Accidental Market Makers*

Matheus Brito † Fernando Chague ‡

March 21, 2026

Abstract

Using comprehensive data from the Brazilian equity market (2017-2022), we document that retail investors act as unintentional liquidity providers on days of high adverse-selection risk — accidental market makers. On such days, professional market makers withdraw while retail investors expand their passive limit-order volume, absorbing institutional demand precisely when it is most costly to supply. Retail investors fail to internalize the winner's curse embedded in passive order execution: day-zero returns on passive retail orders reach −59 basis points on buying days and −58 basis points on selling days, with no subsequent reversal, implying an annual wealth transfer of approximately USD 1.0 billion. A shift-share instrument based on regional flows into Brazil's government bond platform yields a 2SLS estimate of 0.97 standard deviations improvement in market quality per one-percent increase in retail volume — more than six times the OLS estimate of 0.15, consistent with retail investors participating disproportionately when market quality is already deteriorated.

*Chague gratefully acknowledges financial support from CNPq. Brito gratefully acknowledges financial support from Fapesp, grant 2024/09763-1.

†Getulio Vargas Foundation - Sao Paulo School of Economics, Brazil. E-mail: matheus.brito@fgv.edu.br

‡Getulio Vargas Foundation - Sao Paulo School of Economics, Brazil. E-mail: fernando.chague@fgv.br.

1 Introduction

While automated market makers are explicitly designed to supply liquidity through algorithms and profit from it, retail investors in equity markets appear to perform a similar function without either the design or the reward: they act as *accidental market makers.* We show that retail net order imbalances systematically move in the opposite direction of institutional liquidity demand, particularly on days with high information asymmetry. Retail trading therefore absorbs order flow precisely when adverse-selection risk is most acute, effectively shielding market-making institutions from this risk. We further corroborate the beneficial effect of retail trading on market quality using a shift-share instrument, showing that exogenous increases in retail trading flows improve overall market quality.

The existing literature offers mixed evidence on the contribution of retail investors to market quality.[1] We contribute to this debate by providing new evidence from comprehensive data along two dimensions that are important to answer this question. First, our data allow us to decompose retail trading flows by the order types that initiate trades, namely market orders and non-marketable limit orders. Second, we classify institutional trading flows into directional institutions and professional market-making institutions based on their observed trading behavior. Observing the full universe of trading flows, together with the order types used by each investor group, allows us to directly document how and when retail order flow interacts with institutions demanding liquidity.

Our data come from the *Comissão de Valores Mobiliários* (CVM), the Brazilian counterpart to the U.S. Securities and Exchange Commission, and cover the universe of daily equity

[1] Papers documenting beneficial effects of retail trading on market quality include Greene and Smart (1999), Kaniel, Saar, and Titman (2008), Kelley and Tetlock (2013), Peress and Schmidt (2020), Jones, Reed, and Waller (2025), Eaton, Green, Roseman, and Wu (2022), and Friedman and Zeng (2024). Papers documenting harmful or destabilizing effects include Black (1986), De Long, Shleifer, Summers, and Waldmann (1991), Foucault, Sraer, and Thesmar (2011), and Barber, Huang, Odean, and Schwarz (2022).

trading flows in Brazil from January 2017 to December 2022. The Brazilian equity market has several institutional features that make it particularly well suited to our analysis. First, payment for order flow and retail order internalization were not allowed during our sample period, so retail trading flows interacted directly with institutional flows on the centralized exchange, making it possible to measure their contribution to market quality without the confounding effects of off-exchange intermediation. Second, trading takes place on a modern, centralized limit order book, with participation by many of the same professional market makers active in the United States[2] and with broadly comparable participation rates.[3]

Our analysis focuses on the most liquid Brazilian stocks, where retail and institutional investors interact most intensively. We first examine how the three investor groups trade across stock-days sorted by abnormal returns. On normal days, retail investors are nearly neutral (−0.10 p.p.), while directional institutions are net buyers (+0.60 p.p.) and market-making institutions are their primary counterparties (−0.51 p.p.). On days with large price movements, however, retail investors become the main offsetting force, absorbing institutional imbalances primarily through passive orders. In the extreme negative bin, retail investors buy +6.91 p.p. net (+7.47 p.p. through passive orders alone) compared with only +2.07 p.p. from market makers. The pattern is symmetric on positive-return days. We find the same structure around earnings announcements and material fact disclosures: retail investors systematically take the opposite side of directional institutions through passive orders, buying after negative news and selling after positive news.

We next test more formally whether retail investors expand their activity when adverse-

[2]Brazil's market infrastructure closely resembles that of other major exchanges, including HFT participation, co-location services, and well-developed derivatives and ETF markets.

[3]We focus on the most liquid stocks in the Brazilian market, which account for more than 95% of total market capitalization. Within this group, retail investors account for 18% of total trading volume during our sample period, directional institutions for 33%, and market-making institutions for 49%. These shares are remarkably similar to those observed in the United States.

selection risk is high, running stock-day panel regressions of each investor group's trading share on five proxies for adverse-selection risk. The results reveal a clear substitution pattern. Retail participation rises by 5.7 p.p. during abnormal-volatility episodes, 1.5 p.p. on earnings-announcement days, 0.7 p.p. on material-fact-release days, 0.8 p.p. on intense-news days, and 1.5 p.p. on abnormal-spread days — economically large effects relative to retail investors' unconditional share of 17.6 p.p. Market-making participation falls by symmetric magnitudes across all five proxies, while directional institutions show little systematic response. When adverse-selection risk rises, professional market makers withdraw and retail investors fill the gap.

Having established that retail investors step in precisely when adverse-selection risk is high, we next ask whether this behavior is economically large enough to affect overall market quality. We construct an index combining four standard microstructure metrics — price impact, return autocorrelation, bid–ask spread, and the volatility ratio — where higher values indicate better quality. A 1% increase in retail volume is associated with a 0.15 standard-deviation improvement in the index. Examining the components of the index separately, we find that the improvement is driven primarily by narrower bid–ask spreads and lower price impact. Moreover, when we decompose retail volume by order type, only passive retail flows improve market quality, while aggressive and other flows have no meaningful effect.

To alleviate concerns about reverse causality—namely, that retail investors may choose to place orders precisely on days when market quality is already better—we adopt an instrumental-variables approach. Our instrument exploits exogenous variation in *regional* retail investment activity in the government bond market.[4] Specifically, we construct a

[4]Brazilian government bonds can be purchased directly by retail investors through their equity brokers using the Tesouro Direto platform. A comparable platform exists in the United States, TreasuryDirect.gov, although an important difference is that the Brazilian platform allows investors to buy and sell government bonds on every trading day.

shift-share instrument by combining one-year-lagged city-level stock portfolio weights with daily city-level bond trading flows. Intuitively, city-level flows into government bonds capture shocks to retail investors' overall propensity to invest on a given day, while the lagged portfolio weights determine how those shocks map into individual stocks. This design isolates variation in retail stock trading that is orthogonal to daily stock-specific shocks. The IV results not only confirm a causal effect but also reveal that OLS substantially *understates* it: in the baseline specification, a 1% increase in retail volume raises the market quality index by 0.97 standard deviations in 2SLS, compared with only 0.15 in OLS—a ratio of more than six to one. This amplification is consistent with retail investors participating disproportionately on days when market quality is already poor, so that the raw correlation between retail activity and market quality is attenuated toward zero. The first stage is also strong in all specifications, with F-statistics and Wald statistics comfortably above conventional weak-instrument thresholds. Taken together, these results indicate that exogenous increases in retail participation lead to meaningful improvements in market quality.

Do retail investors benefit financially from their liquidity-provision role? No. Retail investors incur large losses on the very days in which they supply liquidity, and these losses are not reversed afterward. Using volume-weighted day-zero returns, we find that retail investors lose about 55.5 basis points on buying days and 48.5 basis points on selling days; on heavy-imbalance days, these losses become even larger, reaching 61.3 and 52.1 basis points, respectively. The losses are especially pronounced for passive orders—the very orders through which retail investors provide liquidity—with day-zero returns of −58.9 basis points on buying days and −58.1 basis points on selling days. These magnitudes are far larger than mechanical trading costs: average bid–ask half-spreads are only about 5.4 to 5.9 basis points in the volume-weighted sample.[5] Nor are these losses offset by future reversals. When we form long–short portfolios based on retail imbalances, cumulative excess returns remain negative and, if anything, become more negative over horizons of up to 40 trading days. Consistent with this pattern, heavy retail buying predicts further losses of roughly 15 basis points over days $t+1$ to $t+5$, 36 basis points over days $t+1$ to $t+20$, and 66 basis points over days $t+1$ to $t+40$. In short, retail investors do not appear to earn compensation for supplying liquidity; instead, their liquidity provision represents a transfer to institutional counterparties.

Why, then, do retail investors behave this way? We interpret this pattern as consistent with information neglect. Passive orders are especially vulnerable to adverse selection because execution itself is informative: a standing buy order is more likely to execute when counterparties possess negative information, and a standing sell order is more likely to execute when counterparties possess positive information. Retail investors may fail to fully internalize this inference, treating execution as benign rather than as a signal that they are being picked off. This mechanism is closely related to the "cursedness" framework of Eyster, Rabin, and Vayanos (2019), in which traders neglect the informational content embedded in market outcomes. It is also consistent with empirical evidence showing that retail investors often overestimate their informational advantage and trade excessively without outperforming (Liu, Peng, Xiong, and Xiong, 2022), as well as with Brazilian evidence linking poor retail trading outcomes to informational neglect (Chague, De-Losso, and Giovannetti, 2018; Birru, Chague, De-Losso, and Giovannetti, 2024). Under this interpretation, retail investors do not intentionally act as market makers; rather, they supply liquidity because they underestimate the informational disadvantage embedded in passive execution.

[5] Moreover, such trading costs apply only to aggressive orders; passive orders should instead earn the spread. This makes the negative returns on passive retail orders even more striking.

Our findings contribute to two main strands of the literature. The first is the literature on retail investors as contrarian liquidity providers. Kaniel, Saar, and Titman (2008), Kaniel, Liu, Saar, and Titman (2012), and Barrot, Kaniel, and Sraer (2016) document that retail buying predicts positive subsequent returns, particularly during periods of market stress, and interpret this as retail investors absorbing institutional demand shocks. Linnainmaa (2010) and Kelley and Tetlock (2013) sharpen this picture by isolating passive limit orders as the operative channel: limit orders earn negative returns because they execute precisely when prices subsequently move against the investor, while aggressive orders earn positive returns. We advance this literature by directly documenting the substitution between retail and institutional market makers at the stock-day level. Using the full universe of Brazilian equity flows, we show that retail passive orders expand precisely when professional market makers contract—on high-adverse-selection days defined by elevated volatility, earnings announcements, and abnormal spreads. This is the mechanism through which retail investors become accidental market makers: they fill the gap left by institutional market makers withdrawing from adverse-selection risk (Nagel, 2012; Peress and Schmidt, 2020), even though they bear the cost of doing so.

The second strand is the causal identification of retail trading effects on market quality. Jones, Reed, and Waller (2025), Eaton, Green, Roseman, and Wu (2022), and Friedman and Zeng (2024) establish, using relief-check and brokerage-outage instruments, that retail trading causally improves market quality, with passive orders as the primary mechanism. Our shift-share instrument—combining lagged city-level stock portfolio weights with daily city-level flows into the Tesouro Direto government bond platform—provides an independent source of exogenous variation at the stock-day level. It reveals that OLS substantially understates the causal effect (0.97 standard deviations in 2SLS versus 0.15 in OLS), consistent with retail investors participating disproportionately on days when market quality is already poor.

Our paper also relates, more broadly, to the literature on retail order internalization and cream-skimming (Hansch, Naik, and Viswanathan, 1999; Grammig and Theissen, 2005; Larrymore and Murphy, 2009), on the heterogeneity of retail investors (Boehmer, Jones, Zhang, and Zhang, 2021; Barardehi, Bernhardt, Da, and Warachka, 2021; Eaton, Green, Roseman, and Wu, 2022), and on payment for order flow as a mechanism for extracting rents from retail order flow (Battalio, Corwin, and Jennings, 2016; Adams, Kasten, and Kelley, 2024; Baldauf, Mollner, and Yueshen, 2024). In Brazil's centralized order book, where there is no formal internalization and no wholesaler routing, the adverse-selection transfer from retail to institutional investors is directly observable—which allows us to quantify the day-zero wealth transfer at approximately USD 1.0 billion per year and to link it to market-wide price efficiency in a way that dealer-market studies cannot.

The paper proceeds as follows. Section 2 describes the data and defines the three investor groups. Section 3 documents the pattern of retail liquidity provision on adverse-selection days. Section 4 discusses some of the potential drivers of retail investors' accidental market-making behavior. Section 5 quantifies the day-zero wealth transfers to institutional counterparties and estimates the causal effect of retail trading on market quality. Section 6 concludes.

2 Data

Our primary dataset comes from the *Comissão de Valores Mobiliários* (CVM), the Brazilian counterpart of the U.S. Securities and Exchange Commission. The data comprise the complete daily equity trading flows for groups of investors from January 2017 to December 2022. The aggregated trading flows are reported separately for purchases and sales, both in financial value and in number of shares. We can therefore infer the average buying and selling prices of each investor group by dividing total financial volume by the total number of shares traded in a given stock–day.

Trading flows are classified into two broad investor groups: retail and institutional.[6] For each group, flows are further separated into ten bins based on whether the investor performed a two-sided trade within a stock–day. At one extreme, a fraction equal to 0 indicates that the investor exclusively bought or exclusively sold the stock on that day. A fraction of 0.1 indicates that the investor both purchased and sold the stock, with an overlap that, rounded, amounts to 10% (i.e., 10% of purchased shares were also sold, or equivalently, 10% of sold shares were also purchased on that stock–day). This classification continues with fractions of 0.2, 0.3, and so forth, up to 1, which indicates that purchases and sales were exactly balanced on that stock–day.[7] Finally, for each stock–day–"investor-type"–"two-sidedness-bin" cell, flows are further decomposed by execution mechanism: non-marketable (passive) limit orders, marketable (aggressive) orders, and call-auction executions.[8]

Firm characteristics and prices, adjusted for corporate actions and dividend payments, are obtained from Econom'atica. Bid-ask spreads, intraday prices, and firm-level news are obtained from Bloomberg.

Our analysis focuses on the 111 most liquid stocks in Brazil (i.e., those with a median daily trading volume above USD 5 million over the sample period) since our objective is to

[6]The institutional group includes foreign investors.

[7]Except for bins 0 and 1, the intermediate bins are fractions rounded to the nearest decimal.

[8]On the Brazilian exchange, call-auction executions include the opening auction at market open, the closing auction at market close, and occasional intraday auctions triggered by large price movements or block orders. In addition, a small fraction of trades is internalized through so-called direct orders.

study retail liquidity provision in the most relevant set of securities.[9] These stocks account for approximately 95% of the total market capitalization of Brazilian listed firms.

2.1 Decomposing Institutional Trading Flows

Based on the binning information, we classify institutional investors into two subgroups. *Directional institutions* are those in bin 0 —investors who either only bought or only sold the stock on a given day. We refer to them as directional because they take a one-way position, whether to speculate on a particular price movement or to meet portfolio rebalancing needs. The complement group, *market-making institutions*, consists of non-directional investors who traded on both sides of the market, though not necessarily in equal amounts.[10] Because institutions that trade on both sides within the day are likely to include traditional liquidity suppliers, this label captures the relevant economic distinction. In Section 3.3, we provide an alternative classification based on investor-level data that corroborates this approach.

Table 1 summarizes trading flows from 2017 to 2022 across the three investor categories: (i) *retail investors*, (ii) *directional institutions*, and (iii) *market-making institutions*. Panel A shows that market-making institutions dominate total trading activity, with an average stock-day volume of USD 40.9 million, followed by directional institutions at USD 22.5 million and retail investors at USD 13.9 million. Taken together, these groups account for an average daily stock volume of USD 77.3 million in the 111 most liquid stocks in our

[9] We restrict the sample to tickers with five characters. This excludes units (tickers that combine preferred and common shares), ETFs, BDRs (Brazilian Depositary Receipts, the equivalent of ADRs in the United States), odd lots, and over-the-counter stocks. We then apply the liquidity filter described above to the remaining securities, yielding a final sample of 111 stocks. For reference, the main Brazilian equity index, the Ibovespa, contains 88 stocks.

[10] We use the term in a broad sense: the category may include dealers and quasi-market-makers as well as strategies such as statistical arbitrage, execution-driven intraday rebalancing, and other high-turnover approaches that may not be intended as market making but can nonetheless supply liquidity by posting and updating limit orders.

sample.

[Table 1 about here]

Panel B shows each group's share of total trading volume relative to all market activity in a stock–day. Market-making institutions are responsible for nearly half of overall trading (49.5%), while directional institutions account for one-third (32.9%), and retail investors for the remaining 17.6%. These fractions highlight the central role of institutional investors in overall trading activity, but also the significant presence of retail investors, consistent with evidence from other markets.[11]

Panel C shows trading imbalances, defined as net buying or selling normalized by total share volume. While average imbalances are close to zero for all groups, the percentiles reveal substantial variation. For instance, retail investors' imbalances range from −4.5% to 4.7% of daily trading volume at the 10th and 90th percentiles. Notably, market-making institutions — which by definition trade both sides of a given stock–day, though not necessarily in equal amounts — display a wider spread, from −8.9% to 8.5%. Directional institutions exhibit the largest variation, from −9.3% to 9.9%. These patterns highlight that although imbalances net to zero on average, each group can hold significant net positions on a given stock–day.

Panel D reports the composition of each group's trading volume by order type. Retail investors are the most passive-order-intensive group: on average, 50.9% of their share volume is executed through non-marketable limit orders, compared with 42.6% for market-making institutions and 36.6% for directional institutions. Aggressive orders account for 40.6% of retail volume and 44.8% of market-maker volume, with directional institutions using aggressive and passive orders in nearly equal proportions (36.6% each). The most striking

[11] Barber et al. (2024) estimate the fraction of retail trading in the US using an algorithmic approach during a sample period comparable to ours. Their results imply a retail share of about 20% of total trading volume, slightly above our estimate for Brazil.

difference concerns other orders — comprising opening and closing auctions, direct trades, and other non-continuous sessions — which account for only 8.5% of retail volume but 26.7% of directional institutional volume, reflecting the greater reliance of large directional traders on block trades and scheduled auctions to execute their positions. The prominence of passive orders in retail trading is central to our analysis: as we show below, it is precisely through these orders that retail investors inadvertently supply liquidity on high-adverse-selection days.

3 Retail Investors as Accidental Market Makers

To examine the role of retail investors as accidental market makers, we study how the trading flows of retail investors, market-making institutions, and directional institutions interact with one another across different order-submission types and market conditions.

First, we group stock–days by abnormal price changes. Specifically, we compute average net trading shares by investor group and order type across bins of standardized excess returns, $StdRet^{ex}_{s,t}$. For stock s on day t, $StdRet^{ex}_{s,t}$ is defined as the stock's excess return relative to the equal-weighted market return (i.e., the cross-sectional average of $Ret_{s,t}$ across all s), standardized by its rolling 252-day standard deviation estimated over days $t-252$ to $t-1$. Stock–day observations are partitioned into seven bins of $StdRet^{ex}_{s,t}$: ≤ -3, $(-3,-2]$, $(-2,-1]$, $(-1,1)$, $[1,2)$, $[2,3)$, and ≥ 3. Observations in the extreme bins are more likely to correspond to periods of elevated adverse-selection risk for market-making investors, since large price changes are often associated with information arrival (Glosten and Milgrom, 1985, Easley and O'Hara, 1987).

Table 2 shows that on a typical day, bin 0—which contains standardized excess returns between −1 and 1 and includes 99,649 stock–day observations—the two institutional groups are on opposite sides: directional institutions hold a small positive average net share of 0.61 percentage points (p.p.), while market-making institutions have a negative average net share of −0.50 p.p. Retail investors are essentially neutral at −0.10 p.p. This pattern indicates that when prices are relatively stable, professional market makers provide most of the immediate counterflow to modest directional imbalances, with very limited net participation from retail.

[Table 2 about here]

When prices change meaningfully (both in negative and positive return bins), retail investors step up and absorb most of the directional imbalances, with market makers filling only part of the gap. On the downside, for example, bins −1 and −2 show directional institutions at −7.16 and −8.98 p.p., respectively, offset primarily by retail at +5.22 and +6.91 p.p., with market makers contributing a smaller +1.94 and +2.07 p.p. On the upside, bins 1, 2, and 3 show directional institutions at +7.96, +8.79, and +8.64 p.p., met mostly by retail at −5.07, −6.51, and −6.88 p.p., while market makers supply a more modest −2.89, −2.28, and −1.76 p.p. The symmetry across signs reinforces that retail investors are the main counterpart to large directional flows when prices move.

On very extreme negative days (≤ -3), bin −3 shows that market makers also trade opposite to retail—with a net share of −0.52 p.p. while retail stands at +8.76 p.p.—so retail remains the primary (indeed, almost the only) group offsetting the heavy selling from directional institutions (−8.24 p.p.). In other words, when price pressure is most severe on the downside, retail investors absorb all of the institutional net selling, with market makers not offsetting (and even slightly reinforcing) the directional flow.

Table 2 also shows net trading shares of aggressive, passive, and other flows, providing further insights into the mechanics of liquidity provision. The results clearly show that retail investors' role as accidental market makers on days with large price changes is driven by passive orders. For example, in the extreme negative bin (≤ -3), retail investors offset nearly all of the heavy net selling by directional institutions through passive orders, contributing +7.47 percentage points of net share volume, while their aggressive orders are only marginally positive (+1.68 p.p.). Similarly, in bin -2, the bulk of retail's counterbalancing flows comes from passive orders (+7.05 p.p.), with aggressive orders contributing just +0.96 p.p. On the upside, the pattern is symmetric: in bin 2, retail passive orders account for -5.88 p.p. of net flows (out of a total of -6.51 p.p.), and in bin 3, retail passive orders account for -6.47 p.p. of net flows (out of a total of -6.88 p.p.). In contrast, aggressive retail orders are near zero in both bins (-0.38 and -0.02, respectively).

Directional and market-making institutions play markedly different roles. For directional institutions, order imbalances are driven mainly by aggressive orders, consistent with their role as active position takers; passive orders are of secondary importance, although they generally have the same sign. Market-making institutions, in contrast, typically exhibit aggressive and passive flows with opposite signs, with passive flows moving against directional institutions' trades. They also account for a larger share of "other" flows, reflecting greater participation in the opening and closing auctions. Importantly, as noted above, in the most extreme negative-return bin (-3), market makers not only fail to offset directional institutions' trades but sometimes move in the same direction, leaving retail passive orders as the main stabilizing force.

3.1 Corporate Announcements

We have shown that retail trading flows step up on days with large price movements. However, large price movements do not necessarily reflect the arrival of meaningful information;

they may instead be driven by liquidity demand from uninformed institutions. To better understand the time-series dynamics of this behavior, we therefore examine retail trading flows around corporate news.

We focus on two categories of firm disclosures that typically convey new information to the market: earnings announcements and material fact releases. A key distinction between them is that earnings announcements occur on predetermined dates and are therefore anticipated by market participants, whereas material fact disclosures are issued on an as-needed basis and are less predictable. Because we study net trading flows, we classify events as positive, neutral, or negative according to firms' abnormal returns relative to the market, using a two-standard-deviation threshold based on the stock's rolling 252-day volatility.

For each event, we compute, for each stock–day, the net trading flow of each investor group as the difference between buy and sell share volume, normalized by total daily share volume. This measure allows us to identify on which side of the market retail investors provide liquidity during informationally intensive periods and to compare their behavior with that of the two groups of institutional investors.

Figure 1 shows that retail investors systematically take the opposite side of news-driven trading around earnings announcements. Following positive announcements, directional institutions become net buyers, whereas retail investors become net sellers; following negative announcements, the pattern reverses, with directional institutions selling and retail investors buying. These imbalances are concentrated on the event day and are much weaker around neutral announcements, consistent with the view that they reflect the arrival of firm-specific information rather than ordinary fluctuations in trading demand. The decomposition by order type further indicates that the retail response is driven primarily by passive orders, suggesting that retail investors supply liquidity through resting limit orders precisely when adverse-selection risk is elevated. Overall, the figure provides direct evidence that retail investors act as accidental market makers around earnings announcements.

[Figure 1 about here]

The magnitudes of these imbalances are in line with those documented for the most extreme return bins in Table 2. On day zero, following a positive earnings announcement, retail investors exhibit a net imbalance of −5.3%, of which −0.8% comes from aggressive orders, −4.2% from passive orders, and −0.4% from other flows; market-making institutions exhibit a net imbalance of 0.3%, with 1.7% coming from aggressive orders, −0.5% from passive orders, and −0.8% from other flows; while directional institutions exhibit a net imbalance of 5.0%, with 2.7% from aggressive orders, 1.1% from passive orders, and 1.2% from other flows.

In contrast, following a negative earnings announcement, retail investors exhibit a net imbalance of 7.2%, of which 1.1% comes from aggressive orders, 6.8% from passive orders, and −0.8% from other flows; market-making institutions exhibit a net imbalance of −3.1%, with −3.7% coming from aggressive orders, −0.03% from passive orders, and 0.6% from other flows; while directional institutions exhibit a net imbalance of −4.1%, with −3.6% from aggressive orders, −0.8% from passive orders, and 0.3% from other flows.

Figure 2 shows that retail investors also take the opposite side of news-driven trading around material fact announcements, despite the fact that these disclosures are unscheduled. Following positive announcements, directional institutions become net buyers while retail investors become net sellers; following negative announcements, the pattern reverses. As in the earnings-announcement case, these imbalances are concentrated on the event day and are driven primarily by retail passive orders. In contrast to Figure 1, however, there is little evidence here of anticipatory positioning by market-making institutions in the days before the announcement, consistent with the lower predictability of material fact disclosures. Moreover, in the case of positive material fact announcements, professional market makers also appear unable to avoid adverse-selection risk, as they provide liquidity to directional institutions on the event day.

[Figure 2 about here]

With respect to the magnitudes of these imbalances, on days with a positive material fact, retail investors exhibit a net imbalance of −4.8%, of which −0.1% comes from aggressive orders, −4.4% from passive orders, and −0.3% from other flows; market-making institutions exhibit a net imbalance of −2.4%, with 0.4% coming from aggressive orders, −1.5% from passive orders, and −1.3% from other flows; while directional institutions exhibit a net imbalance of 7.2%, with 3.4% from aggressive orders, 2.3% from passive orders, and 1.6% from other flows.

On days with a negative material fact, retail investors exhibit a net imbalance of 7.2%, of which 0.1% comes from aggressive orders, 6.8% from passive orders, and −0.6% from other flows; market-making institutions exhibit a net imbalance of −2.7%, with −1.8% coming from aggressive orders, 0.6% from passive orders, and 0.9% from other flows; while directional institutions exhibit a net imbalance of −6.9%, with −4.5% from aggressive orders, −2.2% from passive orders, and −0.3% from other flows.

3.2 Accidental Market Marking on High-Adverse Selection Risk Days

To more formally evaluate the interplay between professional market-makers and retail investors during periods when adverse-selection risk is high, we proceed as follows. We run stock-day panel regressions of the ratio of the group's total buying plus selling volume to the stock's total share volume on a given day on a number of indicator variables for periods of heightened adverse-selection risk: (i) *Abnormal Volatility*, equal to one when the average daily percentage price range over days $t-5$ to $t-1$ exceeds twice the corresponding average over days $t-60$ to $t-1$; (ii) *Earnings Announcement*, equal to one on days when quarterly results are disclosed, with after-hours announcements attributed to the following trading day; (iii) *Material Fact Release*, equal to one on days when firms issue mandatory corporate disclosures, with after-hours releases similarly shifted to the following trading day; (iv) *Intense News Activity*, equal to one when the number of Bloomberg news items over days $t-5$ to $t-1$ is more than twice the average over days $t-60$ to $t-1$; and (v) *Abnormal Bid–Ask Spread*, equal to one when the stock's bid–ask spread on day t exceeds its own trailing 60-day mean by more than two standard deviations. The regressions include stock fixed effects, and standard errors are clustered at the stock level.

Panel A of Table 3 reports the results for retail investors. Retail participation increases significantly for all five proxies of adverse-selection risk. In the univariate specifications, retail trading share is 2.9 p.p. higher during periods of abnormal volatility, 1.6 p.p. higher on earnings-announcement days, 0.9 p.p. higher on material-fact-release days, 0.9 p.p. higher when news activity is unusually intense, and 2.0 p.p. higher on days with abnormally high bid–ask spreads. When all five proxies are included jointly (column 6), the estimates remain positive and statistically significant: retail activity rises by about 5.7 p.p. during abnormal-volatility episodes, 1.5 p.p. on earnings-announcement days, 0.7 p.p. on material-fact-release days, 0.8 p.p. on intense-news days, and 1.5 p.p. on abnormal-spread days. These effects are also economically large relative to baseline retail participation: in the joint specification, the abnormal-volatility coefficient alone amounts to roughly one-third of the constant, while the other coefficients correspond to about 4–9% of the baseline trading share. These results indicate that retail investors expand their trading activity precisely when multiple indicators point to elevated adverse-selection risk.

[Table 3 about here]

Panel B turns to market-making institutions. Their trading share moves in the opposite direction of retail activity and declines sharply when adverse-selection risk rises. In the univariate specifications, market-making participation falls by 11.9 p.p. during abnormal-volatility periods, 1.5 p.p. on earnings-announcement days, 0.8 p.p. on material-fact-release days, 1.4 p.p. during periods of intense news activity, and 1.9 p.p. on abnormal-spread days. When all five indicators are included jointly, these negative effects remain large and statistically significant: the market-making share is lower by 11.3 p.p. during abnormal volatility, 1.4 p.p. on earnings-announcement days, 0.7 p.p. on material-fact-release days, 1.0 p.p. on intense-news days, and 0.9 p.p. on abnormal-spread days. This pattern suggests that market-making institutions withdraw from liquidity provision precisely when adverse-selection concerns are most severe.

Panel C presents the results for directional institutions. Unlike retail investors, these institutions do not systematically expand their participation in response to most information events. The only robust pattern is for abnormal volatility, under which their trading share rises by about 5.4 p.p. in the univariate regression and 5.6 p.p. in the joint specification. The coefficients on earnings announcements and material fact releases are small and statistically insignificant, while the positive effect of intense news activity is modest—0.4 p.p.—and only significant in isolation. The abnormal-bid-ask-spread proxy is uninformative in the univariate specification, but significantly negative in the joint regression, with a coefficient of about -0.6 p.p. Taken together, the results reveal a substitution pattern: when adverse-selection risk rises, market-making institutions scale back their activity, directional institutions do not materially step in, and retail investors expand their participation.

3.3 Alternative Decomposition of Institutional Trading Flows

Our baseline classification of institutions into market-making and directional is simple and transparent: an institution is labeled as market-making if it trades on both sides of the book on the same day, reflecting the assumption that market makers will balance their buying and selling within a trading session. While this approach is directly grounded in observable behavior, it has two limitations. First, our main dataset does not contain the additional trading characteristics — such as intraday trading frequency or inventory dynamics — that would allow us to further corroborate the classification. Second, and more importantly, the classification is based on ex post outcomes rather than predetermined institutional characteristics. An institution that would ordinarily trade directionally might appear two-sided on a particular day simply because it revised its views intraday or executed a large order in multiple tranches. Conversely, a market maker facing unusually one-sided order flow might be misclassified as directional. If such misclassification is systematic and correlated with retail trading activity, it could bias our estimates of how different investor groups interact.

To address these issues, and bring more information about the trading behaviors of market-making institutions, as a robustness exercise, we construct an alternative, ex ante classification of institutional investors based on trading frequency. We employ a dataset, also employed by Chague and Kimura (2026), that allows us to track institutions over time and measure trading frequency directly covering the same period (2017–2022). Observations are recorded at the daily investor–stock level, and we observe the total quantity and financial value of shares purchased and sold, as well as the number of buy and sell transactions.[12] A limitation of this dataset, however, is that it does not allow us to decompose trading flows by order type.

Because investors are anonymized in the data—we observe only whether an investor is retail or institutional, together with the investor's trading activity—we cannot directly identify market-making institutions. We therefore adopt an ex ante classification based on past trading frequency. The underlying idea is that institutions engaged in market making should trade much more frequently than directional investors, since their business model relies on repeatedly earning small margins throughout the day. In this sense, the group we identify can also be interpreted as high-frequency institutions, with substantial overlap between the two concepts. This classification is constructed out of sample: for each semester τ, we use trading behavior observed from the beginning of the sample through semester $\tau - 1$ to assign institutions to frequency groups, and then apply these classifications when analyzing trading patterns in semester τ. This procedure avoids look-ahead bias and allows us to relate ex ante trading intensity to subsequent market behavior.

To maintain comparability with our earlier analysis, we focus on the same set of highly liquid stocks in the Brazilian equity market. When computing investor-frequency metrics, we exclude U.S. bank holidays to ensure consistency with international market participation patterns, although all trading days are retained in the final aggregated dataset. For each institutional investor i in the cumulative classification window up to semester $\tau - 1$, we

[12] The identifier corresponds to the investor's tax ID. Consequently, if an institution or individual trades through multiple brokerage accounts, we observe their activity aggregated across all accounts.

construct two measures of trading frequency: the proportion of trading days on which the investor is active, $\text{PropActive}_{i,\tau-1} = \text{ActiveDays}_{i,\tau-1}/\text{TotalTradingDays}_{\tau-1}$, and the average number of trades per active day, $\text{AvgTrades}_{i,\tau-1}$. We then rank investors in ascending order according to these measures and partition them into two groups such that each group accounts for approximately one-half of total institutional trading volume.

Table 4 presents the distribution of institutional investors across frequency halves for each semester in our sample, along with their trading characteristics. Panel A shows a highly skewed distribution: the vast majority of institutions fall into the low-frequency half (i.e., the directional institutions), while only 26 to 251 institutions per semester are in the higher frequency half (i.e., the market-making institutions). A fraction of the investors cannot be classified because they did not trade in prior periods; these represent new market entrants or institutions returning after a period of inactivity, and we track them separately.

The classification exhibits high persistence over time: 99.9% of the directional institutions remain in the same half after one semester, as do 94.5% of market-making institutions. This persistence is consistent with the view that trading frequency reflects an institutional technology in which firms specialize and a stable organizational characteristic rather than a transient response to market conditions (Chague and Kimura, 2026). Indeed, the Spearman rank correlation of investor frequency rankings across consecutive semesters averages 0.97, confirming that relative trading intensity captures deep and persistent differences across institutions.

[Table 4 about here]

Panel B of Table 4 reports trading characteristics for each group across all semesters. Following Aït-Sahalia and Brunetti (2020), we examine investors along dimensions beyond trading intensity, including two-sided trading behavior and trading concentration. Because our data are recorded at the daily investor–stock level rather than intraday, we construct daily proxies designed to capture these underlying economic mechanisms. To measure within-day two-sidedness, we compute

$$\text{TwoSided}_{i,t} = \frac{2 \times \min(\text{BuyValue}_{i,t}, \text{SellValue}_{i,t})}{\text{BuyValue}_{i,t} + \text{SellValue}_{i,t}},$$

which approaches one when daily buying and selling are balanced and zero when trading is predominantly one-directional. To measure portfolio concentration, we compute the Herfindahl index across stocks at the investor-day level.

Directional institutions are active on only 7% of trading days, on average, and execute 151.9 trades per active day. These institutions display several characteristics consistent with directional trading: low two-sidedness (0.06), indicating predominantly one-directional order flow, and concentrated trading activity, as reflected in a high Herfindahl index (0.84) and a small number of stocks traded per day (2.1). By contrast, market-making institutions trade on 96% of all trading days and execute more than 6,500 trades per day. They exhibit substantially higher two-sidedness (0.60), indicating more balanced buying and selling, and more dispersed trading across securities, as reflected in a lower Herfindahl index (0.27) and a larger number of stocks traded per day (18.9).

Table 5 reports descriptive statistics for each of the four investor groups across 124,908 stock-day observations from July 2017 to December 2022. As expected given how we constructed the groups, Panel A shows that directional institutions and market-making institutions contribute similar trading volumes, while not-classified institutions contribute only US$1.5 million. Panel B shows that directional and market-making institutions together dominate market activity, each accounting for approximately 40% of total shares traded on a given stock-day, followed by retail investors at 18%. Not-classified institutions represent less than 2% of total share volume. Panel C indicates that net trading imbalances are close to zero for all groups, consistent with market clearing, although the dispersion is wider for institutional investors (standard deviations of 7.3% for directional and 6.6% for market-making institutions) than for retail (4.4%). Panel D highlights the concentrated nature of market-making activity: while retail investors average 1,168 distinct participants per stock-day, market-making institutions average only 34 investors, underscoring that a small number of institutions generate half of all institutional trading volume in the Brazilian equity market.

[Table 5 about here]

Appendix A.1 replicates our main findings using the ex ante frequency-based classification of market makers. The patterns are remarkably similar to those obtained under the baseline grouping, which is reassuring on two counts. First, the high persistence of the frequency classification confirms that market making is a stable institutional characteristic rather than a transient behavioral response. Second, the near-identical results across classifications rule out the possibility that our findings are mechanically driven by the way institutions are sorted — the substitution between retail investors and market makers on adverse-selection days is a feature of the data, not of the classification scheme.

Overall, the evidence in this section establishes two key facts about liquidity provision. First, market-making institutions withdraw precisely on days when adverse-selection risk is most acute. Second, retail investors step in on those same days, absorbing institutional demand predominantly through passive limit orders. This substitution explains why retail investors systematically end up on the opposite side of institutional imbalances when information asymmetry is high — not by design, but as the unintended consequence of limit-order strategies that fail to account for the adverse-selection cost of passive execution. We turn next to the potential drivers of this behavior.

4 Drivers of Accidental Market-Making

Why do retail investors behave as accidental market makers? We discuss two sets of candidate explanations: we first rule out the most natural non-behavioral alternatives — intentional market-making and rational liquidity provision — before arguing that the pattern is most coherently explained by behavioral biases that are specifically linked to the passive-order channel.

4.1 Intentional Market-Making

Because retail investors provide liquidity precisely when market conditions are most favorable to their counterparties, their same-day trading performance is naturally poor. However, subsequent price reversals may compensate for these day-zero losses, which would be consistent with an intentional market-making strategy. In this section, we assess whether such reversals plausibly offset initial losses, recognizing that retail investors—unlike professional market makers—may provide liquidity with longer investment horizons in mind.

To examine this possibility, we first construct a precise measure of day-zero returns and then track prices beyond day 0. We compute day-zero returns as follows. For each stock–day with positive net retail buying, the day-zero return is $(P_t^{close} - P_t^{buy})/P_t^{buy}$, where P_t^{close} is the closing price on day t and P_t^{buy} is the volume-weighted average price paid by retail investors for purchases on that day. For stock–days with negative net retail buying (i.e., net retail selling), the day-zero return is $(P_t^{sell} - P_t^{close})/P_t^{sell}$, where P_t^{sell} is the volume-weighted average price received by retail investors for sales. We classify a stock–day as a buying day if retail net flows are positive—that is, if the number of shares purchased by retail investors exceeds the number sold—and as a selling day if retail net flows are negative.

Table 6 reports the volume-weighted results, where each daily observation is weighted by the dollar value of the corresponding retail imbalance, following Barber, Lin, and Odean, 2023. This weighting scheme is useful because it places more emphasis on the stock–days in which retail liquidity provision is economically more important. Table IA.3 in the Internet Appendix reports the equally weighted analog. The equally weighted results point in the same direction, but the volume-weighted estimates are more negative.

[Table 6 about here]

Panel A of Table 6 shows that retail investors incur sizable same-day losses on both buying and selling days. Aggregating across all orders, the average day-zero return is about −55.5 basis points on buying days and −48.5 basis points on selling days. These losses are even more pronounced on heavy-imbalance days, reaching roughly −61.3 basis points for heavy buying days and −52.1 basis points for heavy selling days. The decomposition by order type is also informative. Passive orders exhibit the worst performance, with average day-zero returns of about −58.9 basis points on buying days and −58.1 basis points on selling days. Aggressive orders also lose money, but by substantially less—around −13.3 basis points on buying days and −10.7 basis points on selling days. By contrast, the residual category of "other" orders displays slightly positive average day-zero returns.[13]

[13] The equally weighted results, reported in Table IA.3 in the Internet Appendix, tell a similar but less dramatic story. Across all orders, the average day-zero return is about −21.9 basis points on buying days and −24.3 basis points on selling days. Consistent with the volume-weighted results, passive orders perform substantially worse than aggressive orders, and losses become markedly larger on heavy-imbalance days. This pattern suggests that poor execution is not driven by a small number of outliers, although the economic magnitude is larger when more weight is placed on economically relevant days with larger imbalances.

These day-zero return figures are remarkable because they cannot be attributed simply to bid–ask spreads. Panel B of Table 6 reports summary statistics for bid–ask half-spreads of the stocks traded by retail investors. On average, equally weighted half-spreads are around 7.5 to 7.9 basis points on buying and selling days, while the corresponding volume-weighted figures are only about 5.9 and 5.4 basis points. Thus, the day-zero losses documented in Panel A are an order of magnitude larger than the mechanical trading cost associated with half-spreads. The contrast is particularly striking because, as we show below, a large share of retail trading volume is executed through passive orders, which mechanically buy at the bid when providing liquidity to sellers and sell at the ask when providing liquidity to buyers. If anything, this should mitigate execution costs. Instead, retail investors still experience strongly negative same-day returns, suggesting that they provide liquidity precisely when prices are moving against them.

Next, to evaluate the longer-horizon returns to a potential liquidity-provision strategy based on price reversals, we construct long–short portfolios sorted on retail net flows. On each day $t = 0$, we classify stock–days as buying days (net retail share flows > 0) or selling days (net retail share flows < 0). Within each category, we further rank stock–days into terciles based on the absolute magnitude of net flows. The portfolio takes a long position in the most heavily purchased stocks and a short position in the most heavily sold stocks on day $t = 0$, with both legs initiated at the closing price that day. We then compute cumulative excess returns from the close of day t to the close of day $t + h$, excluding day 0, relative to the equal-weighted market return. Panel (a) reports equal-weighted results, while Panel (b) reports volume-weighted results, where each stock–day return is weighted by the absolute value of its net retail flow.

Figure 3 shows that retail investors' day-zero losses are not compensated by subsequent price reversals. Both equal and volume weighted portfolios exhibit persistently negative cumulative returns for the long-short strategy. Rather than reverting, losses deepen slightly over horizons of up to 40 trading days. The 95% confidence intervals rule out any economically meaningful recovery, indicating that retail investors' liquidity provision is not rewarded by future price reversals, but instead represents a transfer to institutional counterparties.

[Figure 3 about here]

To further examine whether retail order imbalances predict subsequent price movements beyond the immediate execution-day effect, we compute future excess returns from the *closing price on day* t to the closing price on day $t+h$, for $h \in \{5, 20, 40\}$. Thus, by construction, these returns *exclude day-zero returns* and capture only post-close performance. We report these future excess returns separately for stock-days with positive retail net flow ("buying") and negative retail net flow ("selling"), and we further split the sample by imbalance intensity (light, moderate, and heavy) and by order type (all, aggressive, passive, and other). To make the buying and selling columns directly comparable, the sign of returns on selling days is already adjusted by multiplying by -1, so that a positive number always indicates that prices move in the same direction as the retail imbalance, whereas a negative number indicates reversal.

[Table 7 about here]

The results point to a *continuation* pattern, especially for retail buying imbalances. Since day-zero returns are already negative for buying stocks, the fact that post-close future returns are also typically negative implies that prices continue to move against retail buyers even after the market closes on day t. This pattern is particularly strong for heavy buying days and becomes more pronounced at longer horizons. For example, in the all-orders specification, heavy buying is followed by about −15 basis points over t+1 to t+5, around −36 basis points over t+1 to t+20, and roughly −66 basis points over t+1 to t+40. Similar patterns appear for aggressive and passive orders. By contrast, once the sign adjustment is applied, selling-side results are generally much weaker and often close to zero. Overall, the evidence suggests that retail investors not only lose on day zero, but also continue to face adverse price movements afterward, particularly when they are net buyers. Therefore, the results do not support the interpretation that retail investors knowingly accept losses on day zero because they anticipate future price reversals.

4.2 Are Intraday Timing Biases Behind Day-Zero Losses?

The negative day-zero returns documented above could stem from suboptimal intraday timing by retail investors. Behavioral biases may induce comovement among retail traders, leading them to concentrate their trades at the same time and push prices away from fundamentals. In this scenario, stocks may become temporarily overpriced when retail investors are buying, or underpriced when they are selling. A plausible mechanism is that retail investors react to common intraday signals that make a stock salient, and therefore more expensive or cheaper, precisely when they choose to trade. Examples include the sudden dissemination of a stock recommendation on social media or chat forums, or the triggering of a widely followed technical indicator such as a moving-average crossover.[14] As the temporary trading pressure subsides, prices partially correct by the end of the day, generating the day-zero losses we observe.

To evaluate whether poor intraday execution explains our findings, we compute day-zero

[14]See for instance Dorn, Huberman, and Sengmueller (2008) and Barber, Lin, and Odean (2023)

returns relative to average daily prices rather than closing prices. This measure captures whether retail investors systematically buy at higher intraday prices (followed by declines) or at lower intraday prices (followed by price increases).

More specifically, Panel C of Table 6 reports the relative execution performance of retail investors compared to institutions trading on the same side of the market. We compute the relative buying return as $(P_t^{AvgBuy} - P_t^{buy})/P_t^{buy})$, where P_t^{buy} is the volume-weighted average price paid by retail investors and P^{AvgBuy} is the corresponding volume-weighted average price paid by institutions on day t. Similarly, the relative selling return is calculated as $(P_t^{sell} - P_t^{AvgSell})/P_t^{sell}$, where P_t^{sell} is the volume-weighted average price received by retail investors and $P_t^{AvgSell}$ is the volume-weighted average price received by institutions. Positive (negative) values therefore indicate that retail investors execute trades at more (less) favorable prices relative to institutions on the same side of the market.

On average, retail investors trade at slightly better prices than institutions on the same side of the market, with a relative buying return of +0.68 bps and a relative selling return of +3.63 bps. Panel B shows the volume-weighted results, which give greater weight to stock-days with larger imbalances. Here, the retail advantage largely disappears. Retail investors underperform institutions on the buy side, with an average relative buying return of −4.13 bps, and earn essentially no advantage on the sell side (−0.12 bps).

Taken together, these findings suggest that retail investors' day-zero underperformance cannot be attributed to intraday execution biases, consistent with the conclusion by Linnainmaa (2010). In fact, for small imbalances retail investors often achieve marginally better execution prices than institutions, indicating that they are not systematically disadvantaged in the timing of their trades within the day. The evidence instead points to a different explanation: retail investors consistently end up on the wrong side of the market on the particular days when they trade.

4.3 Discussion of Behavioral Mechanisms

We do not provide direct evidence of behavioral bias in this section; the following discussion evaluates candidate mechanisms consistent with our findings.

A natural question raised by these findings is why retail investors place passive orders to absorb institutional order flow rather than withdrawing when adverse selection risk is elevated. Eyster, Rabin, and Vayanos (2019) develop a model of "cursed" traders who fail to fully infer the informational content of prices. In their framework, a cursed trader does not condition demand on the price's informational signal, and therefore does not appreciate that a favorable execution may itself be bad news—the market microstructure analog of the winner's curse. In a limit-order setting, this failure is particularly consequential: a passive order executes precisely when a counterparty has found it profitable to trade against it, meaning execution is an adverse signal about fundamentals. A cursed retail investor, however, does not internalize this inference and continues to supply liquidity even when the probability of being picked off by informed order flow is highest.

While Eyster, Rabin, and Vayanos (2019) establish the theoretical possibility of cursed trading, Liu, Peng, Xiong, and Xiong (2022) provide direct empirical support[15]: in a nationwide survey of Chinese retail investors matched to transaction records, they find that "perceived information advantage"—the belief that one holds a superior signal relative to one's counterpart—is one of the strongest predictors of excessive trading. Crucially, investors who report a perceived information advantage do not outperform, confirming that the belief is a misperception rather than a genuine signal. For our purposes, this bias has a direct

[15]See also Chague, De-Losso, and Giovannetti (2018) for empirical evidence consistent with information neglect also using Brazilian retail investors data.

implication for passive order placement: a retail investor who believes she is better informed than her counterpart has little reason to retreat from the limit-order book when prices move against her. Instead, she may interpret an adverse price move as a temporary mispricing that makes her standing limit orders even more attractive. Birru, Chague, De-Losso, and Giovannetti (2024) further show, using Brazilian data, that attentional failures compound informational misperceptions: retail investors who are inattentive to salient cost signals trade more aggressively, suggesting that the passive-order channel is reinforced when investors are not paying close attention to the informational environment in which they operate.

These biases are not mutually exclusive, and other behavioral mechanisms may also contribute: Eyster, Rabin, and Vayanos (2019) formally show that cursedness and overconfidence are complements, and a retail investor who both overestimates her information and neglects the informational content of execution will supply liquidity far more aggressively than either bias alone would predict. More broadly, the passive-order pattern documented here is likely the product of multiple overlapping biases—and their interactions—that we have not fully mapped. Whether these mechanisms operate additively or interactively is an empirical question that lies beyond the scope of this paper; but the convergence of behavioral theory with our microstructure evidence suggests that retail investors' role as accidental market makers is not a statistical artifact.

A potential concern with this interpretation is that our data capture executed passive orders rather than the broader population of submitted limit orders. If retail investors systematically withdraw or reprice their limit orders when adverse selection risk rises, our evidence of passive liquidity provision could reflect a compositional artifact rather than a failure of inference. The available evidence, however, suggests this concern is limited — and in fact reinforces the cursedness interpretation. Linnainmaa (2010) shows, using Finnish trading records, that unexecuted limit orders earn positive short-run returns, implying that the submission decision itself is not systematically distorted; it is precisely the conditioning on execution that generates the adverse outcome, consistent with the winner's curse mechanism. Kelley and Tetlock (2013) similarly find, using data covering roughly one-third of all self-directed retail trading in the United States, that submitted limit orders carry a positive end-of-day price impact—a standard measure of informed trading—whereas executed limit orders exhibit the contrarian, liquidity-provision pattern. Together, these results suggest that retail investors are not naive at the order-submission stage: the distortion operates through a failure to internalize that execution itself is an adverse signal, which is precisely what the cursedness framework predicts.

Finally, one might interpret the passive order evidence as reflecting pre-committed price-target trading rather than informational neglect: an investor who has decided to buy if the stock falls to a given price will mechanically execute at that level with no active inference about adverse selection. We find this explanation incomplete on both conceptual and empirical grounds. Conceptually, price-target trading is itself a form of information neglect: a rational investor who anticipates that her limit order will execute precisely when adverse-selection risk is elevated should adjust her limit price or cancel the order entirely. The failure to do so is exactly the mechanism Eyster, Rabin, and Vayanos (2019) formalize — the cursedness label applies equally whether neglect occurs at order submission or at order entry into the book. Even setting this aside, price-target trading cannot explain why retail passive volume rises on earnings-announcement days, material-fact-release days, and abnormal-spread days, which carry no implication for where a pre-committed price target would be hit. Nor can it explain the asymmetry we document: passive retail buying concentrated on negative-return days and passive retail selling on positive-return days is the signature of liquidity provision against informed institutional flow, not of price-target execution.

5 How Relevant is Retail Investor's Accidental Market Making?

Having established that retail investors act as accidental market makers — stepping in precisely when professional liquidity provision withdraws — we now turn to the economic consequences of this behavior. We address two distinct questions. First, how large are the losses that retail investors incur by absorbing institutional order flow? We quantify these losses in dollar terms and scale them by institutional trading volume to assess their magnitude relative to overall market activity. Second, is retail liquidity provision large enough to meaningfully affect market quality? To answer this, we examine how several measures of market quality drawn from the microstructure literature respond to retail trading flows, employing an instrumental variables strategy based on regional flows into Brazil's government bond platform to address reverse causality. We close the section with an assessment of external validity, comparing our findings with the broader literature on retail investor performance across different markets.

5.1 Day-Zero Wealth Transfers

We define dollar profits as the day-zero dollar return on buying stock–days multiplied by the number of shares purchased, plus the day-zero dollar return on selling stock–days multiplied by the number of shares sold. To express these profits in relative terms, we scale them by institutional trading activity and report them as a fraction of institutional volume (in basis points). Specifically, institutional volume is measured as the average of total institutional buying and selling (dollar) volume on that day. This scaling yields a measure of daily retail profits per stock–day relative to institutional trading volume, expressed in basis points.

The estimates indicate that retail investors lose, on average, USD 3.9 million per stock–day to their institutional counterparts. When scaled by total institutional trading activity, this corresponds to approximately 13 bps of institutional volume. Aggregated over the 252 trading days in a typical year, this implies total retail losses of roughly USD 1.0 billion annually.[16] These estimates suggest that retail liquidity provision represents a significant annual transfer from retail to institutional investors.

These magnitudes help explain the commercial viability of recent payment-for-order-flow (PFOF) arrangements, in which brokers route retail orders to wholesalers in exchange for compensation: the day-zero losses borne by retail investors are, from the perspective of the institutional counterparty, the economic rents that make retail order flow worth paying for (see, for instance, Battalio, Corwin, and Jennings, 2016, Adams, Kasten, and Kelley, 2024, and Baldauf, Mollner, and Yueshen, 2024).

5.2 The Effects on Market Quality

Is retail liquidity provision meaningful enough to impact overall market quality? To address this question, we compute four measures of market quality used in the literature and relate them to retail trading activity. Our first measure is price impact, defined as the stock–day average (across ten-minute intervals) of the absolute return in basis points divided by traded volume in USD millions. This measure captures price impact per unit of volume, in the

[16]Projecting these magnitudes to the U.S. market, if the average daily trading volume for NYSE-listed stocks is approximately USD 80 billion and institutional investors account for roughly 85% of this activity, 13 bps implies aggregate retail losses of roughly USD 90 million per day, or over USD 20 billion annually.

spirit of Amihud (2002), with lower values indicating more liquid trading and smaller price responses to order-flow shocks.

Our second measure is the absolute value of the first-order autocorrelation of ten-minute returns, computed at the stock-day level, which captures short-horizon return predictability arising from microstructure frictions. If stock prices quickly incorporate information, returns should be serially uncorrelated; thus, higher absolute autocorrelation indicates greater noise and slower information incorporation.

Our third measure is the average bid–ask spread at the stock-day level. Because we do not have access to the full order book, we extracted the average daily bid–ask spread from the Bloomberg terminal. A narrower bid–ask spread implies that transaction prices are closer to the midpoint and thus more likely to reflect fundamental values.

Finally, our fourth measure is based on the volatility ratio, realized variance over a one-hour window divided by realized variance over a ten-minute window, with both scaled to the same horizon. Under a random walk, this ratio equals one. We use the absolute deviation from one, so that lower values indicate volatility scaling more linearly across horizons, consistent with lower noise.

Our benchmark results are based on an market quality index that combines all four measures. Specifically, we compute the first principal component of the four daily measures and orient it so that higher values correspond to better market quality, that is, lower price impact, tighter spreads, lower absolute autocorrelation, and a volatility ratio closer to the random walk benchmark. The first principal component explains slightly more than one-third of the total variance of the four efficiency measures. For ease of interpretation, we standardize this component to have mean zero and unit variance.

Panel A of Table 8 reports descriptive statistics for the market quality index and each of its four components. On average, price impact is approximately 2.52 basis points per million of trading volume, the absolute return autocorrelation (scaled by 100) is 15.07, the bid–ask spread is 15.40 basis points, and the volatility ratio is 0.85. Panel B reports pairwise correlations among the components and the index. All individual components are positively correlated with one another, though not perfectly so. The index loads most heavily on price impact (0.80) and the bid–ask spread (0.81), and less so on the absolute autocorrelation (0.34) and the volatility ratio (0.23), indicating that it primarily captures liquidity conditions rather than informational efficiency. Results for the individual components are reported in the Internet Appendix.

[Table 8 about here]

To examine the contribution of retail trading flows to market quality, we estimate the following panel regressions:

$$MarketQuality_{s,t} = \beta \log(Retail)_{s,t} + \delta\, AdverseDays_{s,t} + \gamma' X_{s,t-1} + \lambda_s + \lambda_t + \varepsilon_{s,t}, \qquad (1)$$

$$\begin{aligned} MarketQuality_{s,t} = {} & \beta^P \log(Passive)_{s,t} + \beta^A \log(Aggressive)_{s,t} + \beta^O \log(Other)_{s,t} \\ & + \delta\, AdverseDays_{s,t} + \gamma' X_{s,t-1} + \lambda_s + \lambda_t + \varepsilon_{s,t}, \qquad (2) \end{aligned}$$

where $MarketQuality_{s,t}$ is the market quality index for stock s on day t. In specification 1, $\log(Retail)_{s,t}$ is the log of total retail trading volume in USD millions across all order types. Specification 2 decomposes this volume by order type: $\log(Passive)_{s,t}$ is the log of total retail passive volume, $\log(Aggressive)_{s,t}$ is the log of total retail aggressive volume, and $\log(Other)_{s,t}$ is the log of retail volume executed in auctions, all measured in USD millions. In both specifications, $AdverseDays_{s,t}$ controls for the adverse-selection environment, $X_{s,t-1}$

is a vector of lagged stock-level controls, and λ_s and λ_t denote stock and date fixed effects. Standard errors are two-way clustered by stock and date.

The vector $X_{s,t-1}$ includes the log of total market trading volume on the previous day, $\log(Volume)_{s,t-1}$; short-run momentum and volatility measured over the previous trading week, $Ret_{s,t-1:t-5}$ and $RetSd_{s,t-1:t-5}$; and intermediate-horizon momentum and volatility measured over the prior three weeks, $Ret_{s,t-6:t-21}$ and $RetSd_{s,t-6:t-21}$. We include stock fixed effects, λ_s, and time fixed effects, λ_t. The coefficients of interest, β captures the impact of a 1% increase in retail volume on price efficiency (in standard deviations). In extension, the specification 2 decomposes this total elasticity into semi-elasticities components β^P, β^A, β^O. We also control for $AdverseDays_{s,t}$, a dummy variable indicating heightened adverse-selection risk constructed as explained in section 3.2.

Table 9 presents the results. Considering aggregate retail trading volume with all controls column (2), we find that a one percentage point increase in retail volume is associated with an increase of 0.145 standard deviations in market quality. When we decompose total retail trading volume into its three components, we find that passive retail volume is responsible for the improvement in market quality. In column (4), a one percentage point increase in passive retail volume is associated with an increase of 0.19 standard deviations in the market quality index. Aggressive and other flows, controlling for passive flows, have no positive, nor negative, effect on market quality.

[Table 9 about here]

5.2.1 Instrumental Variables Approach

Endogeneity is a concern in specification 1. On the one hand, reverse causality could bias estimates upward: retail investors may choose to place passive orders on stock-days when market quality is already better, anticipating that prices more closely reflect fundamentals. We consider this channel unlikely, as it would require retail investors to forecast stock-day market quality — a degree of sophistication inconsistent with the evidence on retail trading performance documented above. On the other hand, the mechanical relationship between retail participation and adverse-selection risk documented in Table 3 biases estimates in the opposite direction: retail investors are most active precisely on high-adverse-selection days, when market quality is poor, causing OLS to understate the true causal effect. The net direction of the bias is therefore ambiguous a priori, and we address both concerns jointly through an instrumental-variables strategy.

The instrument we employ is designed to isolate variation in retail stock trading flows that is plausibly orthogonal to the market conditions of a particular stock-day. Specifically, we implement a shift-share instrument based on *regional* retail investment flows into a different market that retail investors also frequently access through their brokerage platforms: government bonds. Government bonds constitute an important asset class for retail investors in Brazil, where the Tesouro Direto (TD) program makes direct bond investment straightforward.[17] Importantly, while aggregate bond flows may correlate with equity market conditions, our instrument exploits variation in flows *across Brazilian regions*, removing any underlying macro trends that could otherwise induce such correlations.

To compute our shift-share instrument, we proceed as follows. First, we determine the *share* of each stock in the portfolio of each Brazilian city. For each city c and preceding year $y-1$, we compute the total retail trading volume (buys plus sells) in stock s and divide it by total retail volume in city c in that same year, obtaining the portfolio weight

[17] The TD program (Treasury Direct in English) was launched in 2002 by the Brazilian Treasury in partnership with B3 to facilitate retail access to government bonds. The only requirement is a brokerage account. A comparable platform exists in the United States (TreasuryDirect.gov), though an important difference is that the Brazilian platform allows buying and selling on every trading day.

$\omega_{s,c,y-1}$.[18] Second, we compute the daily aggregate *investment shifts* (IS) to the TD platform for each Brazilian city, $IS_{c,t}$, by summing the dollar volume of all purchases and all sales executed on the platform by investors in city c on day t. The idea is that aggregate flows to TD capture moments when overall investment decisions are being made (e.g., because funds became available when wages or bonuses were paid, an advertisement on TV made investing a salient topic, or for any idiosyncratic reason) and therefore should also correlate with stock investment decisions. Finally, we aggregate these stock-city-day observations to the stock-day level to construct our shift-share instrument,

$$Z_{s,t}^{Bonds} = \sum_{c} \omega_{s,c,y-1} \times IS_{c,t}^{Bonds}$$

for stock s on day t. By construction, variation in $Z_{s,t}^{Bonds}$ does not arise from investors' decisions to invest in a particular stock on a particular day, since it uses lagged portfolio weights. Because we are dealing with dollar volumes, we employ $\log(Z_{s,t}^{Bonds})$ as our instrument.

For the instrument be valid, two conditions must hold: (i) *relevance* and (ii) *exclusion restriction*. Since the first can be empirically tested. We discuss the second first. Our instrument is framed in a shift-share panel structure, where the shares, $\omega_{s,c,y-1}$, are predetermined city-stock portfolio weights from the previous year and the shifters, $IS_{c,t}^{Bonds}$, are city-level TD program aggregate funding flows. As emphasized by Goldsmith-Pinkham, Sorkin, and Swift (2020) and Borusyak, Hull, and Jaravel (2025), the exclusion restriction requires that, conditional on controls and fixed effects, the investment shifters $IS_{c,t}^{Bonds}$ are as-good-as-random with respect to the stock-day error term $\varepsilon_{s,t}$, and that the shares $\omega_{s,c,y-1}$ are predetermined and not chosen in anticipation of market quality shocks.

[18]Because we only observe trading flows and not investor portfolios, we infer each city's portfolio weights from relative buying and selling flows.

The construction of the instrument makes the exclusion restriction plausible. On any given day, the cross-city distribution of TD flows is dispersed: some cities experience positive funding shocks while others experience negative ones, and these deviations are aggregated to the stock level using portfolio weights that are predetermined from the prior year. The resulting variation in $Z_{s,t}^{Bonds}$ therefore reflects an idiosyncratic reallocation of funds across cities rather than a directed response to conditions in any particular stock. Beyond this, the two confounders most naturally associated with bond flows are absorbed by the fixed-effect structure. Aggregate macroeconomic shocks: monetary policy, risk appetite, or the aggregate business cycle affect TD flows and equity market quality simultaneously, therefore being absorbed by date fixed effects λ_t. Persistent cross-sectional differences in the relative attractiveness of particular stocks to bond investors are similarly absorbed by stock fixed effects λ_s.

Relevance can be assessed from the first-stage regression in the IV system below (Equations 3:

$$\log(Retail_{s,t}) = \pi \log(Z_{s,t}^{Bonds}) + \theta' X_{s,t-1} + \lambda_s + \lambda_t + u_{s,t}, \tag{3}$$

$$MarketQuality_{s,t} = \beta \log(\widehat{Retail}_{s,t}) + \gamma' X_{s,t-1} + \lambda_s + \lambda_t + \varepsilon_{s,t}, \tag{4}$$

where $MarketQuality_{s,t}$ is our market quality index, $\log(Retail_{s,t})$ is the log of retail stock trading volume (including all type of orders) for stock s on day t, $X_{s,t-1}$ is the vector of lagged controls (including *AdverseRisk* and the controls used in Equation 1 and λ_s and λ_t are stock and date fixed effects.

To identify which order type drives the market-quality improvement, we decompose total retail volume into passive (non-marketable limit), aggressive(marketable), and auction-executed orders. For each channel $k \in \{Pass, Agg, Oth\}$, we instrument $\log(Retail_{s,t}^{k})$ with the shift-share bond platform instrument $\log(Z_{s,t}^{Bonds})$, while the remaining two channels enter as OLS controls:

$$\log(Retail_{s,t}^{k}) = \pi^{k}\log(Z_{s,t}^{Bonds}) + \sum_{j\neq k}\delta^{j}\log(Retail_{s,t}^{j}) + \theta' X_{s,t-1} + \lambda_s + \lambda_t + u_{s,t}^{k}, \quad (5)$$

$$MarketQuality_{s,t} = \beta^{k}\widehat{\log(Retail_{s,t}^{k})} + \sum_{j\neq k}\delta^{j}\log(Retail_{s,t}^{j}) + \gamma' X_{s,t-1} + \lambda_s + \lambda_t + \varepsilon_{s,t}^{k}, \quad (6)$$

where $\log(Retail_{s,t}^{Pass})$, $\log(Retail_{s,t}^{Agg})$, and $\log(Retail_{s,t}^{Oth})$ are the log retail volumes (USD millions) for passive, aggressive, and auction-executed orders in stock s on day t. Controls $X_{s,t-1}$ and fixed effects λ_s, λ_t are as in equations (3)–(4); standard errors are two-way clustered by stock and date.

The coefficient β^{k} captures the effect of a one-percent exogenous increase in channel-k volume on the market quality, holding the other order types fixed. Conditioning on the remaining channels means the instrument isolates a compositional shift toward channel k rather than a scale increase in total retail activity. First-stage validity requires $\log(Z_{s,t}^{Bonds})$ to retain sufficient power after part out the other channels. Thus, the instrument power test in each of the channels provides an evidence of what is the main channel by the instrument affect retail total volume.

Table 10 reports the IV estimates without and with time-varying controls in columns (1) and (2), respectively. A one-percentage-point increase in retail volume raises the market quality by 0.86 standard deviations in the baseline specification and by 0.97 standard deviations once controls are included. Both estimates substantially exceed OLS coefficient of 0.15, consistent with a downward bias: retail participation tends to rise precisely when adverse selection is most severe, so OLS understates the true effect.[19]

We next examine which order type drives the market-quality improvement. Column (3) presents the second-stage estimates of Equation 6 for passive orders, yielding an effect of 2.68 standard deviations, well above the aggregate IV estimate of 0.97. This attenuation is expected: the aggregate retail volume pools passive orders with other order types that do not improve market quality, as shown in columns (3) and (4) of Table 9 and columns (4) and (5) of Table 10, diluting the overall estimate. In the internet appendix, Table IA.6 reports the components of the market quality index as outcome variables, estimated both using total retail volume as in equation 4 and using order type as in the second-stage specification 6. Consistent with the main channel operating through passive orders, the results show that price impact and bid-ask spread are the two components most affected. Specifically, price impact decreases by 16 basis points per million and bid–ask spread narrows by 20 basis points. Moreover, restricting the analysis to passive orders yields larger estimated effects. In the internet appendix, Table IA.6 reports the components.

[Table 10 about here]

5.2.2 Robustness

One plausible concern is that using raw TD flows as shifters might be correlated with common macroeconomic shocks that are shared across all cities, (e.g., monetary policy, risk appetite). These variables simultaneously might affect bond trading by retail investors and stock-market efficiency. While we consider this unlikely, since we exploit the variation in the TD flows

[19] Table IA.5 in the Appendix reports first-stage estimates under stepwise additions of time-varying controls; Figure IA.3 plots the first-stage relationship between the instrument and the endogenous regressor. Across all specifications, the first-stage F-statistic and the corresponding Wald statistic comfortably exceed conventional weak-instrument thresholds (Stock and Yogo, 2002; Lee, McCrary, Moreira, and Jack, 2022), so instrument relevance is not a concern.

across location and the shift-share design uses lagged city-stock portfolio weights to remove contemporaneous correlation with these shocks, nonetheless we purge the common shocks using two ways to be conservative.

First, we residualize TD aggregate funding flows by regressing $IS^{Bonds}_{c,t}$ on city and time fixed effects, thereby purging level differences and common shocks across cities and focusing on deviations from each city's own trajectory. Specifically, we estimate:

$$IS^{Bonds}_{c,t} = \lambda_c + \lambda_t + \nu_{c,t}, \tag{7}$$

where $\nu_{c,t}$ is the local funding shock net of common time shocks and permanent city factors.

Our second approach to remove common macroeconomic shocks from our TD flow instrument is to estimate a latent factor model using principal components. Galaasen, Jamilov, Juelsrud, and Rey (2023) employ a similar approach to construct idiosyncratic shocks for IV purposes. Specifically, we estimate a latent-factor specification with city fixed effects:

$$IS^{Bonds}_{c,t} = \lambda^F_c F_t + \lambda_c + u_{c,t}, \tag{8}$$

where F_t is a common factor extracted as the first principal component of TD flows across cities (which explains approximately 80% of the variance) and λ^F_c is the city-specific loading. The residual $u_{c,t}$ captures city-level innovations after removing both time-invariant heterogeneity and factor-driven aggregate dynamics. We use $u_{c,t}$ as a second measure of the local funding shock net of common time shocks and permanent city factors.

Figure IA.2 in the Appendix plots the idiosyncratic city-level funding shocks used to construct the instrument. We obtain these shocks by purging the raw TD series $IS^{Bonds}_{c,t}$ of common aggregate movements and persistent city-level differences. This residualization step is important for identification: aggregate TD activity may co-move with economy-wide financial conditions (e.g., monetary policy, risk appetite) that can directly affect stock-market efficiency, while time-invariant city heterogeneity may reflect structural differences in wealth, financial access, or investor composition. By extracting residuals, we isolate variation that is more plausibly orthogonal to stock-specific microstructure shocks and that provides cleaner cross-city shifters for the shift-share design.

[Table 11 about here]

Columns (3) and (4) of Table 11 report the corresponding 2SLS estimates. The coefficients are very close to the baseline 2SLS estimate (0.972): the two-way fixed-effects residual shock yields 0.907, and the latent-factor residual shock yields 0.885. This similarity indicates that our instrument is not driven by common macro shocks but by local funding shocks that are plausibly orthogonal to stock-level changes in market quality.

Another potential concern is that our identification strategy exploits local variation in Tesouro Direto (TD) flows that could be correlated with the local diffusion of information. A large literature documents *local bias*: individual investors disproportionately hold and trade stocks of firms headquartered nearby (e.g., Coval and Moskowitz, 1999, 2001; Huberman, 2001; Seasholes and Zhu, 2010; Doskeland and Hvide, 2011; see also Chague, Giovannetti, and Paiva, 2026 for evidence from Brazil). If investors who live closer to a firm's headquarters both (i) allocate more to that firm's stock and (ii) possess superior local information, then regional funding shocks may covary with firm-level information arrival. This concern can persist even when we construct weights using lagged regional portfolios.

To address this, we re-estimate the IV specification after excluding from the construction of $Z_{s,t}^{Bonds}$: (i) the three largest cities that host the headquarters of the most traded firms in our sample: Rio de Janeiro (Petrobras), São Paulo (Itau), and Belo Horizonte (Vale), and (ii) all state capital cities. Column (6) of Table 11 reports the results excluding the three main cities, and column (7) reports the results excluding all state capitals. The 2SLS coefficients remain very similar to the baseline estimate: 0.823 and 0.809, respectively, compared to 0.972 in the baseline. These robustness checks suggest that our results are not driven by local information advantages of investors located near firm headquarters.

5.3 External validity

Our analysis relies on Brazilian data, raising the question of whether these findings generalize to other markets. The U.S. literature on retail traders' short-term performance has reached mixed conclusions, which may partly reflect differences in data coverage across studies. Our findings align most closely with the small set of papers that, like ours, observe (nearly) complete retail trading activity within a single market.[20]

We examine two empirical findings for external validity: (i) the magnitude of *day-zero* returns (execution-to-close) and (ii) the extent to which prices subsequently move in a way that benefits—or harms—retail investors. Table 12 summarizes the main findings across the literature.

[Table 12 about here]

Most studies document negative day-zero returns for retail investors, although magnitudes vary considerably across samples and countries. The evidence is strongest in markets with centralized trading and no payment for order flow—Finland, France, Australia, and

[20]Findings from Taiwan, while based on comprehensive data, are less directly comparable because of dominance of retail investors. According to Barber, Lee, Liu, and Odean (2009), retail investors accounted for 89.5% of all trading volume from 1995 to 1999. This contrasts sharply with Brazil and the United States, where institutional investors dominate trading activity.

Brazil—where researchers can observe the (near-)universe of retail trading activity. In these settings, estimated day-zero losses range from −25 bps to −90 bps, depending on the sample and methodology.

As an external benchmark for the United States in a setting with complete retail data, Table IA.4 in the Appendix reports day-zero returns using the LDB sample from 1991–1996 from Barber and Odean (2000).[21] Retail investors exhibit sizable execution-to-close losses on both sides of the market. In the equally weighted aggregation (Panel A), the average day-zero return is −22.6 bps on net-buying days and −53.6 bps on net-selling days, with losses becoming more severe as trading intensity increases (e.g., heavy buying: −30.2 bps; heavy selling: −60.6 bps). In the volume-weighted aggregation (Panel B), which places more weight on stock-days where retail trades are larger, losses are even larger: −34.5 bps for net buying and −61.0 bps for net selling, again steepening with intensity (heavy buying: −37.2 bps; heavy selling: −62.9 bps).

The evidence on returns beyond day zero is more mixed. Some studies find that retail order imbalances positively predict short-horizon returns, consistent with a liquidity-provision interpretation, but disagree on whether subsequent price dynamics are sufficient to offset contemporaneous losses. Kaniel, Saar, and Titman (2008) and Barrot, Kaniel, and Sraer (2016) document negative day-zero (week-zero for the former paper) performance alongside subsequent short-run reversals, yet the potential gains are not large enough to fully compensate for day-zero losses. Linnainmaa (2010) also finds negative day-zero returns and substantial short-run reversal, but shows that longer-horizon outcomes are unfavorable, with losses dominating over 126 days. Kelley and Tetlock (2013) likewise find negative day-zero returns but find subsequent short-run reversals that could potentially offset day-zero losses.

[21] We thank Terry Odean for providing us this dataset.

Boehmer, Jones, Zhang, and Zhang (2021) find that week-zero returns are positive (they do not report day-zero returns) and that prices continue to move in the direction of retail flows. Overall, however, even when prices move in a manner consistent with retail liquidity provision, these papers do not find evidence that retail investors systematically exploit these dynamics by timing their exits in the manner of a successful liquidity supplier or market maker.

In contrast, Fong, Gallagher, and Lee (2014) find that day-zero losses worsen over the subsequent few days, although the effect dissipates at longer horizons. Also, Barber, Lin, and Odean (2023) find little evidence that retail order imbalances predict short-run returns. Importantly, Barber, Lin, and Odean (2023) emphasize that apparent return predictability can disappear once positions are appropriately dollar-weighted: equal-weighted tests may overstate the economic significance of predictability because they place disproportionate weight on small trades and small stocks. Consistent with this, we do find that retail investors' poor performance is concentrated on stock-days with large dollar imbalances.

6 Conclusion

This paper documents that retail investors act as accidental market makers in the Brazilian equity market. On days of elevated adverse-selection risk — when professional market makers withdraw — retail investors systematically expand their passive limit-order volume, absorbing institutional demand for liquidity precisely when it is most costly to supply. This substitution is not incidental: it is robust across five independent proxies for adverse-selection risk, survives an alternative ex ante classification of institutional investors based on historical trading frequency, and is corroborated by event-study evidence around earnings announcements and material fact disclosures.

Together, these findings carry implications for the regulation of retail participation in equity markets. Policies that restrict retail order flow — through transaction taxes, minimum tick sizes that discourage limit orders, or the channeling of retail orders away from the lit exchange — may reduce market quality precisely on the days when it is most fragile, because they remove a source of liquidity that professional market makers are themselves unwilling to supply under adverse-selection risk. The evidence is consistent with retail investors failing to internalize the winner's curse embedded in passive execution, treating limit-order execution as benign rather than as a signal that they are on the wrong side of an informed trade. At the same time, the large and persistent losses that retail investors incur through passive execution suggest that investor protection policies should pay particular attention to the informational disadvantage embedded in limit-order trading, which is less visible than the commissions and spreads that have historically attracted regulatory attention.

References

Adams, Samuel W., Connor Kasten, and Eric K. Kelley, 2024, How free is free? Retail trading costs with zero commissions, *Journal of Banking & Finance* 165, 107226.

Aït-Sahalia, Yacine, and Celso Brunetti, 2020, High frequency traders and the price process, *Journal of Econometrics* 217, 20–45.

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Baldauf, Markus, Joshua Mollner, and Bart Zhou Yueshen, 2024, Siphoned apart: A portfolio perspective on order flow segmentation, *Journal of Financial Economics* 154, 103807.

Barardehi, Yashar H., Dan Bernhardt, Zhi Da, and Mitch Warachka, 2021, Internalized Retail Order Imbalances and Institutional Liquidity Demand.

Barber, Brad M., Xing Huang, Philippe Jorion, Terrance Odean, and Christopher Schwarz, 2024, A (Sub)penny for Your Thoughts: Tracking Retail Investor Activity in TAQ, *Journal of Finance* 79, 2403–2427.

Barber, Brad M., Xing Huang, Terrance Odean, and Christopher Schwarz, 2022, Attention-induced trading and returns: Evidence from Robinhood users, *Journal of Finance* 77.

Barber, Brad M., Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean, 2009, Just How Much Do Individual Investors Lose by Trading?, *Review of Financial Studies* 22, 609–632.

Barber, Brad M., Shengle Lin, and Terrance Odean, 2023, Resolving a Paradox: Retail trades positively predict returns but Are Not Profitable, *Journal of Financial and Quantitative Analysis* 1–35.

Barber, Brad M., and Terrance Odean, 2000, Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors, *Journal of Finance* 55, 773–806.

Barrot, Jean-Noel, Ron Kaniel, and David Sraer, 2016, Are retail traders compensated for providing liquidity?, *Journal of Financial Economics* 120, 146–168.

Battalio, Robert, Shane A. Corwin, and Robert Jennings, 2016, Can Brokers Have It All? On the Relation between Make-Take Fees and Limit Order Execution Quality, *Journal of Finance* 71.

Birru, Justin, Fernando Chague, Rodrigo De-Losso, and Bruno Giovannetti, 2024, Attention and Biases: Evidence from Tax-Inattentive Investors, *Management Science* 70, 7101–7119.

Black, Fischer, 1986, Noise, *The Journal of Finance* 41, 528–543.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang, and Xintong Zhang, 2021, Tracking retail investor activity, *The Journal of Finance* 76, 2249–2305.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, 2025, A Practical Guide to Shift-Share Instruments, *Journal of Economic Perspectives* 39, 181–204.

Chague, Fernando, Rodrigo De-Losso, and Bruno Giovannetti, 2018, Individuals neglect the informational role of prices: Evidence from the stock market.

Chague, Fernando, Bruno Giovannetti, and Guilherme Paiva, 2026, Familiarity breeds day trade, *Journal of Banking & Finance* 185.

Chague, Fernando, and Lucas Kimura, 2026, Trading frequencies, *Working Paper* .

Coval, Joshua D., and Tobias J. Moskowitz, 1999, Home Bias at Home: Local Equity Preference in Domestic Portfolios, *Journal of Finance* 54, 2045–2073.

Coval, Joshua D, and Tobias J Moskowitz, 2001, The Geography of Investment: Informed Trading and Asset Prices, *Journal of Political Economy* 109.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1991, The survival of noise traders in financial markets, *Journal of Business* 64, 1–19.

Dorn, Daniel, Gur Huberman, and Paul Sengmueller, 2008, Correlated Trading and Returns, *Journal of Finance* 63, 885–920.

Doskeland, Trond M., and Hans K. Hvide, 2011, Do Individual Investors Have Asymmetric Information Based on Work Experience?, *Journal of Finance* 66, 1011–1041.

Easley, David, and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69–90.

Eaton, Gregory W., T. Clifton Green, Brian S. Roseman, and Yanbin Wu, 2022, Retail Trader Sophistication and Stock Market Quality: Evidence from Brokerage Outages, *Journal of Financial Economics* 146, 502–528.

Eyster, Erik, Matthew Rabin, and Dimitri Vayanos, 2019, Financial Markets Where Traders Neglect the Informational Content of Prices, *Journal of Finance* 74, 371–399.

Fong, Kingsley Y. L., David R. Gallagher, and Adrian D. Lee, 2014, Individual Investors and Broker Types, *Journal of Financial and Quantitative Analysis* 49, 431–451.

Foucault, Thierry, David Sraer, and David J. Thesmar, 2011, Individual investors and volatility, *Journal of Finance* 66, 1369–1407.

Friedman, Henry L., and Qinlin Zeng, 2024, Retail investor trading and market reactions to earnings announcements.

Galaasen, Sigurd, Rustam Jamilov, Ragnar Juelsrud, and Helene Rey, 2023, granular credit risk, *Nber Working Paper* .

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, 2020, Bartik Instruments: What, When, Why, and How, *American Economic Review* 110, 2586–2624.

Grammig, Joachim, and Erik Theissen, 2005, Is Best Really Better? Internalization of Orders in an Electronic Stock Exchange.

Greene, Jason T., and Scott B. Smart, 1999, Liquidity provision and noise trading: Evidence from the "Investment Dartboard" column, *Journal of Finance* 54, 1885–1899.

Hansch, Oliver, Narayan Y. Naik, and S. Viswanathan, 1999, Preferencing, Internalization, Best Execution, and Dealer Profits, *Journal of Finance* 54, 1799–1828.

Huberman, Gur, 2001, Familiarity Breeds Investment, *Review of Financial Studies* 14.

Jones, Charles M., Adam V. Reed, and William Waller, 2025, When Price Discovery and Market Quality Are Most Needed.

Kaniel, Ron, Shuming Liu, Gideon Saar, and Sheridan Titman, 2012, Individual Investor Trading and Return Patterns around Earnings Announcements, *Journal of Finance* 67, 639–678.

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2008, Individual investor trading and stock returns, *Journal of Finance* 63, 273–310.

Kelley, Eric K., and Paul C. Tetlock, 2013, How wise are crowds? insights from retail orders and stock returns, *Journal of Finance* 68, 1229–1265.

Larrymore, Noreen L., and Austin J. Murphy, 2009, Internalization and market quality, *Journal of Financial Research* 32, 337–363.

Lee, David S, Justin McCrary, Marcelo J Moreira, and Porter Jack, 2022, Valid T-ratio inference for IV, *American Economic Review* 112, 3260–3290.

Linnainmaa, Juhani T., 2010, Do limit orders alter inferences about investor performance and behavior?, *Journal of Finance* 65, 1473–1506.

Liu, Hongqi, Cameron Peng, Wei A. Xiong, and Wei Xiong, 2022, Taming the bias zoo, *Journal of Financial Economics* 143, 716–741.

Nagel, Stefan, 2012, Evaporating Liquidity, *Review of Financial Studies* 25, 2005–2039.

Peress, Joel, and Daniel Schmidt, 2020, Glued to the TV: Distracted noise traders and stock market liquidity, *Journal of Finance* 75, 1083–1134.

Seasholes, Mark S., and Ning Zhu, 2010, Individual Investors and Local Bias, *Journal of Finance* 65, 1987–2010.

Stock, James H., and Motohiro Yogo, 2002, Testing for Weak Instruments in Linear IV Regression, *NBER Technical Working Paper* 284.

Figures

Figure 1: Retail and institutional trading imbalances around earnings announcements

This figure plots average net flows (difference between buying and selling share volumes) by investor group from day −5 to day +5 relative to earnings announcements and material fact disclosures. *Retail investors* represents net flows from retail investors. *Market-making institutions* are institutional investors that both bought and sold shares of the same stock on the disclosure day, indicating liquidity provision rather than directional trading. *Directional institutions* include all other institutional investors with unidirectional trading patterns. This figure presents event-study estimates of average net trading imbalances around earnings announcements, separately for positive, neutral, and negative events. Events are classified according to the stock's abnormal return on the announcement day relative to the market, using a two-standard-deviation threshold based on the stock's rolling 252-day volatility. For each stock–day in the event window from day −5 to day +5, we compute each investor group's net trading flow as buy volume minus sell volume, normalized by total daily share volume. The top row reports total net trading imbalances for retail investors, market-making institutions, and directional institutions. The bottom row decomposes retail imbalances by execution mechanism into passive, aggressive, and auction-related flows. The plotted series show averages across events by event time.

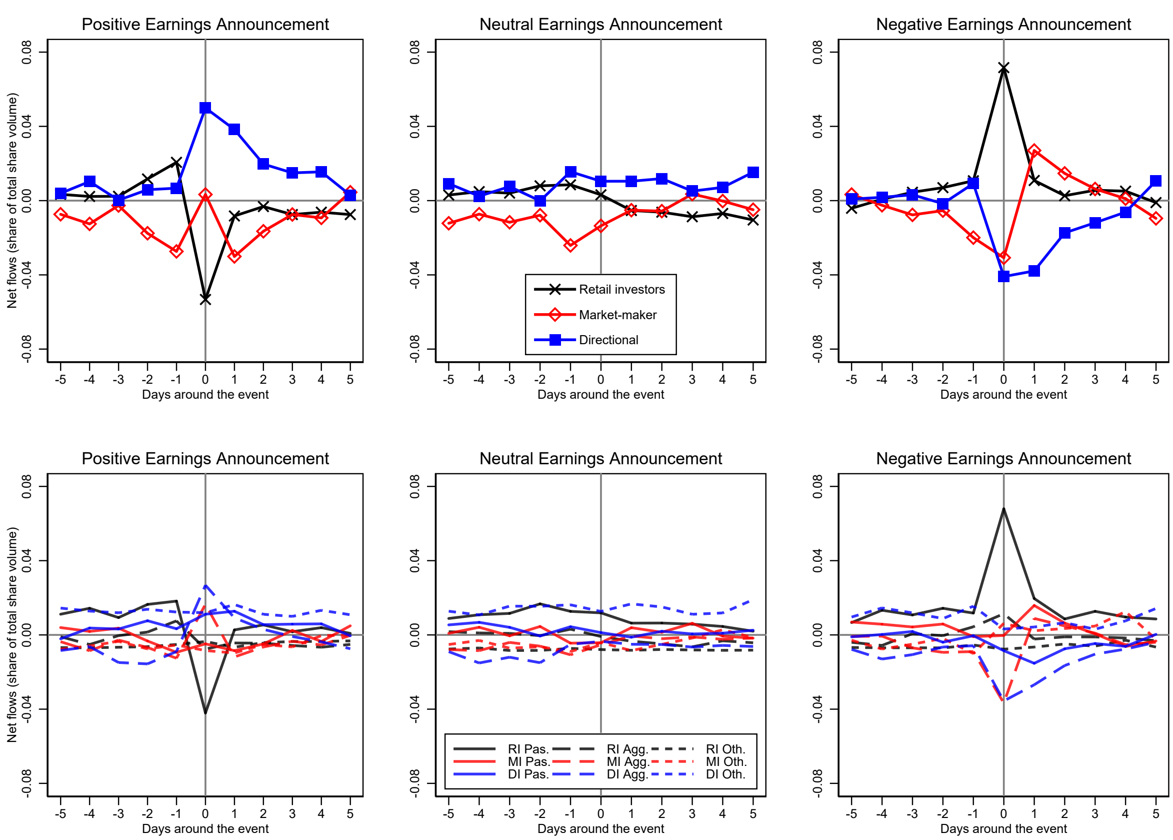


Figure 2: Retail and institutional trading imbalances around material fact announcements

This figure plots average net flows (difference between buying and selling share volumes) by investor group from day −5 to day +5 relative to earnings announcements and material fact disclosures. *Retail investors* represents net flows from retail investors. *Market-making institutions* are institutional investors that both bought and sold shares of the same stock on the disclosure day, indicating liquidity provision rather than directional trading. *Directional institutions* include all other institutional investors with unidirectional trading patterns. This figure presents event-study estimates of average net trading imbalances around material fact announcements, separately for positive, neutral, and negative events. Events are classified according to the stock's abnormal return on the announcement day relative to the market, using a two-standard-deviation threshold based on the stock's rolling 252-day volatility. For each stock–day in the event window from day −5 to day +5, we compute each investor group's net trading flow as buy volume minus sell volume, normalized by total daily share volume. The top row reports total net trading imbalances for retail investors, market-making institutions, and directional institutions. The bottom row decomposes imbalances by investor group and execution mechanism into passive, aggressive, and auction-related flows. The plotted series show averages across events by event time.

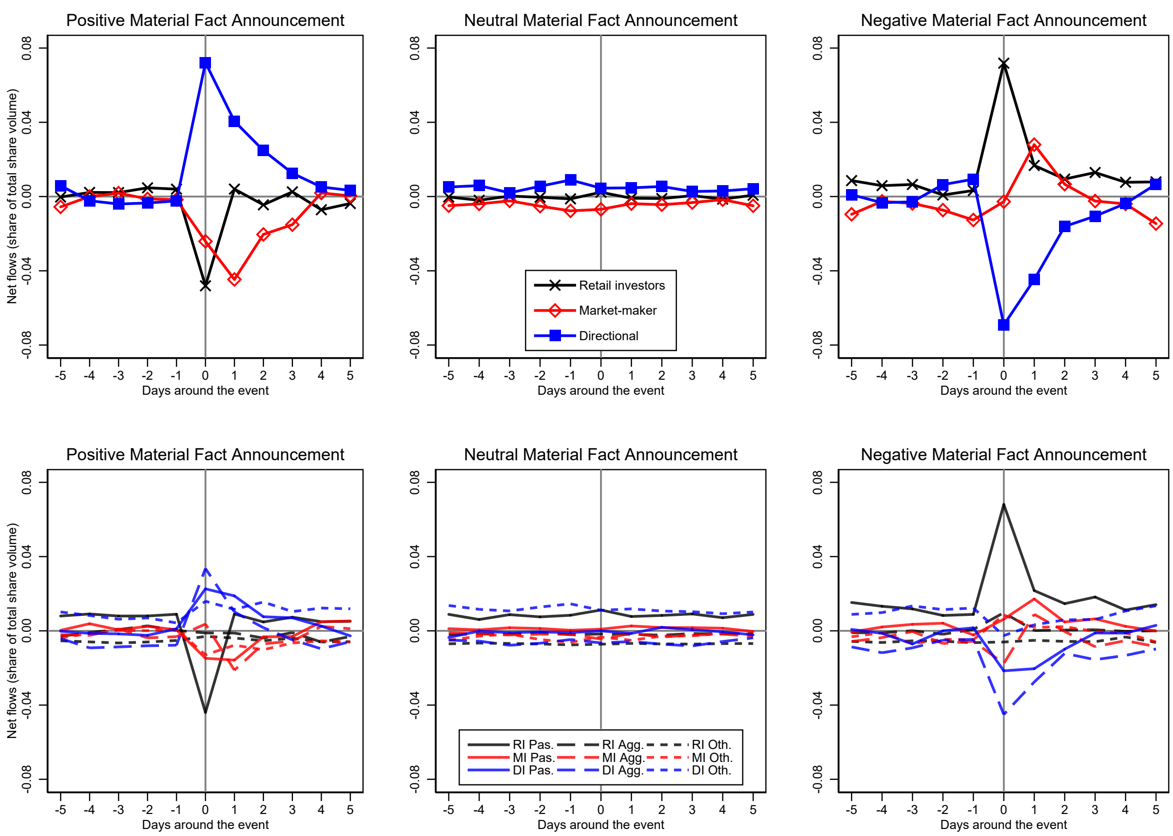
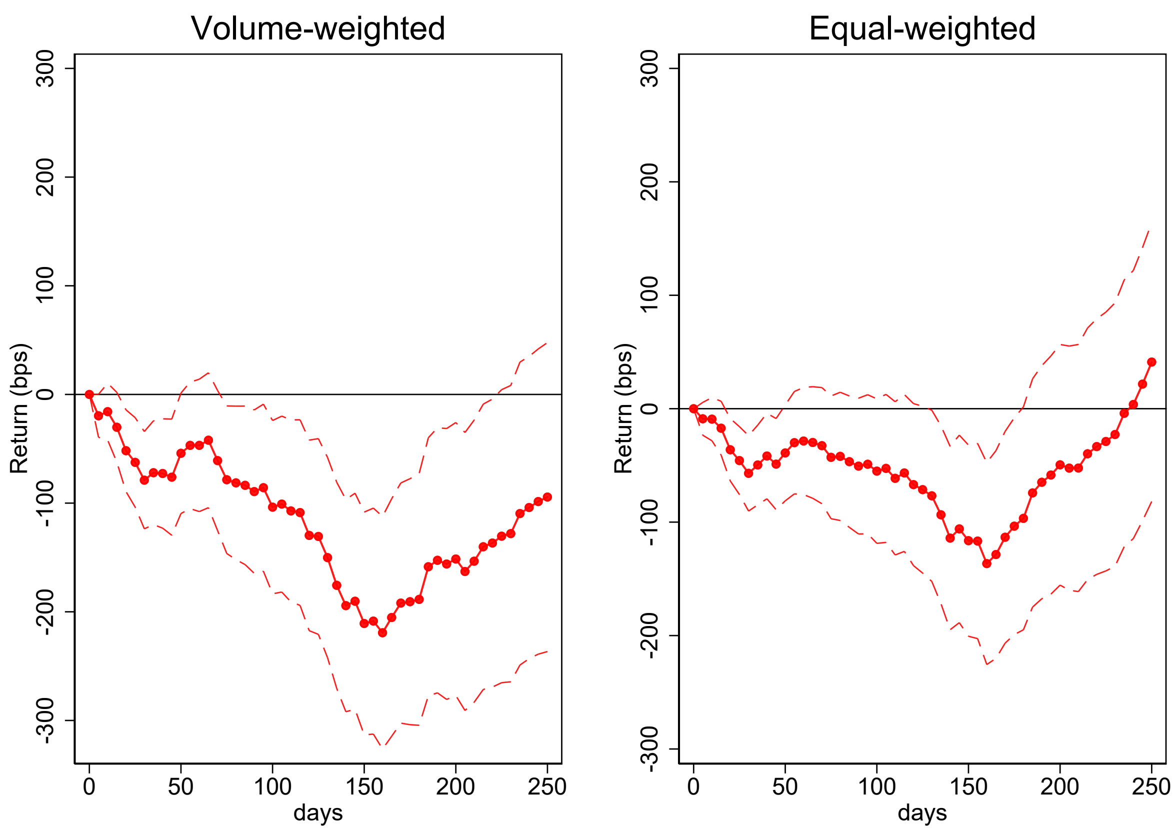


Figure 3: Future returns of long–short portfolios based on retail trading flows

This figure shows cumulative future returns for long–short portfolios constructed from retail net flows over the period 2017–2022. On each day $t = 0$, stock–days are first classified into buying days (net retail share flows > 0) and selling days (net retail share flows < 0). Within each category, stock–days are ranked into terciles according to the magnitude of net absolute flows. The portfolio strategy goes long in the most heavily purchased stocks on day $t = 0$ (top buying tercile) and short in the most heavily sold stocks on day $t = 0$ (top selling tercile). Future excess returns are computed as cumulative returns from the close of day t to the close of day $t + h$, excluding day zero. Panel (a) reports equally weighted results across stock–days, while Panel (b) reports volume-weighted results, where each stock–day return is weighted by the absolute value of its net retail flow in U.S. dollars. Dashed lines represent 95% confidence intervals based on portfolio-level standard errors.



Tables

Table 1: Descriptive statistics

The table presents descriptive statistics of stock-day trading flows from 2017 to 2022 for three groups of investors: retail investors, directional institutions (those who take one-sided positions), and market makers (institutions that both buy and sell on the same day, but not necessarily the same amount). Panel A reports stock-day "Total volume," expressed in USD million, calculated as the sum of daily buy and sell volumes, using the average BRL/USD exchange rate of 4.415. Panel B shows "(Buying + Selling) / Total share volume," defined as the total number of shares bought and sold by each group divided by the total number of shares traded in the stock on that day. Panel C presents "(Buying − Selling) / Total share volume," which captures the net trading imbalances of each group normalized by the total number of shares traded in the stock on that day. Panel D reports the composition of each group's trading volume by order type, expressed as a percentage of the group's total share volume. Passive orders are non-marketable limit orders (agredido); aggressive orders are marketable orders (agressor); other orders include opening and closing auctions, direct trades, and other non-continuous sessions. All fractions are multiplied by 100.

	N	Mean	St. Dev.	Min.	Pct. 10	Pct.25	Pct.50	Pct.75	Pct.90	Max.
Panel A: Total volume (in US$ mi)										
Retail investors	133,885	13.93	33.42	0.00	0.72	1.63	4.06	11.96	31.05	1,456.91
Market-making inst.	133,885	40.87	83.90	0.00	2.65	7.26	17.68	40.06	89.17	2,283.82
Directional inst.	133,885	22.53	41.39	0.00	1.89	4.44	10.30	23.66	51.95	3,083.42
All investors	133,885	77.33	149.23	0.00	6.75	15.33	34.89	76.61	167.82	4,972.73
Panel B: (Buying + Selling) / Total share volume (in percentage)										
Retail investors	133,885	17.59	13.26	0.00	4.85	7.97	13.70	23.57	36.23	100.00
Market-making inst.	133,885	49.52	13.29	0.00	32.60	42.10	50.89	58.59	65.05	100.00
Directional inst.	133,885	32.88	14.41	0.00	14.85	22.86	31.96	41.74	51.51	100.00
Panel C: (Buying − Selling) / Total share volume (in percentage)										
Retail investors	133,885	-0.04	4.51	-50.00	-4.54	-1.80	-0.05	1.77	4.69	50.00
Market-making inst.	133,885	-0.25	7.15	-49.33	-8.91	-4.65	-0.24	4.13	8.49	47.15
Directional inst.	133,885	0.28	7.88	-50.00	-9.31	-4.64	0.19	5.11	9.92	50.00
Panel D: Order-type composition (% of group's total share volume)										
Passive orders										
Retail investors	133,883	50.86	9.48	0.00	38.76	45.51	51.63	56.95	61.78	100.00
Market-making inst.	133,555	42.56	8.80	0.00	32.52	37.62	42.68	47.58	52.32	100.00
Directional inst.	133,824	36.65	13.73	0.00	19.15	27.44	36.44	45.43	53.96	100.00
Aggressive orders										
Retail investors	133,883	40.61	9.09	0.00	30.00	35.04	40.24	45.69	51.54	100.00
Market-making inst.	133,555	44.83	9.19	0.00	33.75	40.04	45.65	50.62	55.01	100.00
Directional inst.	133,824	36.64	13.41	0.00	20.32	27.62	35.90	44.79	53.91	100.00
Other orders										
Retail investors	133,883	8.53	7.32	0.00	2.65	4.14	6.50	10.46	16.56	100.00
Market-making inst.	133,555	12.61	11.55	0.00	3.16	5.25	9.00	15.86	26.53	100.00
Directional inst.	133,824	26.71	17.87	0.00	7.78	13.40	22.61	35.92	52.06	100.00

Table 2: Net share volume by investor group and order type across stock return bins

This table reports average net trading shares by investor group across bins of standardized excess returns, $StdRet^{ex}_{s,t}$. For stock s on day t, $StdRet^{ex}_{s,t}$ is defined as the excess return relative to the equal-weighted market return (i.e., the cross-sectional average $Ret_{s,t}$ across all s), standardized by its rolling 252-day standard deviation estimated over days $t-252$ to $t-1$. Stock–day observations are partitioned into seven bins of $StdRet^{ex}_{s,t}$: ≤ -3, $(-3,-2]$, $(-2,-1]$, $[-1,1)$, $[1,2)$, $[2,3)$, and ≥ 3. For each bin, we compute the average net-flow share of three investor groups: "Directional institutions" (DI), "Market-making institutions" (MM), and "Retail investors" ($Retail$). Net shares are defined as the group's total share purchases minus total share sales, divided by total trading volume in shares, and multiplied by 100. We further decompose each group's net shares into aggressive orders (A), passive orders (P), and other (O) flows, where "other" corresponds to trades executed in the opening auction, closing auction, or intraday large block auctions.

Bin	# Obs.	$StdRet^{ex}$	Directional inst.				Market-making inst.				Retail investors			
			Total	Agg.	Pass.	Other	Total	Agg.	Pass.	Other	Total	Agg.	Pass.	Other
bin -3	695	-4.31	-8.24	-5.33	-2.19	-0.72	-0.52	-2.23	0.60	1.12	8.76	1.68	7.47	-0.40
bin -2	2,075	-2.34	-8.98	-5.55	-3.08	-0.35	2.07	-0.92	1.19	1.80	6.91	0.96	7.05	-1.10
bin -1	14,083	-1.36	-7.16	-4.70	-2.39	-0.07	1.94	-0.62	1.26	1.29	5.22	0.36	5.89	-1.03
bin 0	*99,649*	*-0.04*	*0.61*	*-0.76*	*0.04*	*1.33*	*-0.51*	*-0.23*	*0.09*	*-0.37*	*-0.10*	*-0.24*	*0.96*	*-0.82*
bin 1	13,187	1.38	7.96	3.61	2.38	1.97	-2.89	0.23	-1.64	-1.47	-5.07	-0.20	-4.52	-0.35
bin 2	2,905	2.38	8.79	4.20	2.50	2.10	-2.28	0.92	-1.68	-1.53	-6.51	-0.38	-5.88	-0.25
bin 3	1,193	4.15	8.64	4.39	2.37	1.88	-1.76	1.09	-1.51	-1.34	-6.88	-0.02	-6.47	-0.39
all	133,787	-0.00	0.57	-0.69	0.03	1.23	-0.49	-0.21	0.01	-0.30	-0.08	-0.15	0.85	-0.78

Table 3: Trading flows on days of high adverse-selection risk

The table reports panel regressions of investor-group trading shares on proxies for adverse-selection risk. The dependent variable is the trading share of each investor group, defined as the ratio of the group's buying plus selling volume to total share volume. The explanatory variables are dummy indicators for periods of heightened adverse-selection risk: (i) *Abnormal Volatility*, equal to one when the average daily percentage price range over days $t-5$ to $t-1$ exceeds twice the average computed over days $t-60$ to $t-1$; (ii) *Earnings Announcement*, equal to one on days when quarterly results are disclosed, with after-hours announcements attributed to the next trading day; (iii) *Material Fact Release*, equal to one on days when firms issue mandatory corporate disclosures, with after-hours releases similarly shifted to the next trading day; (iv) *Intense News Activity*, equal to one when the number of Bloomberg news items over days $t-5$ to $t-1$ is more than twice the average over days $t-60$ to $t-1$; and (v) *Abnormal Bid–Ask Spread*, equal to one when the stock's bid–ask spread on day t exceeds its own trailing 60-day mean by more than two standard deviations. Columns (1)–(5) report univariate specifications, while column (6) includes all adverse-selection proxies jointly. All regressions include stock fixed effects. Standard errors are clustered at the stock level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Retail investors						
Abnormal Volatility	2.907***					5.657***
	(0.867)					(1.515)
Earnings Announcement		1.579***				1.476***
		(0.166)				(0.175)
Material Fact Release			0.915***			0.699***
			(0.246)			(0.233)
Intense News Activity				0.949***		0.789***
				(0.185)		(0.173)
Abnormal Bid–Ask Spread					2.039***	1.526***
					(0.302)	(0.319)
Constant	17.530***	17.550***	17.543***	17.516***	17.410***	17.257***
	(0.014)	(0.003)	(0.009)	(0.012)	(0.025)	(0.035)
Fixed effects	yes	yes	yes	yes	yes	yes
Adj. R2	55.19	59.52	55.13	59.52	59.61	59.82
Obs.	133845	133845	133845	133845	133845	133845

Continuous next page...

... Continuous from previous

Panel B: Market-making institutions						
Abnormal Volatility	-11.944***					-11.296***
	(1.163)					(1.162)
Earnings Announcement		-1.476***				-1.414***
		(0.195)				(0.207)
Material Fact Release			-0.770***			-0.655***
			(0.200)			(0.199)
Intense News Activity				-1.372***		-1.035***
				(0.227)		(0.199)
Abnormal Bid–Ask Spread					-1.944***	-0.948***
					(0.356)	(0.271)
Constant	49.731***	49.562***	49.566***	49.625***	49.696***	49.915***
	(0.019)	(0.003)	(0.008)	(0.015)	(0.029)	(0.031)
Fixed effects	yes	yes	yes	yes	yes	yes
Adj. R2	51.91	51.24	51.23	51.27	51.32	52.00
Obs.	133845	133845	133845	133845	133845	133845
Panel C: Directional institutions						
Abnormal Volatility	5.422***					5.639***
	(1.512)					(1.524)
Earnings Announcement		-0.103				-0.062
		(0.189)				(0.193)
Material Fact Release			-0.231			-0.043
			(0.194)			(0.196)
Intense News Activity				0.424**		0.247
				(0.213)		(0.198)
Abnormal Bid–Ask Spread					-0.095	-0.578**
					(0.312)	(0.224)
Constant	32.798***	32.888***	32.895***	32.859***	32.894***	32.828***
	(0.025)	(0.003)	(0.007)	(0.014)	(0.026)	(0.034)
Fixed effects	yes	yes	yes	yes	yes	yes
Adj. R2	47.73	47.61	31.54	47.61	47.61	47.73
Obs.	133845	133845	133845	133845	133845	133845

Table 4: Institutional Investor Frequency Classification: Counts and Characteristics

This table characterizes the classification of institutional investors into frequency halves based on cumulative trading activity. Panel A reports the number of institutional investors assigned to each frequency half by semester. Halves are defined using cumulative trading activity from the start of the sample through the prior semester, with volume-weighted thresholds ensuring each group captures approximately 50% of total trading volume. "Not Classified" refers to investors who traded in semester t but were not present in any prior semester and therefore could not be assigned to a frequency group. The Low Frequency (LFT) half contains investors with the lowest trading intensity, while the High Frequency (HFT) half contains the most active traders. Panel B reports trading characteristics for each frequency half. Statistics are computed at the investor-semester level and then averaged across all semesters from 2017S2 to 2022S2, weighted by the number of investors in each semester. *Prop. Active Days* is the fraction of trading days on which the investor traded at least once. *Active Days* is the total number of days with at least one trade. *Avg. Daily Trades* is the mean number of trades per active day. *Two-Sidedness* equals $2 \times \min(\text{BuyValue}, \text{SellValue})/(\text{BuyValue}+\text{SellValue})$, approaching one when buying and selling are balanced and zero when trading is one-directional. *Herfindahl Index* measures concentration of trading volume across stocks (lower values indicate broader activity). *Stocks Traded* is the average number of distinct stocks traded per day. The classification exhibits high persistence: 99.9% of LFT investors remain in the same half after one semester, as do 94.5% of HFT investors. Spearman rank correlations of investor frequency rankings between consecutive semesters average 0.97, confirming the stability of relative trading intensity over time.

Panel A: Number of Institutional Investors by Semester and Frequency Group

	Frequency				
Semester	Low	High	Not Classified	Total Classified	Total Trading
2017 S2	9,692	26	2,388	9,718	10,066
2018 S1	12,047	34	2,417	12,081	10,988
2018 S2	14,428	46	2,099	14,474	10,998
2019 S1	16,483	55	3,652	16,538	13,309
2019 S2	20,008	82	4,469	20,090	14,980
2020 S1	24,378	94	5,143	24,472	17,284
2020 S2	29,393	197	4,880	29,590	19,255
2021 S1	34,204	251	4,829	34,455	21,598
2021 S2	39,007	250	3,839	39,257	21,374
2022 S1	42,791	229	2,839	43,020	20,157
2022 S2	45,564	249	2,928	45,813	19,589

Panel B: Trading Characteristics by Frequency Group (Averages)

	Frequency Half	
Characteristic	Low	High
Trading Frequency		
Prop. Active Days	0.07	0.96
Active Days	52.5	910.2
Avg. Daily Trades	151.9	6,595.4
Market-Making Proxies		
Two-Sidedness	0.06	0.60
Herfindahl Index	0.84	0.27
Stocks Traded	2.1	18.9

Table 5: Descriptive statistics by investor group

The table presents descriptive statistics of stock-day trading flows from July 2017 to December 2022 for four groups of investors: retail investors, market-making institutions (higher frequency institutions), and directional institutions (lower frequency institutions), and not-classified institutions (investors who traded in period t but were not present in period $t-1$). Institutional investors are assigned to frequency halves using volume-weighted thresholds based on cumulative trading activity from the start of the sample through the prior semester (out-of-sample). Each half accounts for approximately 50% of total institutional trading volume. Panel A reports total stock-day volume in USD million, computed as the sum of daily buy and sell values converted at the average BRL/USD exchange rate of 4.415. Panel B shows (Buying + Selling) / Total share volume, defined as the group's total shares traded divided by total shares traded in that stock on that day. Panel C presents (Buying − Selling) / Total share volume, capturing net trading imbalances normalized by total shares traded. Panel D reports the number of distinct investors active in a given stock on a given day, conditional on the group trading that day. The fractions in Panels B and C are multiplied by 100. Only stock-days satisfying market clearing (total shares bought = total shares sold across all investors) are included.

	N	Mean	St. Dev.	Min.	Pct. 10	Pct. 25	Pct. 50	Pct. 75	Pct. 90	Max.
Panel A: Total volume (in US$ mi)										
Retail investors	124,908	14.58	34.37	0.00	0.81	1.77	4.32	12.63	32.55	1,456.91
Market-making inst.	124,908	30.40	54.58	0.00	2.75	6.53	14.80	31.49	66.76	1,546.03
Directional inst.	124,908	34.14	70.01	0.00	2.66	6.24	14.46	33.17	75.72	3,664.82
Not classified inst.	124,908	1.51	6.57	0.00	0.00	0.03	0.20	0.93	3.13	599.47
All investors	124,908	80.63	153.53	0.00	7.57	16.56	37.00	80.06	175.30	4,981.38
Panel B: (Buying + Selling) / Total share volume (in percentage)										
Retail investors	124,908	17.55	12.97	0.00	4.92	8.06	13.82	23.57	35.88	100.00
Market-making inst.	124,908	39.75	10.81	0.00	26.57	33.19	39.96	46.70	53.03	90.37
Directional inst.	124,908	40.91	12.73	0.00	24.44	32.46	41.13	49.21	56.64	97.63
Not classified inst.	124,908	1.79	3.23	0.00	0.00	0.10	0.56	1.95	5.09	79.19
Panel C: (Buying − Selling) / Total share volume (in percentage)										
Retail investors	124,908	-0.00	4.40	-50.00	-4.50	-1.79	-0.04	1.80	4.73	50.00
Market-making inst.	124,908	-0.34	6.58	-48.34	-8.07	-4.15	-0.31	3.48	7.44	48.19
Directional inst.	124,908	0.03	7.33	-50.00	-8.71	-4.33	0.02	4.35	8.70	50.00
Not classified inst.	124,908	0.31	2.43	-50.00	-0.61	-0.03	0.04	0.38	1.62	50.00
Panel D: Number of distinct investors per stock-day										
Retail investors	124,908	1,168	1,994	0	107	233	540	1,272	2,778	112,639
Market-making inst.	124,908	34	17	0	16	21	30	44	58	144
Directional inst.	124,908	128	85	0	47	73	109	160	234	2,007
Not classified inst.	124,908	9	10	0	1	2	6	11	19	247

Table 6: Volume-weighted day-zero measures by order type and imbalance intensity

This table reports daily summary statistics for retail order imbalances from 2017 to 2022. Columns under "Buying" use stock-days with positive retail net flow; columns under "Selling" use stock-days with negative retail net flow. "Light," "Moderate," and "Heavy" correspond to terciles of the absolute retail net flow computed within day. Panel A reports day-zero returns (in basis points) by order type: all orders, passive orders, aggressive orders, and other orders. Panel B reports the half bid–ask spread (in basis points). Panel C reports day-zero returns relative to non-retail investors (in basis points), computed using the average transaction price of non-retail investors on the same stock-day.

		Buying				**Selling**			
	N	Mean	Pct.25	Pct.50	Pct.75	Mean	Pct.25	Pct.50	Pct.75
Panel A: Day-zero returns									
All orders									
All	1,485	-55.54	-94.16	-48.47	-9.78	-48.45	-83.42	-45.29	-7.87
Light	1,485	-3.34	-35.94	-2.98	28.20	-13.21	-39.69	-11.84	15.37
Moderate	1,485	-19.19	-52.08	-17.35	15.10	-22.25	-53.31	-21.65	11.78
Heavy	1,485	-61.31	-102.21	-55.22	-11.84	-52.13	-89.25	-49.20	-9.52
Passive orders									
All	1,483	-58.92	-96.56	-50.29	-13.01	-58.08	-95.69	-56.59	-15.93
Light	1,483	10.33	-21.41	7.71	38.23	-10.84	-40.82	-7.86	21.85
Moderate	1,483	-11.36	-43.38	-9.27	20.17	-21.70	-55.81	-22.00	11.64
Heavy	1,483	-66.23	-105.35	-58.19	-18.29	-63.07	-103.70	-61.29	-18.99
Aggressive orders									
All	1,483	-13.32	-56.41	-8.76	33.21	-10.67	-53.97	-10.51	29.54
Light	1,483	0.82	-32.97	0.47	35.52	-12.92	-45.81	-12.75	21.12
Moderate	1,483	-5.62	-39.00	-3.76	31.89	-12.63	-48.47	-11.90	25.12
Heavy	1,483	-14.66	-59.93	-10.33	35.38	-10.30	-57.16	-10.53	31.51
Other orders									
All	1,483	3.96	-14.04	4.82	23.08	1.60	-9.66	3.25	16.46
Light	1,483	4.68	-10.90	2.32	18.84	-2.60	-11.55	-0.63	9.65
Moderate	1,483	3.77	-10.68	2.44	17.64	0.99	-7.84	1.05	10.49
Heavy	1,483	3.97	-15.42	4.70	24.93	1.89	-10.27	3.51	17.96
Panel B: Bid–ask spread									
All	1,482	5.90	4.26	5.12	6.48	5.39	4.26	5.12	6.13
Light	1,482	8.31	4.99	6.29	9.03	9.11	5.75	7.39	11.36
Moderate	1,482	7.18	4.85	6.11	7.94	7.17	5.48	6.65	7.96
Heavy	1,482	5.62	3.91	4.73	6.04	4.98	3.85	4.68	5.69
Panel C: Day-zero returns relative to non-retail investors									
All	1,485	-4.17	-10.71	-3.01	3.95	-0.10	-6.24	0.86	7.14
Light	1,485	3.81	-4.31	3.79	11.72	5.61	-0.94	5.60	12.34
Moderate	1,485	0.41	-6.92	0.84	7.83	3.61	-2.92	3.50	9.65
Heavy	1,485	-4.95	-12.08	-3.76	3.84	-0.60	-7.08	0.34	7.27

Table 7: Retail net flows and future excess returns

This table shows the average future excess returns, $Ret^{ex}_{t+1,t+5}$, $Ret^{ex}_{t+1,t+20}$, and $Ret^{ex}_{t+1,t+40}$, for buying stocks (i.e., stock-days with positive net flow) by retail investors in columns (1), (2), and (3), and for selling stocks (i.e., stock-days with negative net flow) in columns (4), (5), and (6). Returns are volume weighted by their net flows and are in excess of market returns. We accumulate returns from the closing price on day t to the closing price on day $t+h$, where $h = 5, 20, 40$. In Panel A, we consider all retail orders; in Panel B only retail aggressive orders; in Panel C only retail passive orders; and in Panel D only retail other orders. In the "All" rows, we consider all buying (or selling) stocks. In the "Light" rows, we consider stocks with light buying (selling), i.e., stocks in the lowest tercile according to their net flow on a given day. In the "Moderate" rows, we consider stocks in the middle tercile. In the "Heavy" rows, we consider stocks in the highest tercile. Average future returns are in basis points. T-statistics are presented in parenthesis and are based on Newey-West standard errors with 10 lags. *, **, and *** indicate significance levels of 0.10, 0.05, and 0.01, respectively.

	Buying			Selling		
	(1) $Ret^{ex}_{t+1,t+5}$	(2) $Ret^{ex}_{t+1,t+20}$	(3) $Ret^{ex}_{t+1,t+40}$	(4) $Ret^{ex}_{t+1,t+5}$	(5) $Ret^{ex}_{t+1,t+20}$	(6) $Ret^{ex}_{t+1,t+40}$
Panel A: All orders						
All	-13.212	-30.287	-58.741**	0.727	2.906	0.508
	(-1.47)	(-1.54)	(-2.20)	(0.10)	(0.16)	(0.02)
Light	20.451**	53.015**	71.780**	-6.424	-38.556	-60.007
	(2.25)	(2.35)	(2.04)	(-0.71)	(-1.56)	(-1.61)
Moderate	0.962	-14.939	-39.391	12.116	5.335	9.773
	(0.11)	(-0.69)	(-1.24)	(1.42)	(0.25)	(0.30)
Heavy	-15.189	-35.926*	-65.759**	0.096	2.800	1.960
	(-1.56)	(-1.65)	(-2.24)	(0.01)	(0.14)	(0.07)
Panel B: Aggressive orders						
All	-8.502	-25.614	-54.943**	7.497	12.573	7.796
	(-1.09)	(-1.47)	(-2.31)	(1.15)	(0.73)	(0.35)
Light	15.898*	30.589	49.419	-5.601	-30.467	-49.646
	(1.80)	(1.41)	(1.47)	(-0.68)	(-1.35)	(-1.38)
Moderate	8.719	7.558	-15.689	4.229	4.066	0.382
	(0.94)	(0.35)	(-0.47)	(0.53)	(0.20)	(0.01)
Heavy	-11.656	-34.603*	-67.022**	9.220	16.422	12.530
	(-1.36)	(-1.79)	(-2.54)	(1.22)	(0.85)	(0.50)

Continuous next page...

... Continuous from previous

Panel C: Passive orders						
All	-12.679	-26.330	-54.260*	-1.601	3.275	-0.060
	(-1.38)	(-1.29)	(-1.86)	(-0.20)	(0.16)	(-0.00)
Light	8.415	48.380**	55.422	-5.807	-23.247	-54.528
	(0.96)	(2.18)	(1.53)	(-0.63)	(-0.95)	(-1.48)
Moderate	6.661	-5.298	-15.706	5.652	1.970	2.853
	(0.77)	(-0.25)	(-0.51)	(0.73)	(0.09)	(0.08)
Heavy	-15.021	-30.603	-62.879*	-1.979	3.087	1.417
	(-1.48)	(-1.34)	(-1.93)	(-0.22)	(0.14)	(0.05)
Panel D: Other orders						
All	-12.679	-26.330	-54.260*	7.951	22.445	45.102*
	(-1.38)	(-1.29)	(-1.86)	(1.07)	(1.26)	(1.78)
Light	10.636	35.432	60.318*	-13.554	-52.959**	-102.978***
	(1.23)	(1.61)	(1.67)	(-1.44)	(-2.14)	(-2.72)
Moderate	-6.221	-15.143	-26.791	-5.315	-9.087	-9.022
	(-0.70)	(-0.76)	(-0.80)	(-0.59)	(-0.42)	(-0.29)
Heavy	-4.707	-27.881	-79.643***	11.334	31.344	61.575**
	(-0.47)	(-1.23)	(-2.68)	(1.41)	(1.61)	(2.20)

Table 8: Descriptive statistics and correlations for the market quality measures

This table reports descriptive statistics (Panel A) and correlations (Panel B) for the market quality index and its underlying components. The market quality index is the standardized first principal component of four daily measures computed from ten-minute interval data: (i) *Price Impact*, the stock-day average of absolute return in basis points divided by traded volume in millions; (ii) *Abs. Auto Corr*, the absolute value of first-order autocorrelation of ten-minute returns; (iii) *B.AS.*, is the daily average extract from Bloomberg; and (iv) *Vol Ratio*, the absolute deviation from one of the ratio of one-hour to ten-minute realized volatility. All measures are averaged at the stock-day level over the 2017–2022 period.

Panel A: Descriptive Statistics

	Mean	S.D.	Pct.5	Pct.25	Pct.50	Pct.75	Pct.95
Market Quality Index	0.000	1.000	-0.766	-0.109	0.164	0.371	0.630
Price Impact	2.523	17.044	0.025	0.117	0.294	0.827	4.914
Abs. Auto Corr.	15.070	11.132	1.197	6.083	12.859	21.818	36.831
B.A.S	15.409	27.290	5.210	7.450	10.010	14.790	37.300
Vol Ratio	0.845	0.083	0.685	0.797	0.859	0.908	0.955

Panel B: Correlation Matrix

	Eff Index	Price Impact	Abs. Auto Corr.	B.A.S	Vol Ratio
Market Quality Index	1.000				
Price Impact	0.803***	1.000			
Abs. Auto Corr.	0.342***	0.051***	1.000		
B.A.S	0.819***	0.449***	0.097***	1.000	
Vol Ratio	0.237***	0.032***	0.196***	0.028***	1.000

Table 9: Retail liquidity provision and market quality

This table reports estimates of (1) columns 1 and 2 and (2) columns 3 and 4. The dependent variable is the market quality index, $MarketQuality_{s,t}$. The main regressors are log retail volumes in USD millions: $\log(Retail)_{s,t}$ all type of orders, and its decomposition into $\log(Passive)_{s,t}$ non-marketable limit orders, $\log(Aggressive)_{s,t}$ marketable orders, and $\log(Other)_{s,t}$ call auctions. Column (2) and Columns (4) adds lagged controls $X_{s,t-1}$: and past returns and volatility over the previous week and the prior three weeks, and also controls for $AdverseDays_{s,t}$, defined in Table 3 and $\log(Volume)_{s,t-1}$ log total market volume on the previous day. Columns (3)–(4) use the order-type decomposition and include the same controls; All specifications include stock and date fixed effects. Standard errors are clustered at the stock-date level. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Market Quality Index			
	(1)	(2)	(3)	(4)
$\log(Retail)_{s,t}$	0.381***	0.151***		
	(0.109)	(0.034)		
$\log(Passive)_{s,t}$			0.278***	0.192***
			(0.098)	(0.062)
$\log(Aggressive)_{s,t}$			0.071	-0.036
			(0.047)	(0.058)
$\log(Other)_{s,t}$			0.037	0.001
			(0.043)	(0.033)
Adverse Days_{jt}		-0.062***		-0.063***
		(0.019)		(0.020)
$\text{Log(Volume)}_{j,t-1}$		0.436***		0.432***
		(0.121)		(0.118)
$\text{Ret}_{j,t-1:t-5}$		-0.243**		-0.232**
		(0.104)		(0.095)
Ret $\text{Sd}_{j,t-1:t-5}$		-3.73***		-3.64***
		(1.00)		(0.955)
$\text{Ret}_{j,t-6:t-21}$		-0.014		-0.0009
		(0.116)		(0.110)
Ret $\text{Sd}_{j,t-6:t-21}$		-1.31		-1.47
		(1.29)		(1.29)
Stock F.E.	Yes	Yes	Yes	Yes
Date F.E.	Yes	Yes	Yes	Yes
Observations	113,582	113,582	113,582	113,582
Adjusted R^2	0.37563	0.43252	0.37922	0.43471

Table 10: Retail and Market Quality

This table reports the second-stage estimates of the IV system given by equations (5)–(6). The dependent variable is the market quality. Each column instruments one order-type volume while the remaining two enter as OLS controls: Column (1) and (2) instruments $\log(Retail)_{s,t}$, Column (3) instruments $\log(Passive)_{s,t}$, Column (4) instruments $\log(Aggressive)_{s,t}$, and Column (5) instruments $\log(Other)_{s,t}$. In all three specifications the instrument is the shift-share bond platform measure $\log(Z^{Bonds}_{s,t})$ defined in equation (3). The Wald statistic is the Kleibergen-Paap rk Wald statistic. All specifications control for cumulative returns and return volatility over the prior week and prior three weeks, lagged market volume, and an adverse-days dummy, and include stock and date fixed effects. Standard errors are two-way clustered by stock and date and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable:	Market Quality Index				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
$\widehat{\log(Retail)}_{s,t}$	0.8645***	0.9718**			
	(0.2780)	(0.4531)			
$\widehat{\log(Passive)}_{s,t}$			2.679**		
			(1.330)		
$\widehat{\log(Aggressive)}_{s,t}$				-12.42	
				(16.68)	
$\widehat{\log(Other)}_{s,t}$					3.754
					(2.893)
Fixed-effects					
date	Yes	Yes	Yes	Yes	Yes
ticker	Yes	Yes	Yes	Yes	Yes
controls	No	Yes	Yes	Yes	Yes
Fit statistics					
Observations	113,582	113,582	113,582	113,582	113,582
Adjusted R^2	0.31424	0.42974	0.43482	0.43792	0.43804
Wald (1st stage), $\widehat{\log(Retail)}_{s,t}$	19.532	14.595			
Wald (1st stage), $\widehat{\log(Passive)}_{s,t}$			11.964		
Wald (1st stage), $\widehat{\log(Aggressive)}_{s,t}$				0.59650	
Wald (1st stage), $\widehat{\log(Other)}_{s,t}$					2.6620

Table 11: Robustness of the IV Specification

This table reports robustness checks for the IV identification. Columns (1) and (2) reproduce, for reference, the baseline OLS and 2SLS estimates from Table **??**. In column (3), the funding shock is given by the absolute residuals from a two-way fixed-effects regression of TD volume on city and time effects. Column (4) uses, as shifters, the residuals from a principal-component model of TD volume, where we first extract the leading component and then residualize TD volume using time-varying loadings and city fixed effects. Column (5) excludes the three largest cities (São Paulo, Rio de Janeiro, and Belo Horizonte) from the construction of the instrument, and column (6) excludes all state capital cities. All specifications controls for past returns, and return standard deviation over the previous week and the previous three weeks and adverse-day dummies. The Wald statistic is the Kleibergen-Paap rk Wald statistic. Also, all specification includes the fixed-effects of ticker and date. Standard errors are clustered at the stock-date level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable:	Market Quality Index					
	OLS	2SLS	TWFE	PCA	WO Major Cities	WO Capitals Cities
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
$\log(Retail)_{s,t}$	0.151***	0.972**	0.907*	0.885*	0.823**	0.809**
	(0.034)	(0.453)	(0.520)	(0.526)	(0.406)	(0.405)
Fixed-effects And Controls						
Ticker F.E	Yes	Yes	Yes	Yes	Yes	Yes
Date F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	113,582	113,582	113,582	113,582	113,582	113,582
Adjusted R^2	0.43252	0.20471	0.23943	0.25025	0.28001	0.28621
F-test (1st stage),$\log(Retail)_{s,t}$		1,776.2	1,204.3	1,134.6	2,261.2	2,065.1
Wald (1st stage),$\log(Retail)_{s,t}$		14.595	28.953	28.400	16.814	15.702

Table 12: Literature Review: Retail Investor Trading Performance

This table summarizes the empirical literature on retail investor trading performance, ordered by publication year. "Day-zero" refers to returns from trade execution to the same-day close (or contemporaneous week/period return when trade-level data are unavailable). "Beyond day-zero" refers to subsequent return predictability or reversal patterns, from the closing price of date t onwards. All returns are in basis points (bps). Sign adjustments convert sell-side returns to a buy-minus-sell framework.

Paper	**Data**	**Sample**	**Day-Zero**	**Beyond Day-Zero**
Kaniel, Saar, and Titman (2008) *JF*	NYSE exchange data	2000–03 USA	Negative. Reported at week-level: −111 bps (for intense buying), −180 bps (for intense selling).	Positive. Reversal pattern: +80 bps (buying), +33 bps (selling) over 20 days.
Linnainmaa (2010) *JF*	Complete records, all individual investors	1995–02 Finland	Mostly negative. Earnings days: −69 bps (expected), −138 bps (unexpected). Normal: +1.6 bps (limit), −36.1 bps (market).	Positive. Reversal pattern at short-run, +53.9 bps (buy-sell; 5 days), but continuation at long-run −148 bps (buy-sell; 126 days).
Kelley and Tetlock (2013) *JF*	Proprietary data, two U.S. wholesalers	2003–07 USA	Negative. −6 bps (market), −1.5 bps (limit).	Positive. Reversal pattern for market order, +66 bps (buy-sell), no pattern for limit orders.
Fong, Gallagher, and Lee (2014) *JFQA*	ASX broker-identified data	1995–07 Australia	Negative. −25 bps (buy-minus-sell); −31 bps (market orders), −20 bps (limit orders).	Negative. Weak continuation pattern at short-term −3.8 bps (day 5); dissipates at longer horizons.
Barrot, Kaniel, and Sraer (2016) *JFE*	Order-level, French online broker	2002–10 France	Negative. −90 bps (simple average across 4.6M trades).	Positive. Reversal pattern; heavily bought outperform heavily sold by +18 bps (16 days).
Boehmer, Jones, Zhang, and Zhang (2021) *JF*	TAQ off-exchange; subpenny algorithm	2010–15 USA	Positive. Reported at week-level: 11 bps (for intense buying), 24 bps (for intense selling)	Positive. Continuation pattern; +57 bps (buying), +30 bps (selling) over 20 days.
Barber, Lin, and Odean (2023) *JFQA*	TAQ off-exchange; subpenny algorithm	2010–19 USA	Negative. −2.60 bps (buy-minus-sell); −1.92 bps (buys); −0.68 bps (sells)	Negative. Continuation pattern at short-term; −30 bps (dollar-weighted, 5 days).

A Internet Appendix

A.1 AMM with an alternative classification of market-making institutions

Next, we replicate the findings of Section 3 using the alternative institutional decomposition described in Section 3.3. The results are qualitatively similar, alleviating concerns that the original classification was spurious or that the retail trading patterns documented above were mechanically driven by the way institutions were grouped.

First, Table IA.1 presents average net trading shares by investor group across bins of standardized excess returns, using the alternative frequency-based classification. The results are broadly consistent with our earlier findings based on ex post two-sidedness. On typical days (bin 0), market-making institutions post a modest negative net share (-0.78 p.p.) while directional institutions and retail investors are close to neutral. When prices move significantly, directional institutions display large imbalances in the direction of price changes: they are heavy net sellers on negative return days (e.g., -7.85 p.p. in bin -3) and heavy net buyers on positive return days (e.g., $+7.16$ p.p. in bin 3). Retail investors absorb the bulk of these directional flows, posting net shares of $+8.87$ p.p. in bin -3 and -6.94 p.p. in bin 3, with the vast majority of this liquidity provision occurring through passive orders ($+7.53$ p.p. and -6.44 p.p., respectively). Market-making institutions, by contrast, provide only limited offsetting flows and occasionally trade in the same direction as directional institutions, particularly in the most extreme negative bin where they post a net share of -1.12 p.p. These patterns confirm our main result: retail investors, not professional market-making institutions, serve as the primary source of liquidity when prices move sharply, absorbing directional institutional order flow predominantly through passive limit orders resting in the book.

[Table IA.1 about here]

Second, Figure IA.1 displays net trading flows around corporate announcements using the alternative frequency-based classification. The patterns are remarkably close to the ones obtained before in Figures 1 and 2. On days with positive earnings or material fact announcements, directional institutions are substantial net buyers, with net flows reaching approximately 4–6 percentage points of total share volume, while retail investors absorb the opposite side as net sellers. On days with negative announcements, the pattern reverses: directional institutions are heavy net sellers and retail investors step in as net buyers, with magnitudes of similar size. Market-making institutions remain close to neutral throughout the event window, providing only modest offsetting flows. The asymmetry is particularly pronounced on day 0, when information is released, and dissipates gradually over the following days. Notably, on neutral announcement days—where the information content is limited—all three groups exhibit minimal net trading imbalances. These results reinforce our main finding: retail investors, not professional market-making institutions, serve as the primary counterparty to informed institutional order flow around corporate news events, consistent with the view that retail limit orders resting in the book are systematically picked off when new information arrives.

[Figure IA.1 about here]

Finally, Table IA.2 replicates the analysis in Table 3 using the alternative decomposition into market-making and directional institutions. The results are very similar and remain so after incorporating the abnormal bid–ask spread proxy. Focusing on column (6), all five proxies for adverse-selection risk are associated with higher retail participation in total market volume: retail trading share is 5.208 p.p. higher on days with abnormal volatility (versus 5.657 p.p. in Table 3), 1.416 p.p. higher on earnings-announcement days (versus 1.476 p.p.), 0.665 p.p. higher on material-fact-release days (versus 0.699 p.p.), 0.860 p.p. higher on days of intense news activity (versus 0.789 p.p.), and 1.500 p.p. higher on days with abnormally high bid–ask spreads (versus 1.526 p.p.).

Market-making institutions, like in the baseline specification, withdraw from the market when adverse-selection risk is high. In column (6), all five proxies are associated with lower market-making participation: trading share falls by 9.489 p.p. during abnormal-volatility episodes (versus 11.296 p.p. in Table 3), 1.228 p.p. on earnings-announcement days (versus 1.414 p.p.), 0.714 p.p. on material-fact-release days (versus 0.655 p.p.), 0.648 p.p. on intense-news days (versus 1.035 p.p.), and 0.785 p.p. on abnormal-spread days (versus 0.948 p.p.).

The results for directional institutions are also broadly similar to those in Table 3. In column (6), directional institution participation increases on days of abnormal volatility, by 3.696 p.p. (versus 5.639 p.p. in Table 3), and decreases on days with abnormal bid–ask spreads, by 0.725 p.p. (versus 0.578 p.p.). The coefficients on earnings announcements, material-fact releases, and intense news activity are all small and statistically insignificant. Overall, the consistency across classification methodologies reinforces our main conclusion: retail investors step in to provide liquidity precisely when professional market-making institutions pull back, regardless of whether these groups are identified ex post based on trading behavior or ex ante based on historical trading frequency.

[Table IA.2 about here]

A.2 Tables

[Figure IA.3 about here]

Figures

Figure IA.1: Net flows around corporate disclosures

This figure plots average net flows (difference between buying and selling share volumes) by investor group from day −5 to day +5 relative to earnings announcements and material fact disclosures. *Retail investors* represents net flows from retail investors. *Market-making institutions* are institutional investors that both bought and sold shares of the same stock on the disclosure day, indicating liquidity provision rather than directional trading. *Directional institutions* include all other institutional investors with unidirectional trading patterns. Net flows are normalized by total daily share volume for each stock. Corporate disclosures are classified as positive, neutral, or negative based on the firm's abnormal return on the disclosure day (or the following trading day for after-hours announcements). An event is positive if the firm's excess return exceeds two standard deviations above the market return, negative if it falls more than two standard deviations below, and neutral otherwise. Standard deviations are computed using daily excess returns over the 252 trading days preceding each event ($t-252$ to $t-1$). The plots show equal-weighted averages across all stock-day events within each classification.

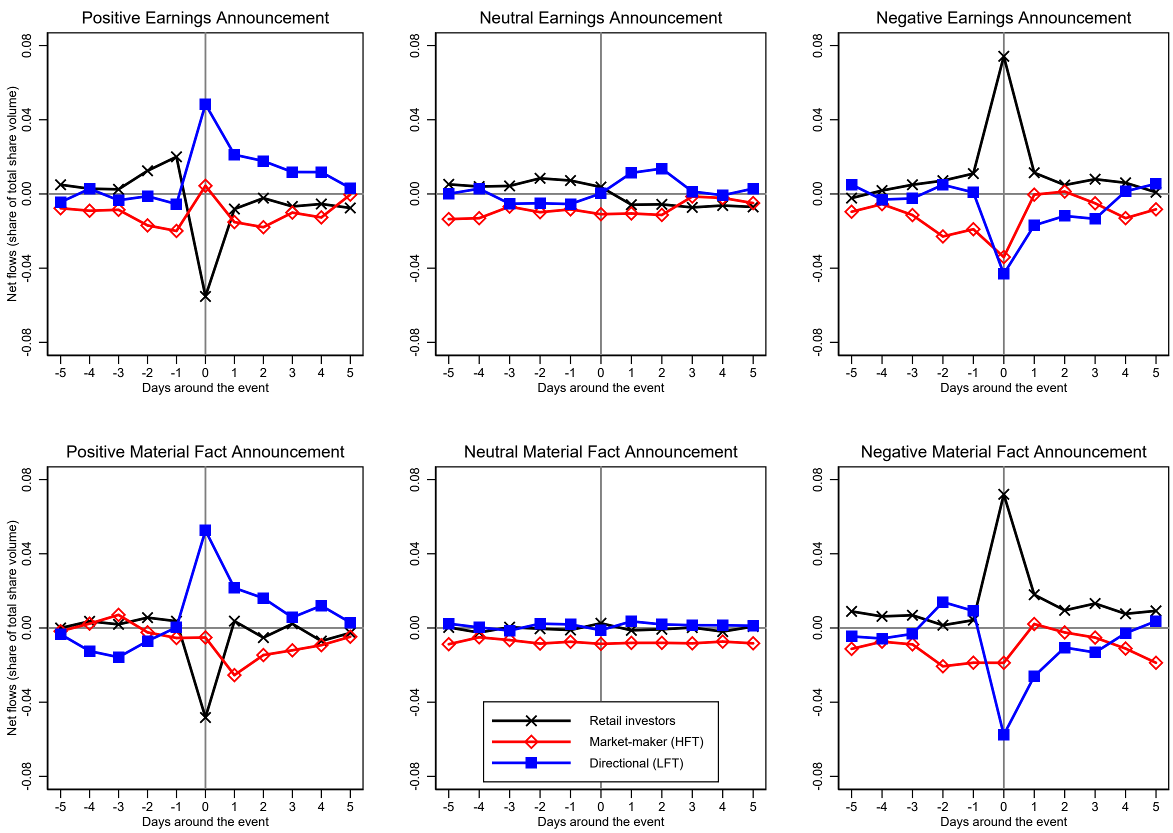


Figure IA.2: First-Stage IV Identification

This figure presents the residuals from the estimation of the two-way fixed effect of the funding of the TD on left-hand-side. On the right-hand-side the residuals from a regression on the first principal component with varying slope and city fixed effect.

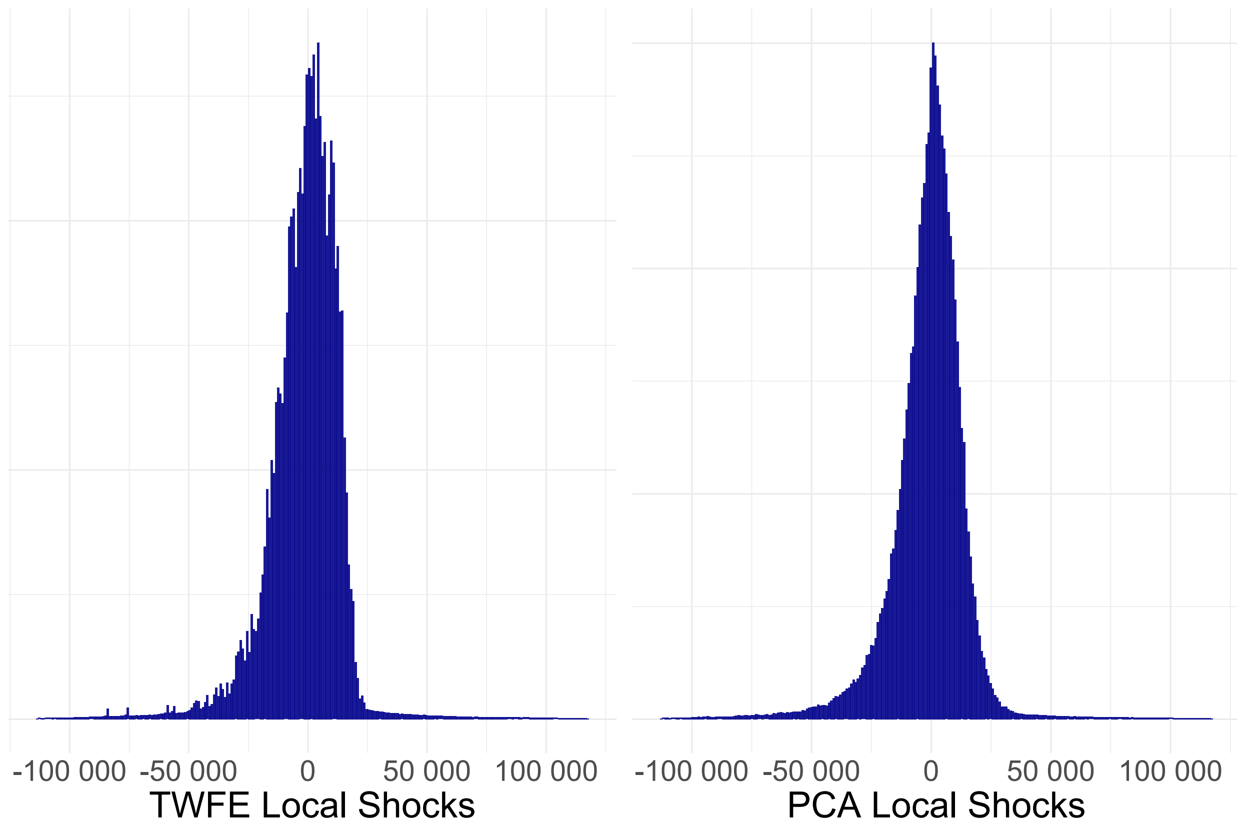
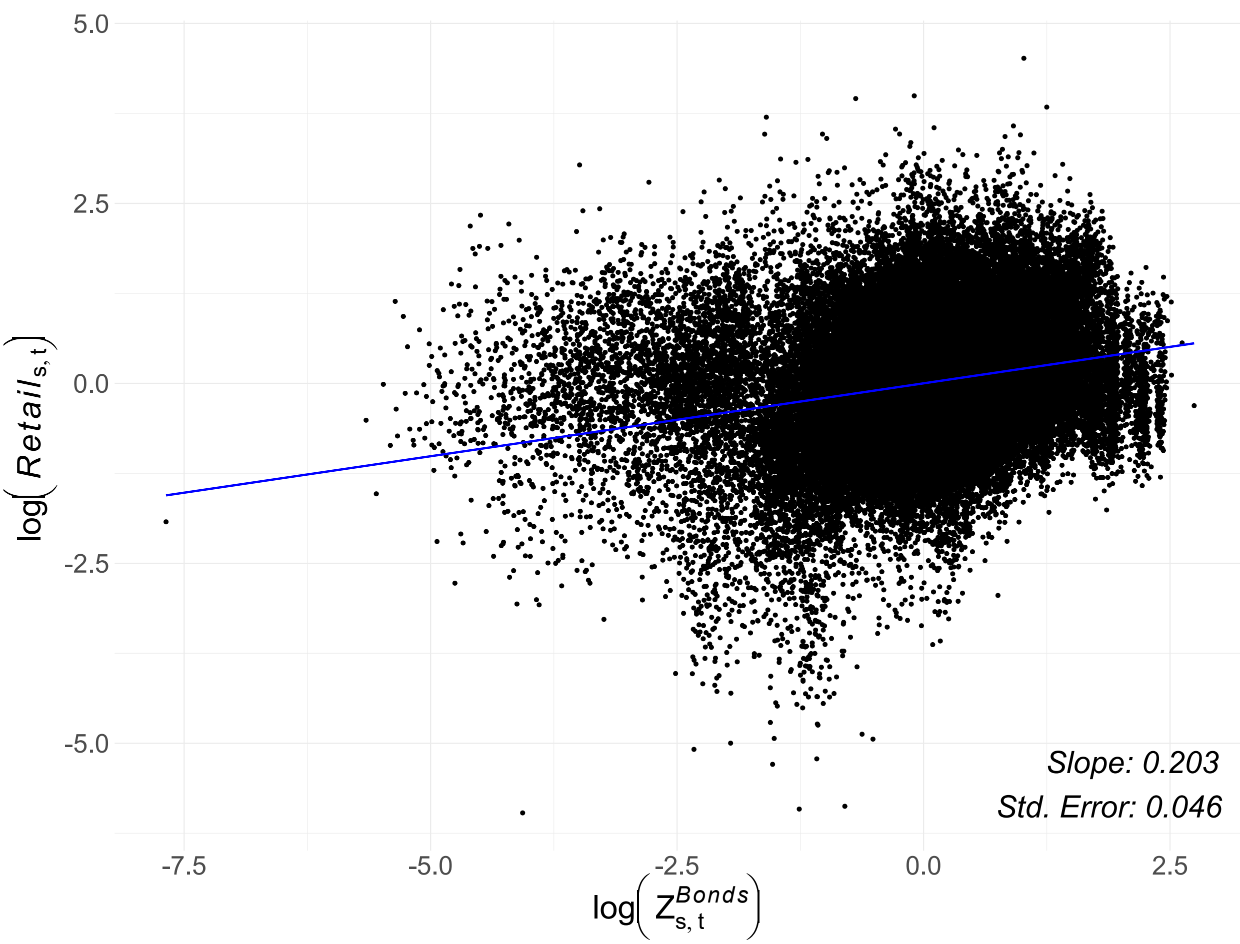


Figure IA.3: First-Stage IV Identification

This figure presents the first-stage relationship underlying the IV identification strategy. The instrument is a shift-share measure constructed as $Z_{s,t}^{Bonds} = \sum_c \omega_{s,c,y-1} \times IS_{c,t}^{Bonds}$, where $\omega_{s,c,y-1}$ are predetermined city-stock portfolio weights from the previous year and $IS_{c,t}^{Bonds}$ are city-level daily aggregate investment flows to the Tesouro Direto (TD) government bond platform. The figure plots residualized $\log(Retail_{s,t})$ against residualized $\log(Z_{s,t}^{Bonds})$, where both variables have been demeaned by stock and date fixed effects. Each point represents a stock-day observation. The solid line depicts the fitted regression line from the first-stage equation, with a slope of 0.203 and a first-stage F-statistic of 6,252.4.



Tables

Table IA.1: Net share volume by investor group and order type across stock return bins

This table reports average net trading shares by investor group across bins of standardized excess returns, $StdRet^{ex}_{s,t}$. For stock s on day t, $StdRet^{ex}_{s,t}$ is defined as the excess return relative to the equal-weighted market return (i.e., the cross-sectional average $Ret_{s,t}$ across all s), standardized by its rolling 252-day standard deviation estimated over days $t-252$ to $t-1$. Stock–day observations are partitioned into seven bins of $StdRet^{ex}_{s,t}$: ≤ -3, $(-3,-2]$, $(-2,-1]$, $(-1,1)$, $[1,2)$, $[2,3)$, and ≥ 3. For each bin, we compute the average net-flow share of four investor groups: "High-frequency traders" (HFT), "Low-frequency traders" (LFT), "Non-classified institutions" (Non), and "Retail investors" ($Retail$). Net shares are defined as the group's total share purchases minus total share sales, divided by total trading volume in shares, and multiplied by 100. For institutional groups (HFT, LFT, Non), we report aggregate net shares only. For retail investors, we further decompose net shares into aggressive orders (A), passive orders (P), and other (O) flows, where "other" corresponds to trades executed in the opening auction, closing auction, or intraday large block auctions.

Bin	# Obs.	$StdRet^{ex}$	*MM-inst.*	*DI-inst.*	*Non*	Retail investors			
			Total	Total	Total	Total	Agg.	Pass.	Other
bin −3	663	−4.31	−1.12	−7.85	0.10	8.87	1.74	7.53	−0.40
bin −2	1,961	−2.34	0.36	−7.71	0.40	6.94	0.99	7.07	−1.11
bin −1	13,173	−1.36	0.33	−6.09	0.44	5.32	0.41	5.92	−1.01
bin 0	*92,798*	−0.04	−0.78	0.14	0.66	−0.02	−0.22	1.00	−0.80
bin 1	12,343	1.38	−1.20	5.56	0.65	−5.02	−0.24	−4.44	−0.33
bin 2	2,762	2.38	−0.14	6.25	0.45	−6.57	−0.40	−5.92	−0.25
bin 3	1,133	4.14	−0.45	7.16	0.23	−6.94	−0.12	−6.44	−0.38
all	124,833	0.00	−0.67	0.05	0.62	0.00	−0.13	0.89	−0.77

Table IA.2: Trading flows on days of high adverse-selection risk

The table reports panel regressions of the trading share of retail investors, high-frequency institutions (HFT), and low-frequency institutions (LFT) on proxies for adverse-selection risk. Trading share is defined as the ratio of each group's buying plus selling volume to total share volume. The explanatory variables are dummy indicators for periods of heightened adverse-selection risk: (i) *Abnormal Volatility*, equal to one when the average daily percentage price range over days $t-5$ to $t-1$ exceeds twice the average computed over days $t-60$ to $t-1$; (ii) *Earnings Announcement*, equal to one on days when quarterly results are disclosed, with after-hours announcements attributed to the next trading day; (iii) *Material Fact Release*, equal to one on days when firms issue mandatory corporate disclosures, with after-hours releases similarly shifted to the next trading day; (iv) *Intense News Activity*, equal to one when the number of Bloomberg news items over days $t-5$ to $t-1$ is more than twice the average over days $t-60$ to $t-1$; and (v) *Abnormal Bid–Ask Spread*, equal to one when the stock's bid–ask spread on day t exceeds its own trailing 60-day mean by more than two standard deviations. Columns (1)–(5) report univariate specifications, while column (6) includes all five adverse-selection proxies jointly. All regressions include stock fixed effects; specifications that use time-varying adverse-selection proxies also include date fixed effects, as indicated by the underlying regressions. Standard errors are clustered at the stock level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Retail investors						
Abnormal Volatility	2.463***					5.208***
	(0.824)					(1.510)
Earnings Announcement		1.511***				1.416***
		(0.146)				(0.146)
Material Fact Release			0.891***			0.665***
			(0.223)			(0.205)
Intense News Activity				0.996***		0.860***
				(0.189)		(0.178)
Abnormal Bid–Ask Spread					1.962***	1.500***
					(0.292)	(0.311)
Constant	17.495***	17.510***	17.501***	17.471***	17.395***	17.238***
	(0.013)	(0.003)	(0.009)	(0.012)	(0.021)	(0.032)
Fixed effects	yes	yes	yes	yes	yes	yes
Adj. R2	56.23	61.00	56.19	61.01	61.09	61.27
Obs.	124884	124884	124884	124884	124884	124884

Continuous next page...

... Continuous from previous

Panel B: Market-making institutions						
Abnormal Volatility	-9.999***					-9.489***
	(0.920)					(0.929)
Earnings Announcement		-1.328***				-1.228***
		(0.183)				(0.186)
Material Fact Release			-0.820***			-0.714***
			(0.195)			(0.194)
Intense News Activity				-0.912***		-0.648***
				(0.214)		(0.200)
Abnormal Bid–Ask Spread					-1.599***	-0.785***
					(0.296)	(0.232)
Constant	39.922***	39.782***	39.791***	39.818***	39.874***	40.060***
	(0.015)	(0.003)	(0.007)	(0.014)	(0.021)	(0.025)
Fixed effects	yes	yes	yes	yes	yes	yes
Adj. R2	32.53	31.86	31.85	31.87	31.93	32.61
Obs.	124884	124884	124884	124884	124884	124884
Panel C: LFT institutions						
Abnormal Volatility	3.298**					3.696***
	(1.290)					(1.328)
Earnings Announcement		-0.145				-0.158
		(0.170)				(0.170)
Material Fact Release			-0.060			0.085
			(0.219)			(0.205)
Intense News Activity				-0.126		-0.235
				(0.173)		(0.170)
Abnormal Bid–Ask Spread					-0.422	-0.725***
					(0.273)	(0.247)
Constant	40.863***	40.919***	40.919***	40.924***	40.947***	40.923***
	(0.021)	(0.003)	(0.008)	(0.011)	(0.020)	(0.026)
Fixed effects	yes	yes	yes	yes	yes	yes
Adj. R2	38.85	38.79	33.06	38.79	38.80	38.86
Obs.	124884	124884	124884	124884	124884	124884

Table IA.3: Equally weighted day-zero measures by order type and imbalance intensity

This table reports daily summary statistics for retail order imbalances from 2017 to 2022. Columns under "Buying" use stock-days with positive retail net flow; columns under "Selling" use stock-days with negative retail net flow. "Light," "Moderate," and "Heavy" correspond to terciles of the absolute retail net flow computed within day. Panel A reports day-zero returns (in basis points) by order type: all orders, passive orders, aggressive orders, and other orders. Panel B reports the half bid–ask spread (in basis points). Panel C reports day-zero returns relative to non-retail investors (in basis points), computed using the average transaction price of non-retail investors on the same stock-day.

		Buying				**Selling**			
	N	Mean	Pct.25	Pct.50	Pct.75	Mean	Pct.25	Pct.50	Pct.75
Panel A: Day-zero returns									
All orders									
All	1,485	-21.90	-51.98	-18.31	9.52	-24.28	-51.70	-24.65	5.31
Light	1,485	0.91	-29.69	1.85	30.10	-10.32	-34.89	-9.36	15.71
Moderate	1,485	-17.35	-50.46	-15.43	15.40	-20.86	-50.14	-20.27	10.13
Heavy	1,485	-48.15	-84.26	-42.56	-7.72	-41.01	-75.96	-40.58	-3.17
Passive orders									
All	1,483	-14.29	-43.83	-11.47	15.44	-25.58	-53.60	-25.48	5.31
Light	1,483	14.37	-16.16	11.57	41.71	-8.11	-35.96	-7.72	22.45
Moderate	1,483	-7.98	-39.01	-5.90	21.87	-19.72	-51.82	-20.56	13.14
Heavy	1,483	-47.93	-83.98	-44.26	-9.07	-47.77	-85.19	-47.18	-9.82
Aggressive orders									
All	1,483	-5.00	-35.05	-1.15	27.16	-12.27	-44.03	-12.84	21.27
Light	1,483	2.28	-28.36	3.21	34.45	-13.85	-42.79	-12.76	18.41
Moderate	1,483	-4.71	-37.92	-3.45	31.12	-12.84	-45.58	-12.49	22.47
Heavy	1,483	-12.18	-50.37	-7.06	30.58	-10.17	-51.20	-11.33	29.30
Other orders									
All	1,483	4.13	-9.84	4.16	18.71	-0.27	-9.10	0.27	10.44
Light	1,483	5.43	-10.95	2.79	21.11	-3.68	-13.07	-1.17	9.16
Moderate	1,483	3.87	-10.77	2.42	17.75	0.88	-7.68	0.92	10.40
Heavy	1,483	3.23	-13.22	4.14	22.29	1.88	-9.27	2.57	14.52
Panel B: Bid–ask spread									
All	1,482	7.53	5.21	6.59	8.51	7.85	5.65	6.77	9.48
Light	1,482	9.50	5.27	6.60	10.30	10.62	5.95	7.55	13.73
Moderate	1,482	7.17	4.93	6.23	8.11	7.33	5.56	6.71	8.21
Heavy	1,482	6.08	4.43	5.23	6.70	5.71	4.74	5.51	6.28
Panel C: Day-zero returns relative to non-retail investors									
All	1,485	0.63	-5.12	1.13	6.69	3.66	-0.66	3.67	8.29
Light	1,485	5.14	-2.95	5.10	12.41	5.56	-0.32	5.76	11.53
Moderate	1,485	0.74	-6.66	1.03	7.99	3.95	-2.16	3.79	9.81
Heavy	1,485	-3.77	-10.55	-2.45	4.09	1.54	-4.79	2.00	8.16

Table IA.4: Day-zero returns: Using the LDB dataset

The table presents descriptive statistics of the aggregated daily day-zero returns (in basis points) from 1991-1996 for the most liquid stocks (i.e., We rank stocks every year based on their median trading volume, and keep only the top 1000.). The day-zero buying return is $Ret^b_{i,0} = (ClosingPrice_{i,0} - BuyingPrice_{i,0})/BuyingPrice_{i,0}$, where $ClosingPrice_{i,0}$ is the closing auction price of stock i at the end of the trading day, and $BuyingPrice_{i,0}$ is the volume-weighted average price paid for stock i by all retail investors during that day. Similarly, the day-zero selling return is $Ret^s_{i,0} = (SellingPrice_{i,0} - ClosingPrice_{i,0})/SellingPrice_{i,0}$, where $SellingPrice_{i,0}$ is the volume-weighted average selling price of stock i by retail investors. In Panel A, the daily day-zero returns are daily equally-weighted average across all stock-day returns; in Panel B, we weight stock-day returns by their corresponding buying or selling volume imbalances. In "Buying with net flow > 0," we include only days with more buying than selling, and equivalently for "Selling with net flow < 0," only days with more selling than buying. Additionally, we group stock-day purchases into terciles according their corresponding net flows to calculate daily day-zero return averages for "Light buying," "Moderate buying," and "Heavy buying" stock-days, with a similar approach applied to stock-day sales. Returns are in basis points.

	N	Mean	Pct25	Pct50	Pct75
Panel A: Equally weighted					
Buying with net flow > 0	1,497	-22.55	-52.45	-21.14	9.58
Light buying	1,497	-16.13	-48.36	-17.33	15.84
Moderate buying	1,497	-21.14	-54.35	-19.21	17.08
Heavy buying	1,497	-30.22	-69.49	-26.32	14.50
Selling with net flow < 0	1,497	-53.63	-73.44	-41.62	-11.74
Light selling	1,497	-48.54	-75.68	-39.31	-5.32
Moderate selling	1,497	-51.67	-73.68	-37.63	-6.19
Heavy selling	1,497	-60.59	-83.59	-42.92	-6.90
Panel B: Volume weighted					
Buying with net flow > 0	1,497	-34.46	-72.12	-28.96	11.83
Light buying	1,497	-17.13	-48.58	-16.90	18.45
Moderate buying	1,497	-22.27	-54.91	-19.51	18.30
Heavy buying	1,497	-37.16	-78.48	-31.23	16.07
Selling with net flow < 0	1,497	-60.96	-83.66	-42.19	-3.53
Light selling	1,497	-46.27	-74.00	-36.68	-2.10
Moderate selling	1,497	-52.90	-74.33	-38.08	-6.04
Heavy selling	1,497	-62.87	-86.95	-40.81	-1.64

Table IA.5: First-Stage IV Specification

This table presents the first-stage regression 3 of the $logZ^{Bonds}_{s,t}$ on the retail aggregate volume $log(Retail)$ without time-varying controls in column (1). Column (2) to (6) sequentially adds the lagged control vector $X_{s,t-1}$: past returns and return volatility over the previous trading week, $Ret_{s,t-1:t-5}$ and $RetSd_{s,t-1:t-5}$, and past returns and return volatility over the prior three weeks, $Ret_{s,t-6:t-21}$ and $RetSd_{s,t-6:t-21}$. Columns (3)–(4) replace total retail volume with its channel decomposition and include the same controls; column (4) additionally controls for $AdverseRisk_{s,t}$ ("Adverse Days"), a dummy that equals one on days with elevated adverse-selection risk as defined in Table 3 (abnormal volatility, abnormal bid-ask spread, earnings announcements, material fact releases, or intense news activity), $\log(Volume)_{s,t-1}$ (log total market volume on the previous day). All specifications include stock and date fixed effects. Standard errors are clustered at the stock-date level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable:	log(*Retail*)						
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables							
$log(Z^{Bonds})_{jt}$	0.203***	0.205***	0.200***	0.202***	0.202***	0.205***	0.087***
	(0.046)	(0.046)	(0.045)	(0.046)	(0.045)	(0.045)	(0.023)
$\text{Ret}_{j,t-1:t-5}$		0.691***	0.466***	0.054	0.156	0.162*	-0.039
		(0.151)	(0.171)	(0.142)	(0.094)	(0.090)	(0.036)
$\text{Ret Sd}_{j,t-1:t-5}$			11.4***	11.6***	7.01***	6.55***	1.01
			(1.56)	(1.57)	(0.903)	(0.930)	(0.709)
$\text{Ret}_{j,t-6:t-21}$				0.417***	0.321**	0.307**	-0.110
				(0.114)	(0.126)	(0.127)	(0.078)
$\text{Ret Sd}_{j,t-6:t-21}$					9.04**	9.07**	7.79***
					(4.29)	(4.32)	(2.82)
Adverse Days_{jt}						0.169***	0.133***
						(0.019)	(0.012)
$\text{Log(Volume)}_{j,t-1}$							0.647***
							(0.026)
Fixed-effects							
date	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ticker	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics							
Observations	113,582	113,582	113,582	113,582	113,582	113,582	113,582
Adjusted R^2	0.75275	0.75339	0.76343	0.76398	0.76687	0.76812	0.84991
F-test (1st stage)	6,277.1	6,394.6	6,294.1	6,456.3	6,519.7	6,754.5	1,776.2
Wald (1st stage)	19.532	19.683	19.429	19.584	19.874	20.461	14.595

Table IA.6: Retail and Market Quality Components

This table reports second-stage estimates of the IV system in equations (5)–(6). The dependent variables are the four components of the market quality index. *Price impact* is the stock-day average of the absolute ten-minute return (in basis points) divided by traded volume (in USD millions). *Bid-ask spread* is the average daily quoted spread extracted from Bloomberg. *Absolute autocorrelation* is the absolute value of the first-order autocorrelation of ten-minute returns at the stock-day level. *Volatility ratio* is the absolute deviation of the ratio of one-hour to ten-minute realized variance (scaled to the same horizon) from one. Panels A–D each instrument a different order-type volume while the remaining two enter as OLS controls: Panel A instruments $\log(Retail)_{s,t}$, Panel B instruments $\log(Passive)_{s,t}$, Panel C instruments $\log(Aggressive)_{s,t}$, and Panel D instruments $\log(Other)_{s,t}$. The instrument in all panels is the shift-share bond platform measure $\log(Z^{Bonds}_{s,t})$ defined in equation (3). The Wald statistic is the Kleibergen-Paap rk Wald statistic. All specifications include cumulative returns and return volatility over the prior week and prior three weeks, lagged market volume, an adverse-days dummy, and stock and date fixed effects. Standard errors are two-way clustered by stock and date and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Price Impact (1)	B.A.S (2)	Abs. Auto Corr. (3)	Vol. Ratio (4)
Panel A: Total Retail				
$\log(\widehat{Retail})_{s,t}$	−16.25*	−21.82**	−1.032	0.0020
	(9.325)	(10.17)	(1.109)	(0.0065)
Observations	113,582	113,582	113,582	113,582
Adjusted R^2	0.3498	0.4123	0.0510	0.1151
Wald (1st stage)	14.595	14.595	14.595	14.595
Panel B: Passive Retail				
$\log(\widehat{Passive})_{s,t}$	−45.13*	−67.98**	−1.181	0.0755**
	(26.61)	(32.26)	(3.383)	(0.0291)
Observations	113,582	113,582	113,582	113,582
Adjusted R^2	0.3544	0.4127	0.0530	0.1480
Wald (1st stage)	11.964	11.964	11.964	11.964
Panel C: Aggressive Retail				
$\log(\widehat{Aggressive})_{s,t}$	211.3	329.6	0.5667	−0.4862
	(285.1)	(452.0)	(16.84)	(0.6492)
Observations	113,582	113,582	113,582	113,582
Adjusted R^2	0.3564	0.4130	0.0539	0.1538
Wald (1st stage)	0.5965	0.5965	0.5965	0.5965
Panel D: Other Retail				
$\log(\widehat{Other})_{s,t}$	−64.42	−98.92	0.0561	0.1456
	(53.52)	(72.82)	(5.153)	(0.0905)
Observations	113,582	113,582	113,582	113,582
Adjusted R^2	0.3561	0.4125	0.0536	0.1546
Wald (1st stage)	2.6620	2.6620	2.6620	2.6620