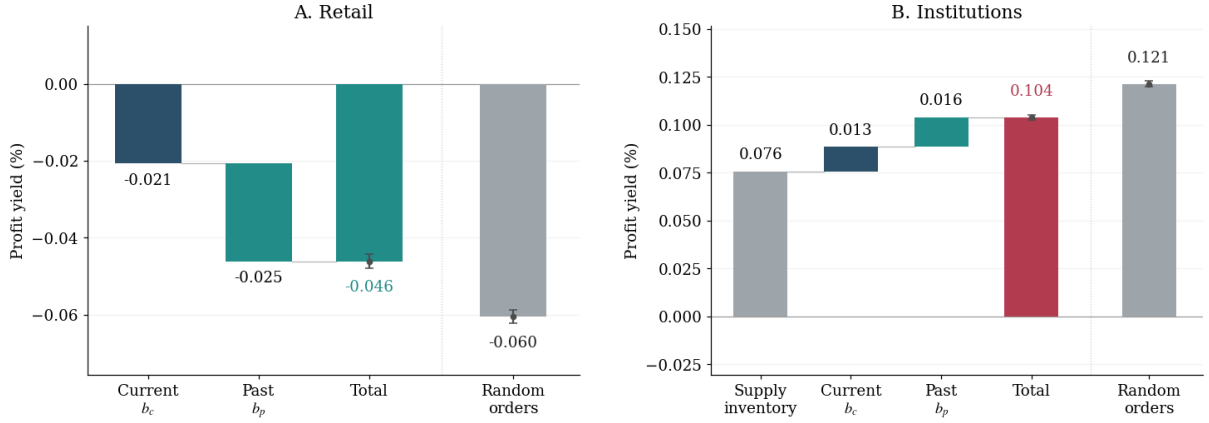


Figure 11. Contrarian Beliefs and Trading Profits Decomposition of profit is based on equations (17) and (18). Simulations use 128 paths with 8,192 retained periods after a 4,096-period burn-in and common dividend and supply shocks across benchmarks. Error bars show 95% confidence intervals.

where N is the number of simulated paths and T is the number of retained periods. This measure reports quarterly adjusted profit per dollar of gross exposure. Its positive denominator preserves the sign of aggregate profit. Each sector and each benchmark uses its own exposure denominator.

Figure 11 provides an accounting decomposition on the estimated equilibrium path. Define the current and past return signals $c_t = -b_c r_t$ and $d_t = -b_p r_{t-1}$. Entrant purchases are $G_t = \tau_E [c_t + d_t - h]_+$, where $h = -\chi > 0$. We allocate the shared entry threshold proportionally to the signed signals:

$$G_t^c = \frac{c_t}{c_t + d_t} G_t, \quad G_t^p = \frac{d_t}{c_t + d_t} G_t \quad \text{if } G_t > 0. \quad (17)$$

Both components are zero otherwise. Holdings follow $H_t^j = (1 - \delta)H_{t-1}^j + G_t^j$ for $j \in \{c, p\}$, yielding the exact identities

$$\Pi_t^R = H_{t-1}^c x_t + H_{t-1}^p x_t, \quad \Pi_t^I = \lambda_{t-1} x_t - H_{t-1}^c x_t - H_{t-1}^p x_t. \quad (18)$$

Applying the same denominator to all components within each sector preserves additivity. The allocation holds the equilibrium path fixed and depends on how the entry threshold is assigned; the components are accounting contributions and may be signed.

Under this allocation, current and past dip buying contribute -0.050 and -0.059 percentage points to the retail yield, accounting for approximately 46% and 54% of its loss. Their institutional counterparts contribute 0.030 and 0.036 percentage points.

The remaining 0.171 percentage points come from the supply-inventory term $\lambda_{t-1}x_t$, capturing institution's profits from earning the risk premium generated by supply fluctuation.

In the end, the last column simulates an economy where retail submit random orders that follow a normal distribution with average and standard deviations matching those in the benchmark calibration. The results show that retail traders further loss profit if they submitting random orders. This is because, due to mean-reverting supply shocks, returns are weakly negatively autocorrelated in the model, with a first-order auto-correlation of -0.09. Therefore, contrarian expectations are qualitatively consistent with the true return process, this results in higher profits than just submitting random orders. However, as the estimated expectations are too contrarian than otherwise implied by the true return process, and institutions always have more precise information about the dividend process, retail traders in the benchmark case are still in the lost side.

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 asymmetry holds across horizons and over time, and it is specific to retail flow. We rationalize it with a model in which retail investors hold contrarian return expectations and face a short-sale constraint, estimated by simulated method of moments. The model reproduces the loss-side slopes and shows that the effect of dip buying depends on its timing. Buying the current dip cushions the initial decline, while buying the past dip amplifies subsequent reversals. A larger retail share combines these demand effects with a loss of institutional risk-bearing capacity, producing a shallow U-shaped relation between retail participation and return volatility.

Several questions remain for future research. First, identifying the supply-pressure channel requires pinning down which institutional friction generates the exogenous pressure that the model takes as a primitive. 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

- Balasubramaniam, Vimal, 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.
- Bali, Turan G., 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.
- Barardehi, Yashar H., 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.
- Barber, Brad M. 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.
- Barber, Brad M. 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.
- Barber, Brad M., 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.
- Barber, Brad M., Terrance Odean, and Ning Zhu. Systematic noise. *Journal of Financial Markets*, 12(4):547–569, 2009b. doi: 10.1016/j.finmar.2009.03.003.
- Barber, Brad M., 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.
- Barber, Brad M., 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.
- Barberis, Nicholas 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.
- Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. X-capm:

- An extrapolative capital asset pricing model. *Journal of Financial Economics*, 115(1): 1–24, 2015. doi: 10.1016/j.jfineco.2014.08.007.
- Barrot, Jean-Noel, 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.
- Bastianello, Francesca and Paul Fontanier. Partial equilibrium thinking, extrapolation, and bubbles. Working paper, SSRN, 2024.
- Bastianello, Francesca and Paul Fontanier. Expectations and learning from prices. *The Review of Economic Studies*, 92:1341–1374, 2025. doi: 10.1093/restud/rdae059.
- Battalio, Robert H, 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>.
- Boehmer, Ekkehart, 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.
- Campbell, John Y., 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.
- Campbell, John Y., 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.
- Cohen, Randolph B., 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.
- Conrad, Jennifer, 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.
- Coval, Joshua 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.
- Cready, William, 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.
- Daniel, Kent, 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.
- Dávila, Eduardo and Cecilia Parlatore. Identifying price informativeness. *The Review of Financial Studies*, 2025. doi: 10.1093/rfs/hhaf051.
- De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Finance*, 45(2):379–395, 1990. doi: 10.1111/j.1540-6261.

1990.tb03695.x.

- Dominitz, Jeff and Charles F Manski. Measuring and interpreting expectations of equity returns. *Journal of applied econometrics*, 26(3):352–370, 2011.
- Dorn, Daniel 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.
- Eaton, Gregory W., 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.
- Friedman, Milton. The case for flexible exchange rates. In *Essays in Positive Economics*, pages 157–203. University of Chicago Press, Chicago, 1953.
- Gelman, Michael, Liron Reiter-Gavish, and Nikolai Roussanov. FOMO economics: External reference-dependence and risk-taking in household portfolios. Working paper, SSRN, 2026.
- Glebkin, Sergei, 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.
- Gorodnichenko, Yuriy and Xiao Yin. Higher-order beliefs and risky asset holdings. Available at SSRN: <https://ssrn.com/abstract=4889206>, 2026.
- Graves, Daniel. Attention and the retail alignment puzzle. Working paper, Yale University, 2024.
- Greenwood, Robin and Andrei Shleifer. Expectations of returns and expected returns. *The Review of Financial Studies*, 27(3):714–746, 2014.
- Han, Bing and Alok Kumar. Speculative retail trading and asset prices. *Journal of Financial and Quantitative Analysis*, 48(2):377–404, 2013. doi: 10.1017/S0022109013000100.
- Hvidkjaer, Soeren. 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.
- Kaniel, Ron, 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.
- Kelley, Eric K. 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.
- Kumar, Alok. Who gambles in the stock market? *The Journal of Finance*, 64(4):1889–1933, 2009. doi: 10.1111/j.1540-6261.2009.01483.x.
- Kumar, Alok 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.

- Laarits, Toomas and Marco Sammon. The retail habitat. *Journal of Financial Economics*, 172:104144, 2025. doi: 10.1016/j.jfineco.2025.104144.
- Laudenbach, Christine, Annika Weber, Rüdiger Weber, and Johannes Wohlfart. Beliefs about the stock market and investment choices: Evidence from a survey and a field experiment. *The Review of Financial Studies*, 39(6):1654–1699, 2026.
- Lee, Charles M. C. 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.
- Lou, Dong. A flow-based explanation for return predictability. *Review of Financial Studies*, 25(12):3457–3489, 2012. doi: 10.1093/rfs/hhs103.
- Luo, Bryan, 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.
- Mitton, Todd 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.
- Nagel, Stefan and Zhengyang Xu. Dynamics of subjective risk premia. *Journal of Financial Economics*, 150(2):103713, 2023.
- Odean, Terrance. Are investors reluctant to realize their losses? *The Journal of Finance*, 53(5):1775–1798, 1998. doi: 10.1111/0022-1082.00072.
- Rabin, Matthew. Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, 117(3):775–816, 2002.
- Rabin, Matthew and Dimitri Vayanos. The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies*, 77(2):730–778, 2010.
- Shefrin, Hersh 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.
- van der Beck, Philippe, Cameron Cohen, and Coralie Jaunin. The equity market implications of the retail investment boom. Working Paper 26-054, Harvard Business School, 2026.
- Vayanos, Dimitri 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.
- Welch, Ivo. The wisdom of the robinhood crowd. *The Journal of Finance*, 77(3):1489–1527, 2022. doi: 10.1111/jofi.13128.
- Yuan, Kathy. 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

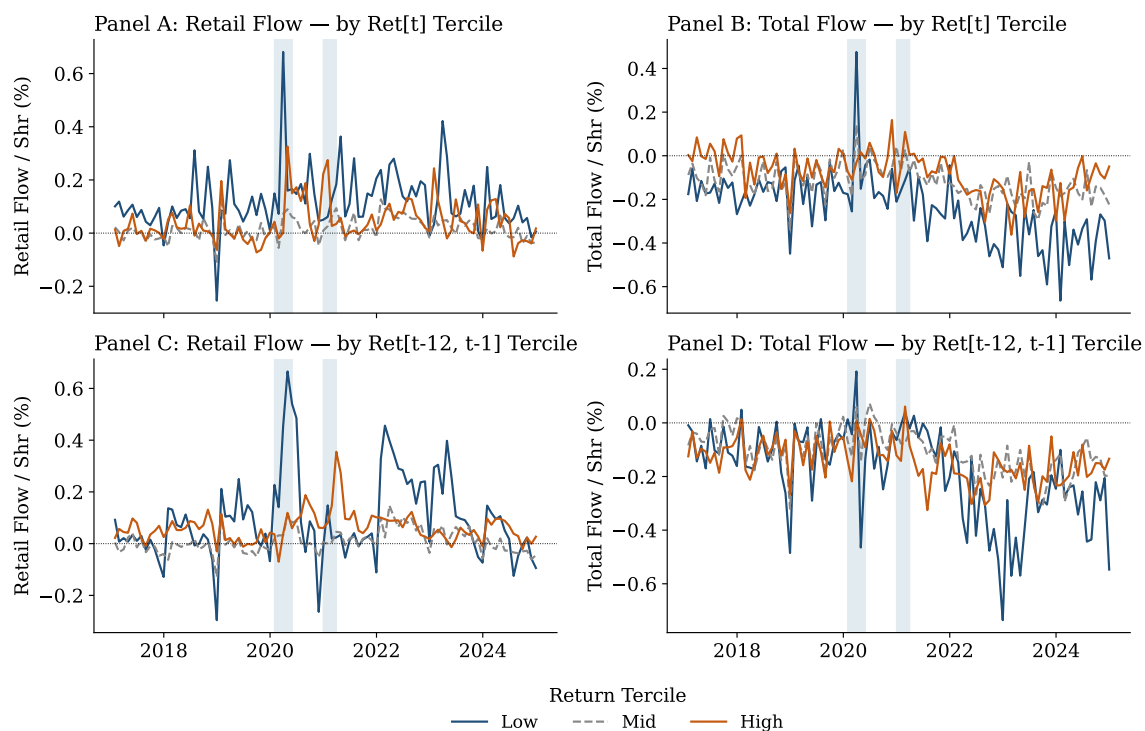


Figure A.1. Retail and Total Flows by Return Tercile, Market-Cap Weighted. Each panel plots the market-cap-weighted cross-sectional average of monthly net flow scaled by shares outstanding (percent) for stocks sorted into terciles within each calendar month. Weights are stock-level market capitalization at month t . Panels A and B sort by contemporaneous return $\text{Ret}[t]$. Panels C and D sort by past twelve-month return $\text{Ret}[t-12, t-1]$. 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 4 in the main text reports the equal-weighted version.

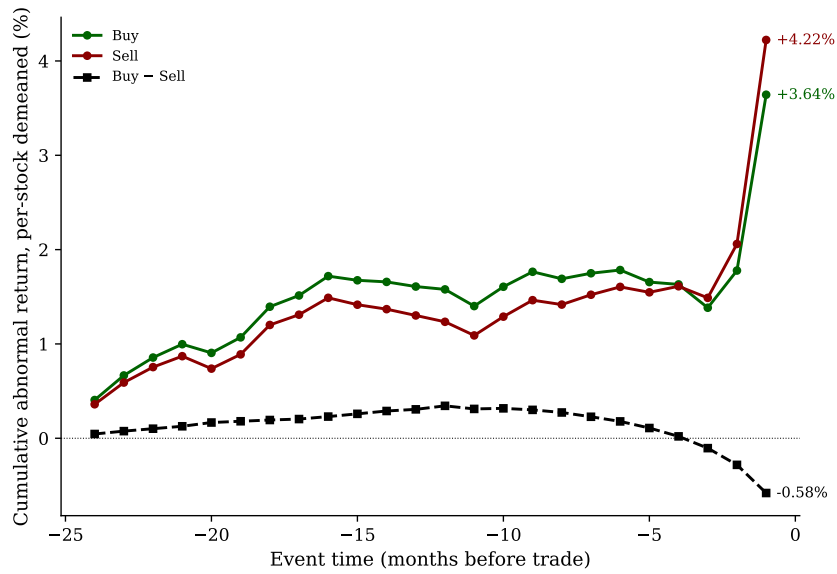


Figure A.2. Pre-Trade Cumulative Market-Adjusted Return: Retail Buys versus Sells. This figure is the panel analog of [Barber et al. \(2009b, Figure 3\)](#). Each stock-month is an event. The buy path weights each event by retail shares bought scaled by shares outstanding; the sell path weights by retail 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.

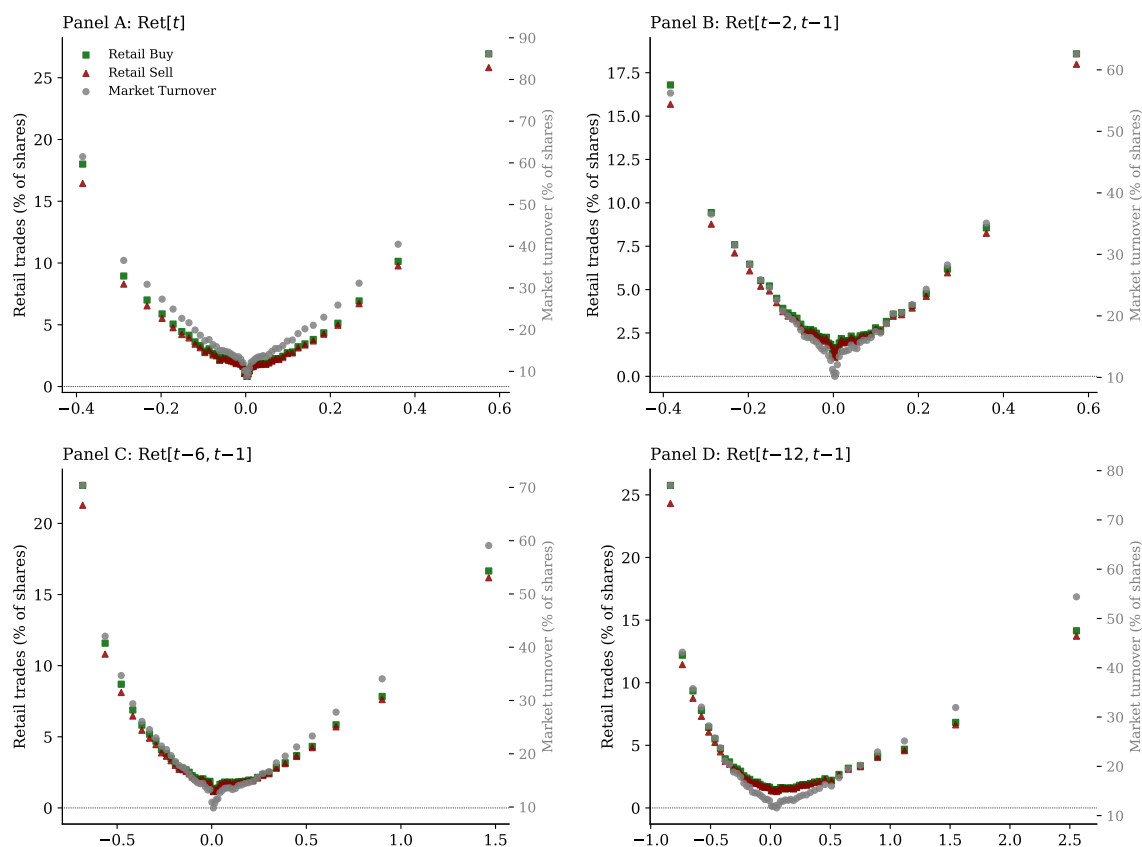


Figure A.3. Retail Buying and Selling Versus Past Returns, with Market Turnover. Each panel is a binscatter with 50 equal-frequency bins. “Retail Buy” is monthly retail shares bought and “Retail Sell” is monthly retail shares sold, each scaled by shares outstanding (percent) and read on the left axis. “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.

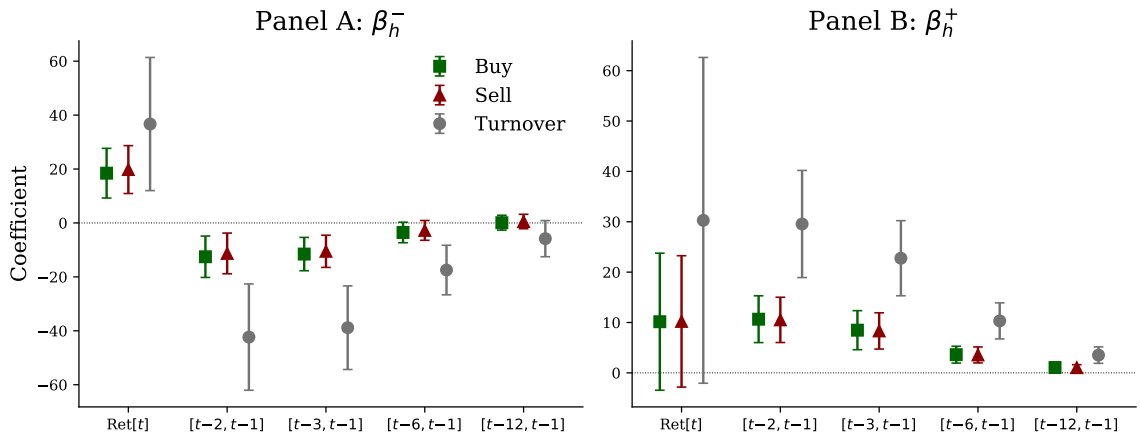


Figure A.4. Loss/Gain-Split Coefficients for Retail Buying, Selling, and Turnover. Markers report the loss-side coefficients β_h^- (Panel A) and gain-side coefficients β_h^+ (Panel B) from single-horizon panel regressions, splitting each past return into its negative and positive parts. Plotted series 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 3. 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.

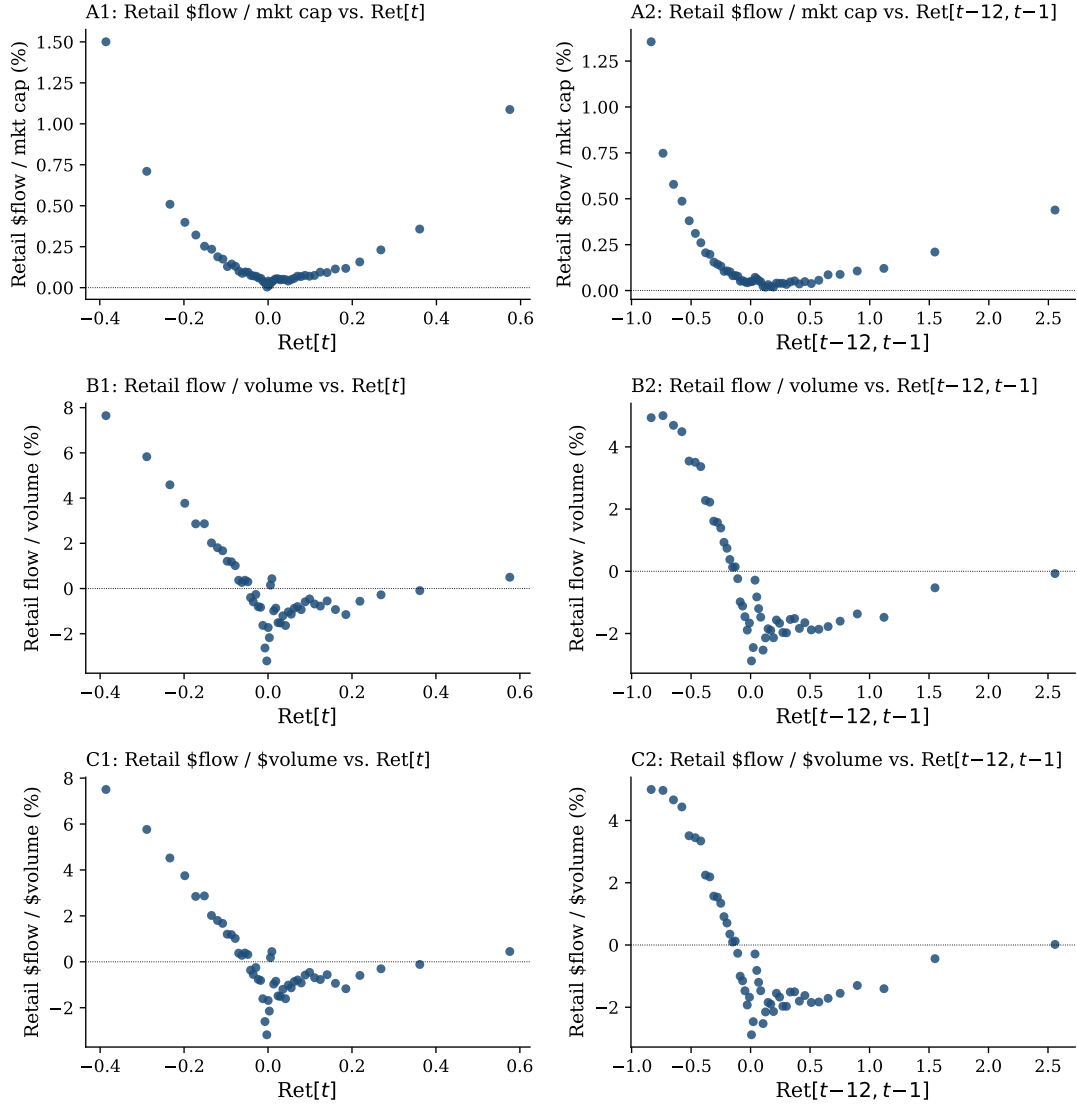


Figure A.5. Buying the Dip Under Alternative Flow Normalizations. Each panel is a binscatter with 50 equal-frequency bins. The rows use three alternative normalizations of retail net flow: retail dollar flow divided by market capitalization (Row A), retail net share flow divided by retail share volume (Row B), and retail dollar flow divided by retail dollar volume (Row C). The columns use the contemporaneous return $\text{Ret}[t]$ and the past twelve-month return $\text{Ret}[t-12, t-1]$. Both axes are winsorized at the 1% and 99% percentiles within each panel. The sample is January 2017 through December 2024.

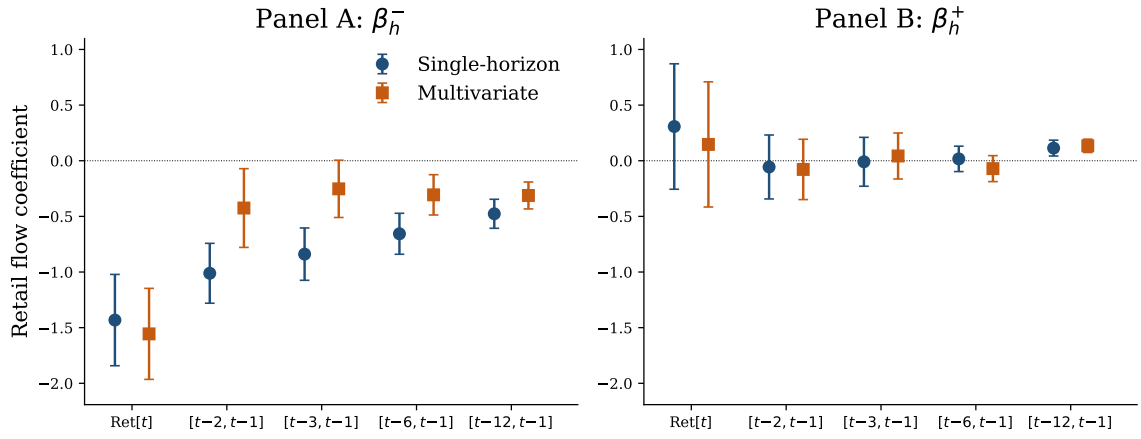


Figure A.6. Buying the Dip With Log Returns. The figure repeats the loss/gain-split coefficient plot of Figure 3 with each past return r_h replaced by its log transform $\ln(1 + r_h)$, defined where $1 + r_h > 0$. The sign split is preserved because $\ln(1 + r_h) < 0$ if and only if $r_h < 0$. Navy circles 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.

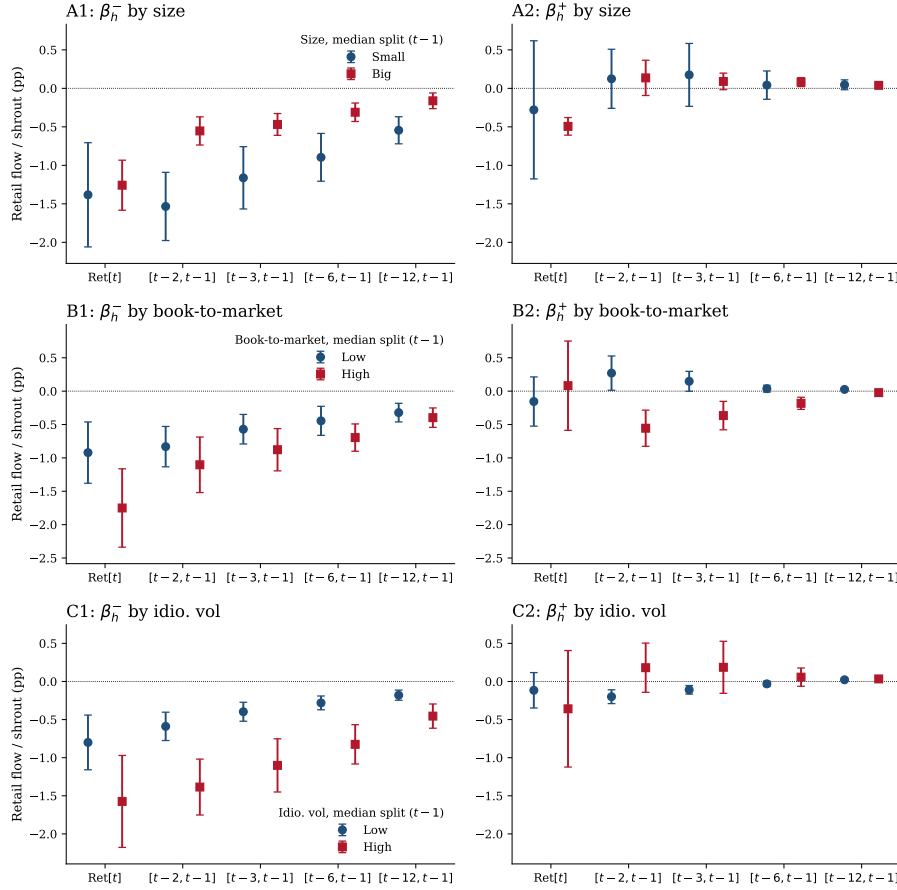


Figure A.7. Loss-Side and Gain-Side Coefficients by Stock Characteristic. Each row estimates the single-horizon specification of equation (1) separately within the two groups of Figure 7, at horizons $h \in \{0, 1, 3, 6, 12\}$ months. Left column: loss-side coefficients β_h^- . Right column: gain-side coefficients β_h^+ . Each regression includes the focal-horizon return splits, the control vector $X_{i,t}$ of column (4) of Table A.1, and stock and time fixed effects. The dependent variable is monthly retail net order flow scaled by shares outstanding (percentage points). Error bars are 95% confidence intervals based on standard errors clustered by stock and by calendar month. The sample is January 2017 through December 2024.

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

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.

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

B Validation and Comparison of the Retail Flow Measure

This appendix validates the S&P Retail Flow measure against the retail trading statistics that Citadel Securities published for GameStop in January 2021, and compares it with the TAQ-inferred retail proxies and the other flow measures defined in Section 2.2.

The cent-fraction rule of BJZZ aligns 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). 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 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. BHJOS uses that rule. 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 $\text{loc} < 0.4$ as retail sells, and trades in $[0.4, 0.6]$ are excluded as ambiguous rather than forced into a buy or sell category. In the example, 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.

B.1 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 B.1 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. 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.

Table B.1. Citadel Securities: GME Retail Flows, January 25–28, 2021

This table reports daily GME retail order flow intermediated by Citadel Securities over January 25–28, 2021. 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, and Net is their difference (Buy – Sell). Mkt. Volume is total GME market volume in shares. Mkt. Share is the retail volume intermediated by Citadel Securities divided by total GME market volume (percent). Net Flow is net retail shares divided by total market volume (percent). Source: Citadel Securities, published no later than January 29, 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

Figure B.1 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 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 B.1 refers to Citadel Securities’ intermediated retail volume relative to *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

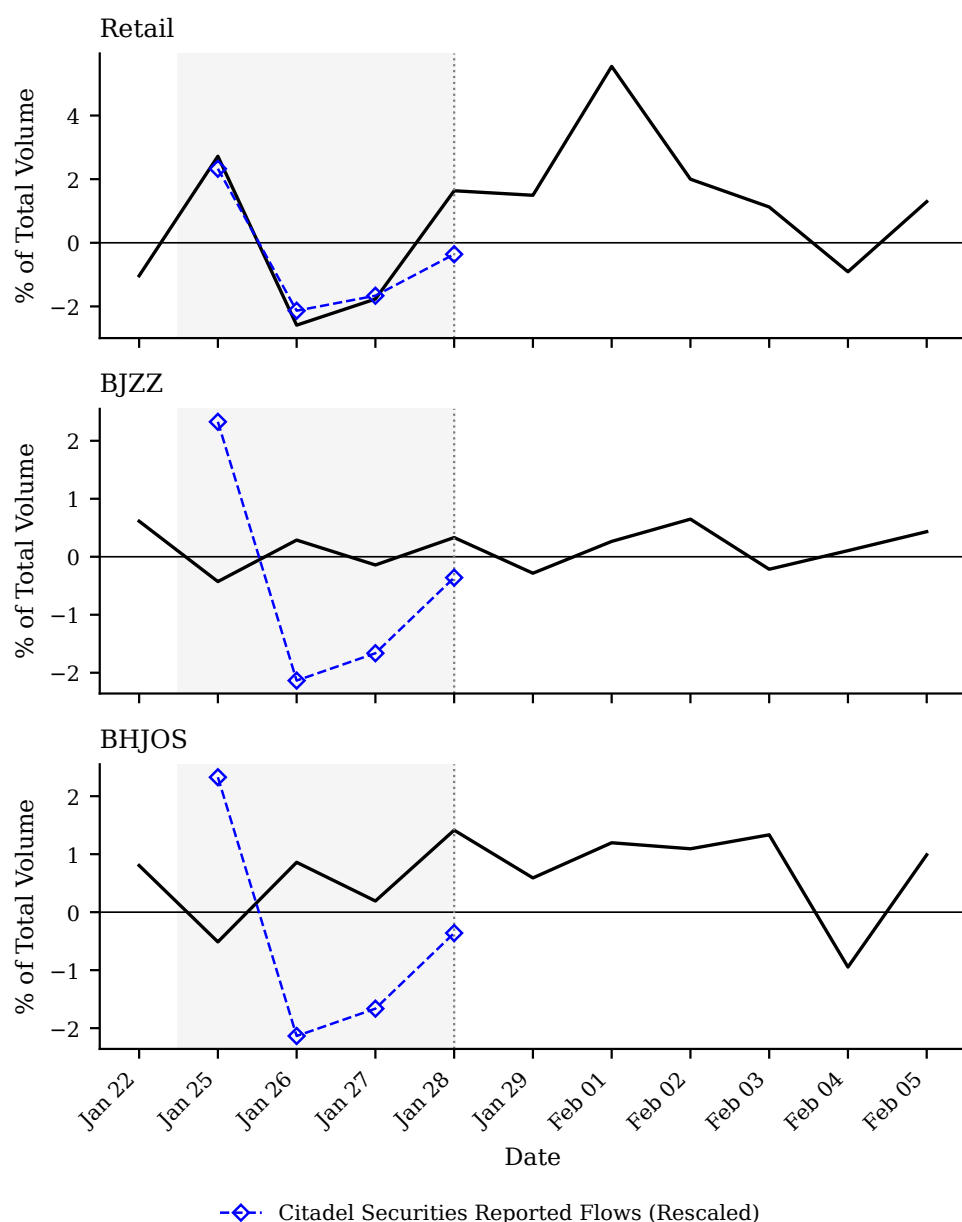


Figure B.1. Net Retail Flows vs. Citadel Securities on GameStop. Each panel plots the daily net retail order flow as a percent of total GME market volume. *S&P Retail Flow*: actual retail flow from the S&P Global Retail Flow Database. *BJZZ*: sub-penny retail flow of [Boehmer et al. \(2021\)](#). *BHJOS*: modified retail flow of [Barber et al. \(2024\)](#). The dashed line with hollow markers shows the Citadel Securities statistics from Table B.1, rescaled by a factor of 2 ($\approx 80\%/39\%$). 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.

episode.⁹

⁹The 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.

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 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.

B.2 Comparison with Flow Measures

We compare the S&P Retail Flow series to the five other flow measures defined in Section 2.2. Table B.2 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 B.2 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 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

Table B.2. Correlations of Retail Flow with Other Flow Measures

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](#)). ΔIO : 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 sample. 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).

	Time-Series Correlation						Cross-Sectional Correlation					
	BJZZ	BHJOS	Inst50k	InstSec	ΔIO	Total	BJZZ	BHJOS	Inst50k	InstSec	ΔIO	Total
<i>Panel A: Full Sample</i> 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
<i>Panel B: By Market Cap Quintile</i>												
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
<i>Panel C: By Momentum Quintile (prior 12-to-1 month return)</i>												
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
<i>Panel D: By Sub-Period</i>												
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

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 ($\rho_{BJZZ} = 0.493$ for the smallest quintile) than among large stocks ($\rho_{BJZZ} = 0.156$ 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.

C Retail Flow and Institutional Trading

This appendix reports the three results behind the statement in Section 3.4 that retail dip-buying is not passive absorption of institutional selling. The first conditions on institutional trading in the same stock-month. The second uses a shock to institutional trading that is unrelated to the stock. The third asks where the shares that retail accumulate come from.

C.1 Conditioning on Institutional Trading in the Same Stock-Month

If retail dip-buying were a by-product of institutional selling, the loss-side response should appear in stock-months in which institutions are net sellers and vanish in stock-months in which they are net buyers. We sort stock-months at the within-month median of each institutional flow proxy in month t , so that institutions are mostly net sellers in the low group and net buyers in the high group.¹⁰ If retail were passive counterparties, the loss-side response should be confined to the low group.

Figure C.1 shows that the two groups have nearly the same shape. Retail flow is higher on average where 13-F ownership fell, as market clearing requires, but the dip-buying slope is about as steep where institutions are buying as where they are selling, and the gain side is flat in both groups of every sort. Figure C.2 reports the loss-side and gain-side coefficients by horizon within each group. Retail therefore buy the dip at close to full strength in stocks that institutions are buying in the same month. Dip buying is not a relabeling of institutional selling; it is a response to the price move, whichever side of the market institutions are on. The result is also consistent with Table A.1, where adding the institutional flow proxies to the controls changes the return coefficients little.

¹⁰The proxy is negative in 90, 99.7, and 81 percent of low-group stock-months for ΔIO , Inst50k, and InstSec, and in 4, 3, and 1 percent of high-group stock-months. The sort variable is a contemporaneous outcome and the proxies are noisy, so the split is descriptive and works against finding a difference between the groups.

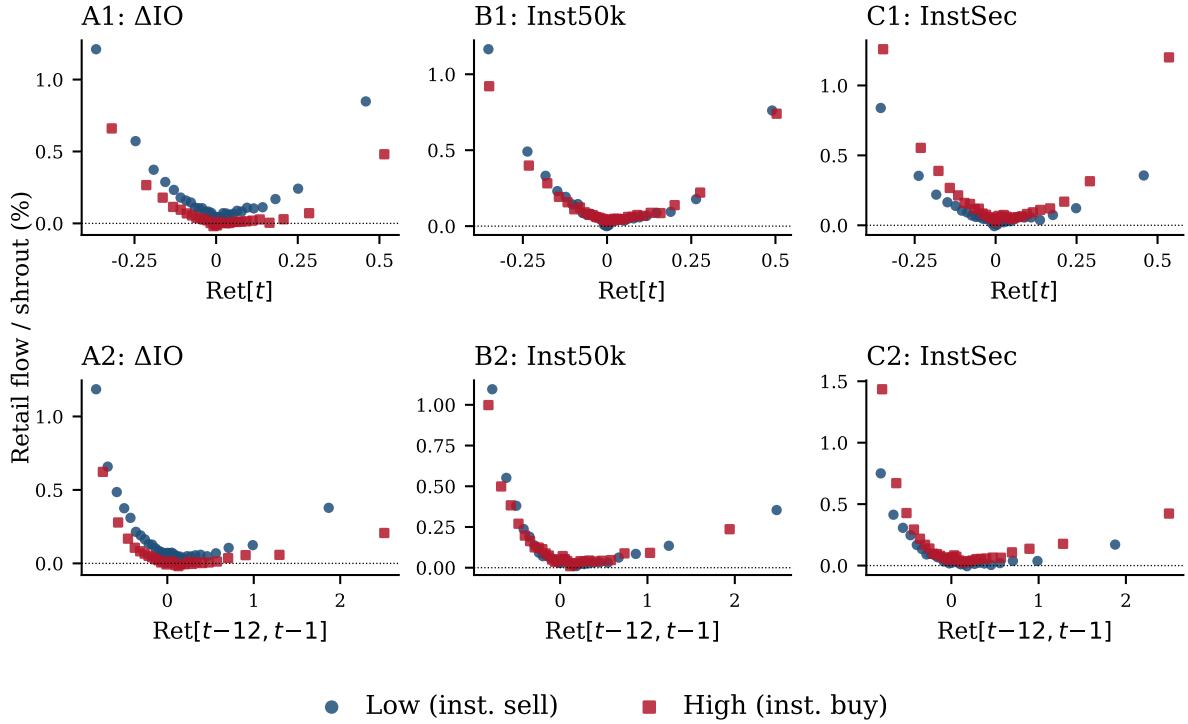


Figure C.1. Buying the Dip by Contemporaneous Institutional Flow. Each panel is a binscatter with 30 equal-frequency bins per group. Columns split stocks at the within-month median of, respectively, the quarterly change in 13-F institutional ownership assigned to month t (ΔIO , Panels A), the large-trade institutional proxy in month t (Inst50k , Panels B), and the change in securities-lending supply in month t (InstSec , Panels C). Variable definitions follow Section 2.2. The vertical axis is monthly retail net order flow scaled by shares outstanding (percent). The horizontal axis is the contemporaneous return $\text{Ret}[t]$ (top panels) or the past twelve-month return $\text{Ret}[t-12, t-1]$ (bottom panels). Both axes are winsorized at the 1% and 99% percentiles within each group and panel. The sample is January 2017 through December 2024.

C.2 A Shock to Institutional Trading

The comparison above conditions on institutional trading that may itself respond to the stock. The second test uses institutional trading that does not. When investors withdraw money from a mutual fund, the fund sells its holdings roughly in proportion, so its trade in any one stock is set by the fund's investors rather than by news about the stock (Coval and Stafford, 2007; Lou, 2012). For each stock-month we add up, across active domestic equity funds, the fund's net flow in the month as a fraction of its assets times the fund's holding of the stock as a fraction of shares outstanding. The sum is flow-induced trading (FIT): the net trade that mutual funds are predicted to make in the stock that month, in percent of shares outstanding.

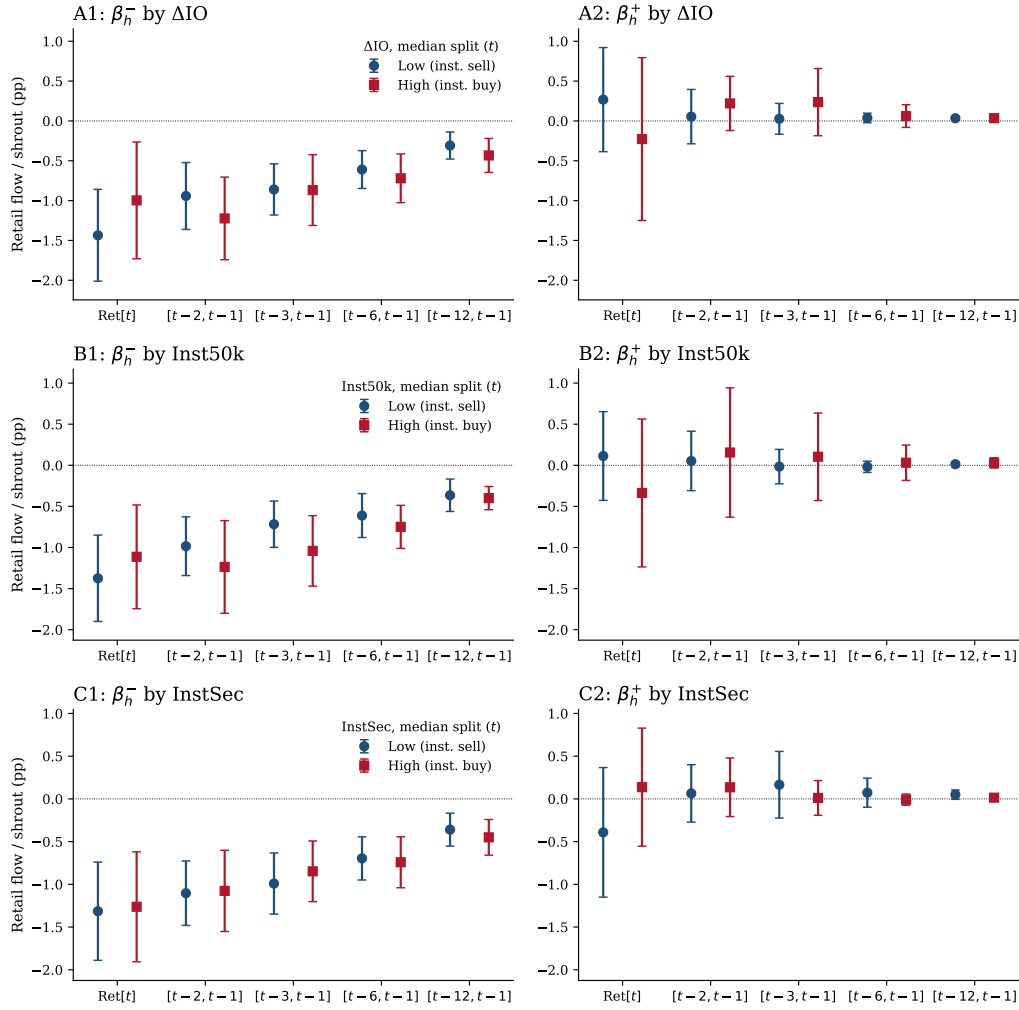


Figure C.2. Loss-Side and Gain-Side Coefficients by Contemporaneous Institutional Flow. Each row estimates the single-horizon specification of equation (1) separately within the two groups of Figure C.1, at horizons $h \in \{0, 1, 3, 6, 12\}$ months. Left column: loss-side coefficients β_h^- . Right column: gain-side coefficients β_h^+ . Each regression includes the focal-horizon return splits, the control vector $X_{i,t}$ of column (4) of Table A.1, which contains the sorting variable's own level, and stock and time fixed effects. The dependent variable is monthly retail net order flow scaled by shares outstanding (percentage points). Error bars are 95% confidence intervals based on standard errors clustered by stock and by calendar month. The sample is January 2017 through December 2024.

Positive values are predicted buying and negative values predicted selling.¹¹

Figure C.3 sorts covered stock-months into five ranges of predicted selling, a near-zero group, and five ranges of predicted buying, and plots the average of four outcomes in each range after removing month effects. The first three panels show that

¹¹Fund flows follow Coval and Stafford (2007); holdings are from the most recent report before the flow month. A stock-month is covered when at least five funds with flows hold the stock and mutual funds own at least one percent of it; 86 percent of stock-months are covered. Predicted selling makes up 83 percent of covered stock-months, and the median predicted trade is a tenth of a percent of shares outstanding. The measure contains no price or volume term. Fund flows persist, so a month of heavy predicted selling is usually the trough of an outflow episode rather than a surprise.

the shock does what institutional trading should do. The month's return (Panel A) rises from the strongest-selling range to the strongest-buying range, monotonically and close to a straight line: the strongest-selling range earns about half a percent less than the near-zero group and the strongest-buying range about one and a half percent more. 13-F institutional ownership (Panel B) falls where funds sell and rises where they buy. Lendable supply (Panel C) moves the same way. Institutions trade, and prices and ownership move with them. Panel D is the test. If retail took a fixed share of whatever funds traded, retail flow would be the mirror image of Panel A: a downward straight line, buying where funds sell and selling where funds buy. It is not. Retail net buying is positive in every range and small, between six and ten hundredths of a percent of shares outstanding. It is highest in the two extreme ranges and lowest through the moderate ranges on both sides, and a line fitted through the eleven ranges has a slope of zero.¹² In the strongest-selling range, retail buy about three percent of the shares that funds are predicted to sell. In the strongest-buying range, retail do not sell; they buy more than in any other range.

Two further facts say what the small selling-side response is. First, adding controls in steps on a fixed sample, the coefficient of retail flow on FIT halves when the stock's prior characteristics, lagged flows, and pressure history are held fixed, halves again with the past twelve months of returns, and is indistinguishable from zero once the current return is in; the institutional flow proxies add nothing, whether they enter before or after the returns.¹³ Given the price path, the fact that funds were selling carries no information about retail trading. Second, retail buy for three months after a stock enters the bottom decile of FIT within a month, forced selling, and do not sell after entries into the top decile, forced buying.¹⁴ A fixed-share counterparty would do both. A price rule does only the first: a selling shock lowers the price and the loss side responds, a buying shock raises the price and the gain side is flat.

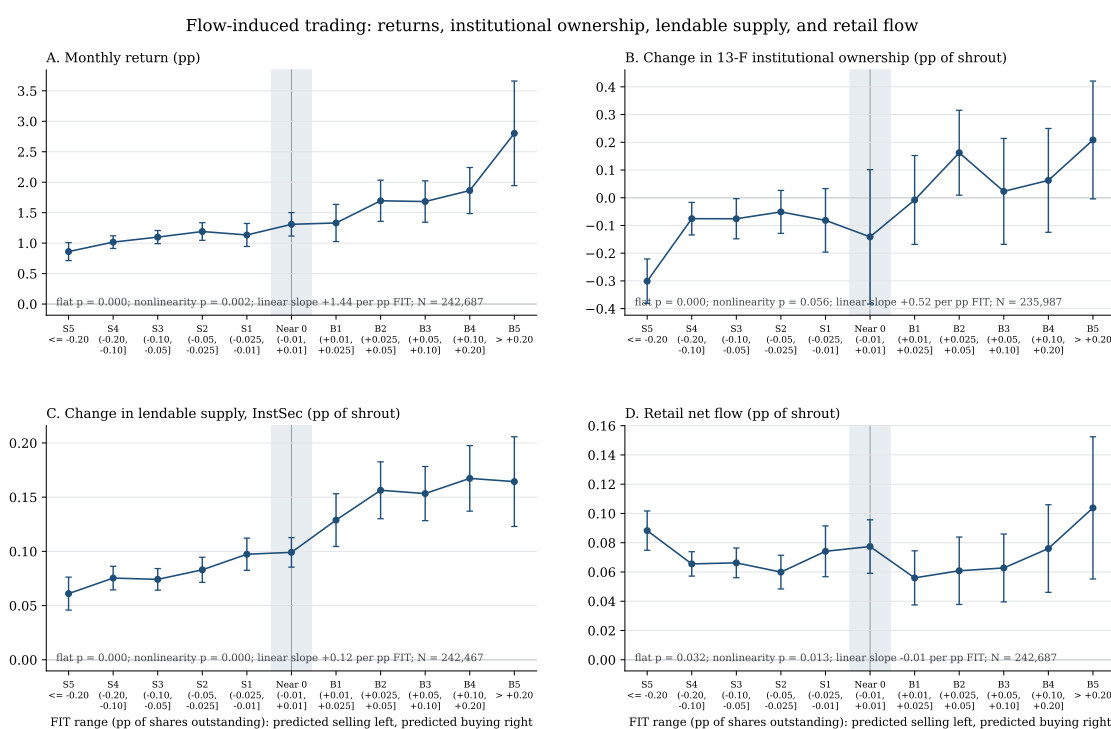
C.3 Where the Shares Come From: Equity Issuance

If retail are not taking the shares institutions sell, where do the shares they accumulate come from? Retail are net buyers throughout the sample, so the shares come from somewhere. Part of the answer is the firm. Table C.1 regresses monthly flow on the one-month change in shares outstanding, split into an issuance part (increases)

¹²Joint test that the eleven range levels are equal: $p = 0.03$. Test of the nine range dummies after a linear FIT term: $p = 0.01$. Linear slope -0.008 percent of shares per percent of FIT ($t = -0.5$). The individual range contrasts with the near-zero group are imprecise.

¹³Retail net flow (percent of shares outstanding) on FIT standardized within month, stock and month fixed effects, two-way clustered standard errors, 2020 to 2024, 141,636 stock-months: -0.0121 (0.0056) with fixed effects only; -0.0058 with lagged characteristics, the prior month's retail flow, and three lags of FIT; -0.0033 adding twelve monthly return lags split at zero; -0.0005 adding the current return split at zero; -0.0009 adding the three institutional flow proxies, with a 95 percent interval of $[-0.010, 0.009]$. The current return is an outcome that retail trading itself affects, so the sequence shows that the price path is sufficient to account for the association, not that no other channel exists.

¹⁴Coefficients on first entries into the bottom and top decile in regressions of retail net flow at $t + h$ on both indicators, with size, Amihud illiquidity, prior return volatility, and institutional ownership as controls, stock and month fixed effects, two-way clustered: bottom decile $+0.028$, $+0.029$, $+0.034$ percent of shares at $h = 0, 1, 2$ (t between 2.3 and 2.9); top decile $+0.005$, $+0.008$, -0.002 , none significant. The test that the two responses sum to zero has $p = 0.08$, 0.02, and 0.06.



Same 242,687 stock-months as Figure 9 (outcome coverage varies by panel). Month FE only, no stock FE or controls.
 Levels at the average month effect; 95% intervals clustered by stock. Equal spacing represents range order.

Figure C.3. Flow-Induced Trading, Returns, Institutional Ownership, and Retail Flow. Stock-months covered by the flow-induced trading measure are sorted into five ranges of predicted selling, five ranges of predicted buying, and a near-zero group, with boundaries at 0.01, 0.025, 0.05, 0.10, and 0.20 percent of shares outstanding on each side. Each panel plots the average level of the outcome in each range from a regression on range indicators and calendar-month fixed effects, evaluated at the average month effect. Panel A: monthly return (percent). Panel B: quarterly change in 13-F institutional ownership assigned to the month (percent of shares outstanding). Panel C: change in securities-lending supply (percent of shares outstanding). Panel D: retail net order flow scaled by shares outstanding (percent). Bars are 95 percent intervals from standard errors clustered by stock. Ranges are equally spaced. The sample is February 2017 through December 2024, 242,687 stock-months; Panels B and C use the stock-months in which the outcome is available.

and a repurchase part (decreases), with the return windows of Section 3.1, firm characteristics, and stock and month fixed effects. Because both sides are in percent of shares outstanding, each coefficient is the fraction of the share change matched by that group's net initiated flow in the month. In a month in which a firm raises its share count by one percent, retail net buying rises by six hundredths of a percent of shares outstanding: retail take about six percent of the new shares in the month, and eight percent over the month and the three after it (Figure C.4). Total initiated flow moves the other way. It is sell-initiated by four percent of the new shares in the month and ten percent over four months. That is the offering itself, the seller hitting bids,

with retail on the other side.¹⁵

Two features separate this from a price response. The return controls remove less than a fifth of the retail coefficient, from 0.072 with no controls to 0.060 with the full set, and the retail dip-buying slopes in the same regression keep their usual size next to it. Retail absorb issuance in addition to buying dips, not instead of it, and they absorb it equally in months in which the stock fell and in months in which it rose.¹⁶ And the response is one-signed. In repurchase months retail supply about one percent of the retired shares; the sell-initiated footprint of buybacks, three percent of retired shares in the month and ten percent over four months, is carried by other traders. Whether or not the price fell, and whichever way the initiative ran, the new shares land with retail.

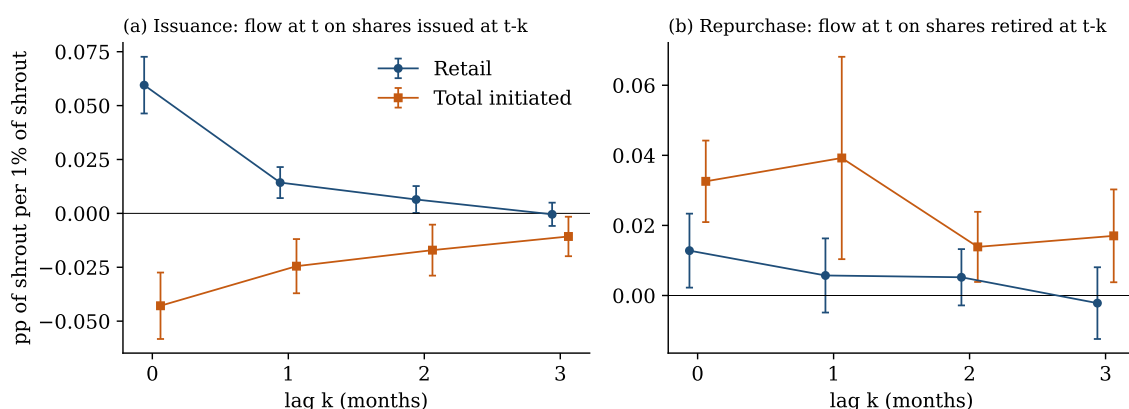


Figure C.4. Retail and Total Initiated Flow Around Equity Issuance and Repurchases. Coefficients on the issuance part (Panel A) and the repurchase part (Panel B) of the one-month percent change in split-adjusted shares outstanding at lags $k = 0$ to 3, from regressions of monthly retail net flow (circles) and total initiated flow (squares), both in percent of shares outstanding, on the current and three lagged share changes, the non-overlapping return windows split at zero, return-distribution moments, firm characteristics, and stock and month fixed effects (the full-controls specification of Table C.1). A coefficient of 0.06 at lag 0 means that the group's net initiated flow in the month matches six percent of the shares issued that month. In Panel B the share change is negative, so a positive coefficient means net selling when shares are retired. Bars are 95 percent intervals from standard errors two-way clustered by stock and month. Stock-months with an absolute share change above 50 percent are excluded. The sample is January 2017 through December 2024.

¹⁵The large-trade institutional proxy shows no response at any lag. Institutions receive their share of new issues, about half, through allocations at the offer price and stock-financed deals that never appear as initiated flow: 13-F holdings rise in issuance quarters where doing nothing would have let them fall, and lendable supply rises with issuance. Stock-months with an absolute share change above 50 percent, which are reverse mergers and placements into shells rather than offerings, are excluded.

¹⁶Retail coefficient on issuance recorded in month t : 0.058 ($t = 7.8$) in months with a negative return and 0.056 ($t = 5.3$) in months with a positive return. Part of the direction runs from retail to the firm: retail inflow in one month predicts a recorded share-count increase the next, prices run up before the recorded date, and firms sell into that demand, as at-the-market programs are designed to do. The coefficient is then an accounting fact about who ends up with the new shares.

Table C.1. Retail and Institutional Flow Around Equity Issuance and Repurchases

Monthly regressions of the indicated net flow (percent of shares outstanding) on the one-month percent change in split-adjusted shares outstanding, split into an issuance part (positive changes, zero otherwise) and a repurchase part (negative changes, zero otherwise), at lags 0 to 3, with the non-overlapping return windows $\text{Ret}[t]$, $\text{Ret}[t-2, t-1]$, $\text{Ret}[t-6, t-2]$, and $\text{Ret}[t-12, t-6]$ split at zero, idiosyncratic volatility and skewness, beta, log size, book-to-market, leverage, profitability, age, and gross profitability, and stock and month fixed effects. Coefficients on the supply parts read as the fraction of the share change matched by the group's net initiated flow in the month; because the repurchase part is negative, a positive coefficient on it means net selling when shares are retired. The return-split coefficients are in percent of shares outstanding per unit return. Panel A reports the full specification for each flow measure; Panel B adds the controls to the retail regression in steps. The last column is the sum of the issuance coefficients at lags 0 to 3. Stock-months with an absolute share change above 50 percent are excluded. t -statistics from standard errors two-way clustered by stock and month are in parentheses. *, **, *** denote significance at the 10%, 5%, and 1% levels. The sample is January 2017 through December 2024.

	Issuance _{t}	Repurchase _{t}	Issuance _{$t-1$}	Repurchase _{$t-1$}	Ret[t] ⁻	Ret[t] ⁺	Ret[$t-2, t-1$] ⁻	Ret[$t-2, t-1$] ⁺	$\sum_{k=0}^3 \text{Iss.}$	N
<i>Panel A: full specification, by flow measure</i>										
Retail flow	0.0595*** (8.87)	0.0128** (2.38)	0.0143*** (3.89)	0.0057 (1.06)	-1.018*** (-4.53)	-0.394 (-1.53)	-0.908*** (-7.15)	-0.112 (-0.84)	0.080	281,627
BJZZ	0.0108** (2.55)	0.0037* (1.69)	0.0055*** (2.63)	-0.0066** (-2.24)	0.703*** (6.94)	-0.354** (-2.03)	-0.072 (-0.84)	0.012 (0.28)	0.018	281,588
BHJOS	0.0117*** (3.17)	0.0014 (0.91)	0.0068*** (3.32)	-0.0066 (-1.64)	0.624*** (5.02)	-0.584** (-2.51)	-0.117 (-1.47)	0.112* (1.94)	0.019	281,623
Inst50k	-0.0027 (-1.55)	-0.0027 (-1.02)	-0.0021* (-1.92)	0.0052* (1.87)	0.327*** (4.09)	0.124 (0.83)	0.171*** (3.02)	-0.169*** (-4.40)	-0.009	254,505
InstSec	0.0293*** (14.60)	0.0234*** (6.72)	0.0110*** (9.40)	0.0194*** (3.09)	1.136*** (10.41)	0.072 (0.57)	0.418*** (6.14)	0.088*** (3.64)	0.060	281,050
Non-retail initiated	-0.1024*** (-9.46)	0.0198*** (2.74)	-0.0388*** (-4.78)	0.0335* (1.85)	2.978*** (6.54)	1.050* (1.80)	1.499*** (6.47)	-0.223 (-0.95)	-0.175	281,627
Total initiated	-0.0429*** (-5.45)	0.0326*** (5.49)	-0.0245*** (-3.82)	0.0392*** (2.66)	1.960*** (5.18)	0.657 (1.47)	0.591*** (3.01)	-0.334* (-1.79)	-0.095	281,627
<i>Panel B: retail flow, controls added in steps</i>										
Supply only	0.0723*** (11.05)	0.0018 (0.52)	0.0168*** (5.34)	0.0056* (1.68)					0.098	371,593
Adding return splits and risk	0.0653*** (8.63)	0.0140*** (2.82)	0.0137*** (3.97)	0.0065 (1.32)	-1.173*** (-4.92)	-0.445* (-1.80)	-0.999*** (-8.04)	-0.163 (-1.31)	0.087	297,915
Adding firm characteristics	0.0595*** (8.87)	0.0128** (2.38)	0.0143*** (3.89)	0.0057 (1.06)	-1.018*** (-4.53)	-0.394 (-1.53)	-0.908*** (-7.15)	-0.112 (-0.84)	0.080	281,627

D Calibration, Measurement, and Inference

This appendix describes the calibration, measurement conventions, inference, and numerical experiments. The return–dividend regression R^2 and price informativeness are not calibration targets.

D.1 Fixed primitives and parameter normalizations

The discount factor is $\beta = 0.987$. The persistent dividend component has persistence $\rho_\theta = 0.960$ and innovation standard deviation $\sigma_\theta = 0.016$. We set supply persistence to $\rho_\lambda = 0.500$ and the reference retail share to $\alpha_0 = 0.250$. The mean dividend $\bar{v} = 1.000$, mean supply $\bar{\lambda} = 0.002$, and individual retail capacity $\eta_R = 5.200$ are normalizations. These choices imply $\tau_R = \alpha_0 \eta_R = 1.300$; the capacity of entering investors is $\tau_E = \delta \tau_R$, rather than τ_R .

Since $v_t = \bar{v} + \theta_t + \epsilon_t$, with independent persistent and transitory components,

$$\text{Corr}(v_t, v_{t-1}) = \frac{\rho_\theta \text{Var}(\theta_t)}{\text{Var}(\theta_t) + \sigma_\epsilon^2}, \quad \text{Var}(\theta_t) = \frac{\sigma_\theta^2}{1 - \rho_\theta^2}. \quad (19)$$

We choose the transitory variance to match dividend autocorrelation $a_v = 0.820$:

$$\sigma_\epsilon^2 = \frac{\sigma_\theta^2}{1 - \rho_\theta^2} \left(\frac{\rho_\theta}{a_v} - 1 \right) = 5.575 \times 10^{-4}. \quad (20)$$

The resulting transitory standard deviation is 0.024. The dividend process therefore contains a persistent state and independent measurement-period variation; its autocorrelation is not obtained by changing the innovation variance of a single AR(1). The loading of institutional value on the persistent state is $\beta\rho_\theta/(1 - \beta\rho_\theta) = 18.055$.

Fixing η_R removes an exact redundancy in retail demand. Multiplying η_R by a positive constant and dividing b_c , b_p , and χ by the same constant leaves the demand schedule unchanged. The flow moments consequently do not separately identify retail risk aversion, subjective payoff variance, and the scale of subjective gains. Mean supply is also a level normalization. Adding a constant to the entire supply path changes the level of prices and institutional positions, but leaves price changes and retail decisions unchanged. No holdings-level moment is used to estimate $\bar{\lambda}$.

D.2 Measurement and return normalization

The equilibrium is solved using the raw price change $r_t = p_t - p_{t-1}$ and raw quantities. Only after simulation do we construct the return measure used in the regression moments:

$$R_t = c_r r_t, \quad c_r = \frac{0.183}{s_r}, \quad Q_t^R = F_t + \bar{F}, \quad (21)$$

where s_r is the simulated standard deviation of raw price changes. The return factor is recomputed at each parameter candidate during estimation. For applications, it is computed once from the larger baseline simulation and then held fixed across shocks, retail shares, and belief counterfactuals. That baseline simulation gives $s_r = 0.279$ and $c_r = 0.655$. The numerical calculations retain the unrounded factor.

This transformation may be written $R_t = r_t/P_*$ for the fixed reference denominator $P_* = 1/c_r = 1.526$. The denominator is chosen to match the return standard deviation; it is not the model's mean price. Reported returns are therefore constant-scale price changes, rather than realized simple or log returns. Figures multiply R_t and Q_t^R by 100. A response labeled 30% refers to $100R_t = 30$ in these reporting units. There is no rescaling within the equilibrium or within an impulse-response calculation.

Retail flow has reporting multiplier one. Its standard deviation of 0.014 is matched by the estimated parameters. This requires a measurement mapping from the empirical signed-flow series to model net purchases in common quantity units. A buyer-initiated volume imbalance or a flow divided by a time-varying denominator is not automatically identical to Q_t^R ; using such a measure would require an explicit additional mapping.

The absolute flow moment fixes a scale that a flow-volatility ratio alone cannot determine. For example, with η_R and the dividend process fixed, multiplying $(b_c, b_p, \chi, \tau_I, \bar{\lambda}, \sigma_\lambda, \bar{F})$ by the same positive constant multiplies holdings and flows by that constant and leaves prices and returns unchanged. Independently normalizing both returns and flows would preserve the regression slopes under this

transformation. The raw-flow target removes this freedom under the maintained measurement mapping.

The aggregate net-flow identity gives

$$M_F \equiv \frac{\text{sd}(F_t + \Delta I_t)}{\text{sd}(F_t)} = \frac{\text{sd}(\Delta \lambda_t)}{\text{sd}(F_t)}, \quad \sigma_\lambda = \sqrt{\frac{1 + \rho_\lambda}{2}} M_F \text{sd}(F_t). \quad (22)$$

The final equality uses the stationary AR(1) supply process. Adding the constant passive inflow to both aggregate and retail net purchases leaves their standard deviations unchanged. Thus the ratio and the absolute retail-flow moment jointly discipline supply innovation volatility. Total flow here denotes aggregate net purchases by the modeled investors, not gross turnover.

D.3 Moment conditions and computation

For $j \in \{0, 1\}$, we estimate separate continuous, common-intercept regressions

$$Q_t^R = a_j + \beta_j^- \min(R_{t-j}, 0) + \beta_j^+ \max(R_{t-j}, 0) + e_{j,t}. \quad (23)$$

The regressions pool observations across simulated paths. Both use the same dependent-variable dates, omitting the first retained observation of each path, and lags never cross path boundaries. The coefficients describe equilibrium associations. They are distinct from the structural coefficients b_c and b_p .

The six targets are

$$m^* = (-0.057, 0.000, -0.031, 0.000, 1.500, 0.014)', \quad m = (\beta_0^-, \beta_0^+, \beta_1^-, \beta_1^+, M_F, \text{sd}(F_t))'. \quad (24)$$

The past loss-side slope and the absolute retail-flow standard deviation are supplied empirical inputs. The current loss-side slope, both zero gain-side slopes, and the total-to-retail flow ratio are imposed calibration targets. Retail-flow autocorrelation, return autocorrelation, price informativeness, the return–dividend R^2 , and the shape of the volatility curve are not targeted.

We estimate the six parameters jointly using coordinates

$$\phi = (\chi / \tau_L, b_c / \tau_L, b_p / \tau_L, \tau_L, \delta, \sigma_\lambda)' \quad (25)$$

and minimize $[m_S(\phi) - m^*]'W[m_S(\phi) - m^*]$, with

$$W = \text{diag}(w_j^{-2}), \quad w = (0.010, 0.010, 0.010, 0.010, 0.100, 0.002)'. \quad (26)$$

The weights specify numerical calibration scales rather than inverse sampling variances. Each objective evaluation uses 64 independent paths with 4,096 retained periods following a burn-in of 4,096 periods. Common innovations are held fixed across parameter candidates. A bounded least-squares search from three dispersed starting values converges to the same interior solution and matches the six targets to numerical precision. These convergence checks establish reproducibility of the reported local solution; they do not establish global identification.

The additional supplied mean-to-standard-deviation target is $E[Q_t^R] / \text{sd}(Q_t^R) = 0.150$. Conditional on the active-economy parameters, the passive inflow is calibrated analytically:

$$\hat{F} = 0.150 s_{F,S} - \bar{F}_S = 2.100 \times 10^{-3}, \quad (27)$$

where the final value is rounded. Because each dip-buying regression has a free

intercept and the remaining moments use standard deviations, adding this constant does not change the six-parameter objective. The passive inflow and its matching trend in outside supply are held fixed in subsequent exercises. Positive average flow does not imply that retail investors are net buyers at every date.

D.4 Conditional standard errors

The available target covariance is

$$\Omega_{\text{target}} = \text{diag}(0, 0, 1.600 \times 10^{-6}, 0, 0, 0). \quad (28)$$

The nonzero entry is the inherited variance of the past loss-side slope. The remaining entries condition on the imposed targets and on the supplied retail-flow standard deviation, whose sampling variance is unavailable. The reported standard errors are therefore conditional local measures of uncertainty; they are not full empirical parameter standard errors. A full calculation requires the joint sampling covariance of the empirical moments and any estimated primitives.

We estimate simulation covariance from 40 independent moment vectors, each computed with the estimation sample dimensions. Their sample covariance estimates the variability of one simulated moment estimator and is not divided by 40. Assuming independence of empirical and simulation errors, write $\Omega = \Omega_{\text{target}} + \Omega_{\text{simulation}}$. With six moments and six parameters, the nonsingular local Jacobian $J = \partial m_S / \partial \phi'$ gives

$$V_\phi = J^{-1} \Omega J^{-1}. \quad (29)$$

For the reported parameter vector $\vartheta = (\chi, \tau_L, b_c, b_p, \delta, \sigma_\lambda)'$, we apply the delta method to $\vartheta = g(\phi)$:

$$V_\vartheta = Dg(\phi) V_\phi Dg(\phi)'. \quad (30)$$

This transformation retains covariances with institutional capacity. The covariance matrices already refer to moment estimators, so there is no additional division by the number of observations. We do not report a standard error for the separately profiled passive inflow because the sampling variance of its mean-to-standard-deviation target is unavailable.

The Jacobian is computed by central finite differences using common innovations and steps $h_j = \iota \max(|\phi_j|, \ell_j)$, where $\ell = (0.010, 0.010, 0.010, 0.010, 0.010, 0.001)'$. The reported calculation uses $\iota = 0.005$. Repeating it at $\iota = 2.500 \times 10^{-3}$ and 0.010 changes no standard error by more than 0.048% relative to the reporting choice. The Jacobian has rank six and scaled condition number 204.274. The especially small conditional standard error for supply volatility reflects the fixed flow-volatility anchors in equation (22); it does not establish comparable empirical precision.