The Retail Habitat

Toomas Laarits[*] Marco Sammon [†]

First version: May 9, 2022
This version: March 10, 2024

Abstract

Retail investors trade hard-to-value stocks. Stocks with a high share of retail-initiated trades are composed of more intangible capital, have longer duration cash-flows and a higher likelihood of being mispriced. Consistent with retail-heavy stocks being harder to value, we document that such stocks are less sensitive to earnings news, more sensitive to retail order flow and are especially expensive to trade around earnings announcements. Additionally, the well-known earnings announcer risk premium is limited to low retail stocks only. Overall, our findings document a new dimension of investor heterogeneity and suggest a comparative advantage of retail in trading hard-to-value stocks.

[*]New York University, Stern School of Business, tl1221@stern.nyu.edu.

[†]Harvard Business School, msammon@hbs.edu. We thank Erik Stafford, Yakov Amihud, Luis Viceira, Robert Battalio and Alexi Savov as well as the seminar audience at Dartmouth and NYU Stern for comments and feedback on the paper. We thank Niels Gormsen, Eben Lazarus, and Andrei Golubov for sharing data.

1 Introduction

The cross-section of asset returns remains the principal venue for discriminating between theories of risk and return. With the well-established failure of the CAPM, a plethora of factors has been been suggested to account for cross-sectional variation in returns instead. A closely related research agenda concerns the portfolios of different types of investors. Reflecting the failure of the CAPM, prior research has documented substantial heterogeneity in investor portfolios going well beyond differences in the share devoted to the market portfolio as prescribed by the CAPM. On a theoretical level, existing work has emphasized various investor characteristics such as risk aversion, elasticity of intertemporal substitution, investment horizon, financial sophistication, and background risk as important drivers of portfolio choice (Curcuru et al., 2010). More practically, an active empirical strand of work starting with Koijen and Yogo (2019) focuses on the identity of the investor—for instance banks, insurance companies, pension funds, investment advisors, and so on—and on each group's demand for assets with various characteristics.

In this paper we argue that the contrasting features of institutional and retail investors offer one particularly useful dimension to get a handle on such cross-sectional differences in trading and holdings. Institutional investors are frequently conceptualized in the finance literature as "smart money": they are typically much larger than any individual investor and therefore have the scale to acquire and process various kinds of data. Their scale also allows them to better overcome the fixed costs of trading, to take on leverage and short assets. Retail investors, by contrast, are often seen as suffering from a litany of behavioral biases and cognitive errors, as less equipped to carry out meaningful research, and hence conceptualized as mere "noise" traders in the sense of Kyle (1985). All the while, retail investors have advantages of their own. As they are investing their own money, retail investors have more control over the investment horizon and they do not face flow sensitivity to recent performance. They are also not constrained by mandates restricting the investable set or tying their compensation to performance relative to a specific benchmark.

Indeed, a number of empirical findings cut against a pure "noise trader" view of retail investors. A prominent strand of work found that retail trades on the aggregate predict returns going forward with a positive sign.[1] The pandemic era surge in retail trading (Welch, 2022), and the response of the stock market to stimulus checks (Greenwood et al., 2023) serve as a reminder of the potentially large effect of retail traders on asset markets. On the other side of the retail-institutional divide, recent work in Di Maggio et al. (2019) documents that institutional investors trade away from stocks that are about to make earnings announcements, at times where their presumptive informational advantages ought to be the strongest.

These findings, taken together, suggest that a conceptualization that goes beyond an informed *versus* uninformed dichotomy is required to account for the holdings and trading patterns of retail

[1] See Kaniel et al. (2008), Barrot et al. (2016), Boehmer et al. (2021), among others.

and institutional investors. Our argument is that the specific nature of the stocks traded by retail and institutional investors plays a critical role. Specifically, we argue that the above empirical findings reflect the tendency of retail investors to trade stocks that are *hard-to-value*: stocks whose value is weakly tied to fundamentals, for instance, because they have long duration cashflows or have high shares of intangible capital.

To motivate this idea, we provide a model in the spirit of Kyle (1985). The model features an informed institutional investor with the capacity to research and generate information about one stock in the economy. The institutional investor is endowed with an attention budget, which they can use to receive a more precise signal about the value of the stock they specialize in. Retail order flow is modelled as unpredictable and unrelated to fundamentals, but retail traders' behavior reflects an element of sophistication in that their trading intensity is allowed to depend on stock characteristics.

Where in the cross-section should this informed investor choose to produce information? At first blush it may seem that a stock with high retail trader presence would be the best bet, as the informed investor will find it easier to hide their trades among the retail investors' order flow. This logic only holds, however, if retail investors' propensity to trade is equal across stocks. Suppose, however, that retail investors tend trade in hard-to-value stocks, defined as stocks for which the informed investors' research would yield the lowest amount of incremental information. Then, even conditional on a high level of noise-like trading activity, the informed investors may want to avoid learning about such stocks, as their expected profits would be lower on account of a weaker informational advantage. Overall, the model illustrates that where informed investors choose to trade depends on the relative strength of these two aspects: the intensity of retail trading, and the propensity of retail traders to focus on hard-to-value stocks.

Our empirical analysis builds on this insight and documents four sets of findings. The first part of our analysis establishes new facts on the distribution of retail trading in the cross-section. We show that retail trading intensity is both concentrated and persistent, consistent with our suggestion that retail traders occupy a specific segment of the market. In terms of magnitudes, almost 90% of stocks in the top 20% of retail trading intensity at any given point in time are still in the top two quintiles of retail trading intensity 12 months later.

Having documented a persistent retail focus on a subset of stocks, we aim to test which of the two forces suggested by the theoretical framework are more important empirically. We find that the concept of difficult-to-value—defined as the incremental price signal gained from researching fundamentals—-is a particularly useful summary characteristic for explaining the cross-sectional heterogeneity in retail trading intensity: retail investors trade hard-to-value stocks. We employ three different proxies of difficult-to-value to quantify this relationship. Firstly, we examine cash-flow duration (Gormsen and Lazarus, 2023). We believe firms with longer duration cashflows are harder to value because investors need to forecast fundamentals further in the future. Secondly, we examine various measures of intangible capital (Peters and Taylor (2017), Kogan et al. (2017)), which by nature is harder to value than physical capital (see Lev and Gu (2016)). Thirdly, we examine two composite measures, the mispricing score of Stambaugh and Yuan (2017) and the valuation uncertainty score of Golubov and Konstantinidi (2021). Across all three sets of measures we find that hard-to-value stocks see higher retail trading intensity: high retail stocks have longer duration cashflows, more intangible capital, more expected mispricing and more valuation uncertainty. Nearly all of these relationships hold both within the bottom and top 20% of firms by market capitalization, suggesting the relationship between retail investor activity and firm size is not driving our results. Combined with the evidence on retail trading persistence, these results establish a new way to capture the trading and ownership patterns of retail *versus* institutional investors.

In the second set of results, we turn to the implications of retail investors' concentration in hard-to-value firms. Our main focus is on earnings announcements, as these are times where fresh value-relevant information is released, which could resolve some of the *ex ante* valuation uncertainty. We find that high retail stocks have more volatile announcement news and returns. The standard deviation of standardized unexpected earnings (SUE) is more than three times as large for high retail stocks than low retail stocks. We also find that the dispersion in analysts' forecasts for high retail stocks is almost five times as large as for low retail stocks. What is more, at the individual analyst level, analysts produce more inaccurate forecasts for high retail stocks, conditioning on a measure of an analyst's attention constraints, and even when controlling for heterogeneity in analyst skill and firm characteristics. All these results are in line with our claim that retail traders concentrate in hard-to-value stocks.

To provide further evidence on the differences between the stocks heavily traded by retail and institutional investors, we argue that because current earnings should be less price relevant for firms with long-duration cashflows or significant amounts of intangible capital, such firms will be less sensitive to fundamental news. To test this prediction, we estimate earnings-response regressions following Kothari and Sloan (1992), and find that for a given magnitude earnings surprise, high retail stocks' prices respond significantly less to earnings news than low retail stocks. Specifically, a stock in the highest quintile in terms of past retail trading share has a roughly 40% lower sensitivity to standardized unexpected earnings news than a stock in the middle quintile. This effect is unchanged by controlling for a litany of characteristics known to be correlated with retail activity and holds at almost every point along the firm size distribution.

In the third set of results, we ask whether retail trades themselves influence stock prices, particularly around earnings announcements. Broadly, retail-heavy stocks see substantial abnormal retail-initiated trading around earnings announcements, where abnormal trading is defined as retail volume minus average retail volume in that stock over the past year. We find that in the pre-announcement period, retail trading makes up about an additional 1.3% of total volume in high retail stocks, relative to the average stock. Retail investors' trading, however, is not larger relative to shares outstanding, evidence that institutional investors tend to avoid trading such stocks before earnings announcements. We also find that retail traders as a group increase their holdings of such stocks in anticipation of scheduled announcements and hence hold a disproportionate share of the earnings news risk of such stocks. This result adds to the phenomenon documented in Di Maggio et al. (2021) by showing that institutions' tendency to exit positions ahead of earnings announcements has significant heterogeneity across the retail sort: it is strongest for the stocks that retail tends to trade the most before the announcement itself. Our results thus offer an explanation for why institutions tend to exit high-retail stocks ahead of earnings announcements: they understand that hard to value stocks have volatile and idiosyncratic earnings-day returns, and want to avoid exposure to these risks.

The earnings announcement setting also provides a venue to study the question of trading costs across the retail sort. Based on the logic of Kyle (1985), if retail investors act as noise traders, one might expect transaction costs to be relatively lower in high retail stocks. In addition, if institutions avoid learning about these stocks – e.g., because they're hard to value – we would also expect lower transaction costs due to lower risk of adverse selection. On the other hand, if retail investors do have an informational advantage, one might expect higher transaction costs in such stocks. In addition, due to the persistence of retail order flow, it may be risky to bet against retail trades, even if these trades are unrelated to firm fundamentals, which would also lead to higher expected transaction costs.

Earnings announcements are a natural laboratory to study these competing forces, as they are a time when adverse selection risk is especially salient (Krinsky and Lee, 1996). Empirically, we find that high retail stocks have abnormally high bid-ask spreads in a 5-day window around earnings announcements, relative to the stock-level average over the past month. In terms of magnitudes, the abnormal effective spread is 4 basis points higher on the earnings announcement day itself for high retail stocks relative to the average stock. For reference, this increase is $2/3^{rds}$ the size of the average value-weighted bid-ask spread in 2021 of 6 basis points (Greenwood and Sammon, 2022). This alone, however, does not pin down the mechanism of why transaction costs are higher around earnings announcement for high retail stocks.

To better understand whether retail investors have information about fundamentals, we aim to understand whether the *direction* of retail order flow predicts earnings surprises. On the announcement level, we find strong evidence that stocks heavily bought by retail investors in anticipation of earnings news releases outperform stocks heavily sold by retail investors. This pattern, previously documented by Kaniel et al. (2012), is particularly pronounced for stocks with high past average retail trading, again emphasizing the active role of retail traders in incorporating information into security prices. In a decomposition exercise we attribute about half of this predictable outperformance to liquidity provision and the other half to potential private information. The strong role for retail investors' private information is evidence that concerns about adverse selection may explain the higher trading costs in high retail stocks – and that this effect dominates the expected effects of retail investors being pure "noise" traders.

In the fourth and final set of results, we link the earnings announcer premium to retail trading intensity. As shown in a long literature staring with Beaver (1968), stocks tend to earn high average returns when they are scheduled to make earnings announcements. A potential explanation for the earnings announcer premium is that announcing firms provide information about non-announcing firms and therefore the premium is compensation for exposure to systematic risk, as argued in Savor and Wilson (2016). Our prior, therefore, is that this premium is unlikely to be earned by high retail stocks, as their earnings announcements are mostly comprised of idiosyncratic information. This is precisely what we find. In a three trading day window starting with the announcement, we document an earnings announcement premium of 18bps for stocks in the top quintile of market capitalization. However, among this set of large stocks, those in the highest retail trading quintile see an average return of negative 18.5bps over the same time window. The earnings announcer premium is also depressed for high retail stocks among all the other size quintiles. In all, we find that unlike on the aggregate, holding high retail stocks through earnings announcements earns no risk premium, in line with these stocks' announcement returns representing idiosyncratic news.

Our work contributes to an active strand of research that has highlighted the importance of investor heterogeneity and less-than-perfect risk-sharing in determining the risk-return trade-off in security prices. One part of this work seeks to estimate demand curves of different investor classes as functions of various characteristics (Koijen and Yogo (2019), Koijen et al. (2020), McLean et al. (2020), Haddad et al. (2021), van der Beck (2022)). Our work documents a new point of distinction in the trading habits of two principal investor classes: retail and institutional investors. Other recent work in Balasubramaniam et al. (2023) and Gabaix et al. (2022) has studied the portfolios of retail investors specifically. Balasubramaniam et al. (2023) use account-level data from India to document the role of characteristics in attracting retail holdings. They find that firm age and nominal price, and, to a weaker degree, turnover and recent returns are the characteristics that best capture the heterogeneity in retail holding intensity. Our aggregate retail trading data is consistent with a retail focus on firm age and nominal price, as well as turnover and past returns, while pointing to a unifying strand underlying these regularities.

Outside of that recent work, the literature on retail investors has devoted surprisingly little attention to the determinants of retail trading and holdings in the cross-section. Most of the existing literature has focused on various behavioral frictions that bring stocks to the attention of retail investors. However, we find that there is substantial and persistent cross-sectional heterogeneity in retail trading intensity, and it can be explained by a metric which is not obvious from looking at past returns, betas, or accounting figures alone. Our results add to this literature by suggesting that difficult-to-value stocks attract particular retail attention or, equivalently, repel institutional investors.

Indeed, this aspect of retail selection allows us to reconcile two broad, seemingly contradictory aspects of retail investing. On one hand, research has repeatedly found that retail trades – on aggregate – tend to positively predict stock returns going forward. For example, Kaniel et al. (2012) show that the direction and magnitude of retail order flow predicts returns on and after earnings announcements. Along the same lines, in more recent work, Welch (2022) documents that Robinhood investors as a group did well in 2020-21.[2] On the other hand, retail traders have been shown to suffer from a litany of behavioral biases including: excessive trading (Barber and Odean (2000), Barber and Odean (2002)), familiarity bias (Huberman (2001), Seasholes and Zhu (2010)), extrapolation (Benartzi, 2001) and the disposition effect (Odean (1998), Dhar and Zhu (2006), Vaarmets et al. (2019)), to name a few. Moreover, relaxation of retail investors' budget constraints sees the prices of retail-heavy stocks rally (Greenwood et al., 2023). Because of the selection by retail traders into hard to value stocks, these biases and predictable errors are particularly hard for professional investors to correct.

More broadly, our results can be used to recast several existing results in the asset pricing literature by emphasizing how the relative importance of two types investors can directly contribute to these phenomena. First, previous literature has shown significant effects of retail investor buying on stock prices (Kumar and Lee (2006), Greenwood et al. (2023)). Our results on the concentration of retail investor trading, as well as the types of stocks preferred by retail, may explain why retail investors can have such a large effect on prices despite their relatively small share of overall stock market wealth. Second, the focus on hard-to-value stocks can explain why retail order flow is a strong predictor of returns going forward, as documented in Kaniel et al. (2012). In fact, we show that such predictability is particularly pronounced within the set of high retail share stocks . Given that retail order flow is persistent, and that retail investors focus on stocks which are relatively more expensive to trade, it may be difficult for institutional investors to maintain bets against retail order flow long enough to benefit from long-run reversion. Finally, we show that the stocks which retail investors tend to favor have high mispricing scores (Stambaugh and Yuan, 2017), suggesting they are often in the extreme ends of anomaly portfolios. This opens the door for retail investors to directly contribute to anomaly returns, as retail investors' trading in these stocks makes it tougher for institutional investors to try to correct any mispricing.

2 Hypothesis development

In this section we outline a model and three predictions that guide our empirical exercises.

[2] The evidence for retail investors' trading performance in options mixed, with Bryzgalova et al. (2023) finding that retail investors lose money on average, while Bogousslavsky and Muravyev (2024) argue average losses are small, and investors may use options as a relatively less expensive way to access leverage.

2.1 Motivation

Consider a model in the spirit of Kyle (1985) with multiple securities and two periods. There are gains to specialization, so an informed insider, representing institutional investors, can only learn about one stock, and each investor has a fixed total attention constraint. The securities themselves are heterogeneous in two ways: (A) the level of noise trading intensity, standing in for differences in the intensity of retail trading, and (B) the amount of effort the insider has to expend to get a signal of a given precision, standing in for the difficulty of valuing the stock. Within this model we ask: where in the cross-section would the institutional investor find it most profitable to produce information?

Reflecting standard intuition, all else equal, the insider's profit will be larger in stocks with higher noise trading intensity. This effect is amplified in the two-period Kyle model we consider, as the insider has an additional opportunity to hide their order flow with noise trading activity before uncertainty is resolved. One might expect, then, that institutional investors expend most of their attention learning about stocks with more retail trading activity.

What this line of argument misses, however, is that retail trading activity need not be equally distributed in the cross-section of stocks. Instead, retail trading may be most concentrated in stocks where insiders have the worst quality of information, which in the model are the stocks for which a given amount of effort results in a relatively less precise signal. If retail investors concentrate their trading in such hard-to-value stocks then high retail stocks might offer worse overall expected profits to the insiders, as they face poor enough signal precision to outweigh the expected benefits of hiding their trade among retail order flow.

The two-period Kyle (1985) model described in Appendix A.1 allows us to make this point explicitly. We simulate the model and plot the insider's profit as a function of signal precision and noise trading intensity. Appendix Figure A1 shows that, unsurprisingly, the insider's profit is monotonically increasing in both signal precision and noise trading activity. The more surprising result is that the insider's profit can be lower in a high noise trading intensity stock than a low noise trading intensity stock, if the precision of their signal is sufficiently higher in the low noise trading intensity stock.

In our baseline version of the model, retail investors order flow is uncorrelated with securities' terminal payoffs. The retail investors exhibit some sophistication, though, in that they pick particular stocks in the cross section.[3] The results from the calibrated two-period model therefore suggest that which of these forces dominates—hiding among noise traders vs. precision of signals—is an empirical question, and in this Section, we outline specific predictions motivated by these two

[3] There is evidence that retail investors' trades may contain information about future fundamentals (Kaniel et al. (2008), Barrot et al. (2016), Boehmer et al. (2021)). In the case where retail investors don't just focus on hard-to-value stocks, but have some signal, the insider would be even less inclined to learn about high retail stocks, as this would further erode their informational advantage (Aase et al., 2011).

competing forces.

2.2 Cross-sectional heterogeneity in retail trading intensity

Motivated by the model, we first seek to establish which of these two forces —hiding among retail order flow vs. precision of signal—dominates. To this end, we employ a number of proxies for difficult-to-value and summarize them across the retail sort. Across various proxies mentioned in the introduction, such as cash-flow duration (Gormsen and Lazarus, 2023), intangible capital (Peters and Taylor (2017), Kogan et al. (2017)), and presence in mispricing portfolios (Stambaugh and Yuan, 2017) or valuation uncertainty (Golubov and Konstantinidi, 2021), we find that stocks with a higher share of retail trading are harder to value. We find this result in unconditional sorts, as well as double sorting on size and the difficulty-to-value proxy. Having shown that stocks heavily traded by retail can be summarized by the concept of difficult-to-value, we turn to specific predictions on earnings announcements, on trading costs, and on risk premia.

2.3 Predictions on Earnings Announcements

Given the difficulty in forecasting the fundamentals of hard to value firms, they are more likely to have large earnings surprises, and therefore larger earnings-day returns (Golubov and Konstantinidi, 2021). Further, we might expect that because such firms are hard to value, there is more dispersion in analysts' earnings forecasts (Diether et al. (2002), Zhang (2006)). This implies the following testable predictions:

Prediction 1A: *High retail stocks should have more volatile earnings-day returns and earnings news. In addition, high retail stocks should have more dispersion in analysts forecasts. Finally, their earnings surprises should be mostly driven by the idiosyncratic component of earnings news.*

A natural concern with testing Prediction 1A is that there is selection in terms of which types of analysts cover high retail stocks and low retail stocks. If, for example, low quality analysts cover high retail stocks, such stocks may have larger earnings surprises even though they are just as hard to value as low retail stocks. In Section 5, we develop a test of prediction 1A which compares accuracy *within the stocks a given analyst covers*, allaying these selection concerns.

Further, if high retail stocks are hard to value, any news about current cashflows will have a relatively smaller effect on prices. The logic is that for firms with long duration cashflows, or a significant amount of their value in intangible capital, current earnings are not as relevant for total present value. Additionally, in hard to value stocks, different investors may focus on different pieces of the news, leading to more disagreement and ultimately to under-reaction (Hong and Stein, 2007). Or, in stocks where prices are not informative, investors may choose to ignore public signals

(Banerjee et al., 2021). Finally, investors may fail to process the news altogether because it requires too much effort to understand (Hirshleifer et al. (2009), Engelberg (2008), Cohen et al. (2020)), which would also manifest as under-reaction. These mechanisms yield the following prediction:

Prediction 1B: *High retail stocks should respond relatively less to earnings news.*

2.4 Retail trading behavior around earnings announcements

Our second prediction regards how retail investors trade around earnings announcements. There is a long literature studying such behavior (Hirshleifer et al. (2008), Kaniel et al. (2012)), but our focus is on differences in retail trading around earnings *conditional on the set of stocks they were previously trading intensely.* If institutional investors are generally unwilling to trade hard to value stocks, they might be especially wary around earnings announcements, given their tendency to have extreme returns and high trading costs. Therefore, we might expect retail investors to become an even larger share of trading volume in such stocks around earnings events. Further, if institutional investors want to reduce their exposure to hard to value firms ahead of earnings announcements, we would expect net buying by retail. Collectively, this implies the following testable hypothesis:

Prediction 2: *High retail stocks should have more abnormal retail trading intensity on and before earnings announcements. During this period, retail investors should be net buyers from institutional investors.*

2.5 Retail trading and the earnings announcement premium

Lastly, we turn to a prediction for the earnings announcer premium. This is motivated by Savor and Wilson (2016), who argue that the premium derives from the information in announcements about non-announcing firms. This mechanism seems unlikely to apply to high retail firms for at least two reasons. If high retail stocks are hard to value, the information contained in a given earnings announcement might not be useful for understanding other firms. Second, if Prediction 1A is true, high retail stocks' earnings news will have a relatively larger idiosyncratic component, which is less useful for valuing non-announcing firms. In either case, we would expect high retail stocks to have a smaller or non-existent announcer premium.[4] This leads to the following testable prediction:

Prediction 3: *High retail stocks should have a lower or non-existent earnings announcement premium.*

[4]Not all evidence, however, points in the same direction. For example, Frazzini and Lamont (2007) shows that the earnings announcer premium is mostly earned in stocks where many small investors are buying. Further, Barber et al. (2013) argues that the announcer premium comes from exposure to idiosyncratic risk to be disclosed and based on Prediction 1A, we expect this risk to be larger in high retail stocks.

3 Data

In this section, we briefly describe our main data sources and variable construction. Our key measure of retail trading activity is $\text{RSVOL}_{i,t}$, the retail share of trading volume, defined as

$$\text{RSVOL}_{i,t} = \frac{\text{RBuy}_{i,t} + \text{RSell}_{i,t}}{\text{Volume}_{i,t}}, \tag{1}$$

where $\text{RBuy}_{i,t}$ and $\text{RSell}_{i,t}$ are the number of shares in retail-initiated buy and sell trades, respectively. $\text{Volume}_{i,t}$ is total daily volume on the TAQ tape. In words, $\text{RSVOL}_{i,t}$ is the fraction of stock i's total trading volume on day t accounted for by retail-initiated buys and sells. We report $\text{RSVOL}_{i,t}$ in percentage terms. In addition to a daily measure of retail-initiated trading, we also construct a monthly counterpart. For each month τ, we sum up the retail-initiated trades $\text{Rbuy}_{i,t}$ and $\text{Rsell}_{i,t}$ as well as total volume $\text{Volume}_{i,t}$ and then construct monthly $\text{RSVOL}_{i,\tau}$ according to Equation 1.

Retail trades are identified using the algorithm proposed in Boehmer et al. (2021) that relies on the regulation of U.S. security markets requiring price improvement for retail-initiated trades that are internalized. Note that $\text{RSVOL}_{i,t}$ will typically be lower than the true fraction of trading coming from retail investors, as the Boehmer et al. (2021) algorithm may fail to classify some retail trades. Indeed, recent work in Barber et al. (2022) argues the Boehmer et al. (2021) algorithm can fail to classify retail-initiated trades, particularly among stocks with large bid-ask spreads. All that matters for most of our findings, however, is that the *ordinal* ranking of stocks on gross retail activity is correct. We construct this measure using the TAQ millisecond data from 2007-2021.[5]

Several papers have questioned the accuracy of the BJZZ algorithm for classifying individual trades (see e.g., Barber et al. (2022) and Battalio et al. (2023)). In most of our applications, we are interested in ranking stocks based on retail trading activity, rather than directly using measures of net or gross retail order flow, mitigating the potential impact of mismeasurement. To further allay these concerns, in Table 2, we show that our rankings based on retail trading intensity are strongly inversely correlated with institutional ownership from 13F data. Yet another way of quantifying retail trading activity at a monthly frequency is by using SEC rule 605 reports filed by wholesalers. In Appendix A.2, we show that our rankings are similar to those based on this regulatory data on internalized retail orders. Finally, in Appendix A.2, we also show our rankings are similar to those based on the number of Robinhood users in Robintrack data.

Our sample consists of all CRSP ordinary common shares that are traded on major exchanges and can be matched to the retail activity data. Specifically, we restrict to share codes 10-11 and

[5] Boehmer et al. (2021) note that from 1/2016-9/2018, the SEC's tick size pilot program likely affected the prevalence of subpenny price improvements.

exchange codes 1-3. For the mapping between TAQ and CRSP identifiers, we use the linking table provided by Wharton Research Data Services (WRDS).

To quantify cross-sectional differences in retail activity, each month, we sort securities into five groups based on retail trading intensity the prior month i.e., $\text{RSVOL}_{i,\tau-1}$. When forming these groups we do not use NYSE breakpoints, as is standard in much of the portfolio formation literature (see e.g., Fama and French (1993)). This is because NASDAQ stocks have more retail activity on average, so by forming NYSE breakpoints, we would be missing an important dimension of retail heterogeneity. Panel A of Figure 1 plots the time series of average $\text{RSVOL}_{i,t}$ in the 1st and 5th quintiles of portfolios sorted on prior month $\text{RSVOL}_{i,\tau}$. This figure shows that there is substantial cross-sectional heterogeneity in retail activity. Specifically, in some stocks, retail investors only account for about 2% of total trading volume while in other stocks they account for over 20%.

For our analysis of how retail investors respond to news, we focus on earnings announcements. To this end, we need to establish the first time investors could have traded on earnings information during normal market hours. We identify these days using the earnings release date and time in IBES. If earnings are released before 4:00 PM Eastern Time on a trading day between Monday and Friday, that day will be labeled as the effective earnings date. If earnings are released on or after 4:00 PM Eastern time between Monday and Friday, over the weekend, or on a trading holiday, the next trading date in CRSP is labeled as the effective earnings date. To be conservative, we instead use the first trading day on or after the release date of quarterly earnings (RDQ) in Compustat if it occurs at least one day before the date identified using IBES (Livnat and Mendenhall, 2006). We use the mapping file from WRDS to link IBES data to CRSP.[6] Because of our focus on earnings announcements we restrict the sample to firms for which we are able to construct earnings expectations.

For a detailed description of all the variables used in our analyses, see Section A.3 of the Appendix.

4 Retail Trading and Stock Characteristics

We document significant cross-sectional dispersion as well as persistence in retail trading activity. We then examine stock-level characteristics that account for this heterogeneity. Our main finding is that stocks favored by retail traders can be characterized as relatively hard to value. Consistent with this, we show that such stocks are more expensive to trade, have more volatile fundamentals and larger magnitude earnings-day returns.

[6] At the start of our sample in 2007, IBES covers 88% of ordinary common shares traded on major exchanges in CRSP. This number declined slightly over time to 84% by 2020. The firms not covered by IBES tend to be smaller and younger on average.

4.1 Retail Trading Intensity in the Cross-Section

There is substantial heterogeneity in the intensity of retail-initiated trading in the cross-section of stocks. In the 2007 to 2021 sample, marketable retail orders identified by the Boehmer et al. (2021) algorithm make up 7.94% of daily total trading volume for the average stock. Our first set of results document that the cross-sectional variability of retail-initiated share of volume, denoted $\text{RSVOL}_{i,t}$, is large relative to its unconditional mean.

To establish this, each month, we sort securities into quintiles based on retail trading intensity the prior month i.e., $\text{RSVOL}_{i,\tau-1}$. In order to quantify differences across the retail sort, we estimate regressions of the form:

$$\text{Outcome}_{i,\tau} = a + \beta_1 1_{i\in Q1_{\tau-1}} + \beta_2 1_{i\in Q2_{\tau-1}} + \beta_4 1_{i\in Q4_{\tau-1}} + \beta_5 1_{i\in Q5_{\tau-1}} + \epsilon_{i,\tau} \tag{2}$$

where $1_{i\in Qj_{\tau-1}}$ are indicator variables for whether stock i was in retail trading intensity quintile j in the previous month $\tau - 1$. The omitted group is the middle quintile of retail trading intensity. Standard errors are double clustered at the stock and month levels and we include monthly fixed effects.

In the first three columns of Table 1 we show the moments of $\text{RSVOL}_{i,t}$, the retail-initated share of trading. The gap in retail trading share between high and low retail stocks is about 13%. The second and third columns restrict the sample to the smallest and largest quintiles in terms of market capitalization, respectively. As the columns show, the gap in retail trading intensity is present for both small and large stocks, with respective sizes of 14% and 10%. In unreported results we confirm that this gap holds at all points in the size distribution.

The other two sets of three columns repeat this analysis for total share turnover and retail-initiated share turnover. Both turnover and retail-initiated turnover are measured as the number of shares traded, normalized by shares outstanding and reported in percentage terms. The gap in share turnover going from low to high retail is about 8%, while the gap in retail-initiated turnover is about 3%.[7] Both of these measures see larger gaps across the retail sort when restricting the sample to small stocks, though the differences across high and low retail stocks are statistically significant in all specifications: we report a formal test of of equality between the coefficients on Q1 and Q5 in the table footer. Again, the differences in turnover and retail initiated turnover are present for both small and large stocks.

The retail sort is persistent over time. In Appendix A.5, Table A4 shows the 12 month transition

[7] This raises the concern, however, that sorting on retail share of trading volume is just another way of sorting on overall turnover. In Appendix A.4 we perform a double sort on overall turnover and retail-initiated turnover. We show that within each portfolio formed on overall turnover, the sub-portfolios formed on retail-initiated turnover look like portfolios sorted on the retail share of trading volume. This implies that the variation in the retail share of trading volume is coming from retail trading itself, not a failure to trade by institutional investors.

probabilities across RSVOL-sorted bins. As Panel A of the Table shows, stocks in the highest quintile in terms of retail share of trading have a 66% probability of remaining in the top quintile 12 months in the future. These same stocks have an almost 90% probability of remaining in one of the top two retail-heavy portfolios.[8]

The time-series dimension of average retail share is illustrated in Panel A of Figure 1. Here we plot the equal-weighted average retail intensity within the top and bottom quintile of past retail intensity. For high retail stocks (Q5), retail investors have become an an increasingly large fraction of trading volume, now at around 20% of total shares traded. For low retail stocks (Q1) retail intensity has been relatively stable at about 2% of total trading.

Retail trading is also more concentrated than trading in general. We illustrate the time-series aspect this tendency of retail trades to be concentrated in Panel B of Figure 1. In this Figure we show the cumulative share of dollar volume stemming from the top 10, 50, and 100 shares in terms of dollar volume in each quarter. The top 100 stocks in terms of retail trading intensity make up over 60% of retail dollar trading volume throughout this period, and close to 80% in the most recent data. The top 10 stocks in terms of retail trading intensity account for over 20% of all retail-initiated volume throughout the sample and over 40% in the most recent years. For reference, the corresponding numbers for all trading volume in the most recent data are roughly 60% and 30%, meaning that retail trading is more concentrated than overall trading.

4.2 Stock Characteristics across Retail Portfolios

The results in the previous section establish substantial heterogeneity as well as a substantial degree of persistence in retail trading intensity.

The first main goal of our paper is to characterize the retail habitat, meaning, to establish which types of stocks tend to attract a lot of retail trading. To this end, we summarize firm characteristics across $\text{RSVOL}_{i,\tau-1}$ quintiles in Tables 2, 3. We group firm characteristics into two thematic groups: fundamentals and valuation (we explore volatility/trading costs in Appendix Appendix A.12).

In Table 2, we present fundamentals across the $\text{RSVOL}_{i,\tau}$ sort. We find that high retail stocks are smaller, younger, have low nominal prices, low recent returns (measured from month -12 to month -2), higher book-to-market ratios and tend to have low or negative earnings yields.[9] Note that the first column reports a median regression to document that the typical firm in the high retail bucket is a small firm. There are a number of very large firms, though, in the high retail quintile and for

[8]Given the persistence of retail activity, a natural question is what leads a stock to transition from the bottom to the top quintile of retail trading intensity. In Appendix A.6, we discuss examples of such stocks like Hertz and First Republic Bank.

[9]These findings are broadly consistent with Kumar and Lee (2006), who find that retail intensity is highest in, "small firms, lower priced firms, firms with lower institutional ownership, and value (high B/M) firms ..." See also Balasubramaniam et al. (2023).

that reason the average firm size is larger in the high retail bucket than in the low retail bucket. As a validation of the Boehmer et al. (2021) algorithm, we include one minus institutional ownership share from Form 13F data in the last column, and show that it is monotonically increasing from the low to high retail quintiles.

In the table footer we formally test for equality between the high retail (Q5) and low retail (Q1) dummy variable coefficients. As the p-values show, all differences bar the CAPM beta are statistically significant. We also report the gap in the Q5 and Q1 dummy values, controlling for size by including dummy variables for five size quintiles. The relationships described above continue to hold, with the with the exception of the difference in B/M ratio that is no longer significant, and past returns that switches signs.

Overall, results in Table 2 establishes firm age and nominal price as important determinants of retail trading interest, reflecting the findings in Kumar and Lee (2006) and Balasubramaniam et al. (2023). That said, there are no substantial differences in the baseline measures of risk: CAPM beta or book-to-market ratio.

In the subsequent Table 3, however, we document substantial differences in various valuation and valuation uncertainty metrics across the retail sort, establishing our first main empirical finding: retail investors tend to more heavily trade stocks that are harder to value. For ease of interpretation we winsorize all measures at the 1% level, and then transform into z-scores, meaning we subtract their mean and divide by their standard deviation (see Appendix A.3 for more details on how these variables are constructed).

The first dimension of difficulty to value is the duration of cash-flows. In the first column of Table 3 we report a proxy for cash-flow duration (CF) constructed after Gormsen and Lazarus (2023). We find that high retail stocks tend to have longer duration cash-flows. Also consistent with high retail stocks being harder to value, high retail stocks have a relatively larger share of their value in intangibles. Specifically, they have more intangible capital (K_{Int}), knowledge capital (K_{Know}), and more organization capital (K_{Org}). The variables are from the Peters and Taylor Total Q dataset (Peters and Taylor (2017)). Further, high retail have more valuable patents, relative to their total market value.[10] Also consistent with high retail stocks being harder to value, they have more valuation uncertainty (denoted VU, from (Golubov and Konstantinidi, 2021)) and higher mispricing scores (Stambaugh and Yuan, 2017).

Just like in the prior table we test formally for the equality of coefficient estimates of high and low retail trading intensity quintiles. In all cases the differences across the retail sort are statistically significant. Also mirroring the prior table, we test for the equality of the Q5 and Q1 dummy coefficients controlling for size by including five size dummy variables. In all cases, the differences

[10]We obtain the market value of patents, PAT, from (Kogan et al., 2017). To compute this metric, we sum the total real dollars of patents over the past 5 years and divide this quantity by a firm's real market capitalization at the end of the current year.

between Q5 and Q1 are statistically significant and remarkably similar to the estimates without size controls.[11]

A potential concern is that the results in Table 3 reflect an industry tilt and in Appendix A.7 we document substantial differences in retail trading intensity across the Fama French 49 industries. The differences in the metrics included in Table 3, however, are robust to controlling for industry. Rather than transforming all the valuation metrics into z-scores across all observations, we instead form z-scores within each Fama-French 49 industry. We find that this only slightly attenuates the differences between high and low retail stocks, suggesting that the retail tilt toward hard to value securities is not driven solely by cross-industry differences.

Overall, the results in Table 3 establish a new fact consistent with Prediction 1: stocks with high shares of retail trading tend to be harder to value.

As an additional test of Prediction 1, in Appendix A.1.4, we present alternative evidence that high retail stocks are hard to value. Specifically, building on the logic of a model with limited attention, we assume that analysts covering more stocks should be relatively less accurate. If a stock is hard to value, however, even conditional on the number of stocks being covered, heterogenity in analyst skill and differences in firm fundamentals, one might expect individual analysts' estimates to be less accurate for high retail stocks. Table A1 shows this is indeed the case – and that the inaccuracy of analysts' estimates is monotonically increasing from the low to high retail portfolios.

So far, we have shown that retail investors tend to favor *trading* hard to value stocks. But, returning to the importance of understanding which types of investors own different stocks (Gabaix et al., 2022), we want to provide evidence that retail investors favor *holding* them as well. First, the last column of Table 2 shows that high retail trading intensity stocks tend to have lower institutional ownership. Because the retail quintiles are formed on trading not ownership, this relationship is not purely mechanical.

To refine this result, we zoom in on one dimension of institutional heterogeneity: total assets under management. To this end, we first sort 13F-filing institutions into quintiles based on the total value of their equity holdings. Then, essentially treating each group as one large fund, we compute the fraction of each stock they hold. In Appendix A.8, Panel A of Table A5 shows that in the top quintile of institution size, there is a monotonic decreasing relationship between fraction of shares held and retail trading intensity. On the other hand, among the three smallest quintiles of institution size, the relationship is flipped, with a tilt toward high retail stocks. Panel B of Table A5 shows that this pattern is mirrored among active mutual funds. These results suggest that

[11] Of course, this is just one set of ways to quantify whether a stock is hard to value. For example, one could imagine that firms with fewer comparable companies, or with more business segments Cohen and Lou (2012) would also be harder to value. Also see Décaire et al. (2023) on drivers if analyst disagreement. In this paper, we focus on the "accounting" definitions in Table 3, as they fit more cleanly with our analysis on earnings announcements throughout the rest of the paper.

small institutions look more like retail investors than large institutions. Some possible reasons for this are that large institutions face different investment rules/mandates (Ma et al. (2019), Beber et al. (2021)) and constraints on owning a large fraction of small stocks (Edmans et al., 2013).

Overall, this evidence suggests that retail investors (and small institutions) are likely the primary holders of hard to value stocks. This has implications for how such stocks respond to demand shocks around news events, which we discuss in the next section.

5 Earnings Announcements

In light of the evidence that high retail firms tend to be harder to value, we turn our attention to the fundamentals, specifically to quarterly earnings announcements. We show that high retail stocks have a wider distribution of both earnings surprises and earnings-day returns. Next, consistent with high retail stocks being harder to value, we show that such stocks are less sensitive to fundamental information revealed in earnings announcements. Finally, we show that retail investors are especially active and trading costs are especially elevated in high retail stocks around earnings announcements. Consistent with these two facts, we show that prices tend to move with retail order flow almost exclusively in high retail stocks and especially around earnings announcements.

5.1 Distribution of Standardized Unexpected Earnings

Prediction 1A argues that high retail stocks should have more volatile stock returns and fundamental news around earnings announcements. As evidence for this hypothesis, we first document that the distribution of earnings news is much wider for stocks with a high share of retail trades the preceding month.

To quantify the nature of earnings news, we use analyst expectations from IBES. Specifically, for our baseline results, we follow DellaVigna and Pollet (2009) and Hartzmark and Shue (2018), defining standardized unexpected earnings (SUE) as:

$$\mathrm{SUE}_{i,t} = \frac{\mathrm{EPS}_{i,t} - \mathrm{E}_{t-1}[\mathrm{EPS}_{i,t}]}{P_{i,t-1}} \tag{3}$$

where $\mathrm{EPS}_{i,t}$ is the value variable in the IBES unadjusted detail file i.e., "street" earnings per share. $\mathrm{E}_{t-1}[\mathrm{EPS}_{i,t}]$ is the mean estimate of earnings per share in the last IBES statistical period before earnings were released and $P_{i,t-1}$ is the last closing price before the earnings announcement.[12] Table 4 contains summary statistics on earnings-day returns, SUE and analyst coverage.

[12]Here, and everywhere else, all results are robust to instead using the earnings value and mean analyst estimate from the main IBES summary file i.e., the adjusted data.

All the measures of earnings in IBES are "street" earnings, which are designed to take out the effect of one-time items (similar to EPSFXQ in Compustat, which excludes extraordinary items). The "unadjusted" terminology means that earnings were not adjusted for stock splits – as is done in the standard IBES summary – which is useful, because in constructing the "adjusted" file, IBES rounds estimates and actual earnings to the nearest penny, which can reduce the precision of our earnings surprise measure (for more details see the description on WRDS).

Before examining fundamentals, we directly look at earnings-day returns. The first three columns of Table 4 show that high retail stocks have systematically higher earnings day return volatility, measured over a 1-, 3-, or 5-day window starting with the first post-announcement trading day. In the table footer we again report a formal test for the equality of Q5 and Q1 coefficients, and repeat the analysis controlling for size dummy variables. High retail stocks see about 10bps more return volatility on announcement day and the gap is statistically significant controlling for size.

Column 4 of Table 4 shows the standard deviation of SUE across the retail sort. We construct firm-level estimates of SUE volatility in a rolling look-back window of 15 announcements. Consistent with our hypothesis, high retail stocks tend to see more volatile earnings surprises. In the table footer we again report the gap between Q5 and Q1 and the p statistic for the null hypothesis that the coefficients on the extreme quintiles are equal.

To further establish why the earnings of these stocks are so hard to predict, column 5 of Table 4 reports differences in SUE volatility, but restricting to the idiosyncratic component of earnings surprises. To decompose earnings news into idiosyncratic and systematic components, we follow the method in Glosten et al. (2021) and regress firm-level SUE on market-wide value-weighted SUE and SIC-2 industry-wide value-weighted SUE in five year rolling windows. The systematic component of earnings is the predicted value from this regression in the last year of the five year rolling window, while the idiosyncratic component is the residual. Column 5 shows that the volatility of SUE is essentially all driven by the idiosyncratic component of SUE, indicating that the larger SUE volatility relates to information that is specific to these firms, rather than larger sensitivity to economy-wide news. This finding is also consistent with Prediction 1A i.e., that the larger fundamental volatility of high retail firms is coming from the idiosyncratic component of earnings news.

Collectively, the evidence in Table 4 is consistent with prediction 1A: high retail stocks both have more volatile earnings news and more volatile earnings-day stock returns. One alternative explanation for why high retail stocks have larger earnings surprises is that such stocks have lower analyst coverage on average, which leads to less accurate forecasts. However, in the last column we report the number of analysts and, conditional on size, find that analyst coverage is increasing in the retail intensity (Martineau and Zoican, 2019). Column 6 of Table 4 summarizes the dispersion of analyst forecast errors. We calculate the standard deviation of firm-quarter-analyst level forecast errors and normalize them by pre-announcement stock price. Recall that prediction 1A also specified that high retail stocks should have more dispersion in analysts forecasts. Column 6 shows that, consistent with this and the broader notion that these are harder to value securities, high retail stocks tend to see a larger dispersion in analyst forecast errors. This result survives controlling for size. In the 7th and final column, we report the mean number of analysts covering each stock. This column shows that the number of analysts is lower by about 3.5 analysts for high retail stocks, but the gap disappears when controlling for size.

Before moving on, we would like to highlight that, as discussed in Section 2, one concern with the results in 4 is that there is something systematically different about analysts which cover high retail stocks and low retail stocks. If Prediction 1A is correct, however, it should be that *among the stocks a given analyst covers*, their estimates will be relatively less accurate for those that are hard to value. In Appendix A.1.4 we aim to test this directly. Specifically, we start with the assumption that analysts have limited attention, and spread their finite effort equally over the stocks they cover. Then, we test whether or not analysts are relatively less accurate in predicting the earnings of high retail stocks and low retail stocks, accounting for the fact that a larger coverage universe will naturally lead to less accurate estimates. In this test, we can include analyst fixed effects, allaying concerns about selection in terms of analyst quality for high versus low retail stocks.

In Appendix Table A1 we find that there is a monotonic decreasing relationship between retail trading intensity and analyst accuracy. This survives including time fixed effects, analyst fixed effects and a battery of control variables we show are correlated with retail trading intensity in Table 2 and Table 3. Further, the magnitude of the estimated effect on analyst is economically large. Specifically, going from the top to the bottom quintile of retail trading intensity is roughly on par with the mean of analyst accuracy over our whole sample.

The results in Table A1 bolsters both prediction 1 and 1A. First, it allays concerns about selection in terms of analyst quality driving the relationship between the magnitude of earnings surprise and retail trading intensity. Second, and more importantly, it provides additional evidence that high retail stocks are hard to value. Even conditional on analyst quality, and how many stocks each analyst covers, analysts are significantly less accurate in forecasting the earnings of high retail stocks – consistent with their fundamentals being hard to forecast.

5.2 Return Sensitivity to Earnings Surprises

Prediction 1B states that high retail stocks should respond less to earnings news than low retail stocks. The logic is that, by nature of having longer duration cashflows and more valuation uncertainly, today's fundamental news is likely less important for today's price. To quantify this, we follow Kothari and Sloan (1992) and estimate earnings response regressions of the form

$$r^i_{t,t+n} = \alpha + \beta \text{SUE}_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t}, \tag{4}$$

where $r^i_{t,t+n}$ is the cumulative market-adjusted return from the first day investors could trade on earnings information to n days later.[13] We include both firm and time (year-quarter) fixed effects. Controls in $X_{i,t}$ include a variety of factors known to be correlated with retail activity: nominal price, returns from month $t-12$ to $t-2$, time since listing, market capitalization, book-to-market, gross profit margin, book long-term leverage, MAX (lottery demand from Bali et al. (2017)) and month $t-1$ returns (Kumar and Lee (2006), Balasubramaniam et al. (2023), Bali et al. (2021), Luo et al. (2021)). Additional controls include idiosyncratic volatility and total volatility, all computed over the past 12 months. Standard errors are double clustered at firm and time level.

In Equation (4), β is the earnings response coefficient. We are interested in how this varies across retail portfolios, so we interact $\text{SUE}_{i,t}$ with dummy variables for each quintile of retail trading intensity in the month before the earnings announcement. The omitted group is the middle bucket of retail activity. Table 5 contains the results. The first row shows that, consistent with Kothari and Sloan (1992), SUE is positively related to earnings-day returns. The four interaction terms of RSVOL quintiles and SUE show that high retail stocks respond less to earnings innovations, while low retail stocks respond more to earnings innovations than the average stock. The gap in this sensitivity to fundamental news is large. Specifically, the difference in coefficients on SUE $\times$ $Q5$ and SUE $\times$ $Q1$ is over .6, compared to an unconditional effect of just over 1. In the second set of three columns we control for a litany of firm characteristics listed in the above paragraph. The weaker sensitivity of high retail stocks to earnings surprises is left virtually unchanged.

To better interpret the magnitudes, we can repeat the exercise in Table 5, except instead of including dummy variables for quintiles of retail trading activity, we can sort firms into quintiles on other characteristics. For example, if we form quintiles based on returns from month $t-12$ to month $t-2$, we find high past return stocks respond more to earnings news than low past return stocks. In terms of the spread between the top and bottom quintile of past returns, it is nearly identical in magnitude to the spread between high and low retail stocks (0.63 vs. 0.61). Another example is that high book-to-market stocks respond less to earnings news than low book-to-market stocks. Again, in terms of the spread in responses between the top and bottom quintile, it is roughly 0.2, about one third the size of the spread for high vs. low retail stocks.

A potential concern with the results in Table 5 is that high retail stocks don't respond less to news, they just respond more slowly. This would be consistent with the results in Luo et al. (2021) that high retail stocks have a stronger post-earnings announcement drift. Columns 2, 3, 5 and 6 show, however, that the differential response of high retail stocks to earnings news is of roughly constant magnitude over horizons of up to 4 days after the announcement. This suggests that our results are not driven by high retail stocks responding more sluggishly to news.[14]

[13]Following Campbell et al. (2001), market-adjusted returns are defined as the difference between firm i's return and the market factor from Ken French's data library.

[14]In Appendix A.9, Table A6 replicates Table 5, except we sort on retail activity in terms of net flows, instead of gross flows. We find that the stocks with the highest and lowest net retail flow have the smallest earnings responses, which suggests that the decreased sensitivity to earnings news occurs both in stocks where retail investors are rushing

A final concern with the results in Table 5 is that our results are driven by differences in the way stocks with different characteristics respond to earnings news. For example, retail investors tend to favor trading highly volatile stocks, and it's possible that such stocks respond differently to earnings news than less volatile stocks. To address this concern, in unreported results, we re-estimate Table 5, adding interaction terms between SUE and all the control variables. we find that this does not substantially reduce the difference in earnings responses between high and low retail firms, allaying this concern.

As discussed above, a number of the characteristics that vary across retail-sorted portfolios reflect a size effect. This implies another potential concern with the results in Table 5: retail investors select into small stocks and such stocks e.g., by nature of being less covered by media outlets (Martineau and Mondria (2022)) respond less to earnings news. We demonstrate, however, that the weaker sensitivity of high retail intensity stocks to earnings news is not subsumed by size. In Table 6 we re-estimate the regression 4 but include dummy variables for quintiles of firm size, as well as their interaction with SUE. As Table 6 shows, high retail share stocks are less responsive to earnings news across the size distribution, and this difference is statistically significant at the 5% level for all but the smallest size portfolios.

5.3 Retail Trading around Earnings Announcements

Having shown that high retail stocks tend to be less responsive to earnings news, the natural next question is whether this is driven by selection i.e., retail tend to pick stocks which don't respond much to news or whether it is directly driven by retail investor trading (Barber and Odean (2008), Hirshleifer et al. (2008), Kaniel et al. (2012), Luo et al. (2021)). While the differences in characteristics across the retail sort are consistent with retail selection into hard-to-value stocks, we also find evidence of retail trading being an important driver of driving the response to earnings news.

To do this, we first establish two facts: (1) retail trading intensity is especially high around earnings announcements (2) high retail intensity stocks are especially illiquid around earnings announcements. Jointly, these facts open the door for retail investors being an important factor in price determination around earnings announcements.

Prediction 2 argues that in high retail stocks (A) retail investor trading activity should be elevated on and before earnings announcements and (B) retail investors should be net buyers ahead of these announcements. To evaluate these claims, in Figure 2 we plot net abnormal retail-originated trading volume around earnings announcements. In the top left panel we show the average abnormal volume (abnormal meaning relative to the unconditional mean in the respective portfolio) in stocks

in and where retail investors are rushing out before earnings announcement.

belonging to the top and bottom retail quintile around earnings announcements.[15] As the red line indicates, high retail stocks see substantial volume from retail buys in the run-up to earnings announcements. Retail investors are also relatively more active in low retail stocks the day before earnings announcements, although the effect is more muted. In terms of magnitudes, the day before an earnings announcement, retail investors make up an additional 0.8 percentage points of total volume in high retail stocks – relative to their unconditional average trading intensity – compared to making up an extra 0.2 percentage points in the low retail stocks.

The bottom panels show the same results but cumulate the daily data. Again, in terms of magnitudes, in high retail stocks, over the 10 days before the earnings announcements, retail make up 2 percentage points more of total trading volume than one would expect given their average trading intensity, while in low retail stocks this effect is less than 50 basis points. The results in Figure 2 are consistent with Prediction 2. Specifically that (1) high retail stocks have especially elevated retail trading intensity around earnings announcements and (2) this is driven by net retail buying behavior in the pre-earnings announcement period.

Retail investors being net buyers is equivalent to institutions exiting high retail stocks ahead of earnings announcements. Di Maggio et al. (2021) argue this is because institutional investors want to avoid exposure to extreme returns around earnings announcements, as this can lead to outflows. Our findings build on their results, showing that there is significant variation in this effect across stocks. In addition, our results on the retail habitat may offer a more fundamental explanation for *why* institutions tend to exit high-retail stocks ahead of earnings announcements: they understand that hard to value stocks have volatile and idiosyncratic earnings-day returns and therefore avoid them.[16]

In Appendix A.10, Table A7 shows that in addition to a directional effect, retail-initiated trades make up a particularly large amount of overall trading around earnings announcements. This finding is driven by two separate phenomena. First, in the pre-earnings period, retail make up an abnormally large share of volume, but retail-initiated turnover is not statistically significantly higher, suggesting institutional investors are trading less rather than retail trading more. Second, in the post-earnings period, retail trade more both on an absolute (i.e., when normalizing by shares outstanding) and relative (i.e., when normalizing by total trading volume) basis, suggesting that such events drive retail activity.

Next, we turn to the question of whether high retail stocks are more expensive to trade around earnings announcements. On one hand, if retail investors act as noise traders, and/or if institutional investors avoid learning about such stocks, one might expect relatively lower transaction costs. As

[15] In Appendix A.11, we argue that past retail trading intensity is a good measure of the persistent component of retail activity. We also show that retail investors' response to earnings news is dominated by the transitory component of retail trading intensity.

[16] These results are also consistent with the findings in de Silva et al. (2023) regarding heightened retail option trading around earnings announcements.

shown in Appendix A.1, the logic is that, in a model like Kyle (1985), larger amounts of noise trading, or a lower information advantage for insiders decrease the market maker's risk of adverse selection and therefore trading costs.

On the other hand, it's possible that retail investors are not noise traders, and have information about future fundamentals (see e.g., Kaniel et al. (2012)). Further, even if retail investors themselves have no information about fundamentals, retail order flow is persistent (see e.g., Boehmer et al. (2021)). This makes providing liquidity to retail trades risky, as it's possible that subsequent retail trades in the same direction will further push prices against the market maker's position. These forces may make providing liquidity to retail investors relatively riskier, and therefore lead to increased trading costs.

In Appendix A.12, we show that high retail stocks have higher average trading costs than low retail stocks. This suggests that retail investors may have a significant informational advantage in high retail stocks – although these results do not clarify whether that advantage is due to information about future fundamentals or the risk of providing liquidity to these trades. In the next two subsections, we use earnings announcements as a laboratory to understand the relationship between retail trading activity and transaction costs.

Earnings announcements are a useful setting for understanding transaction costs, because they are a time when adverse selection risk is especially salient. The logic is that an investor who is willing to trade right before the public information release may have superior information, suggesting any trade is likely a bad deal (Krinsky and Lee, 1996). Institutional investors' desire to exit before earnings announcements (Di Maggio et al., 2021) may be because they are aware of such adverse selection risk, while retail investors are not.[17]

Given the results in Appendix A.12, a natural question is if high retail stocks are especially expensive to trade around earnings announcements. Given the results in Table A9, however, we need to account for the higher average level of trading costs in high retail stocks. In addition, given the results in Table 4, we also need to account for the more volatile nature of high retail stocks' earnings news, as firms with extreme news might be more expensive to trade on average (Kim and Verrecchia, 1994). To address both these concerns, we estimate the following regression:

$$\begin{aligned}\text{DM Effective Spread}_{i,t} = \alpha + \beta_1 1_{i\in Q1_{\tau-1}} + \beta_2 1_{i\in Q2_{\tau-1}} + \beta_4 1_{i\in Q4_{\tau-1}} + \beta_5 1_{i\in Q5_{\tau-1}} + \\ \theta_1 1_{i\in \text{Q1SUE}_t} + \theta_2 1_{i\in \text{Q2SUE}_t} + \theta_4 1_{i\in \text{Q4SUE}_t} + \theta_5 1_{i\in \text{Q5SUE}_t} + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t}\end{aligned} \tag{5}$$

where DM Effective Spread$_{i,t}$ is the "demeaned" effective spread, defined as the effective bid-ask spread from the WRDS intraday indicators suite minus the average effective spread for that stock

[17] Retail investors' trading in the face of adverse selection could be the result of overconfidence (Statman et al., 2006) which may be especially prevalent among the retail population (Peng and Xiong (2006), Barber et al. (2020)).

in the month before the earnings announcement. $1_{i \in QkSUE_t}$ is an indicator for whether firm i's SUE is in the kth quintile of SUE among all firms that released that quarter.

Table 7 shows that, even conditional on the nature of the news and differences in average trading costs, high retail stocks are especially expensive to trade before, on, and after earnings announcements. In terms of magnitudes, in the pre- and post- earnings period, high retail stocks are about 2 basis points more expensive to trade, while on the earnings day itself they are about 4 basis points more expensive to trade.[18]

Overall, Table 7 provides suggestive evidence that market makers are especially concerned about adverse selection risk in high retail stocks around earnings announcements. One explanation for this is that retail investors have an informational advantage in such stocks. In the next subsection, we aim to understand if the *direction* of retail investors' order flow has different predictive power for future returns in high and low retail stocks. This will shed light on the mechanism behind the results in Table 7, as this exercise will help us quantify retail investors' information about future fundamentals.

5.4 Retail Imbalances and Returns

A natural next question is whether prices move with or against retail order flow around earnings announcements. Given retail investors' relatively small share of the market, for prices to move with their order flow, it must be that institutions are actively trading in the same direction as retail or, at a minimum, are not trading against retail orders. Either explanation would be surprising, however, as there are reasons to believe that retail investors don't have private information around earnings announcements. For example, retail tend to trade against news (i.e., buying on negative earnings surprises and selling on positive surprises), which leads them to underperform (Luo et al. (2021), Kogan et al. (2023)). The question of why more institutions don't bet against retail is especially salient in our setting, as the algorithm we use to track retail activity (Boehmer et al. (2021)) could be run in real time by any investor with access to TAQ data.

One reason investors may hesitate to trade against retail is that, as shown in Boehmer et al. (2021), retail order flow is auto-correlated. This persistence makes betting against retail orders risky, as more orders in the same direction may arrive and force early liquidation at a loss (De Long et al., 1990). More broadly, one could view betting against retail as a type of liquidity provision, which has been shown to earn high risk-adjusted returns (Nagel, 2012). Given that, as shown

[18]While these magnitudes seem small, these regressions control for a host of firm-level characteristics and fixed effects. Further, because this is in terms of the demeaned effective spread, it accounts for the higher average trading costs for high retail firms. In addition, because it is demeaned over the month before the earnings announcement, rather than the unconditional firm level average, it accounts for the run up in trading costs that occur before the announcement itself. Finally, this 2-4 basis point increase is large relative to the unconditional value-weighted bid-ask spread in 2021, which was 6 basis points.

in Appendix A.12 and Table 7, high retail stocks are expensive to trade, we may expect that institutions are unwilling to provide liquidity in such stocks. Finally, there may be frictions which prevent institutional investors from trading the type of stocks that retail investors favor (Haddad et al., 2021). Collectively, these channels imply that retail order flow is more likely to move prices in high retail stocks, as other investors may avoid trading in the opposite direction.

So far, we have been sorting firms into quintiles based on gross retail activity i.e., retail buys plus retail sells. So, to test this hypothesis, we construct a measure of retail order imbalance as

$$\text{mroibvol}_{i,t} = \frac{\text{RBuy}_{i,t} - \text{RSell}_{i,t}}{\text{RBuy}_{i,t} + \text{RSell}_{i,t}}. \tag{6}$$

This measure is useful for determining whether retail investors are taking or providing liquidity. Another way to say this is that mroibvol speaks to whether or not other investors tend to be trading with or against retail investors.

To test whether returns tend to move with or against net retail order flow we calculate $\text{mroibvol}_{i,t}$ on the firm-week level and construct weekly mroibvol quintiles. We refer to the week in which the mroibvol quintiles are calculated as the "focal week'. We then regress excess returns in weeks surrounding the focal week on said dummy variables. Specifically, we estimate:

$$rx_{i,t-\tau} = a + \sum_{k=1}^{5} b_k 1_{i \in Qmroibvol_{k,t}} + \epsilon_{i,t} \tag{7}$$

where t is the focal week and τ is either -1, 0 or 1. In words, the left-hand-side of Equation 7 is a weekly excess return either in the week when the mroibvol imbalance was calculated, or the preceding or succeeding week. We estimate these regressions separately for all stocks, and for stocks in the low and high retail trading intensity buckets. We calculate excess returns with respect to an equal-weighted return of all stocks in the sample in a given week. Further, we separately estimate these regressions in all weeks, and in weeks where the a given stock sees an earnings announcement.

Table 8 contains the results. Panel A uses all weeks for the analysis, Panel B restricts to focal weeks that contain an earnings announcement. In all columns, *mroibvol* is measured in the focal week, meaning week $\tau = 0$. Returns are measured in the week indicated in the table header.

Let's start with the first set of three columns, meaning week -1. The coefficient on retail sells is .15, indicating that stocks that were most sold by retail in week 0 saw positive returns in the prior week; the coefficient on retail buys is $-.65$ meaning that the stocks that saw the most retail buying in week 0 had large negative returns in the prior week. Therefore the first column indicates that retail behaves in a contrarian manner with respect to prior week returns. In the table footer we report a test of the equality of returns between extreme retail buys and sells. In this case this gap is negative 80 basis points statistically significant at the 1% level.

The second and third columns repeat the same analysis, but restrict the sample to low and high retail trading stocks. In both subsets the same conclusion holds: retail traders as a group behave in a contrarian manner with respect to prior week returns. Note, though, that the effect size is more than three times as strong for stocks with high past retail presence.

The second set of three columns documents the contemporaneous relationship between retail order imbalance and returns. Here the results tend to be quite muted, the return differential between heavily bought and heavily sold stocks are not statistically significant. Only for the high retail stocks do we see a gap in the returns between extreme retail buys and sells, but the coefficients do not show a consistent pattern.

The third set of three columns in Panel A of Table 8 studies the predictive power of retail imbalances. Looking across all stocks (Column 7) we see that intense retail buying has predictive power over returns in the subsequent week, consistent with the results in Kaniel et al. (2008). Stocks heavily sold by retail see a negative excess return of 8 basis points in the subsequent week while heavily bought stocks see a positive excess return of 10 basis points. Such a gap in next week returns is particularly pronounced for stocks with a high retail presence: as the final column shows the gap between heavily bought and heavily sold sotcks is on average 37 basis points.

Overall, these results emphasize the importance of recognizing cross-sectional heterogeneity in retail trading intensity. The stocks with high retail presence see more contrarian buying after poor returns and stronger return predictability from retail flows.

Panel B of Table 8 repeats the above analysis but restricts to focal weeks during which an earnings announcement took place. The overall results are strikingly similar to the ones reported in Panel A: retail buying tends to be contrarian with respect to prior week returns and retail buying has predictive power over next week returns. Focusing on the high retail quintile, the table reports a return differential of 51bps for heavily bought over heavily sold stocks.

The return predictability documented in Table 8 leaves open at least two possibilities: either retail trades as a group are providing liquidity to the market, buying stocks at depressed prices (and selling at elevated prices) or they possess private information or skill. We report the results from one approach that tries to discriminate between these two possibilities. In Table 9 we follow the methodology of Kaniel et al. (2012) in order to decompose these average returns into liquidity provision, and into private information.

In each day of the sample we estimate a cross-sectional regression of long horizon (60 trading day) cumulative abnormal returns on retail order imbalance (mroibvol) dummy variables, formed over the the past ten days, controlling for past cumulative abnormal returns over the past ten days and limiting the sample to stocks not within 20 trading days of an earnings announcement. We then use these estimated coefficients to predict the expected cumulative abnormal return (ECAR) for firms making an earnings announcement on that day.

In the first three columns of Table 9 we regress firm-announcement level cumulative abnormal returns on pre-announcement retail imbalance indicators. The first column shows, stocks that were heavily bought by retail traders in the run-up to an earnings announcement see 1.15% returns over the next 61 trading days, starting with the day the announcement content was first tradeable. This effect is stronger for stocks that are in the high retail trading quintile, as shown in the third column: the gap between returns of heavily bought and heavily sold stocks is 2.3%.

The second set of three columns reports the same analysis with ECAR—expected cumulative abnormal return—-as the left hand side variable. ECAR is estimated contemporaneously from the relationship between retail imbalance and future CAR of non-announcing firms. The ECAR returns, therefore, can be interpreted as the current level of returns from liquidity provision because they reflect the current relationship between retail imbalances and returns in non-announcing firms. As the fourth column shows, across all stocks the predicted return differential from liquidity provision is about .92%. The estimated returns from liquidity provision among the highest retail stocks is 1.37% leaving over .9% of return differential to private information.

In all, we find a stronger return predictability from retail imbalances for stocks that see a higher level of retail trading. About half of this return predictability can be attributed to private information. This is important, because it sheds light on the mechanism for why high retail stocks have higher trading costs around earnings announcements than low retail stocks. Specifically, even if retail investors' signals are not particularly precise, their *informational advantage* may be relatively large. The logic is that if institutional investors totally avoid learning about such stocks, institutional investors face significant adverse selection risk and thus set high transaction costs.

Table 9 suggests the other half of the predictive power of retail investors' order flow for future returns is due to liquidity provision. In Appendix A.12, we offered several explanations for why institutional investors may avoid betting against retail order flow (and thus why prices may tend to move with net retail demand), even if they believe it is not information-driven. The results in Table 8 provide further evidence for these claims. Specifically, retail order flow predicts returns in the following week, with low returns after retail selling and high returns after retail buying. This effect is roughly twice as strong in high retail stocks as the average stock, and is even stronger around earnings announcements. So, betting against retail order flow in high retail stocks is especially risky, because returns may continue to move with that flow in the following week.

6 The Earnings Announcer Premium across Retail Portfolios

The results in the prior section establish that high retail stocks are less sensitive to earnings news. In this section we show that this gap in terms of sensitivity translates into a return differential in portfolios that take exposure to announcing stocks as a function of their retail sort. Our analysis is motivated by the finding in Savor and Wilson (2016) that announcing firms outperform those with no scheduled announcements, and that the aggregate announcer portfolio has alpha with respect to the buy-and-hold portfolio. We aim to refine this result and test Prediction 3, which argues that the earnings announcer premium should be lower, or non-existent, among high retail stocks.

To test this hypothesis, in Table 10, we decompose average returns around earnings announcements into pre- and post- announcement components as a function of size and retail trading intensity.

The first three columns focus on a narrow window: the last trading day before the earnings announcement, and the first trading day on which the announcement could have been traded. The second set of three columns focuses on a 6-day announcement window, containing three trading days prior to the announcement, the day the earnings news could have been first traded on, as well as the next two trading days. In both sets of announcement windows, "Pre" refers to the portion prior to the announcement, "Post" refers to the portion after the announcement. Each panel restricts the sample to the indicated size quintile and Q5 is the dummy variable for the 20% of stocks with the highest share of retail trading within that size bucket. All regressions contain month dummies and standard errors are clustered by day and firm.

The first takeaway from Table 10 is the presence of the earnings announcer premium. Specifically, in the first column and the fourth column, the coefficient on the size dummies is always positive, suggesting that, on average, announcing firms have positive returns. Further, the coefficient on the interaction term with low retail is always higher than the interaction term with high retail. This suggests the announcer premium is systematically higher among low retail stocks than high retail stocks.

Secondly, Table 10 shows that high retail stocks see lower announcement time returns, consistent with Prediction 3's implication of a smaller announcer premium for high retail securities. Let's first focus on the third column, representing the first trading day on which earnings announcement is tradeable. The bottom panel shows that among the largest quintile of stocks, the average announcement time return is 14 bps. Similarly, the average announcement time returns are 30, 41, 30, and 10 basis points for the remaining size quintiles. In all cases, the announcer premium is considerably smaller for the high retail share stocks—compared to stocks in Q3—and in all cases this difference is statistically significant. In fact, the coefficient on Q5 is in all cases larger than the unconditional return.

The second set of three columns repeats the same analysis over a six-day event window straddling the earnings announcement. Again the same pattern emerges: high retail stocks underperform others in the earnings announcement window, and this gap is present across the size quintiles, representing mostly lower post-announcement returns. Overall, the findings replicate the known result that average stock returns are high around earnings announcements but find a substantial amount of heterogeneity across the retail sort: the announcement risk premium is negligible among high retail stocks.

Note too that the summary statistics reported in Table 4 provide additional support to the view that these return differentials represent risk premia: the gap in average SUE is small relative to the difference in average post-announcement returns. Among the small stocks, the gap in average SUE is 45 basis points, but average returns differ by more than 1.5 %. Among the largest stocks, the gap in average SUE is 6 basis points, but the difference in average returns is above 30 basis points, requiring an SUE sensitivity of close to 5 to account for the return gap.

In sum, the results in Table 10 are consistent with Prediction 3: high retail stocks do not earn the earnings-announcer premium. Savor and Wilson (2016) argue the earnings announcer premium is compensation for exposure to systematic news. As shown in Table 4, however, high retail stocks' SUE is mostly composed of idiosyncratic news. Therefore, one explanation for the findings in Table 10 is that high retail stocks do not earn the announcer premium because their SUE is mostly composed of idiosyncratic information.

7 Conclusion

In this paper, we establish a new fact: retail investors tend to favor trading stocks which are hard to value. Consistent with this, such stocks have more volatile realizations of both fundamental news and earnings-day returns. Further, these stocks tend to respond less to earnings news of a given size, and are relatively more expensive to trade around earnings announcements.

We additionally show how retail investors trade around earnings announcements. Retail are abnormally active in the pre-earnings announcement period, acting as net buyers from institutional investors, particularly in the stocks they favor generally. Intense buying by retail in the run-up to announcements predicts positive excess post-announcement returns, particularly for stocks with a large retail presence.

Finally, we link the fact that retail investors favor hard to value stocks to the earnings announcer premium. Past literature has argued that this premium is earned as compensation for exposure to the systematic risk contained in earnings news. We find that high retail stocks have a small systematic component in their earnings news and that any news about these firms is hard to interpret. So, consistent with the systematic risk-based explanation of the earnings announcer premium, it is not earned in high retail stocks.

Overall, our findings document a new dimension of investor heterogeneity. Retail investors have a comparative advantage relative to institutional investors in trading hard-to-value stocks. Further, we find that stocks with significant retail trading activity have low institutional ownership, suggesting retail also have a comparative advantage in holding such stocks. This pattern is especially stark when comparing retail to large investment managers, suggesting institutional constraints are an important determinant of aggregate risk sharing. Therefore, our results speak to an novel dimension of cross-sectional heterogeneity in which groups of investors bear different types of stocks' risk.

8 Figures

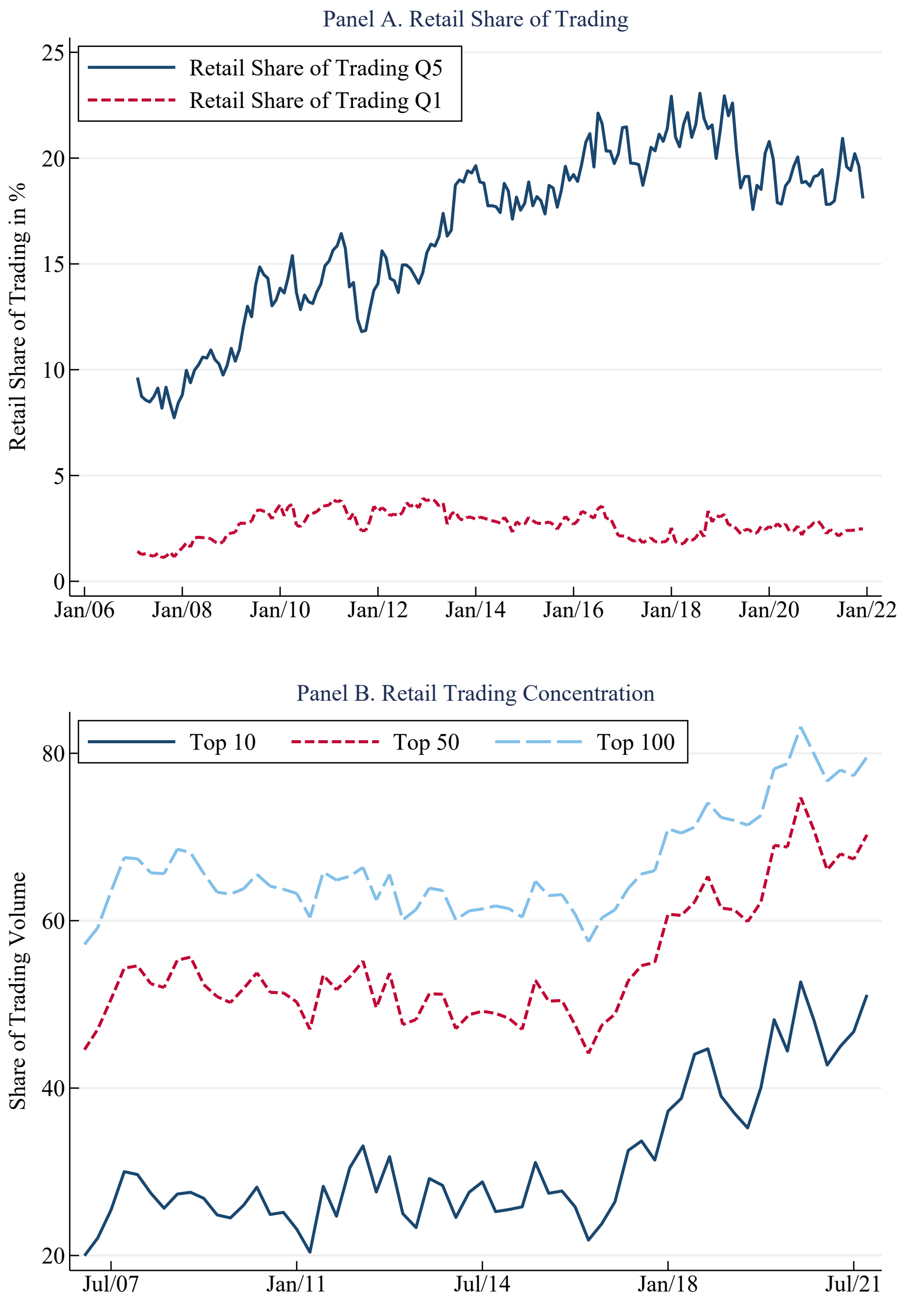

Figure 1: Retail share of trading volume and retail trading concentration. Panel A shows the average retail share of trading volume in the top and bottom quintile sorted on previous month's retail trading intensity. Panel B shows the cumulative share of total retail dollar volume stemming from the top 10, 50, and 100 stocks sorted by retail dollar volume.

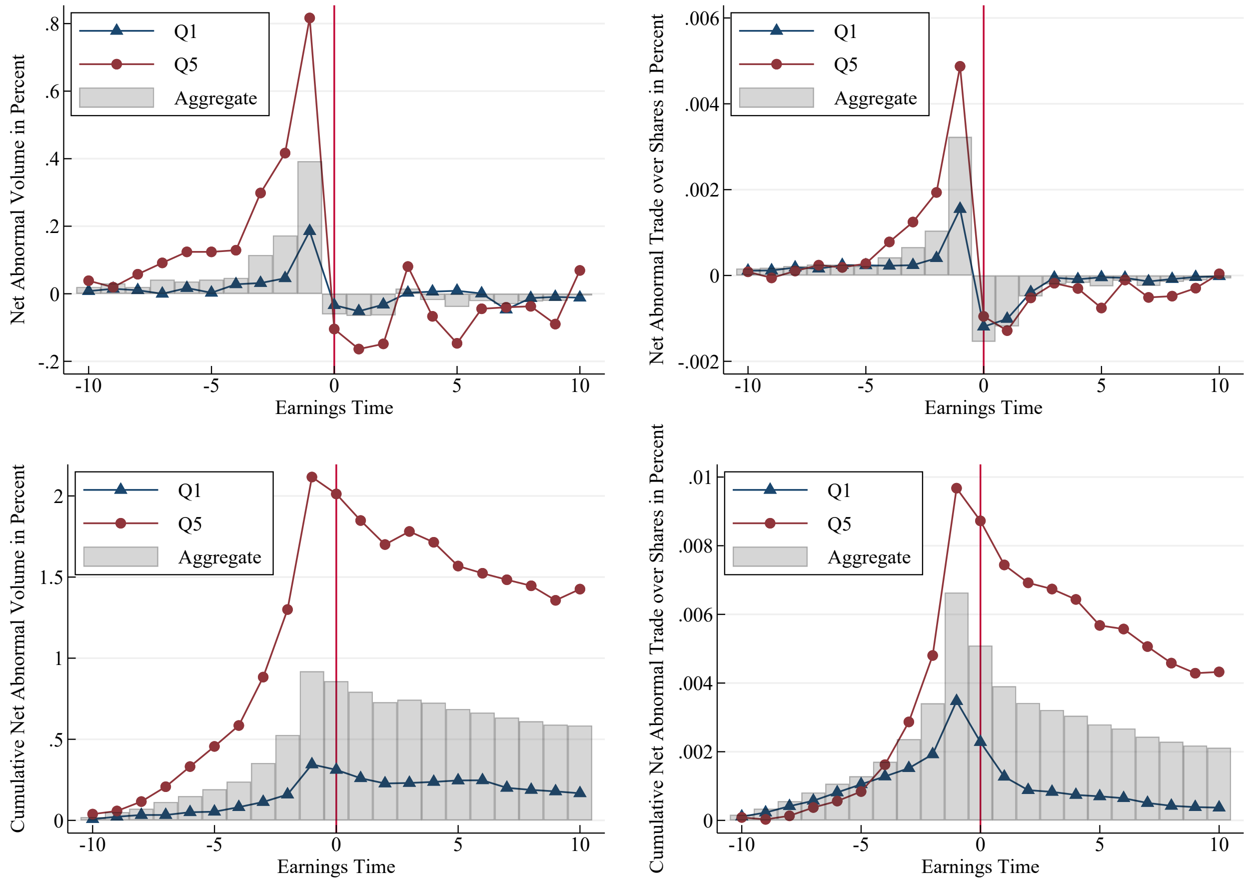

Figure 2: Abnormal Trading Volume around Earnings Announcements. Daily abnormal net retail share of volume and abnormal net retail-initiated turnover, in percent units. Q1 represents the bottom quintile of retail intensity, while Q5 represents the top quintile. We subtract out the unconditional means in respective series to construct abnormal volume/turnover and take an equal-weigthed average within each quintile. Bottom panels cumulate the values in top panels staring at time -10 relative to earnings announcement day at time 0.

9 Tables

	Retail Shr. Volume			Turnover			Retail-initiated TO		
	All	Small	Large	All	Small	Large	All	Small	Large
Low	-2.42***	-2.88***	-2.31***	-0.82***	-1.45***	0.33	-0.42***	-0.20***	-0.40***
	(-66.58)	(-64.07)	(-69.32)	(-3.31)	(-5.12)	(0.67)	(-37.26)	(-5.48)	(-21.73)
2	-1.29***	-1.36***	-1.30***	-0.52***	-0.77***	-0.65**	-0.24***	-0.12***	-0.24***
	(-65.83)	(-52.59)	(-57.40)	(-3.12)	(-3.51)	(-2.02)	(-30.15)	(-14.37)	(-18.59)
4	2.51***	2.74***	2.50***	1.80***	1.58***	2.97***	0.57***	0.33***	0.64***
	(52.95)	(48.60)	(48.90)	(6.45)	(7.43)	(4.42)	(23.32)	(17.98)	(13.62)
High	10.84***	12.46***	8.21***	7.10***	9.46***	26.50***	2.91***	2.52***	4.98***
	(49.44)	(50.35)	(25.51)	(7.84)	(12.00)	(8.96)	(20.03)	(17.84)	(10.52)
Average	6.37	12.38	4.51	20.68	14.30	22.06	1.32	1.90	1.02
Q5-Q1	13.26	15.34	10.52	7.93	10.91	26.17	3.34	2.72	5.37
p(Q1=Q5)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N	462,558	92,581	92,439	462,558	92,581	92,439	462,558	92,581	92,439
R^2	0.72	0.54	0.77	0.02	0.04	0.12	0.21	0.10	0.38

Table 1: Trading in five retail share of trading sorted portfolios. Firm-month level regressions of retail share of trading, turnover, and retail-initiated turnover on prior month retail trading intensity. Low refers to the quintile with least retail-initiated trading in the prior month; high refers to the quintile with most retail-initiated trading in the prior month. Small and Large refer to the first and fifth quintile in terms of firm size, respectively.Standard errors clustered on the firm and month level.

	Cap	Age	Prc	Past R	B/M	E/P	β_{CAPM}	100-Inst.
Low	0.53***	-0.69	1.73	-1.89***	0.01	0.01***	-0.05***	-5.74***
	(38.24)	(-1.31)	(1.55)	(-4.27)	(1.14)	(9.28)	(-6.48)	(-13.41)
2	0.64***	0.24	3.42***	-0.59*	-0.03***	0.01***	-0.02***	-4.77***
	(39.53)	(0.71)	(4.86)	(-1.88)	(-4.53)	(11.81)	(-4.93)	(-19.51)
4	-0.73***	-1.71***	-9.06***	-1.01**	0.09***	-0.03***	0.02***	8.49***
	(-65.26)	(-4.18)	(-9.74)	(-2.11)	(8.39)	(-17.65)	(3.22)	(26.32)
High	-1.16***	-7.87***	-27.62***	-7.58***	0.23***	-0.15***	-0.05***	31.49***
	(-111.65)	(-16.15)	(-19.49)	(-8.76)	(9.19)	(-23.90)	(-3.68)	(49.48)
Average	7.26	20.67	34.78	9.47	0.65	0.00	1.11	31.04
Q5-Q1	-1.69	-7.18	-29.35	-5.68	0.22	-0.16	0.00	37.24
p(Q1=Q5)	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.00
Q5-Q1, size	-0.03	-1.39	-6.61	10.07	-0.03	-0.10	0.24	19.86
p(Q1=Q5), size	0.00	0.02	0.00	0.00	0.29	0.00	0.00	0.00
N	416,432	416,432	416,432	416,432	416,432	416,432	416,432	416,432
R^2		0.03	0.08	0.00	0.02	0.11	0.00	0.26

Table 2: Fundamentals in five retail share of trading sorted portfolios. Firm-month level regressions on dummy variables representing retail trading intensity quintiles formed the prior month. Cap is market cap; Age is time since listing; Prc is nominal price; Past R is the returns from month $t = -12$ to $t = -2$ i.e., the returns used to form momentum portfolios (Jegadeesh and Titman (1993)); B/M is book-to-market; E/P is the earnings-to-price ratio; β_{CAPM} is the market beta computed over the previous 252 trading days; Inst. is institutional ownership from 13F data and 100-Inst. is therefore an alternative proxy for retail ownership. Note the first column estimates a median regression. Standard errors clustered on the firm and month level.

	CF	K_{Int}	K_{Know}	K_{Org}	PAT	VU	Mispric.
Low	-0.10***	-0.05***	-0.10***	-0.06***	-0.17***	-0.11***	-0.13***
	(-3.87)	(-4.02)	(-10.67)	(-4.25)	(-7.69)	(-4.42)	(-5.68)
2	-0.08***	-0.05***	-0.07***	-0.07***	-0.07***	-0.09***	-0.10***
	(-4.95)	(-6.98)	(-10.15)	(-7.93)	(-4.78)	(-5.80)	(-7.46)
4	0.12***	0.18***	0.16***	0.16***	0.08***	0.25***	0.17***
	(5.98)	(12.25)	(13.04)	(11.75)	(4.00)	(12.97)	(9.81)
High	0.26***	0.56***	0.72***	0.43***	-0.01	0.77***	0.46***
	(8.62)	(18.46)	(19.61)	(13.71)	(-0.23)	(24.50)	(14.73)
Q5-Q1	0.36	0.61	0.82	0.48	0.16	0.87	0.59
p(Q1=Q5)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q5-Q1, size	0.39	0.21	0.53	0.06	0.44	0.31	0.40
p(Q1=Q5), size	0.00	0.00	0.00	0.06	0.00	0.00	0.00
N	348,122	460,598	460,598	460,598	462,558	216,981	256,831
R^2	0.02	0.05	0.09	0.03	0.01	0.10	0.04

Table 3: Valuation across five retail share of trading sorted portfolios. Firm-month level regressions on dummy variables representing retail trading intensity quintiles formed the prior month. All valuation metrics transformed into z scores for ease of interpretation. CF is cashflow duration, computed after Gormsen and Lazarus (2023); K_{Int}, K_{Know}, and K_{Org} are measures of intangible capital (total, knowledge, and organizational capital, respectively) normalized by market capitalization from Peters and Taylor (2017); PAT is the real market value of patents over the past five years (data obtained from Kogan et al. (2017)) divided by market capitalization; VU is valuation uncertainty from Golubov and Konstantinidi (2021); Mispricing is the mispricing score from Stambaugh and Yuan (2017). Standard errors clustered on the firm and month level.

	SD(Returns)			SD(SUE)		Analysts	
	(0, 0)	(0, 2)	(0, 4)	Full	Idiosyn.	SD	Number
Low	-0.01*	-0.03*	-0.04**	-0.02**	-0.08**	-0.10***	-0.53***
	(-1.82)	(-1.97)	(-1.99)	(-2.10)	(-2.04)	(-10.67)	(-2.61)
2	-0.01**	-0.02**	-0.03**	-0.02***	-0.06**	-0.08***	0.35***
	(-1.98)	(-2.37)	(-2.48)	(-2.81)	(-2.34)	(-11.03)	(2.87)
4	0.04***	0.07***	0.10***	0.05***	0.19***	0.21***	-1.19***
	(5.48)	(5.83)	(6.00)	(6.47)	(5.81)	(13.43)	(-7.42)
High	0.12***	0.23***	0.31***	0.16***	0.62***	1.01***	-4.48***
	(8.20)	(8.92)	(9.28)	(10.07)	(8.92)	(16.22)	(-21.74)
Average	7.72	9.43	10.43	2.04	3.95	0.42	8.38
Q5-Q1	0.13	0.25	0.34	0.18	0.70	1.11	-3.95
p(Q1=Q5)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q5-Q1, size	0.08	0.15	0.20	0.10	0.41	0.76	1.25
p(Q1=Q5), size	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N	149,204	149,204	149,204	149,204	149,204	135,028	149,204
R^2	0.01	0.02	0.02	0.02	0.02	0.05	0.06

Table 4: Announcement-Level Summary Statistics. Firm-announcement level regressions on dummy variables representing retail trading intensity quintiles formed the prior month. Announcement returns are measured starting on day 0, the first day on which the announcement is tradeable. SD(SUE) measures the standard deviation of SUE in a look-back window of 15 earnings announcements, for the full SUE, as well as separately for the idiosyncratic component of SUE. Analysts SD and Number refer to the standard deviation of analyst forecasts and the number of analysts, respectively. Quarterly earnings announcements from 2007 to 2021. Standard errors clustered on the firm and quarter level.

	Market-Adjusted Return					
	(0, 0)	(0, 2)	(0, 4)	(0, 0)	(0, 2)	(0, 4)
SUE	1.056***	1.158***	1.173***	1.087***	1.189***	1.229***
	(14.79)	(11.25)	(11.66)	(14.27)	(11.02)	(11.49)
SUE x Q1	0.158	0.180	0.204	0.210*	0.240	0.261*
	(1.93)	(1.47)	(1.72)	(2.38)	(1.89)	(2.03)
SUE x Q2	0.339***	0.352**	0.394**	0.389***	0.431***	0.446***
	(4.07)	(3.00)	(3.02)	(4.23)	(3.51)	(3.46)
SUE x Q4	-0.181*	-0.141	-0.0962	-0.183*	-0.123	-0.0743
	(-2.57)	(-1.40)	(-0.96)	(-2.53)	(-1.15)	(-0.69)
SUE x Q5	-0.474***	-0.471***	-0.446***	-0.420***	-0.394***	-0.388***
	(-6.85)	(-4.78)	(-4.69)	(-5.87)	(-3.96)	(-3.96)
Controls				Yes	Yes	Yes
N	148934	148934	148934	138195	138195	138195
R^2	0.09	0.09	0.09	0.10	0.10	0.10

Table 5: Post-announcement return sensitivity to realized standardized earnings surprise. Regression of post-announcement returns on standardized unexpected earnings. Post-announcement return period indicated in column header. 0 refers to the first day announcement information is tradeable during normal market hours. SUE is standardized unexpected earnings, defined in Equation 3 and Qk is an indicator variable for whether stock i was in retail intensity quntile k at the end of the month before the earnings announcement. Quarterly earnings announcements from 2007 to 2021. SUE and returns winsorized at the 1st and 99th percentile. Control variables include nominal price, returns from month $t-12$ to $t-2$, time since listing, market capitalization, book-to-market, gross profit margin, book long-term leverage, MAX (lottery demand) and month $t-1$ returns. Additional controls include betas on the three Fama-French factors as well as the momentum factor, idiosyncratic volatility and total volatility, all computed over the past 12 months.

.

	Standardized Unexpected Earnings					
	(0, 0)	(0, 2)	(0, 4)	(0, 0)	(0, 2)	(0, 4)
Size 1 x SUE	0.544***	0.654***	0.680***	0.577***	0.683***	0.702***
	(13.13)	(12.43)	(10.27)	(13.10)	(12.14)	(10.10)
Size 1 x SUE x Q5	-0.0619	-0.0911	-0.0775	0.0162	0.0212	0.0414
	(-1.13)	(-1.37)	(-0.87)	(0.29)	(0.31)	(0.46)
Size 2 x SUE	1.168***	1.327***	1.366***	1.178***	1.357***	1.413***
	(13.20)	(11.08)	(9.04)	(11.90)	(10.28)	(8.48)
Size 2 x SUE x Q5	-0.485***	-0.551***	-0.482**	-0.404***	-0.480**	-0.416*
	(-5.48)	(-4.37)	(-2.99)	(-3.80)	(-3.28)	(-2.17)
Size 3 x SUE	1.476***	1.532***	1.547***	1.591***	1.686***	1.716***
	(7.60)	(6.64)	(6.13)	(7.05)	(6.27)	(5.99)
Size 3 x SUE x Q5	-0.645**	-0.539*	-0.582*	-0.654**	-0.573*	-0.606*
	(-3.30)	(-2.31)	(-2.34)	(-2.93)	(-2.10)	(-2.17)
Size 4 x SUE	1.477***	1.719***	1.627***	1.478***	1.755***	1.703***
	(6.77)	(6.02)	(5.23)	(6.60)	(6.01)	(5.37)
Size 4 x SUE x Q5	-0.699**	-0.899**	-0.676	-0.684**	-0.989**	-0.728*
	(-2.72)	(-2.70)	(-1.88)	(-2.63)	(-2.94)	(-1.99)
Size 5 x SUE	1.467***	1.874***	2.200***	1.583***	1.888***	2.298***
	(3.77)	(3.90)	(5.12)	(3.36)	(3.46)	(4.63)
Size 5 x SUE x Q5	-0.760	-0.994	-1.309**	-0.858	-0.989	-1.417*
	(-1.64)	(-1.84)	(-2.88)	(-1.53)	(-1.61)	(-2.59)
Controls				Yes	Yes	Yes
N	30359	30363	30363	28780	28784	28784
R^2	0.03	0.03	0.04	0.03	0.03	0.04

Table 6: Post-announcement return sensitivity to realized standardized earnings surprise. Regression of post-announcement returns on standardized unexpected earnings. Post-announcement return period indicated in column header. 0 refers to the first day announcement information is tradeable during normal market hours. SUE is standardized unexpected earnings, defined in Equation 3, Qk is an indicator variable for whether stock i was in retail intensity quntile k at the end of the month before the earnings announcement and Size j is an indicator for whether firm i was in size quintile j at the end of the month before the earnings announcement. SUE and returns winsorized at the 1st and 99th percentile. Quarterly earnings announcements from 2007 to 2021.

	Demeaned Effective Spread (basis points)					
	(-5,-1)	(-3,-1)	-1	0	(0,2)	(0,4)
Low Retail	-0.402**	-0.359	-0.124	-0.745**	-0.341*	-0.126
	(0.184)	(0.239)	(0.308)	(0.317)	(0.203)	(0.216)
2	-0.132	-0.0814	-0.098	0.0557	0.105	0.0814
	(0.135)	(0.163)	(0.238)	(0.263)	(0.167)	(0.147)
4	0.666**	0.680**	0.782*	0.747*	0.641**	0.576**
	(0.258)	(0.288)	(0.392)	(0.447)	(0.283)	(0.248)
High Retail	1.538***	1.431**	0.861	2.927***	2.390***	2.164***
	(0.502)	(0.609)	(0.844)	(0.944)	(0.679)	(0.529)
Observations	137,141	137,141	137,141	137,141	137,141	137,141
R-squared	0.107	0.105	0.099	0.108	0.118	0.124
Time FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES

Table 7: Retail activity and demeaned trading costs around earnings announcements. Left-hand-side variables are average demeaned effective spread computed over various windows around earnings announcements. Demeaned effective spread is effective bid-ask spread minus average effective spread over the calendar month before the earnings announcement. Control variables include nominal price, returns from month $t-12$ to $t-2$, time since listing, market capitalization, book-to-market, gross profit margin, book long-term leverage, MAX (lottery demand) and month $t-1$ returns. Additional controls include betas on the three Fama-French factors as well as the momentum factor, idiosyncratic volatility and total volatility, all computed over the past 12 months. Standard errors are double clustered at the permno/year-month level.

Panel A.

	Week -1			Week 0			Week 1		
	All	Low	High	All	Low	High	All	Low	High
Retail Sells	0.15***	0.14***	-0.03	-0.30***	-0.13***	-0.79***	-0.08***	-0.04	-0.08
	(4.82)	(4.36)	(-0.39)	(-8.71)	(-4.23)	(-10.19)	(-2.79)	(-1.48)	(-1.12)
2	0.05***	0.10***	-0.20***	-0.17***	-0.02	-0.67***	-0.03*	-0.01	-0.04
	(3.01)	(3.77)	(-3.65)	(-9.03)	(-0.88)	(-12.33)	(-1.68)	(-0.49)	(-0.82)
4	-0.27***	-0.11***	-0.53***	-0.11***	-0.12***	0.26***	-0.00	0.04	0.01
	(-13.57)	(-3.96)	(-9.49)	(-4.98)	(-4.01)	(4.51)	(-0.28)	(1.57)	(0.23)
Retail Buys	-0.65***	-0.24***	-1.38***	-0.35***	-0.16***	-0.40***	0.10***	0.04	0.29***
	(-20.74)	(-7.72)	(-19.30)	(-10.68)	(-4.99)	(-5.44)	(3.64)	(1.27)	(4.18)
Constant	0.16***	0.09**	0.35***	0.20***	0.16***	0.17**	0.02	0.06*	-0.18***
	(10.23)	(2.55)	(4.94)	(12.41)	(4.86)	(2.47)	(1.07)	(1.96)	(-2.67)
Q5-Q1	-0.80	-0.38	-1.35	-0.06	-0.02	0.39	0.18	0.08	0.37
Q1=Q5 p	0.00	0.00	0.00	0.10	0.36	0.00	0.00	0.00	0.00
N	1,907,139	382,104	361,259	1,907,139	382,104	361,259	1,907,139	382,104	361,259
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Panel B.

	Week -1			Week 0			Week 1		
	All	Low	High	All	Low	High	All	Low	High
Retail Sells	0.14**	0.01	0.12	-0.66***	-0.58***	-1.05***	-0.21***	-0.12	-0.24
	(2.11)	(0.08)	(0.70)	(-5.23)	(-3.33)	(-3.79)	(-2.90)	(-1.38)	(-1.30)
2	0.09*	0.01	-0.09	-0.41***	-0.43***	-0.98***	-0.12**	-0.12	0.01
	(1.91)	(0.10)	(-0.63)	(-4.34)	(-2.66)	(-3.78)	(-2.55)	(-1.57)	(0.05)
4	-0.22***	-0.12	-0.42***	-0.95***	-0.42***	-1.04***	-0.11**	-0.13*	-0.07
	(-4.56)	(-1.60)	(-2.60)	(-10.43)	(-2.80)	(-3.93)	(-2.39)	(-1.73)	(-0.42)
Retail Buys	-0.72***	-0.34***	-1.23***	-1.76***	-0.88***	-2.05***	0.02	-0.04	0.27
	(-10.56)	(-4.34)	(-6.84)	(-15.46)	(-5.71)	(-7.14)	(0.28)	(-0.47)	(1.57)
Constant	0.20***	0.24***	0.24	0.63***	0.74***	0.22	0.20***	0.23***	-0.06
	(4.27)	(3.74)	(1.52)	(8.61)	(6.37)	(0.98)	(4.58)	(3.19)	(-0.40)
Q5-Q1	-0.86	-0.34	-1.35	-1.10	-0.29	-1.00	0.23	0.08	0.51
Q1=Q5 p	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.36	0.01
N	146,161	30,542	27,095	146,161	30,542	27,095	146,161	30,542	27,095
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 8: Mroibvol bins and weekly returns in excess of the market. Week 0 refers to the focal week, -1 and 1 to the week before and after, respectively. Mroibvol is the marketable retail order imbalance, measured in the focal week. Dependent variable is return in excess of the equal-weighted market return. Panel A contains all weeks, while Panel B restricts to instances where week 0 contains an earnings announcement.

	CAR[0, 60]			ECAR[0, 60]		
	All	Low	High	All	Low	High
Retail Sells	-0.46***	-1.77***	1.06**	-0.35**	-0.67***	0.12
	(-2.88)	(-6.20)	(2.33)	(-2.50)	(-5.14)	(0.42)
2	-0.63***	-1.06***	1.10	-0.23**	-0.54***	0.64*
	(-4.97)	(-5.73)	(1.58)	(-1.96)	(-5.38)	(1.74)
3	-0.65***	-1.29***	0.77	-0.11	-0.34***	0.96*
	(-5.49)	(-7.70)	(1.05)	(-0.98)	(-3.12)	(1.80)
4	-0.15	-1.04***	2.26***	0.05	-0.26**	1.32***
	(-1.13)	(-5.12)	(3.15)	(0.40)	(-2.35)	(3.38)
Retail Buys	1.15***	-0.78***	3.34***	0.57***	-0.05	1.49***
	(7.19)	(-2.65)	(7.04)	(2.64)	(-0.33)	(2.92)
Q5-Q1	1.61	1.00	2.28	0.92	0.62	1.37
p(Q1=Q5)	0.00	0.01	0.00	0.00	0.00	0.01
N	140,437	31,860	18,274	140,437	31,860	18,274
R^2	0.00	0.01	0.01	0.00	0.01	0.00

Table 9: Return predictability decomposition. Firm-announcement level regressions of Cumulative Abonrmal Return or Expected Cumulative Abnormal Return on dummies representing quintiles of retail order imbalance (mroibvol) formed over the last ten trading days prior to the announcement. ECAR estimated using the contemporaneous relationship between returns and order imbalance in non-announcing firms. Day 0 refers to the first day on which the earnings news was tradeable. Columns Low and High restrict the sample to the first and fifth quintile of retail trading intensity. Standard errors clustered by trading day. First two rows of table footer report the difference between the high and low retail buying quintile and test against the null of equal coefficients.

	(-1, 0)			(-3, 2)		
	All	Pre	Post	All	Pre	Post
Size 1 x Q1	-0.226	-0.233**	-0.00925	-0.539	-0.414**	-0.165
	(-1.07)	(-3.04)	(-0.05)	(-1.83)	(-3.13)	(-0.70)
Size 1 x Q5	-1.279***	0.0966	-1.305***	-1.766***	0.492**	-2.163***
	(-4.77)	(0.88)	(-5.86)	(-4.63)	(2.61)	(-7.01)
Size 1	0.377*	0.312***	0.102	0.827***	0.496***	0.369*
	(2.54)	(5.76)	(0.76)	(3.81)	(5.45)	(2.04)
Size 2 x Q1	0.208	-0.0297	0.227	0.227	-0.0331	0.259
	(1.21)	(-0.50)	(1.49)	(1.08)	(-0.31)	(1.46)
Size 2 x Q5	-0.852***	0.0706	-0.902***	-0.710*	0.241	-0.924***
	(-3.90)	(0.90)	(-4.59)	(-2.34)	(1.42)	(-3.73)
Size 2	0.359**	0.0749	0.297**	0.391**	0.0956	0.306*
	(3.16)	(1.70)	(2.83)	(2.80)	(1.24)	(2.42)
Size 3 x Q1	0.0404	0.00508	0.0396	-0.00815	-0.0493	0.0538
	(0.26)	(0.09)	(0.28)	(-0.05)	(-0.57)	(0.33)
Size 3 x Q5	-0.783***	0.0978	-0.877***	-0.666**	0.186	-0.824***
	(-4.10)	(1.19)	(-4.81)	(-2.77)	(1.32)	(-3.80)
Size 3	0.425***	0.0221	0.408***	0.597***	0.0846	0.510***
	(4.05)	(0.53)	(4.03)	(4.72)	(1.30)	(4.34)
Size 4 x Q1	0.0279	0.0382	-0.0166	0.229	0.100	0.110
	(0.22)	(1.08)	(-0.14)	(1.38)	(1.45)	(0.78)
Size 4 x Q5	-0.567**	0.0696	-0.635***	-0.500*	0.113	-0.624***
	(-3.28)	(1.15)	(-4.04)	(-2.37)	(0.97)	(-3.49)
Size 4	0.340***	0.0419	0.302***	0.418***	0.135**	0.302**
	(3.74)	(1.71)	(3.54)	(4.08)	(2.72)	(3.22)
Size 5 x Q1	0.0425	-0.0136	0.0560	0.200	0.0758	0.129
	(0.39)	(-0.32)	(0.54)	(1.62)	(1.28)	(1.16)
Size 5 x Q5	-0.216	0.0583	-0.270**	-0.195	0.163	-0.361**
	(-1.75)	(0.94)	(-2.73)	(-1.15)	(1.87)	(-2.63)
Size 5	0.257***	0.125***	0.138*	0.372***	0.202***	0.176*
	(3.65)	(3.67)	(2.22)	(4.69)	(5.32)	(2.56)
Observations	30355	30360	30359	30363	30364	30363

Table 10: Cumulative returns around earnings announcements. Column headers refer to first and last days in return window. 0 is the first trading day on which announcement information is tradeable. Qk is an indicator variable for whether stock i was in retail intensity quintile k at the end of the month before the earnings announcement and Size j is an indicator for whether firm i was in size quintile j at the end of the month before the earnings announcement. Pre refers to trading days prior to announcement, post refers to trading days after announcement. Five panels sorted on size. Monthly fixed effects. Standard errors clustered by firm and month. Earnings announcements in 2007-2021.

10 References

References

Aase, K. K., Bjuland, T., and Oksendal, B. (2011). Insider trading with partially informed traders. *NHH Dept. of Finance & Management Science Discussion Paper*, (2011/21).

Appel, I. R., Gormley, T. A., and Keim, D. B. (2016). Passive investors, not passive owners. *Journal of Financial Economics*, 121(1):111–141.

Balasubramaniam, V., Campbell, J. Y., Ramadorai, T., and Ranish, B. (2023). Who owns what? a factor model for direct stockholding. *The Journal of Finance*, 78(3):1545–1591.

Bali, T. G., Brown, S. J., Murray, S., and Tang, Y. (2017). A lottery-demand-based explanation of the beta anomaly. *Journal of Financial and Quantitative Analysis*, 52(6):2369–2397.

Bali, T. G., Hirshleifer, D., Peng, L., and Tang, Y. (2021). Attention, social interaction, and investor attraction to lottery stocks. Working Paper.

Banerjee, S., Davis, J., and Gondhi, N. (2021). Choosing to disagree: Endogenous dismissiveness and overconfidence in financial markets. *Available at SSRN 3335257*.

Barber, B. M., De George, E. T., Lehavy, R., and Trueman, B. (2013). The earnings announcement premium around the globe. *Journal of Financial Economics*, 108(1):118–138.

Barber, B. M., Huang, X., Jorion, P., Odean, T., and Schwarz, C. (2022). A (sub) penny for your thoughts: Tracking retail investor activity in taq. Working Paper.

Barber, B. M., Huang, X., Ko, K. J., and Odean, T. (2020). Leveraging overconfidence. *Available at SSRN 3445660*.

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

Barber, B. M. and Odean, T. (2002). Online investors: do the slow die first? *The Review of Financial Studies*, 15(2):455–488.

Barber, B. M. and Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, 21(2):785–818.

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

Battalio, R. H., Jennings, R. H., Saglam, M., and Wu, J. (2023). Identifying market maker trades as' retail'from taq: No shortage of false negatives and false positives. *Available at SSRN 4579159*.

Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of Accounting Research*, pages 67–92.

Beber, A., Brandt, M. W., Cen, J., and Kavajecz, K. A. (2021). Mutual fund performance: Using bespoke benchmarks to disentangle mandates, constraints and skill. *Journal of Empirical Finance*, 60:74–93.

Benartzi, S. (2001). Excessive extrapolation and the allocation of 401 (k) accounts to company stock. *The Journal of Finance*, 56(5):1747–1764.

Boehmer, E., Jones, C. M., Zhang, X., and Zhang, X. (2021). Tracking retail investor activity. *The Journal of Finance*, 76(5):2249–2305.

Bogousslavsky, V. and Muravyev, D. (2024). An anatomy of retail option trading. Working Paper.

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

Bushee, B. J. (1998). The influence of institutional investors on myopic r&d investment behavior. *Accounting Review*, pages 305–333.

Campbell, J. Y., Lettau, M., Malkiel, B. G., and Xu, Y. (2001). Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk. *The Journal of Finance*, 56(1):1–43.

Chacko, G. C., Jurek, J. W., and Stafford, E. (2008). The price of immediacy. *The Journal of Finance*, 63(3):1253–1290.

Cohen, L. and Lou, D. (2012). Complicated firms. *Journal of Financial Economics*, 104(2):383–400.

Cohen, L., Malloy, C., and Nguyen, Q. (2020). Lazy prices. *The Journal of Finance*, 75(3):1371–1415.

Curcuru, S., Heaton, J., Lucas, D., and Moore, D. (2010). Heterogeneity and portfolio choice: Theory and evidence. In *Handbook of financial econometrics: Tools and techniques*, pages 337–382. Elsevier.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, 98(4):703–738.

de Silva, T., Smith, K., and So, E. C. (2023). Losing is optional: Retail option trading and expected announcement volatility. Working Paper.

Décaire, P. H., Sosyura, D., and Wittry, M. D. (2023). Model input choice and valuation disagreement. Working Paper.

DellaVigna, S. and Pollet, J. M. (2009). Investor inattention and friday earnings announcements. *The Journal of Finance*, 64(2):709–749.

Dhar, R. and Zhu, N. (2006). Up close and personal: Investor sophistication and the disposition effect. *Management Science*, 52(5):726–740.

Di Maggio, M., Franzoni, F., Kermani, A., and Sommavilla, C. (2019). The relevance of broker networks for information diffusion in the stock market. *Journal of Financial Economics*, 134(2):419–446.

Di Maggio, M., Franzoni, F. A., Kogan, S., and Xing, R. (2021). Avoiding volatility: Institutional trading before scheduled news. working Paper.

Diether, K. B., Malloy, C. J., and Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance*, 57(5):2113–2141.

Edmans, A., Fang, V. W., and Zur, E. (2013). The effect of liquidity on governance. *The Review of Financial Studies*, 26(6):1443–1482.

Engelberg, J. (2008). Costly information processing: Evidence from earnings announcements. In *AFA 2009 San Francisco meetings paper*.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Frazzini, A. and Lamont, O. A. (2007). The earnings announcement premium and trading volume. Working Paper.

Gabaix, X., Koijen, R. S. J., Mainardi, F., Oh, S., and Yogo, M. (2022). Asset demand of u.s. households. Working Paper.

Glosten, L., Nallareddy, S., and Zou, Y. (2021). Etf activity and informational efficiency of underlying securities. *Management Science*, 67(1):22–47.

Golubov, A. and Konstantinidi, T. (2021). Valuation uncertainty. Working Paper.

Gormsen, N. J. and Lazarus, E. (2021). Duration-driven returns. *Available at SSRN 3359027*.

Gormsen, N. J. and Lazarus, E. (2023). Duration-driven returns. *The Journal of Finance*, 78(3):1393–1447.

Greenwood, R., Laarits, T., and Wurgler, J. (2023). Stock market stimulus. *Review of Financial Studies*.

Greenwood, R. and Sammon, M. C. (2022). The disappearing index effect. Working Paper.

Haddad, V., Huebner, P., and Loualiche, E. (2021). How competitive is the stock market? theory, evidence from portfolios, and implications for the rise of passive investing. Working Paper.

Hartzmark, S. M. and Shue, K. (2018). A tough act to follow: Contrast effects in financial markets. *The Journal of Finance*, 73(4):1567–1613.

Hirshleifer, D., Lim, S. S., and Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance*, 64(5):2289–2325.

Hirshleifer, D. A., Myers, J. N., Myers, L. A., and Teoh, S. H. (2008). Do individual investors cause post-earnings announcement drift? direct evidence from personal trades. *The Accounting Review*, 83(6):1521–1550.

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

Hong, H. and Stein, J. C. (2007). Disagreement and the stock market. *Journal of Economic*

Perspectives, 21(2):109–128.

Huberman, G. (2001). Familiarity breeds investment. *The Review of Financial Studies*, 14(3):659–680.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91.

Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. (2016). A rational theory of mutual funds' attention allocation. *Econometrica*, 84(2):571–626.

Kaniel, R., Liu, S., Saar, G., and Titman, S. (2012). Individual investor trading and return patterns around earnings announcements. *The Journal of Finance*, 67(2):639–680.

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

Kim, O. and Verrecchia, R. E. (1994). Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics*, 17(1-2):41–67.

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, 132(2):665–712.

Kogan, S., Makarov, I., Niessner, M., and Schoar, A. (2023). Are cryptos different? evidence from retail trading. *Evidence from Retail Trading (March 20, 2023)*.

Koijen, R. S., Richmond, R. J., and Yogo, M. (2020). Which investors matter for equity valuations and expected returns? Working Paper.

Koijen, R. S. and Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, 127(4):1475–1515.

Kothari, S. P. and Sloan, R. G. (1992). Information in prices about future earnings: Implications for earnings response coefficients. *Journal of Accounting and Economics*, 15(2-3):143–171.

Krinsky, I. and Lee, J. (1996). Earnings announcements and the components of the bid-ask spread. *The Journal of Finance*, 51(4):1523–1535.

Kumar, A. and Lee, C. M. (2006). Retail investor sentiment and return comovements. *The Journal of Finance*, 61(5):2451–2486.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, pages 1315–1335.

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

Lev, B. and Gu, F. (2016). *The end of accounting and the path forward for investors and managers*. John Wiley & Sons.

Li, J. and Lin, Z. (2023). Price multipliers are larger at more aggregate levels. Technical report. Working Paper.

Livnat, J. and Mendenhall, R. R. (2006). Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research*, 44(1):177–205.

Luo, C., Ravina, E., Sammon, M., and Viceira, L. M. (2021). Retail investors' contrarian behavior around news and the momentum effect. Working Paper.

Ma, L., Tang, Y., and Gomez, J.-P. (2019). Portfolio manager compensation in the us mutual fund industry. *The Journal of Finance*, 74(2):587–638.

Martineau, C. and Mondria, J. (2022). News selection and asset pricing implications. Working Paper.

Martineau, C. and Zoican, M. (2019). Retail trading and analyst coverage. Working Paper.

McLean, R. D., Pontiff, J., and Reilly, C. (2020). Taking sides on return predictability. Working Paper.

Nagel, S. (2012). Evaporating liquidity. *The Review of Financial Studies*, 25(7):2005–2039.

Odean, T. (1998). Are investors reluctant to realize their losses? *The Journal of Finance*, 53(5):1775–1798.

Peng, L. and Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3):563–602.

Peters, R. H. and Taylor, L. A. (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics*, 123(2):251–272.

Savor, P. and Wilson, M. (2016). Earnings announcements and systematic risk. *The Journal of Finance*, 71(1):83–138.

Seasholes, M. S. and Zhu, N. (2010). Individual investors and local bias. *The Journal of Finance*, 65(5):1987–2010.

Stambaugh, R. F. and Yuan, Y. (2017). Mispricing factors. *The Review of Financial Studies*, 30(4):1270–1315.

Statman, M., Thorley, S., and Vorkink, K. (2006). Investor overconfidence and trading volume. *The Review of Financial Studies*, 19(4):1531–1565.

Vaarmets, T., Liivamägi, K., and Talpsepp, T. (2019). How does learning and education help to overcome the disposition effect? *Review of Finance*, 23(4):801–830.

van der Beck, P. (2022). Identifying elasticities in demand-based asset pricing. Working Paper.

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

Zhang, X. F. (2006). Information uncertainty and stock returns. *The Journal of Finance*, 61(1):105–137.

11 Online Appendix

A.1 Trade-off between signal precision and noise trading intensity in Kyle (1985)

In this section, we illustrate the trade-off between signal precision and noise trading intensity in the context of a 2-period Kyle (1985)-style model.[19] We would like to be clear that the model only has one asset, and our comparative statics are comparing investor outcomes *across equilibria*. That being said, our preferred interpretation is that, in practice, there are gains to specialization. So, one could view the stock market as many versions of the one-asset model running in parallel, and based on differences in nose trading intensity and difficulty to value, informed investors pick a single stock to specialize in.

Another important assumption in this model is that retail investors are pure noise traders, in the sense that their order flow is uncorrelated with securities' terminal payoffs. It may be, however, that retail investors have information about future fundamentals(Kaniel et al. (2008), Barrot et al. (2016), Boehmer et al. (2021)). We believe this baseline assumption of retail investors as classical noise traders is conservative, in the sense that it would not naturally discourage informed investors from learning about stocks with a large retail trader presence. As shown in Aase et al. (2011), if noise traders (i.e., retail investors in our setting) have order flow which is correlated with fundamentals, the insider's informational advantage would be relatively smaller, making informed investors less inclined to learn about high retail stocks. In other words, we would naturally bias informed investors from learning about high retail stocks if we modeled retail investors as having information.

A.1.1 Model setup

The model has two trading periods, $t = 1$ and $t = 2$. There is a single risky asset whose value is distributed:

$$v \sim N(0, \sigma_v^2) \tag{A1}$$

There is a *strategic* risk-neutral informed investor who receives an unbiased signal before the first trading period:

$$s = v + \epsilon \tag{A2}$$

where v is the true value of the asset and ϵ is signal noise. ϵ is independent of v and normally distributed with mean zero and standard deviation σ_ϵ. This implies that $s \sim N(v, \sigma_\epsilon^2)$.

The informed investor submits demands to a set of *competitive* risk-neutral market makers at times

[19] This section borrows heavily from Alex Chinco's "Two Period Kyle (1985) Model" notes.

1 and 2, y_1 and y_2. To prevent prices from being fully revealing, there are a group of noise traders who submit random demands z_1 and z_2, where the z_t are independent and normally distributed with mean zero and standard deviation σ_z.

The set of competitive market makers observe total order flow x_t each period:

$$x_t = y_t + z_t \tag{A3}$$

There is perfect competition among market makers, so they must set prices equal to the expected fundamental value of the asset given total demand:

$$p_1 = E[v|x_1] \quad \text{and} \quad p_2 = E[v|x_1, x_2] \tag{A4}$$

In period 1, the informed investor chooses demand y_t to solve:

$$H_0 = max_{y_1} E\left[(v - p_1)\, y_1 + H_1|s\right] \tag{A5}$$

where H_{t-1} is the informed investor's value function entering period t.

In period 2, they choose y_2 to maximize:

$$H_1 = max_{y_2} E\left[(v - p_1)\, y_2|s, p_1\right] \tag{A6}$$

An equilibrium is made up of two components: (1) a linear demand rule for the informed investor in each period:

$$y_t = \alpha_{t-1} + \beta_{t-1} s \tag{A7}$$

And (2) a liner pricing rule for the market makers in each period:

$$p_t = \kappa_{t-1} + \lambda_{t-1} x_t \tag{A8}$$

The informed investor updates their beliefs about v after observing s. Their posterior beliefs about the mean and variance are:

$$\mu_{v|s} = \left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_\epsilon^2}\right) \times \quad \text{and} \quad \sigma_{v|s}^2 = \left(\frac{\sigma_\epsilon^2}{\sigma_v^2 + \sigma_\epsilon^2}\right) \times \sigma_v^2 \tag{A9}$$

where going forward, I will use θ in place of $\left(\frac{\sigma_v^2}{\sigma_v^2+\sigma_\epsilon^2}\right)$.

The market makers extract an unbiased signal about v from total demand. Substituting in the informed trader's demand rule, the $t = 1$ signal is:

$$v = \frac{x_1}{\beta_0} - \epsilon - \frac{z_1}{\beta_0} \tag{A10}$$

This implies that the market makers' posterior beliefs after observing x_1 are:

$$\mu_{v|x_1} = \left(\frac{\beta_0^2\sigma_v^2}{\beta_0^2\sigma_s^2 + \sigma_z^2}\right) \times x_1 \quad \text{and} \quad \sigma^2_{v|x_1} = \left(\frac{\beta_0^2\sigma_\epsilon^2 + \sigma_z^2}{\beta_0^2\sigma_s^2 + \sigma_z^2}\right) \times \sigma_v^2 \tag{A11}$$

Another way to think about this is that the total order flow x_1 is a signal about the informed trader's signal s rather than the fundamental value of the asset v. This would imply the $t = 1$ signal is:

$$s = \frac{x_1}{\beta_0} - \frac{z_1}{\beta_0} \tag{A12}$$

which gives posterior beliefs:

$$\mu_{s|x_1} = \left(\frac{\beta_0^2\sigma_s^2}{\beta_0^2\sigma_s^2 + \sigma_z^2}\right) \times x_1 \quad \text{and} \quad \sigma^2_{s|x_1} = \left(\frac{\sigma_z^2}{\beta_0^2\sigma_s^2 + \sigma_z^2}\right) \times \sigma_s^2 \tag{A13}$$

A.1.2 Solving the model

Given the market makers' zero profit condition, $\kappa_0 = 0$ and

$$\kappa_1 = E[v|x_1] - \lambda_1 E[x_2|x_1] = p_1 - \left(\theta\mu_{s|x_1} - p_1\right) = p_1 \tag{A14}$$

where the last equality comes from $\theta\mu_{s|x_1} = p_1$.

Substituting in the market makers' linear pricing rule into H_1

$$H_1 = max_{y_2} E\left[(v - \kappa_1 - \lambda_1 x_2)\, y_2 | s, p_1\right] \tag{A15}$$

Taking the first order condition with respect to y_2 yields optimal demand:

$$y_2 = -\frac{p_1}{2\lambda_1} + \frac{\theta}{2\lambda_1} s \tag{A16}$$

so $\alpha_1 = -\frac{p_1}{2\lambda_1}$ and $\beta_1 = \frac{\theta}{2\lambda_1}$.

With this, we can partially solve for the market makers' price impact coefficient, λ_t, in period 2:

$$\lambda_1 = \frac{Cov[x_2, v|x_1]}{Var[x_2|x_1]} = \frac{\beta_1\sigma^2_{v|x_1}}{\beta_1^2\sigma^2_{s|x_1} + \sigma_z^2} \tag{A17}$$

Now, turning to the period one solution, we start by taking a guess at at the informed investors' value function which we will verify later:

$$E[H_1|s] = \phi_1 + \omega_1 \left(\mu_{v|s} - p_1\right)^2 \tag{A18}$$

Substituting in the price impact and demand coefficients into H_0 yields:

$$H_0 = max_{y_1} E\left[\left(v - p_1\right) y_1 + \phi_1 + \omega_1 \left(\theta s - p_1\right)^2 |s\right] \tag{A19}$$

Taking the first order condition with respect to y_1 implies:

$$y_1 = \frac{\theta}{2\lambda_0}\left(\frac{1 - 2\omega_1\lambda_0}{1 - \omega_1\lambda_0}\right) s \tag{A20}$$

With all this, we can now solve for the time 1 price impact coefficient:

$$\lambda_0 = \frac{Cov[x_1, v]}{Var[x_1]} = \frac{\beta_0\sigma_v^2}{\beta_0^2\sigma_s^2 + \sigma_z^2} \tag{A21}$$

To verify the guess about H_1, substitute the equilibrium coefficients for demands and prices into Equation A18:

$$H_1 = \left[\frac{1}{2\lambda_1}\left([v - \theta s] + \frac{1}{2}[\theta s - p_1] - \lambda_1 z_2\right)(\theta s - p - 1)\,|s\right] \tag{A22}$$

which simplifies to:

$$H_1 = \text{Constant} + \frac{1}{4\lambda_1}\left(\mu_{v|s} - p_1\right)^2 \tag{A23}$$

This reveals that $\omega_1 = \frac{1}{4\lambda_1}$ and that H_1 is consistent with the original guess.

To solve the model, start with some initial guess for $\hat{\beta_0}$, and use this to compute the other equilibrium coefficients. This can be done in stages, first computing $\hat{\sigma^2_{v|x_1}}$ and $\hat{\sigma^2_{s|x_1}}$, and then using these to compute $\hat{\lambda_1}$:

$$\begin{aligned}
\hat{\lambda_0} &= \frac{\hat{\beta_0}\sigma_v^2}{\hat{\beta_0^2}\sigma_s^2 + \sigma_z^2} \\
\hat{\sigma^2_{v|x_1}} &= \frac{\hat{\beta_0^2}\sigma_\epsilon^2 + \sigma_z^2}{\hat{\beta_0^2}\sigma_s^2 + \sigma_z^2}\sigma_v^2 \\
\hat{\sigma^2_{s|x_1}} &= \frac{\sigma_z^2}{\hat{\beta_0^2}\sigma_s^2 + \sigma_z^2}\sigma_s^2 \\
\hat{\lambda_1} &= \frac{1}{\sigma_z}\sqrt{\frac{\theta}{2}\left(\hat{\sigma^2_{v|x_1}} - \frac{\theta}{2}\hat{\sigma^2_{s|x_1}}\right)}
\end{aligned} \tag{A24}$$

A solution has been found when you have minimized the distance between the guess $\hat{\beta_0}$ and $\frac{\theta}{2\hat{\lambda_0}}\left(\frac{1-2\hat{\omega_1}\hat{\lambda_0}}{1-\hat{\omega_1}\hat{\lambda_0}}\right)$, which is a condition $\hat{\beta_0}$ has to satisfy in equilibrium.

A.1.3 Simulation Results

For each set of parameters, we simulate the economy 10,000 times and compute averages of the insider's total profit, defined as $x_1 \times (v - p_1) + x_2 \times (v - p_2)$. Figure A1 plots the insider's profit against the imprecision of their signal (σ_ϵ) for several values of noise trading intensity σ_z. In all simulations, we set fundamental volatility, σ_v, to one.

Unsurprisingly, the insider's profit is monotonically decreasing in signal imprecision (i.e., moving from left to right), and is monotonically increasing in noise trader intensity. The more interesting result is that the insider's profit can be lower in a high noise trading intensity stock (e.g., the stock represented by the yellow line) than a low noise trading intensity stock (e.g., the stock represented by the blue line) if the precision of their signal is sufficiently higher in the low noise trading intensity stock.

Mapping this back to our empirical results, our preferred interpretation of a hard to value stock is one that, for a given amount of learning energy expended, investors receive a relatively less precise signal. Specifically, consider the multi-asset noisy rational expectations equilibrium with endogenous learning of Kacperczyk et al. (2016). Suppose that an investor i can allocate attention K_{ij} to asset j to receive a signal with variance $1/\left(c_j K_{ij}\right)$ i.e., with a precision of $c_j K_{ij}$ (the model features a transformation which makes the assets are uncorrelated). Further, suppose that each investor i has an overall attention budget of K, so $\sum_{j=1}^{J} K_{ij} < K$.

In this setting, c_j is an asset j-specific parameter that governs how easy it is to learn about that asset. Lower values of c_j imply that for a given amount of learning (K_{ij}), investors will receive a less precise signal i.e., lower values of c_j imply that a stock is harder to value. So, mapping this back to the example in Figure A1, suppose that the red line represents a low retail stock ($\sigma_z = 1$) while the yellow line represents a high retail stock ($\sigma_z = 2$). Suppose that the total attention constraint $K = 4$, $c_{red} = 1$ and $c_{yellow} = 0.1$. If an investor allocates all their attention to the low retail stock, their signal will have variance $1/4$ (standard deviation 0.5), while if the investor allocates all their attention to the high retail stock, their signal will have variance 2.5 (standard deviation of about 1.58). With these parameters, the informed investor would make more money allocating all their attention to the low retail stock than to the high retail stock, even though the high retail stock has double the noise trading intensity.

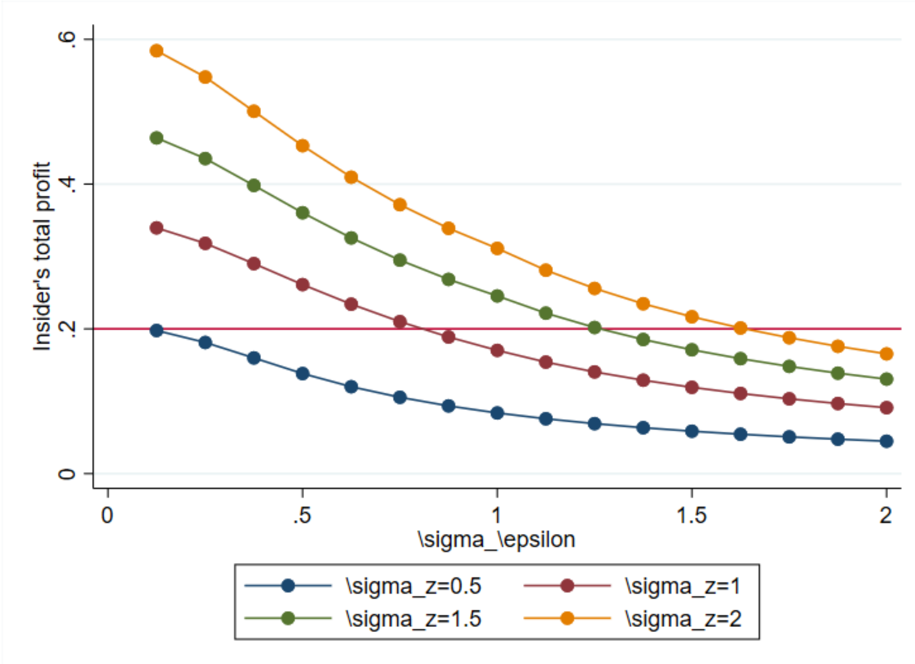

Figure A1: Insider's profits as a function of noise trading intensity and signal precision. Each point represents the average of the insider's total profit in periods 1 and 2 across 10,000 simulations. Fundamental volatility (σ_v) is fixed at 1.

A.1.4 Mapping the model to the data

To map the model to our empirical setting, we would like to develop a measure which captures both (1) limited attention and (2) heterogeneity in ease of learning about different stocks. While neither of these quantities are directly observable, in this subsection, we aim to show empirical evidence that high retail stocks are harder to learn about than low retail stocks.

We focus on the forecasts of sell-side analysts. These forecasts give us a way of measuring the accuracy of signals received about a given stock. Suppose, as a starting premise, that analysts have limited attention. Then we would expect that as a given analyst covers more stocks, the accuracy of their forecasts *on all the stocks they cover* should decrease.

To quantify this, for each analyst j, we can compute how many stocks (is) that analyst is covering at a given point in time t: Num. Stocks Covered$_{i,j,t}$. We believe that the simplest way of capturing limited attention is to divide measures of analyst forecast errors by Num. Stocks Covered$_{i,j,t}$. The intuition here is that analysts equally spread their attention over the stocks they cover, and therefore their prediction accuracy is expected to decrease at rate 1/Num. Stocks Covered$_{i,j,t}$. There may be other reasons that analysts produce inaccurate forecasts, including the nature of the firm, the ease of learning about the underlying fundamentals and analyst skill, which we will try to account for in our empirical design.

We use this logic to construct two measures of scaled analyst (in)accuracy. First, we define scaled analyst forecast inaccuracy for earnings per share (EPS) as:

$$\text{Accuracy}^{earnings}_{i,j,t} = \frac{|\text{Est}_{i,j,t} - \text{Actual}_{i,t}|/Prc_{i,t}}{\text{Num. Stocks Covered}_{i,j,t}} \tag{A25}$$

where $\text{Est}_{i,j,t}$ is analyst j's estimate for stock i's EPS in quarter t, $\text{Actual}_{i,t}$ is the realized EPS in quarter t and $Prc_{i,t}$ is the last closing price before earnings were released. To avoid using stale forecasts, we only use forecasts from the last IBES statistical period before earnings are actually released i.e., we are only using one statistical period for each earnings announcement. In words, $\text{Accuracy}^{earnings}_{i,j,t}$ is the measure of standardized unexpected earnings from Hartzmark and Shue (2018) constructed at the analyst level, and scaled by Num. Stocks $\text{Covered}_{i,j,t}$.

Similarly, we define scaled analyst forecast inaccuracy for prices as:

$$\text{Accuracy}^{price}_{i,j,t} = \frac{|\text{Est}_{i,j,t} - \text{Actual}_{i,t}|/\text{Actual}_{i,t}}{\text{Num. Stocks Covered}_{i,j,t}} \tag{A26}$$

where $\text{Est}_{i,j,t}$ is analyst j's estimate for stock i's price at time t and $\text{Actual}_{i,t}$ is the realized price at time t. When computing $\text{Accuracy}^{price}_{i,j,t}$ we exclusively use 12-month ahead price forecasts. Note that there are many potentially stale price forecasts (because price forecasts are not associated with a particular event like earnings forecasts), so we discard any forecast made 90 days before the associated IBES statistical period – which correspond to each calendar month. As in the main body of the paper, all the inputs used to compute $\text{Accuracy}^{earnings}_{i,j,t}$ and $\text{Accuracy}^{price}_{i,j,t}$ (except the pre-earnings announcement price) come from the IBES unadjusted files – and we manually apply the cumulative factor to adjust price from CRSP.

To test whether high retail stocks are harder to value than low retail stocks, we run the following regression:

$$100\times\text{Accuracy}_{i,j,t} = \alpha+\beta_1 Q1_{i,t-1}+\beta_2 Q2_{i,t-1}+\beta_4 Q4_{i,t-1}+\beta_5 Q5_{i,t-1}+\gamma X_{i,t-1}+\phi_1\text{staleness}_{i,j,t}+a_t+b_j+\epsilon_{i,j,t} \tag{A27}$$

where $Q1$, $Q2$, $Q4$ and $Q5$ are quintiles of retail trading intensity in month $t-1$ (the middle quintile, $Q3$, is the omitted group). $X_{i,t-1}$ are a set of stock-specific control variables which includes essentially all the variables in Tables 3, 2 and A9. a_t are a set of time fixed effects, while b_j are a set of fixed effects for each analyst. The b_j fixed effects should capture differences in skill across analysts. Finally, $\text{staleness}_{i,j,t}$ is the time (in days) between the IBES statistical period and the date the forecast was made – capturing the staleness of a given forecast. Ex-ante, one would expect ϕ_1 to be positive, as more stale forecasts are likely less accurate on average. The left hand side variables are multiplied by 100 to ease the interpretation of the coefficients. Standard errors are triple clustered at the stock, analyst and time level.

Given analysts' limited attention, if high retail stocks are harder to value, one might expect analysts' accuracy to be relatively lower for such stocks *conditional on how many stocks they cover*, their skill and the nature of the firm's they're covering. Therefore, we might expect β_5 to be positive, and β_1 to be negative.

Table A1 contains the results. In columns 1-3, the left-hand-side variable is $100 \times \text{Accuracy}_{i,j,t}^{earnings}$. The unit of observation is analyst-stock-quarter to match the frequency of earnings announcements. Column 1 shows that analysts' relative accuracy is significantly lower (recall that higher values of $\text{Accuracy}_{i,j,t}^{earnings}$ denote *less* accurate forecasts) for high retail stocks than low retail stocks. In fact, the relationship between forecast inaccuracy and quintile of retail trading intensity is monotonic. To better frame the magnitudes of the estimated effects, the mean of $100 \times \text{Accuracy}_{i,j,t}^{earnings}$ is 0.058, so the difference in accuracy between the top and the bottom retail quintiles is more than two times the mean.

Column 2 adds analyst fixed effects to account for heterogenity in analyst skill. The point estimates are hardly changed by including these fixed effects, suggesting differences in analyst skill are not a key driver of these results. Finally, column 3 adds a large suite of variables which we show in the main body of the paper are correlated with retail trading activity: market capitalization, firm age, returns from month $t-12$ to month $t-2$, the MAX factor of Bali et al. (2017) (a measure of lottery demand), book-to-market, earnings-to-price, market beta, standard deviation of daily returns, intraday idiosyncratic volatility computed from trades, Kyle's lambda, the effective bid-ask spread, intangible capital divided by market capitalization and the dollar value of patents divided by market capitalization. We omit cashflow duration (Gormsen and Lazarus, 2021), valuation uncertainty (Golubov and Konstantinidi, 2021) and mispricing (Stambaugh and Yuan, 2017) because that would dramatically shrink our sample of observations based on data availability. While including all these firm level controls does shrink the estimated magnitudes – the difference between the 1st and 5th quintile is still statistically significant, and is on par with the magnitude of the unconditional mean.

In columns 4-6, the left-hand-side variable is $100 \times \text{Accuracy}_{i,j,t}^{price}$. The unit of observation is analyst-stock-month, to match the frequency of IBES statistical periods associated with price forecasts. Column 4 shows a large difference in price forecast accuracy for high and low retail stocks. Again to better frame the magnitudes of the estimated effects, the mean of $100 \times \text{Accuracy}_{i,j,t}^{price}$ is 8.06. So the difference between the top and bottom quintiles of retail activity is economically large.

In column 5, we include analyst fixed effects, which shrink the point estimates by slightly less than a factor of 2, but the difference between the high and low retail quintiles remains statistically significant. Finally in column 6, we also include all the firm-level controls, and again the relationship between relative inaccuracy is monotonically increasing from low to high retail.

So, to summarize, Table A1 shows that conditional on the number of stocks an analyst is covering, time-varying average analyst accuracy, heterogeneity in analyst skills and differences in fundamen-

	Forecasts of Next Quarter Earnings (|Est-Actual|/Prc)/Num Stocks Covered			Forecasts of price in 12 months (|Est-Actual|/Actual)/Num Stocks Covered		
	(1)	(2)	(3)	(4)	(5)	(6)
Low Retail	-0.0138***	-0.0166***	-0.00947***	-1.438***	-0.974***	-0.824***
	(0.003)	(0.004)	(0.002)	(0.218)	(0.175)	(0.193)
2	-0.0127***	-0.0120***	-0.00557***	-0.982***	-0.598***	-0.423***
	(0.003)	(0.003)	(0.002)	(0.145)	(0.122)	(0.137)
4	0.0334***	0.0315***	0.0126***	2.825***	1.965***	1.308***
	(0.007)	(0.007)	(0.004)	(0.348)	(0.264)	(0.205)
High Retail	0.132***	0.118***	0.0421***	11.43***	6.356***	3.526***
	(0.020)	(0.021)	(0.008)	(0.914)	(0.529)	(0.450)
Staleness	0.0000857	0.000215***	0.000191**	0.0172***	0.0116***	0.0135***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.002)
Observations	1,024,786	1,024,339	1,024,339	1,939,687	1,939,631	1,939,631
R-squared	0.059	0.128	0.179	0.097	0.28	0.291
Fixed Effects	YQ	YQ/Analyst	YQ/Analyst	YM	YM/Analyst	YM/Analyst
Firm-Level Controls	NO	NO	YES	NO	NO	YES
Clustering		Stk/YQ/Analyst			Stk/YM/Analyst	

Table A1: Individual analyst accuracy and retail trading intensity. Results from the following regression:

$$100{\times}\text{Accuracy}_{i,j,t} = \alpha+\beta_1 Q1_{i,t-1}+\beta_2 Q2_{i,t-1}+\beta_4 Q4_{i,t-1}+\beta_5 Q5_{i,t-1}+\gamma X_{i,t-1}+\phi_1 \text{staleness}_{i,j,t}+a_t+b_j+\epsilon_{i,j,t}$$

where Qk denote quintiles of retail trading activity. The firm-level controls in X are market capitalization, firm age, returns from month $t-12$ to month $t-2$, the MAX factor of Bali et al. (2017) (a measure of lottery demand), book-to-market, earnings-to-price, market beta, standard deviation of daily returns, intraday idiosyncratic volatility computed from trades, Kyle's lambda, the effective bid-ask spread, intangible capital divided by market capitalization and the dollar value of patents divided by market capitalization. a_t are time fixed effects, while b_j are analyst fixed effects. The unit of observation for columns 1-3 is stock-quarter-analyst, while the unit of observation for columns 4-6 is stock-month-analyst. All left-hand-side variables have been multiplied by 100 to ease the interpretation of the associated coefficients. Standard errors are triple clustered at the stock, analyst and time levels.

tals across firms, analysts produce relatively less accurate forecasts for high retail stocks than low retail stocks. Further, these effects are economically large – on par with the magnitude of unconditional average accuracy. This is consistent with high retail stocks being harder to value – because for a given (expected) amount of attention spent – the analysts' estimates are further from the truth.

A.2 Validation

As discussed in Barber et al. (2022) and Battalio et al. (2023), the algorithm described in Boehmer et al. (2021) (BJZZ) may have both false positives and false negatives when identifying retail-initiated trades. For most of our applications, we are not interested in individual trades, but rather the ranking of stocks on retail trading intensity. To validate the use of the BJZZ algorithm in our setting, we compare our ranking of stocks on retail trading intensity to rankings based on other measures of retail trading activity in Table A2.

In panel A, we form 5 quintiles based on the number of Robinhood users from Robintrack, and compare these to 5 quintiles formed on retail trading intensity, defined as $\text{RSVOL}_{i,t} = \frac{\text{RBuy}_{i,t}+\text{RSell}_{i,t}}{\text{Volume}_{i,t}}$. Although we expect these two quantities to be related, they may not be perfectly correlated, as the number of users holding a stock is not necessarily a measure of trading intensity.[20] We find that almost 70% of stocks in the bottom quintile of retail trading intensity are in the bottom two quintiles of Robintrack users. Similarly, over 60% of stocks in the top quintile of retail trading intensity are in the top two quintiles of Robintrack users.

In panels B and C of Table A2, we form 5 quintiles based on fraction of total volume coming from internalized orders at Citadel and Virtu, two of the largest wholesalers for retail order flow. Owing to SEC rule 605, wholesalers need to make available on their websites data with statistics on price improvement for their internalized orders. We define wholersaler internalization intensity as total shares from internalized trades at each wholesaler divided by total volume in CRSP. We show that the overlap between these wholesaler-based measures and retail trading intensity calculated using BJZZ is even higher than the overlap with Robintrack activity in panel A. For example, almost 100% of stocks in the top quintile of retail trading intensity are in the top two quintiles of wholesaler internalization intensity for both Citadel and Virtu. Similarly, nearly all stocks in the bottom quintile of retail trading intensity are in the bottom two quintiles of internalization intensity.

Part of this relationship is mechanical. The BJZZ algorithm is designed to identify internalized orders that receive price improvement. And wholesalers report orders they *choose* to internalize (there is always the option to send the order directly to the exchange), which may be those that

[20] As discussed in Luo et al. (2021), in their dataset from a large discount retail brokerage, a small number of day-traders make up the majority of dollar trading volume.

receive price improvement, as offering price improvement is part of satisfying rule 605. That being said, at a high level, the results in Table A2 show that raking stocks on our BJZZ-based measure is consistent with ranking stocks on other measures of retail trading activity.

Panel A: Quintiles of Robintrack Users

		Low	2	3	4	High
Quintile of Retail Trading Intensity	Low	36.3%	32.6%	19.3%	9.3%	2.5%
	2	21.8%	26.8%	25.5%	18.4%	7.5%
	3	17.4%	17.5%	21.7%	25.5%	17.9%
	4	14.5%	10.4%	16.6%	22.3%	36.2%
	High	9.8%	12.4%	16.8%	24.5%	36.6%

Panel B: Quintiles of Virtu 605 Trades

		Low	2	3	4	High
Quintile of Retail Trading Intensity	Low	69.6%	24.5%	5.2%	0.6%	0.1%
	2	26.2%	49.8%	21.4%	2.5%	0.1%
	3	3.6%	24.6%	54.2%	17.0%	0.5%
	4	0.3%	1.1%	19.2%	65.7%	13.7%
	High	0.1%	0.1%	0.3%	14.4%	85.2%

Panel C: Quintiles of Citadel 605 Trades

		Low	2	3	4	High
Quintile of Retail Trading Intensity	Low	75.6%	21.4%	2.5%	0.2%	0.3%
	2	22.0%	57.0%	19.8%	1.1%	0.0%
	3	2.3%	21.0%	61.0%	15.3%	0.3%
	4	0.1%	0.5%	16.9%	70.6%	11.8%
	High	0.1%	0.1%	0.0%	12.9%	86.9%

Table A2: Validation. Each month, we form 5 quintiles on retail traing intensity, defined as $\text{RSVOL}_{i,t} = \frac{\text{RBuy}_{i,t}+\text{RSell}_{i,t}}{\text{Volume}_{i,t}}$. In each panel, we compute the percentage of firms in each retail trading intensity quintile that fall into quintiles formed on other measures of retail investor activity. In panel A, we form quintiles based on the number of Robintrack users. In panels B and C, we compute quintiles based on wholsaler internalization intensity, defined as internalized order volume divided by total CRSP volume, for Virtu and Citadel.

A.3 Variable definitions

This section contains a list of all the variables used in our empirical exercises and their associated definitions.

- **Retail Trading:** the sum of retail-initiated buys and retail-initiated sells divided by total trading volume in CRSP, multiplied by 100.
- **Turnover:** total trading volume in CRSP divided by shares outstanding, multiplied by 100.

- **Retail Turnover:** retail initiated buys plus retail initiated sells divided by shares outstanding, multiplied by 100.
- **Cap:** market capitalization, calculated as price times shares outstanding.
- **Age:** time in months since listing in CRSP.
- **Prc:** nominal price.
- **Past R:** cumulative return from month $t - 12$ to month $t - 2$.
- **B/M:** book to market ratio from the WRDS ratios suite.
- **E/P:** earnings to price ratio from the WRDS ratios suite.
- $\beta_{\mathbf{CAPM}}$**:** beta with respect to the market factor from the WRDS beta suite.
- **100-Inst.:** 100 minus the percentage of the stock's shares owned by institutional investors, defined as all 13F-filing institutions.
- **CF:** cashflow duration from Gormsen and Lazarus (2023). A composite score with high duration firms having high investment, low profitability, low beta, and low payout.
- $K_{\mathbf{Int}}$**:** the sum of capitalized SG&A and R&D over market capitalization from Peters and Taylor (2017).
- $K_{\mathbf{Know}}$**:** capitalized R&D over market capitalization from Peters and Taylor (2017).
- $K_{\mathbf{Org}}$**:** capitalized SG&A over market capitalization from Peters and Taylor (2017).
- **PAT:** market value of patents over the past 5 years, divided by market capitalization, as constructed in Kogan et al. (2017).
- **VU:** valuation uncertainty from Golubov and Konstantinidi (2021). A score calculated using the distribution of possible valuations based on multiples analysis, with wider distributions implying more valuation uncertainty.
- **Mispricing:** a composite score from Stambaugh and Yuan (2017) based on rankings on various asset pricing anomalies, with higher rankings denoting more mispricing.
- **SD:** standard deviation of daily returns.
- **Ivol**$_t$**:** trade-based intraday volatility, computed as the sum of squared (centered) 1-second returns, following Holden and Jacobsen (2014).
- λ**2:** the coefficient from a regression of percentage change in price on the square-root of signed dollar order imbalance (Holden and Jacobsen, 2014). Higher values denote more price impact for a given order imbalance. For more detail, see the *WRDS Intraday Indicators Formula Note.*

- **Espread:** percentage effective bid-ask spread, defined as the weighted average percent difference between the trade price and the midpoint, where the average is taken across all trades in a given day, and the weights are proportional to the dollar value of each trade.

- **Rspread:** Percentage realized bid-ask spread, defined as the weighted average percent difference between the trade price and the midpoint five minutes after the trade, where the average is taken across all trades in a given day, and the weights are proportional to the dollar value of each trade.

- **Ann. Return:** the return on the first trading day the earnings information could have been traded on during normal market hours. For example, if earnings were released at 2pm on a trading day, that day's return will be the announcement day return. If earnings were released at 5pm on a trading day, the next trading day's return will be the announcement return.

- **SUE:** standardized unexpected earnings. Actual earnings minus mean expected earnings, divided by the price the day before the earnings announcement date. Both the actual and expected earnings are from the IBES unadjusted summary file.

- **Idio. SUE:** idiosyncratic standardized unexpected earnings. Following Glosten et al. (2021), we estimate a regression of SUE on market-wide value-weighted average SUE and SIC-2 value-weighted average SUE in 5-year rolling windows. The idiosyncratic component of SUE are the residuals from this regression in the last year of the 5-year rolling window.

- **Analysts Disp.:** standard deviation of analyst estimates from the IBES unadjusted summary file.

- Market-adjusted return: Return minus return on the value-weighted market portfolio, following Campbell et al. (2001).

- **mroibvol:** marketable retail order imbalance. The ratio of retail-initiated buys minus retail-initiated sells to retail-initiated buys plus retail-initiated sells (Boehmer et al., 2021).

A.4 Retail favoring versus institutional avoidance

As discussed in the main body of the paper there are multiple reasons why a stock could be classified as having a high retail share of total trading volume. One way is that the stock has high retail-initiated turnover (RTO), defined as retail buys plus retail sells, divided by shares outstanding. For example, in 12/2019, when sorting stocks into 5 portfolio on RTO, the bottom portfolio has an average RTO of 13 basis points, while the top portfolio has an average RTO of 4.14%. To put this last number in perspective, across all stocks in 12/2019, the average total monthly turnover is 18%. And for reasons discussed in the main body of the paper, our estimate of retail-initiated turnover may understate true trading by retail investors due to the the methodology in Boehmer et al. (2021) only identifying internalized retail market orders.

Another way a stock could be classified as high retail, however, is through institutional investor avoidance. Specifically, one could imagine a stock that has relatively low RTO, but also low overall turnover, thus making the retail *share* of trading volume relatively larger. This is because the retail share of trading of trading volume is the ratio of RTO to overall turnover. This raises the concern that a sort on the retail share of trading volume is actually a sort on overall turnover.

To determine whether a stock is high retail because retail investors favor the stock, or institutions avoid the stock, we perform a double sort. First, we sort stocks into quintiles on overall turnover and then, within each of these quintiles, we sort into 5-sub portfolios based on retail-initiated turnover. The idea is that a stock in one of the relatively low RTO portfolios may still have a high retail share of trading volume if it is in a low overall turnover portfolio. Similarly, a stock with relatively high retail-initiated turnover may have a low retail share of trading volume if institutions heavily trade those stocks i.e., overall turnover is relatively high.

Table A3 contains the results. Panel A reports overall turnover, defined as total shares traded in a month divided by shares outstanding. By construction, this is increasing from left to right, as this is the first dimension of the sort. Panel B reports retail-initiated turnover, defined as total retail buys plus total retail sells in a month, divided by shares outstanding. Again, by construction, this is increasing from top to bottom, as this is the second dimension of the sort.

Panel C reports the retail share of trading volume, defined as total retail buys plus total retail sells in a month, divided by total trading volume. Note that this is not one of the variables used in the double sort (although it is the ratio of retail-initiated turnover to overall turnover). That being said, it's unsurprising that we see this quantity monotonically increasing from top to bottom because high RTO stocks are expected to have a higher retail share of trading volume.

Another feature of this panel is that the relationship with retail share of trading volume and turnover is U shaped, in the sense that it is highest both among very low and very high turnover stocks. The U-shape, however, is not very steep. In fact, within each quintile formed on overall turnover, there is a similar gap in the retail share of trading volume between high and low RTO stocks. This suggests that our sort in the main body of the paper on retail share of trading volume is indeed its own dimension of cross-sectional heterogeneity, and not an indirect sort on e.g., turnover.

A.5 Persistence of Retail Trading

As discussed in the main body of the paper, retail trading activity is persistent. Table A4 shows that stocks in the highest quintile in terms of retail share of trading have a 66% probability of remaining in the top quintile, and an almost 90% probability of remaining in the top two quintiles

Panel A: Overall Turnover

		Quintile of Turnover (1st sort)				
		1	2	3	4	5
RTO Quintile (2nd Sort)	1	3.00	9.38	14.50	21.43	36.44
	2	4.14	9.81	14.91	22.27	40.03
	3	4.56	10.00	15.16	22.88	44.89
	4	4.69	10.02	15.33	23.36	52.86
	5	5.09	9.98	15.31	23.64	114.66

Panel B: Retail-Initiated Turnover

		Quintile of Turnover (1st sort)				
		1	2	3	4	5
RTO Quintile (2nd Sort)	1	0.06	0.16	0.24	0.37	0.77
	2	0.13	0.25	0.36	0.57	1.43
	3	0.21	0.34	0.48	0.77	2.38
	4	0.33	0.49	0.68	1.13	4.34
	5	0.84	1.17	1.62	2.61	17.44

Panel C: Retail Share of Trading Volume

		Quintile of Turnover (1st sort)				
		1	2	3	4	5
RTO Quintile (2nd Sort)	1	3.81	2.19	2.09	2.18	2.77
	2	4.99	3.10	2.88	3.00	4.33
	3	6.53	4.09	3.71	3.96	6.38
	4	9.39	5.83	5.17	5.62	9.72
	5	16.21	12.58	11.49	12.04	15.47

Table A3: Results of double sort on overall turnover and retail-initiated turnover. Panel A reports overall turnover, defined as total shares traded in a month divided by shares outstanding. Panel B reports retail-initiated turnover, defined as total retail buys plus total retail sells in a month, divided by shares outstanding. Panel C reports the retail share of trading volume, defined as total retail buys plus total retail sells in a month, divided by total trading volume.

12 months in the future.

Panels B and C of Table A4 repeat the same transition-probability analysis, but also condition on the market capitalization of the stock at time $t = -12$. Again we see substantial persistence in portfolio assignments over time. Among small stocks (those in the bottom 20% of market capitalization) with the highest share of retail trading, over 70% are in the top two quintiles 12 months later. Among large stocks (those in the top 20% of market capitalization) this persistence is considerably stronger, a full 90% of stocks in the high retail quintle are in the top two quintiles 12 months later, with over 66% staying in the top bin.

A.6 Examples of stocks that went from low to high retail trading activity

A natural question is how stocks go from having low retail trading intensity to high retail trading intensity. One such company is Hertz, which declared bankruptcy in May 2020. Figure A2 shows gross retail activity in Hertz steadily climbing before the bankruptcy, and continuing to increase to almost 30% of trading volume thereafter. Despite being in bankruptcy, Hertz issued 29 million dollars of common stock before the SEC stopped further sales. And the Financial Times noted the prevalence of retail investors in Hertz at the time. The line in Figure A2 breaks because Hertz was delisted in October 2020, and emerged from bankruptcy in July 2021.

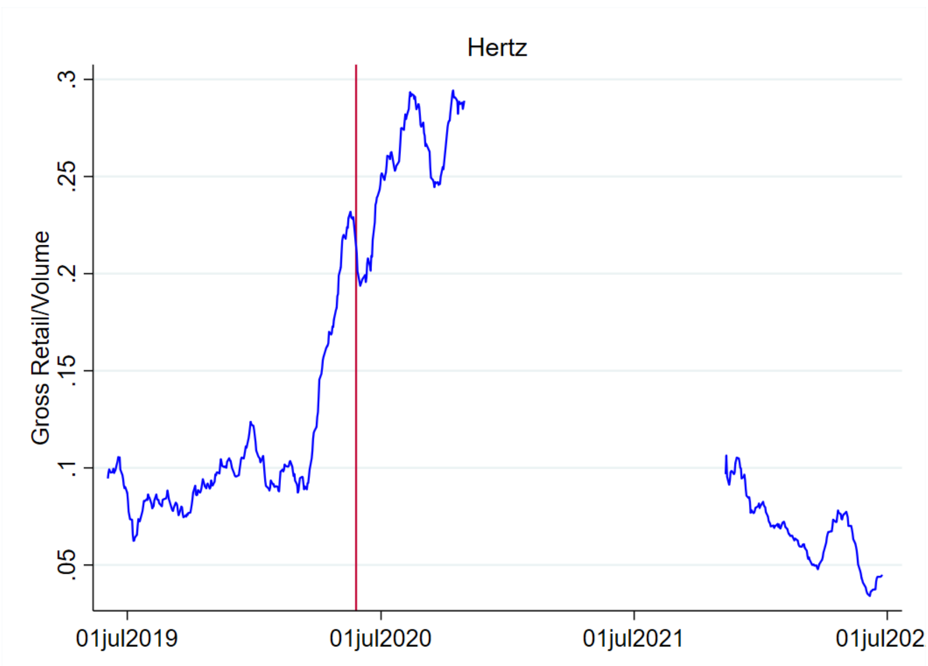

Figure A2: Gross retail activity in Hertz. 22-day moving average of gross retail activity, defined as retail buys plus retail sells over trading volume, in Hertz stock between 7/2019 and 12/2021.

Panel A.

$t=-12$	Retail Portfolio at $t=0$ 1	2	3	4	5
1	53.44	27.65	12.31	4.88	1.72
2	28.48	35.50	23.94	9.67	2.40
3	13.08	25.58	34.22	21.43	5.69
4	5.11	10.68	24.32	40.29	19.60
5	1.73	2.29	6.16	23.86	65.96

Panel B. Small stocks only.

$t=-12$	Retail Portfolio at $t=0$ 1	2	3	4	5
1	37.83	24.79	17.08	11.79	8.51
2	22.86	24.88	21.72	17.51	13.03
3	13.83	19.48	23.25	22.68	20.77
4	7.98	14.13	20.78	26.72	30.38
5	4.99	8.84	14.98	25.87	45.32

Panel C. Large stocks only.

$t=-12$	Retail Portfolio at $t=0$ 1	2	3	4	5
1	52.83	27.58	13.41	4.63	1.55
2	27.25	33.94	24.81	11.06	2.95
3	12.38	25.09	31.98	23.44	7.10
4	3.69	10.69	24.06	38.53	23.02
5	0.83	2.54	6.09	23.74	66.81

Table A4: Transition Matrix across Retail Portfolios. Panel A shows the probability (in percentage points) that a stock in retail intensity portfolio i at time $t=-12$ ends up in the indicated retail portfolio 12 months later at time $t=0$. Panels B and C repeat the analysis, but additionally condition on the stock being in the bottom or top quintile in terms of market cap at time $t=-12$, respectively.

A more recent example is the (now closed) First Republic Bank. As discussed in an article from CNN, "... In January and February, trading in First Republic stock was outright sleepy. Retail investors averaged just $20,000 in daily net purchases. But after the collapse of SVB, that daily average trading of the company's stock exploded to $10.3 million, according to data through April 10 from VandaTrack. TD Ameritrade's Investment Movement Index, which tracks retail traders, found that its clients were net buyers of First Republic Bank in March even as the company's shares plummeted more than 88% over worries about uninsured deposits and the overall health of the banking system."

Although these are selected examples, we believe they are broadly consistent with prediction 1. Intuitively, car rental companies become harder to value in pandemics, and banks become harder to value in banking crises. So, once a stock becomes hard to value, retail investors tend to more heavily trade that stock.

A.7 Industry tilt of retail trading intensity

In this subsection, we show there is substantial heterogenity in retail trading intensity across industries. To this end, by year, we plot the share of total dollar trading volume in each Fama French 49 industry which comes from retail-initiated buys and sells against each industry's share of total market capitalization at the end of each year. The top panel of Figure A3 uses data from 2010, and shows that retail tend to favor mining, gold and alcohol firms, while avoiding insurance, utilities and banks. The bottom panel of Figure A3 uses data from 2019, showing that retail continued to avoid trading banks and insurance companies, while favoring sin stocks and biotech firms.

A.8 Retail trading intensity and institution size

As we show in Table 2, stocks with more retail trading intensity have less institutional ownership. One explanation for this is that institutions face constraints, either through mandates against holding e.g., unprofitable firms (Ma et al. (2019), Beber et al. (2021)) or a desire to avoid crossing the 5% ownership threshold which triggers additional regulatory scrutiny (Edmans et al., 2013). These constraints, however, are not likely to affect all institutions equally. For example, one could imagine that small institutions are less likely to own 5% of any given firm's shares outstanding, even if they hold a very concentrated position in the stock.

To measure this, each quarter, we compute total dollar holdings of equities (restricting to ordinary common shares traded on major exchanges) for every 13F-filing institution and then sort institutions into quintiles based on this quantity. Then, we compute the percentage of each stock held by each of these groups of institutional owners, setting the value to zero if no institutions in a given quintile hold the stock. Finally, we compute the average of this quantity within each quintile formed on retail

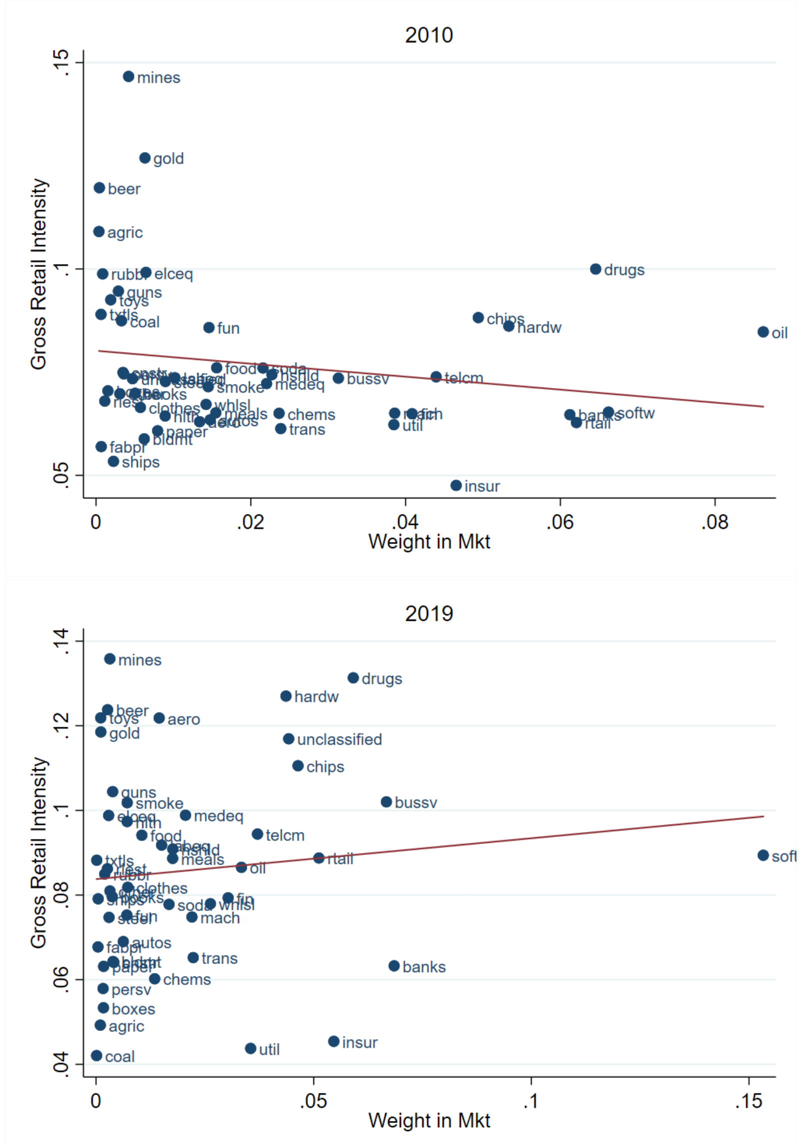

Figure A3: Retail trading intensity across Fama French 49 industries. The x-axis represents each industry's share of total market capitalization at the end of each year. The y-axis represents the share of total dollar trading volume in each industry each year which comes from retail-initiated buys and sells. The top panel uses data from 2010 and the bottom panel uses data from 2019.

trading intensity. Panel A of Table A5 shows that, among the largest institutions, the percentage of a given company held is monotonically decreasing in retail trading intensity. Conversely, among the smallest institutions, this quantity is monotonically increasing. This suggests that small institutions look more like retail investors, while large institutions look less like retail investors. In unreported results, we find this pattern holds within each type of institutional investor, as classified by Bushee (1998).

One concern with this methodology is that institutions can be composed of many funds, so using the 13F data may be muddling constraints at the fund level with constraints at the institution level. To address this, we perform a similar exercise with active mutual funds. We start with the universe of funds in the S12 database, and then use the procedure in Appel et al. (2016) to identify and drop passive funds. We then add up the holdings of equities at the fund level, and then sort funds into 5 groups each quarter based on total dollar holdings. Finally, as above, we compute the percentage of each firm owned by each group of active funds, and then take the average of this quantity within the quintiles of retail trading intensity. Panel B of Table A5 shows a result similar to panel A: large mutual funds tend to avoid high retail stocks, while small mutual funds tend to favor them.

	Panel A: 13F Institution Size				
Retail Quintile	Small	2	3	4	Large
Low	0.44%	0.84%	1.77%	5.03%	72.61%
2	0.44%	0.88%	1.79%	5.11%	72.16%
3	0.61%	1.19%	2.15%	5.46%	65.79%
4	0.97%	1.63%	2.81%	5.87%	55.06%
High	1.46%	1.89%	2.90%	5.04%	31.11%
	Panel B: S12 Active Fund Size				
Retail Quintile	Small	2	3	4	Large
Low	0.002%	0.03%	0.19%	1.38%	23.64%
2	0.003%	0.03%	0.21%	1.41%	23.23%
3	0.003%	0.03%	0.22%	1.43%	20.20%
4	0.006%	0.04%	0.24%	1.39%	15.67%
High	0.011%	0.05%	0.21%	1.04%	7.35%

Table A5: Institutional ownership and retail trading intensity. Each month, we form 5 quintiles on retail traing intensity, defined as $\text{RSVOL}_{i,t} = \frac{\text{RBuy}_{i,t}+\text{RSell}_{i,t}}{\text{Volume}_{i,t}}$. Quintiles of institution and active fund size are formed each quarter based on total dollar holdings of equities. Table entries represent the average percentage of each stock in each retail trading intensity bucket's shares outstanding collectively held by each bucket of 13F or S12 size.

A.9 Earnings Sensitivity and Pre-Announcement Retail Flows

In Section 5 we show that high retail stocks have both an especially high retail trading intensity and especially high trading costs around earnings announcements. In this appendix we re-visit our results on the responsiveness of high retail stocks to earnings news. To this end, we estimate a modified version of Equation 4:

$$r^i_{t,t+n} = \alpha + \beta \text{SUE}_{i,t} + \beta_1 1_{i\in Q1_{\tau-1}} + \beta_2 1_{i\in Q2_{\tau-1}} + \beta_4 1_{i\in Q4_{\tau-1}} + \beta_5 1_{i\in Q5_{\tau-1}} + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \quad \text{(A28)}$$

where $1_{i \in Qk_{\tau-1}}$ are indicators for quintiles of gross or net retail flows, formed over the 22 trading days before the earnings announcements. In Table A6, Columns 1, 3 and 5 show that stocks with high pre-announcement gross retail trading intensity are less responsive to earnings news. This is consistent with Table 5, which is sorting on gross retail trading intensity in the previous calendar month, rather than the previous 22 trading days.

Columns 2, 4 and 6 replicate these results, but using net flows ahead of the earnings announcement instead of gross flows. The coefficients on the "Low Flow" (i.e., most retail selling) and "High Flow" (i.e., most retail buying) interaction terms are consistently negative. Although the coefficient for the high retail inflow bucket is slightly more negative, it is not statistically significantly different from coefficient for the high retail outflow bucket. These results suggest that in terms of responsiveness to earnings news, it doesn't seem to matter whether retail are rushing into the stock or rushing out of the stock before earnings announcements.

A.10 Retail trading intensity around earnings announcements

In addition to a directional effect, retail-initiated trades make up a particularly large amount of overall (gross) trading around earnings announcements. To quantify this, we estimate regressions of the form:

$$\text{Retail Intensity}_{i,t} = \alpha + \beta_1 1_{i \in Q1_{\tau-1}} + \beta_2 1_{i \in Q2_{\tau-1}} + \beta_4 1_{i \in Q4_{\tau-1}} + \beta_5 1_{i \in Q5_{\tau-1}} + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t}, \tag{A29}$$

where Retail Intensity is retail's share of total trading volume or fraction of shares outstanding. $1_{i \in Qk_{t-1}}$ are dummy variables for quintiles of retail trading intensity, formed over the previous month $\tau - 1$, where the middle quintile is the omitted group. $X_{i,t}$ are the same controls as Equation 4. To account for level differences in retail trading across quintiles of past retail intensity, we subtract the mean Retail Intensity$_{i,t}$ at the stock level over the previous 252 trading days before $t = -5$.[21] Table A7 contains the results. It shows that leading up to, on, and after earnings announcements, retail investors make up a higher share of trading volume, relative to their average past intensity in the stock. In terms of magnitudes, retail investors make up a 1.3 percentage point larger share of trading volume in the pre-announcement period than they do over the past year.

The bottom panel shows that this finding is in fact driven by two separate phenomena. First, in the pre-earnings period, the coefficients in the top panel are positive, while the coefficients in the bottom panel are near zero.[22] This suggests that, consistent with the refinement of prediction 2 discussed in the main body of the paper, institutional investors are trading less in the pre-earnings

[21] All results are stronger when not subtracting average past retail activity, but without demeaning, the results would not speak to prediction 2 which is about *abnormal* retail intensity around earnings announcements.

[22] Even though the regression only using data from $t = -1$ has a positive and statistically significant coefficient on the indicator variable for the top retail quintile, its magnitude is about $1/10^{th}$ as large as the same coefficient using only data from $t = 0$.

	Cumulative post-earnings announcement return					
Return Window:	(0, 0)		(0, 2)		(0, 4)	
	(1)	(2)	(3)	(4)	(5)	(6)
SUE	1.113***	1.022***	1.227***	1.077***	1.291***	1.121***
	(0.142)	(0.148)	(0.171)	(0.157)	(0.182)	(0.176)
SUE x Low Flow	0.394**	-0.330**	0.367*	-0.22	0.264	-0.25
	(0.151)	(0.128)	(0.190)	(0.138)	(0.194)	(0.161)
SUE x 2 Flow	0.450***	-0.000998	0.488***	0.0713	0.425***	0.00241
	(0.120)	(0.120)	(0.154)	(0.137)	(0.153)	(0.165)
SUE x 4 Flow	-0.280***	-0.0287	-0.206	0.0647	-0.213	0.0974
	(0.089)	(0.091)	(0.125)	(0.096)	(0.146)	(0.108)
SUE x High Flow	-0.590***	-0.384***	-0.616***	-0.333***	-0.678***	-0.361***
	(0.109)	(0.093)	(0.143)	(0.106)	(0.143)	(0.127)
Obs	110,331	110,331	110,331	110,331	110,331	110,331
R-Sq	0.104	0.100	0.108	0.104	0.109	0.107
Flow	Gross	Net	Gross	Net	Gross	Net
Time FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES

Table A6: Pre-earnings retail flow share and earnings-announcement returns. Left-hand-side variables are cumulative market-adjusted earnings-announcement returns from $t = 0$ to $t = n$ where $n = 0, 2, 4$. Quintiles of retail flow share are formed each quarter using the cumulative flow share over the 22 trading days before the earnings announcement. In columns 1, 3 and 5, these are based on gross flows i.e., (retail buys + retail sells)/(retail buys + retail sells + non-retail buys and sells). In columns 2, 4 and 6, these are based on net flows i.e., (retail buys - retail sells)/(retail buys + retail sells + non-retail buys and sells). Time fixed effects are for year-quarter. Control variables include nominal price, returns from month $t - 12$ to $t - 2$, time since listing, market capitalization, book-to-market, gross profit margin, book long-term leverage, MAX (lottery demand) and month $t - 1$ returns. Additional controls include betas on the three Fama-French factors as well as the momentum factor, idiosyncratic volatility and total volatility, all computed over the past 12 months. Standard errors are double clustered at the permno/year-month level.

period rather than retail trading more. Second, and also consistent with this refinement, in the post-earnings period, retail trades more both on an absolute (i.e., when normalizing by shares outstanding) and relative (i.e., when normalizing by total trading volume) basis, suggesting that such events drive retail activity. On the earnings day itself, retail investors make up almost 50 basis points more of total volume and 10 basis points more of shares outstanding than their past average.

	Demeaned Retail as % of Trading Volume (percentage points)					
Timing:	(-5,-1)	(-3,-1)	-1	0	(0,2)	(0,4)
Q1	-0.176***	-0.171***	-0.180***	-0.303***	-0.194***	-0.154***
	(0.032)	(0.033)	(0.042)	(0.049)	(0.037)	(0.034)
Q2	-0.0740***	-0.0837***	-0.0628**	-0.151***	-0.0948***	-0.0862***
	(0.020)	(0.022)	(0.030)	(0.037)	(0.026)	(0.022)
Q4	0.186***	0.193***	0.208***	0.205***	0.199***	0.163***
	(0.035)	(0.037)	(0.045)	(0.046)	(0.034)	(0.033)
Q5	1.290***	1.262***	1.338***	0.471***	0.602***	0.649***
	(0.087)	(0.090)	(0.101)	(0.100)	(0.084)	(0.079)
Observations	137,927	137,927	137,927	137,927	137,927	137,927
R-squared	0.109	0.086	0.053	0.058	0.076	0.088
Time FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES

	Demeaned Retail as % of Shares Outstanding (basis points)					
Timing:	(-5,-1)	(-3,-1)	-1	0	(0,2)	(0,4)
Q1	0.320***	0.169**	-0.228**	-4.507***	-2.086***	-1.283***
	(0.081)	(0.081)	(0.092)	(0.399)	(0.195)	(0.133)
Q2	0.212***	0.117**	-0.0617	-2.370***	-1.160***	-0.729***
	(0.053)	(0.055)	(0.073)	(0.317)	(0.156)	(0.106)
Q4	-0.529***	-0.394***	0.0992	4.453***	2.203***	1.409***
	(0.115)	(0.118)	(0.134)	(0.474)	(0.244)	(0.172)
Q5	-0.712*	-0.361	1.052**	10.56***	5.557***	3.761***
	(0.409)	(0.393)	(0.437)	(1.074)	(0.659)	(0.525)
Observations	137,927	137,927	137,927	137,927	137,927	137,927
R-squared	0.07	0.061	0.052	0.099	0.085	0.076
Time FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES

Table A7: Retail activity and trading intensity around earnings announcements. Cross-sectional regression where left-hand-side variables are measure of retail trading intensity around earnings announcement. In the top panel, retail trading intensity is defined as (retail buys + retail sells)/(retail buys + retail sells + non-retail buys and sells) while in the bottom panel, retail trading intensity is defined as (retail buys + retail sells)/(shares outstanding). In all columns, we subtract the mean of these quantities computed over the previous 252 trading days. Qk is an indicator variable for whether stock i was in retail intensity quntile k at the end of the month before the earnings announcement. Time fixed effects are for year-month. Control variables include nominal price, returns from month $t-12$ to $t-2$, time since listing, market capitalization, book-to-market, gross profit margin, book long-term leverage, MAX (lottery demand) and month $t-1$ returns. Additional controls include betas on the three Fama-French factors as well as the momentum factor, idiosyncratic volatility and total volatility, all computed over the past 12 months. Standard errors are double clustered at the permno/year-month level.

A.11 Persistent and transitory components of retail trading

In the main body of the paper, we argue that "hard to value" is a unifying characteristic that explains a significant share of cross-sectional heterogeneity in retail trading intensity. Of course, retail trading intensity is likely both a function of persistent firm characteristics (e.g., size, nominal price, difficulty to value) as well as more transitory factors (e.g., news that grabs retail investors' attention (Barber and Odean, 2008)). In this subsection, we propose a method for decomposing retail trading into a transitory component and persistent component. As a validation exercise for this decomposition, we show that retail investors' response to earnings news is coming mostly from the transitory component of retail trading intensity.

As a starting point, we believe a simple method for estimating the persistent component of retail trading intensity is to examine past retail trading intensity. Specifically, for each stock, each month, we compute the average gross retail share of total trading volume over the previous 12 months, and label this as the persistent component. We then compute the difference between realized retail trading and past retail trading intensity, and call this the transitory component.

Another method is to regress retail trading on a set of firm-level characteristics known to be correlated with retail trading intensity. Specifically, we focus on the variables in Tables 3, 2 and A9 shown to be correlated with retail trading intensity: market capitalization, firm age, returns from month $t-12$ to month $t-2$, the MAX factor of Bali et al. (2017) (a measure of lottery demand), book-to-market, earnings-to-price, market beta, standard deviation of daily returns, intraday idiosyncratic volatility computed from trades, Kyle's lambda, the effective bid-ask spread, intangible capital divided by market capitalization and the dollar value of patents divided by market capitalization. We omit cashflow duration (Gormsen and Lazarus, 2021), valuation uncertainty (Golubov and Konstantinidi, 2021) and mispricing (Stambaugh and Yuan, 2017) because that would dramatically shrink our sample of observations based on data availability. In unreported results, however, we find that including these variables does not substantially change the empirical conclusions.

We then run a regression of retail trading intensity in month $t+1$ on this set of control variables as of time t. While this partially alleviates concerns about a look-ahead bias, we note that we run this regression with the full sample of years – rather than in rolling windows – so there is some look-ahead bias in terms of the estimated coefficients relating firm characteristics to retail trading intensity.

To validate these measures, we run a regression of retail trading intensity on earnings news:

$$\text{Retail Intensity}_{i,t} = \beta_1 \text{Positive Earnings News}_{i,t} + \beta_2 \text{Negative Earnings News}_{i,t} + \alpha_t + \gamma_i + \epsilon_{i,t} \tag{A30}$$

where the left-hand-side variable is either overall retail trading intensity, defined as retail buys plus retail sells divided by overall trading volume, or the transitory component of retail trading intensity, computed using both of the methods described above. For positive and negative earnings news, we either use the positive and negative components of standardized unexpected earnings (SUE) (Hartzmark and Shue, 2018) or the positive and negative components of earnings-day returns (Frazzini and Lamont, 2007). α_t are a set of year-month fixed effects, while γ_i are a set of firm fixed effects. The unit of analysis is firm-year-month, and the regression only includes firm-year-months with earnings announcements. Finally, standard errors are double clustered at the firm and year-month level.

If the decomposition is working as expected, we would expect a significant share of the variation in retail trading intensity to be coming from the transitory component – as the earning news is likely not known ahead of time. It should not explain all the variation, however, as the unit of analysis is year-month, and there are many other factors within that time which may be related to the persistent component of retail trading.

Panel A of Table A8 presents results using the simple decomposition based on a past moving-average of retail trading activity. Columns 1-6 show that, as expected, overall retail trading intensity is higher when there is larger earnings news. Columns 7-12 show that there is also a large increase in the transitory component of retail trading intensity. The highlighted orange cells show that the magnitudes of the estimated increase in retail trading are similar when using the overall and transitory components of retail trading intensity. In these columns, we are including both sets of fixed effects, which is a high bar – as it accounts for both time-variation in average retail trading intensity and the fact that retail investors may tend to favor some stocks more than others for reasons we can't observe.

Panel B of Table A8 presents results using the residualization based on firm fundamentals shown to be correlated with retail trading activity. Broadly, the results are similar to those in Panel A, both in terms of estimated magnitudes and statistical significance.

In summary, Table A8 shows that our proposed decompositions of retail trading activity into persistent and transitory components work as expected around earnings announcements. Specifically, that there is a large surge in our estimated transitory component of retail trading activity after large earnings news – positive or negative. Further, it shows that both decompositions yield similar results, suggesting that past retail trading activity, which is straightforward to compute for every stock in our sample, is a reasonable proxy for the persistent component of retail trading activity.

A.12 Retail trading intensity and average trading costs

A natural question is whether stocks with more retail trading intensity have higher or lower average trading costs than stocks with less retail trading intensity. One one hand, if retail investors act as noise traders a-la Kyle (1985), one might expect trading costs to be relatively lower in high retail

Panel A: Decomposition based on moving averages

	Retail Share of Trading Volume						Transitory Retail Share of Trading Volume					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\|SUE\|$ positive	134.3***	125.4***	52.86***				15.56***	15.56***	19.36***			
	(7.747)	(7.389)	(2.639)				(2.564)	(2.133)	(1.953)			
$\|SUE\|$ negative	79.81***	73.29***	30.39***				10.39***	9.983***	12.51***			
	(7.820)	(7.151)	(2.186)				(1.594)	(1.408)	(1.454)			
$\|Ret\|$ positive				13.28***	11.00***	7.791***				5.864***	5.862***	6.435***
				(0.689)	(0.651)	(0.357)				(0.380)	(0.341)	(0.334)
$\|Ret\|$ negative				21.93***	19.05***	9.454***				7.738***	8.078***	8.396***
				(1.005)	(0.845)	(0.435)				(0.559)	(0.422)	(0.382)
Observations	120,808	120,808	120,598	120,808	120,808	120,598	119,430	119,430	119,210	119,430	119,430	119,210
R-squared	0.115	0.169	0.724	0.038	0.102	0.72	0.008	0.057	0.128	0.024	0.075	0.143
FE	None	Time	Firm/Time	None	Time	Firm/Time	None	Time	Firm/Time	None	Time	Firm/Time

Panel B: Decomposition based on firm-level characteristics

	Retail Share of Trading Volume						Transitory Retail Share of Trading Volume					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\|SUE\|$ positive	138.7***	130.8***	56.24***				51.76***	46.55***	15.20***			
	(8.560)	(8.176)	(2.824)				(5.076)	(4.780)	(2.750)			
$\|SUE\|$ negative	81.56***	75.47***	31.62***				25.49***	21.69***	6.523***			
	(8.224)	(7.444)	(2.357)				(4.248)	(3.840)	(1.686)			
$\|Ret\|$ positive				13.15***	11.07***	7.888***				8.100***	7.056***	6.098***
				(0.671)	(0.634)	(0.357)				(0.571)	(0.526)	(0.365)
$\|Ret\|$ negative				21.42***	18.76***	9.598***				14.35***	13.11***	9.101***
				(0.972)	(0.824)	(0.444)				(0.770)	(0.638)	(0.408)
Observations	111,129	111,129	110,977	111,129	111,129	110,977	111,129	111,129	110,977	111,129	111,129	110,977
R-squared	0.11	0.162	0.723	0.039	0.099	0.719	0.02	0.058	0.538	0.026	0.064	0.546
FE	None	Time	Firm/Time	None	Time	Firm/Time	None	Time	Firm/Time	None	Time	Firm/Time

Table A8: Validation of decomposotion of retail trading intensity into persistent and transitory components. Results from the following regression

$$\text{Retail Intensity}_{i,t} = \beta_1 \text{Positive Earnings News}_{i,t} + \beta_2 \text{Negative Earnings News}_{i,t} + \alpha_t + \gamma_i + \epsilon_{i,t}$$

where the left-hand-side variable is either overall retail trading intensity, defined as retail buys plus retail sells divided by overall trading volume, or the transitory component of retail trading intensity, computed using the simple moving average decomposition, or the residualization on firm fundamentals. For positive and negative earnings news, we either use the positive and negative components of standardized unexpected earnings (SUE) (Hartzmark and Shue, 2018) or the positive and negative components of earnings-day returns (Frazzini and Lamont, 2007). α_t are a set of year-month fixed effects, while γ_i are a set of firm fixed effects. The unit of analysis is firm-year-month, and the regression only includes firm-year-months with earnings announcements. Standard errors are double clustered at the firm and year-month level.

stocks. More broadly, for the reasons outlined in Appendix A.1, suppose that fewer investors are learning, and/or investors are learning less about the fundamentals of high retail stocks because they are hard to value. Then, again in the framework of Kyle (1985), trading costs are expected to be lower in the stocks retail investors tend to favor, as the market maker faces a smaller risk of adverse selection.

As we discuss in Section 5.4, however, it's possible that that retail investors are not truly noise traders, and have information about high retail stocks' fundamentals (see e.g., Kaniel et al. (2012)). This could lead to relatively higher trading costs in high retail stocks even if their signals are not particularly precise. The logic is that if institutional investors totally avoid learning about such stocks, the *informational advantage* of retail investors would be relatively large. And, as a consequence, this could create significant risk of adverse selection and thus high expected transaction costs.

Another reason trading costs may be relatively higher in high retail stocks is that betting against retail order flow itself is risky. The logic is that – as shown in e.g., Boehmer et al. (2021) – retail order flow is persistent. When an initial retail order arrives, market makers may not want to provide liquidity, as it's possible that subsequent retail trades in the same direction will further push prices against the market maker's position. In Table 8, we provide evidence that prices tend to move in the same direction as retail order flow in high retail stocks. This suggests that this mechanism may be especially strong in the stocks retail investors tend to focus on.

More broadly, there may be information other than signals about fundamentals which is relevant for transaction costs. For example, another possibly important source of information are signals about future demand (see e.g., Li and Lin (2023)). Specifically, suppose that retail order flow has no information about fundamentals, but, as shown in Boehmer et al. (2021), retail order flow is positively autocorrelated. In other words, retail order flow contains information about future retail demand.

Then, consider a market-maker's decision after observing a retail-initiated buy order. One option is to lean against the order by providing liquidity to the retail investor – betting on reversion – because the order is known to be unrelated to fundamentals. This can be risky, however, if the retail buy is followed by more retail buy orders, as it will push prices further against the market maker's position. As a result, when trading against a retail order, a market maker may decide to set a larger spread, as compensation for risk of prices – at least in the short-run – continuing to move in the same direction.

Given these competing forces, it's an empirical question as to whether high retail stocks have relatively lower or higher trading costs. Table A9 reports summary statistics on volatility and trading costs across retail portfolios. The first two columns report measures of stock price volatility. In the first column, we show that that high retail stocks tend to have higher overall volatility, as measured by the standard deviation of daily returns each month. In the second column, we report averages of trade-based intraday volatility, computed by averaging the squared 1-second returns each day.[23] These measures are also elevated for high retail stocks, though in the case of intraday volatility the differences mostly reflect a size effect.

The remaining three columns summarize measures of liquidity. λ_2 stands for Kyle's lambda, the coefficient from a regression of returns on the signed square root of dollar order imbalance. This measure of illiquidity is higher for high retail stocks, and a substantial gap remains controlling for size.

Finally, Espread and Rspread stand for the percent effective and realized bid-ask spread, respectively.[24] Both are higher among high retail stocks, but including the size dummies makes accounts for a large part of both differences.

Overall, the evidence in Table A9 is consistent with high retail stocks being relatively more expensive to trade. This suggests that retail investors may have a significant informal advantage in high retail stocks – although these results do not clarify whether that advantage is due to information about future fundamentals or information about future retail demand. In the main body of the paper, in Table 9, we aim to further disentangle these components, providing information that, at least around earnings announcements, retail investors seem to be compensated both for liquidity provision and information about future fundamentals.

Before moving on, we want to highlight an apparent tension between the results in Tables 1 and Table A9. Specifically, high retail stocks seem to simultaneously have (A) higher trading volume and (B) lower liquidity, contrary to standard intuition that stocks with more trading volume should be more liquid. This relationship, however, may depend on the source of trading volume. If, for example, there is significant volume due to trading on information or a particular group of investors demanding liquidity one could believe that both volume and trading costs would be simultaneously elevated (Chacko et al., 2008). We explore this more in Appendix **??**, showing that trading costs are initially decreasing in turnover, but after a certain point, higher levels of turnover are associated with higher bid-ask spreads and Kyle's λ.

[23]This measure, as well as all the measures of trading costs in Table A9, are from the WRDS intraday indicators suite, which is built on the millisecond TAQ data.

[24]Specifically, following Holden and Jacobsen (2014), the percent effective spread for any trade k is defined as: Percent Effective Spread$_k$ $= (2D_k(P_k - M_k))/M_k$ where D_k is equal to 1 if trade k is a buy, and -1 if trade k is a sell, classified using the algorithm in Lee and Ready (1991). M_k is the midpoint of NBBO quotes and P_k is the price that trade k occurred at. For each stock, each day, WRDS takes a value-weighted average of this quantity, where the weights are proportional to the dollar size of each trade k. In words, the percent effective spread is the percent distance away from the midpoint that the (value-weighted) average trade occurs at. The realized spread is defined as Percent Realized Spread$_k$ $= (2D_k(P_k - M_{k+5}))/M_k$ where M_{k+5} is the midpoint 5 minutes after trade k. The realized spread is designed to capture how far the midpoint moves 5 minutes after trade k occurs.

	SD	Ivol t	$\lambda 2$	Espread	Rspread
Low	-0.39***	-0.01**	-0.83***	-0.04***	-0.01***
	(-20.24)	(-2.04)	(-12.94)	(-7.06)	(-4.07)
2	-0.26***	-0.03***	-0.61***	-0.06***	-0.03***
	(-21.05)	(-10.04)	(-12.61)	(-16.04)	(-14.21)
4	0.49***	0.06***	1.45***	0.14***	0.07***
	(23.77)	(14.08)	(15.96)	(21.37)	(19.82)
High	1.59***	0.30***	6.53***	0.58***	0.31***
	(38.18)	(18.09)	(26.80)	(31.04)	(28.61)
Q5-Q1	1.98	0.30	7.37	0.62	0.32
p(Q1=Q5)	0.00	0.00	0.00	0.00	0.00
Q5-Q1, size	1.63	-0.01	4.35	0.04	0.01
p(Q1=Q5), size	0.00	0.40	0.00	0.00	0.05
N	453,554	453,554	453,554	453,554	453,554
R^2	0.44	0.10	0.21	0.19	0.13
Month FE	Yes	Yes	Yes	Yes	Yes

Table A9: Liquidity in five retail share of trading sorted portfolios. Firm-month level regressions on dummy variables representing retail trading intensity quintiles formed the prior month. SD is the standard deviation of daily stock returns in a given month; Ivol t is intraday volatility computed from trades; λ_2 is Kyle's lambda, estimated with an intercept; Espread and Rspread are the effective and realized spread, computed using the methodology in Holden and Jacobsen (2014). Monthly fixed effects. Standard errors clustered on the firm and month level.