The Retail Habitat

Toomas Laarits* Marco Sammon †

First version: May 9, 2022
This version: May 10, 2023

Abstract

Retail investors trade hard-to-value stocks. Controlling for size, 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, the findings document a new dimension of investor heterogeneity and suggest a comparative advantage of retail in holding 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, and Alexi Savov as well as the seminar audience at NYU Stern for comments and feedback on the paper.

1 Introduction

Stock market behavior during the pandemic has served as a reminder of the importance of retail investors for financial markets. Throughout 2020 and 2021, the popular press was preoccupied with a small group of stocks that attracted a particularly large amount of retail trading. However, the set of stocks that experience a high share of retail activity goes beyond the set of "meme" stocks. In the top panel of Figure 1 we show the average share of retail-initiated trades across five portfolios sorted on the prior month share of retail-initiated trades. The bottom quintile has about 2% of trades originating from retail investors, whereas this number is close to 20% for the top quintile. Despite this large heterogeneity in retail trading intensity, the academic literature on retail investors has not devoted significant attention to the determinants of retail trading in the cross-section.

In this paper we document a new fact that explains the distribution of retail trading activity: retail investors make up a large share of volume in stocks that are hard to value. Specifically, retail investors favor trading in firms that have longer-duration cashflows, have high shares of intangible capital and higher likelihood of mispricing. Or, equivalently, institutional investors tend to shun trading in such stocks. We rule out our results being driven pure firm-size effect, showing that these cross-sectional regularities continue to hold at all points along the distribution of market capitalization.

In line with high retail stocks being hard to value, we show that they have larger earnings news volatility, stemming mostly from the firm-specific component of earnings news, and, correspondingly, analysts' estimates of these firms' fundamentals are more dispersed. Consistent with current cashflows being less value-relevant for hard to value firms, we show that high retail stocks' prices are less sensitive to standardized unexpected earnings (SUE). Again, this finding is not driven by a firm size effect, as these patterns are true even within the top 20% of firms by market capitalization.

The results described so far are consistent with selection into hard-to-value stocks on behalf of retail investors, or selection out of such stocks by institutional investors. We also seek to establish, however, how retail investors' trading behavior itself affects security prices. To this end, we use earnings announcements as a laboratory, documenting an active role of retail investors. Specifically, we find that high retail stocks have substantial abnormal retail trading before and on announcements, and returns move in the same direction as net retail flow.

Finally, we document that the well-known earnings announcer risk premium – the average return earned in response to scheduled earnings news – is positive only for low retail stocks. Our proposed mechanism is that this premium is compensation for exposure to systematic risk (as argued in Savor and Wilson (2016)) and high retail stocks' earnings announcements are mostly composed of idiosyncratic news.

Overall, the results in this paper document a novel dimension of heterogeneity in the cross-section of equities. We show that retail investors like to trade precisely the stocks where institutional investors, who presumably are better equipped to research and act upon fundamentals, are least advantaged.

The first step of our analysis documents new facts on the distribution of retail trading in the cross-section. We show that retail trading intensity is persistent, meaning that stocks with high shares of retail trading continue to see a large share of retail trades in the subsequent year. This finding suggests that retail trading intensity is a function of persistent underlying characteristics. 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.

In addition to being persistent, retail traders as a group tend to concentrate on a relatively small subset of stocks. We find that the distribution of retail-initiated dollar volume is more heavy-tailed than that of overall dollar trading volume, and this concentration has increased in the recent past. In 2020, for example, the top 10 most heavily traded stocks by retail investors made up over 40% of total retail dollar volume. This combination of persistence and concentration opens the door for retail trading itself to have a direct effect on prices, even though retail investors control less than 30% of total stock market wealth.

Having documented that retail investors tend to persistently focus on a subset of stocks, we aim to identify firm-level characteristics which explain this cross-sectional heterogeneity. Previous literature has focused on fundamentals e.g., size, value, price and past returns (Kumar and Lee (2006), Balasubramaniam et al. (2021), Bali et al. (2021), Luo et al. (2021)). We show that the concept of difficulty to value is a particularly useful summary characteristic for explaining cross-sectional heterogeneity in retail trading intensity. Firstly, we examine cashflow duration (Gormsen and Lazarus (2021)). 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 (Lev and Gu (2016)). Thirdly, we examine two composite measures, specifically the mispricing score of Stambaugh and Yuan (2017) and the valuation uncertainty score of Golubov and Konstantinidi (2021). We find that high retail stocks have longer duration cashflows, more intangible capital, more expected mispricing and more valuation uncertainty. Many of these relationships hold both within the bottom and top 20% of firms by market capitalization, suggesting the known relationship between retail investor activity and firm size is not driving our results.

Building on the newly documented cross-sectional regularity, 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 large amounts of information are released, which could resolve some of the *ex ante* valuation uncertainty. We find that high retail stocks have more volatile earnings announcement news and returns. In terms of magnitudes, the standard deviation of standardized unexpected earnings (SUE) is more than 3 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.

We then 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, we estimate earnings-response regressions following Kothari and Sloan (1992), and find that for a given magnitude of earnings surprise, high retail stocks' prices respond significantly less to earnings new 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.

So far, we've summarized evidence on a new dimension of retail *selection*: retail investors tend to favor trading hard-to-value stocks. Next, we ask whether retail trades themselves influence stock prices, especially around earnings announcements.

Broadly, retail-heavy stocks see large 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. Further, on the earnings announcement day itself, retail investors make up an additional 0.5% of trading volume in high retail stocks.

Retail traders also appear to have a material impact on returns of the stocks they favor. We measure retail investors' effect on prices by examining whether returns tend to move with or against net retail order flow. The logic of this test is that if retail investors' trading itself is driving price changes, returns should move in the same direction as net retail trades.[1] We find that, in the average week, prices tend to move with net retail order flow only in high retail stocks. Further, consistent with the increased importance of retail investors around the release of earnings news, this effect is twice as strong in weeks with earnings announcements. Specifically, for a high retail stock in a week that contains an earnings announcement, a one standard deviation increase in net retail order imbalance corresponds to an additional return of almost 1%.

We examine the mechanism behind the tendency of prices to move in the direction of net retail order flow in high retail stocks. One explanation is that institutions are especially sensitive to

[1]With our data, we can not conclusively show that the retail trades themselves are moving prices *per se*. We are only arguing that prices moving in the same direction as retail order flow is consistent with retail trades moving prices.

trading costs (Di Maggio et al. (2022)) and high retail stocks are expensive to trade. Such stocks may be difficult to trade because e.g., assets which are harder to value often have larger adverse selection costs. Consistent with this, we find that high retail stocks have effective and realized bid-ask spreads (measured using the method in Holden and Jacobsen (2014)) three times as large as low retail stocks.

Trading costs can also explain the impact of retail investors on prices around earnings announcements. Specifically, 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. Although the magnitude of this effect may seem small, this estimate is controlling for the extreme nature of news for high retail stocks and a host of firm-level controls and fixed effects. Further, 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).

Finally, 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 such fundamental, 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 surprises are mostly comprised of idiosyncratic news.

Our test examines the returns to individual high and low retail firms around earnings announcements. In a three trading day window starting with the announcement, we find 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 a robustness check, we show that results are similar when constructing calendar-time portfolios that go long announcing firms and short non-announcing firms.

Overall, our results highlight an economically motivated determinant of retail trading, and, by extension, of holdings in the cross section.

An active strand of research 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), Haddad et al. (2021)). 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. (2021) and Gabaix et al. (2022) has studied the demand of retail investors specifically. Balasubramaniam et al. (2021) 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 top characteristics that 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 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. 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)), to name a few. Because of the selection by retail traders into hard to value stocks, these biases and predictable errors are particularly hard to correct by professional investors.

In Section 2, we lay out predictions which motivate our main empirical exercises. In Section 3, we explain our main empirical measure of retail trading activity and other key data sources. In Section 4, we explore the relationship between retail trading intensity and firm-level characteristics. In Section 5 we examine how retail investors trade around earnings announcements and under what conditions prices move with net retail order flow. Finally, in Section 6 we link retail trading activity to the earnings announcer premium and in Section 7 we conclude.

2 Hypothesis development

In this section we outline three sets of predictions which motivate our empirical exercises. The first set pertain to the characteristics that lead retail investors to trade some stocks relatively more intensely than others. The second set regard retail investors' trading patterns around earnings announcements. The last set examine the link between retail investor activity and the earnings announcer premium.

2.1 Cross-sectional heterogeneity in retail trading intensity

One way to view cross-sectional heterogeneity in retail trading intensity is that there are some stock-level features that tend to that attract retail investors. Past work has shown that retail investors select into stocks based on observable fundamentals e.g., small market capitalization, high book-to-market and low nominal price (Kumar and Lee (2006)). Retail investors are also attracted to stocks based on past performance e.g., those which have had extreme individual past returns (Bali et al. (2021)) or which have had low cumulative returns over the past 12 months (Luo et al. (2021)). A common theme among these characteristics is that such firms are more likely to have extreme fundamental realizations going forward. This could be because small firms with poor past returns are effectively highly leveraged (Daniel and Moskowitz (2016)) or because retail investors are selecting into firms with lottery-ticket-like payoffs (Bali et al. (2017)).

Accounting metrics and past returns are not the only characteristics that explain the cross-sectional distribution of retail activity. A second strand of literature has shown that retail investors are drawn to stocks which grab their attention (Barber and Odean (2008)). Attention can be stimulated by extreme news events (Hirshleifer et al. (2008)), analyst coverage (Martineau and Zoican (2019)), social media (Bali et al. (2021)) and news media coverage (Engelberg and Parsons (2011)). Which stocks end up being covered in the financial press, however, can be a function of the underlying fundamentals. For example, Martineau and Mondria (2022) argue the media is more likely to cover firms with more fundamental uncertainty, especially ahead of extreme news events. Further, events during the pandemic serve as a reminder that the trading of retail investors can itself launch a stock into the orbit of the financial press.

Yet another way to view cross-sectional heterogeneity in retail trading intensity is that there are some stocks which non-retail investors (that is, institutional investors) decide to avoid. For example, large investors might shun small stocks because once they own more than 5% of a firm's shares, they will have to file a 13D or 13G form with the SEC, subjecting them to more regulatory scrutiny (Edmans et al. (2013)). Alternatively, institutional investors may face mandates against owning unprofitable firms or volatile securities which prevent them from trading particular types of stocks (Beber et al. (2021)).

A common theme in these two views is that either retail are attracted to, or institutions avoid, stocks which are likely to experience extreme news events or returns in the near future. The question remains, however, as to whether there is a deeper shared characteristic between such firms that explains why they have systematically larger absolute earnings surprises and earnings-day returns.[2] One explanation is that some stocks are harder to value e.g., because they are composed of more intangible capital (Lev and Gu (2016)), have longer-duration assets Gormsen and Lazarus (2021), are more opaque (Bhattacharya et al. (2003)) or are more complicated (Cohen and Lou (2012)). This leads to our first testable prediction.

Prediction 1: Stocks with more retail trading intensity (high retail) should be relatively harder to value.

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)). Finally, this logic may not apply equally to all types of volatility. For example, among the small, lottery-ticket-like stocks favored by retail investors, idiosyncratic news explains a relatively larger share of both stock price and fundamental volatility than systematic news (Hou and Loh (2016)). This implies the following additional testable prediction.

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.

Building on prediction 1, 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, today's earnings are not as relevant for total present value. Alternatively, 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] In unreported results, we find that the absolute magnitude of earnings surprises is persistent at the stock level. Between 2010 and 2021, 75% of firms which had standardized unexpected earnings (SUE, defined in equation 2) in the top 20% of absolute surprises four quarters ago are in the top two quintiles of surprises in the current quarter, evidence that fundamental volatility is persistent at the stock level. See also e.g., Foster (1977).

Also building on prediction 1, if high retail stocks have more valuation uncertainty, we might expect them to be more expensive to trade, especially around earnings announcements. 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.[3] One way to measure adverse selection is through transaction costs, which implies the following prediction:

Prediction 1C: High retail stocks should be relatively more expensive to trade, especially around earnings announcements.

2.2 Retail trading behaviour around earnings announcements

The predictions in the last subsection are about how retail *select* different stocks. Our second set of predictions regard how retail investors trade around earnings announcements. There is a long literature studying such behaviour (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. Building on prediction 1, 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. This implies the following testable hypothesis:

Prediction 2: High retail stocks should have more abnormal retail trading intensity around earnings announcements.

The qualifier *abnormal* is included to account for level differences in retail trading intensity between high and low retail stocks. The mechanism discussed above implies that retail investors should be an especially large share of trading volume around earnings announcements in the stocks they have previously been trading intensely.

As a refinement of prediction 2, we might expect retail to be an even larger fraction of trading volume *before* earnings announcements. The logic is that either because they are inattentive (Liu et al. (2019)), or overconfident (Peng and Xiong (2006)), retail may keep trading at their usual intensity in the pre-announcement period even though institutional investors getting out. If this is true, then retail should become a larger share of trading volume, but not trade more as a fraction of shares outstanding.

Further, if institutional investors want to reduce their exposure to hard to value firms ahead of

[3] 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)).

earnings announcements, we would expect net buying by retail. The retail buying may be the result of the channels discussed above e.g., upcoming earnings announcements are covered by financial journalists or are discussed on social media, which stimulates retail demand.

Relatedly, we might expect significant trading on the announcement day itself by retail investors. This is motivated by findings that individual investors buy stocks with extreme past returns (Odean (1999)), which may occur at the time of an earnings announcement and that retail investor trading is stimulated by attention (Da et al. (2011)) and earnings announcements are attention-grabbing events. Jointly, these channels imply the following refinement of prediction 2:

Prediction 2A: Gross retail activity should be high in the pre-announcement period as a function of total trading volume, but not as a function of shares outstanding. The increased retail trading should be driven by retail in ows. Gross retail activity should also be especially high on the day of the earnings announcement itself.

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). 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 prediction 1C above, however, high retail stocks may be hard to trade, and therefore institutions may be unwilling to provide liquidity in this instance.[4] 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. This leads to the following refinement of prediction 2.

Prediction 2B: For high retail stocks, prices should move in the direction of net retail order

[4]If prediction 1C holds in the data, it may create a self-reinforcing cycle, in the sense that it will push even more institutions to avoid trading high retail stocks around earnings announcements. This is because, as shown in Di Maggio et al. (2022), institutional investors are especially sensitive to transaction costs.

imbalance, especially around earnings announcements.

2.3 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. First, if prediction 1 is true, high retail stocks should be hard to value, so 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.[5] This leads to the following testable prediction:

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

3 Data

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

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

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

[5] 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.

internalized. Note that $\text{RSVOL}_{i;t}$ is a lower bound on 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.[6]

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 exchange codes 1-3. By doing this, we miss out on some stocks popular with retail investors which are ADRs e.g., Nokia. 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., $\text{RSVOL}_{i; \quad 1}$. 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 $\text{RSVOL}_{i;}$. 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.[7] Because of our focus on earnings announcements we restrict the sample to firms for which we are able to construct earnings expectations.

[6] 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.

[7] 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 firms on average.

4 Retail Trading and Stock Characteristics

In this section, we document significant cross-sectional dispersion in retail trading activity. We then examine stock-level characteristics which explain this heterogeneity. Our main finding is that high retail stocks are relatively harder to value. Consistent with this, we show that such stocks are more expensive to trade, have more volatile fundamentals and larger (in magnitude) earnings-day returns.

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.

In Panel A of Table 1 we show the moments of $\text{RSVOL}_{i;t}$ in five equal-weighted portfolios constructed on the value of $\text{RSVOL}_{i;}$ in the prior month. The share of retail-initiated trades ranges from 2% in the bottom quintile, to 20% in the top quintile. The 90th percentile stock in the fifth quintile has more than a quarter of total share volume originate from retail traders. Table 1 also illustrates the heavy tail in retail trading: the share of retail-initiated volume doubles going from the second to fourth quintile, and doubles again going from the fourth to fifth quintile.

The retail sort is persistent over time. In Table 2 we show the 12 month transition 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.

Panels B and C of Table 2 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.

One concern with the results in Table 1 is that there is a significant size effect contributing to this gap in retail share of trading. The logic is that retail investors select into small stocks (Kumar and Lee (2006), Balasubramaniam et al. (2021)) which are less heavily traded on average. To illustrate the variation in retail trade intensity separate from a size effect, we first sort stocks into five market capitalization (size) quintiles, and then sort on prior month $\mathrm{RSVOL}_{i;}$ within each size portfolio. We show the resulting average retail-initiated trading in the bottom two sections of Panel A. Among those stocks in the bottom quintile of firm size, retail make up over 5 as much of trading in the top quintile of past retail intensity than the bottom quintile. Among large stocks, the corresponding magnitude is similar at about 4 .

The time-series dimension of average retail share is illustrated in the top panel 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.

In Panels B and C of Table 1, we repeat the same analysis as Panel A for turnover, and retail-initiated turnover. Both turnover and retail-initiated turnover are measured as the number of shares traded, normalized by shares outstanding and reported in percentage terms. Panel B shows that average turnover tends to be larger among stocks with high prior month retail share of trading. The monthly turnover ranges from 18.2% to 26% going from the low to high retail intensity quintile. The bottom parts of Panel B show that this gap in turnover is evident controlling for size: both within the smallest and largest size quintile the turnover is larger for high retail stocks. Further, the gap is sizeable: over 10 percentage points in both cases.

Finally, Panel C of Table 1 shows that a considerable amount of the gap in turnover across $RSVOL$ portfolios stems from retail-initiated trades themselves. The gap between high and low retail stocks is over 3 percentage points, and about the same within size sorts. In light of the results in the middle panel, this is a useful metric, because it is not sensitive to the higher average trading volume in high retail stocks. This makes the differences between high and low retail even more dramatic, with stocks in the top quintile having over 10 as much trading (in turnover terms) relative to those in the bottom quintile.

The difference between the median and the 90th percentile in Panel A of Table 1 suggests a tail-heavy distribution of retail trading. We quantify the time-series aspect this tendency of retail trades to be concentrated in Figure 2. 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. We do this calculation separately for all trades and well as for retail-initiated trades. In both panels, we collapse day-stock level dollar volumes down to the quarter-stock level, and then find the share of total quarterly dollar volume stemming from the indicated number of stocks with the highest dollar volumes in that quarter.

The right panel of Figure 2 shows that in the late 2000s, the top 10 retail-traded stocks made up over 20% of retail trading volume. This measure climbed to nearly 50% during the pandemic, reflecting an increased concentration of retail trading. It is not, however, a pandemic-specific phenomenon, as retail trading concentration was rising from 2016-2019. The left panel shows that total volume is less concentrated than retail trading volume, with the percent of total trading volume attributed to each group of stocks being below the corresponding lines in the right panel. Further, although overall trading volume also became more concentrated over the past 15 years, the trend was not as strong for total volume as it was for retail volume.[8]

4.2 Stock Characteristics across Retail Portfolios

The aim of our paper is to identify the retail habitat i.e., which types of stocks tend to attract a lot of retail trading intensity. To this end, we summarize firm characteristics across $RSVOL$ quintiles in Tables 3, 4 and 5. We group firm characteristics into three thematic groups: fundamentals, valuation and volatility/trading costs.

In Table 3, we present fundamentals across the $\text{RSVOL}_{i;}$ and size sort. We find that high retail stocks are smaller, younger, have low nominal prices, low recent returns, higher book-to-market ratios and tend to have low or negative earnings yields.[9] 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.

To document whether these patterns are driven by selection on firm size, we re-compute these summary statistics within quintiles of market capitalization. Some of the patterns e.g., the differences in book-to-market and age are driven largely by firm size differences. Other patterns, however, exhibit heterogeneity between large and small stocks. For example, among large stocks, retail tend to have the highest trading intensity among the mega caps.

In Table 4 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.

One dimension of difficulty to value is the duration of cash-flows. In Table 4 we report a proxy for cash-flow duration (CF) constructed after Gormsen and Lazarus (2021). We find that high retail stocks tend to have longer duration cash-flows, and this result holds both within small firms and large firms. 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

[8]As an alternative way to visualize the concentration of retail trading, in the Appendix Figure A1, we compare the cumulative share of dollar trading volume across all stocks sorted from low to high volume, and the cumulative share of retail-initiated dollar trading volume across all stocks sorted from low to high retail volume.

[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 e.g., Balasubramaniam et al. (2021).

(K_{Int}), knowledge capital (K_{Know}), and more 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).

Overall, the results in Table 4 establish a new fact consistent with prediction 1: stocks with high shares of retail trading tend to be harder to value.

Table 5 reports summary statistics on volatility and trading costs across retail portfolios. The first three 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 bottom two panels we again double sort on size and $\text{RSVOL}_{i;t}$, and these conditional averages show that retail heavy stocks tend to have higher volatility, even controlling for a size effect. In the second column, we report averages of trade-based intraday volatility, computed by averaging the squared 1-second returns each day.[11] These measures are also elevated for high retail stocks, though the difference mostly reflects a size effect. The remaining three columns summarize measures of liquidity. λ_2 stands for Kyle's lambda, estimated with an intercept.[12] This measure of illiquidity is higher for high retail stocks, and this gap survives controlling for size. Finally, Espread and Rspread stand for the percent effective and realized spread, respectively.[13] Both are higher among high retail stocks, but controlling for size this relationship only continues to hold for large stocks. Broadly, this evidence is consistent with prediction 1C, that retail stocks should be relatively more expensive to trade.

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

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

[11] This measures, as well as all the measures of trading costs in Table 5, are from the WRDS intraday indicators suite, which is built on the millisecond TAQ data.

[12] Motivated by Kyle (1985), λ_2 is computed by estimating a regression of returns on log dollar volume. For more details, see the *WRDS Intraday Indicators Formula Note*.

[13] 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.

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. This finding holds unconditionally, as well as within size buckets.

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:

$$\mathrm{SUE}_{i,t} = \frac{\mathrm{EPS}_{i,t} - \mathrm{E}_{t-1}[\mathrm{EPS}_{i,t}]}{P_{i,t-1}} \tag{2}$$

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.[14] Table 6 contains summary statistics on earnings-day returns, SUE and analyst coverage.

Before examining fundamentals, we directly look at earnings-day returns. The first column of Table 6 shows that high retail stocks have systematically lower earnings-day returns, and this is true both among stocks in the top and bottom 20% of the distribution of market capitalization. Further, column 2 shows that high retail stocks have a wider distribution of earnings day returns. Specifically, the standard deviation of these returns is almost 40% larger for high retail stocks than low retail stocks. Again, this pattern holds both among large and small stocks.

The next two columns of Table 6 show summary statistics on SUE. Consistent with having more negative and volatile earnings day returns, high retail stocks have both more negative SUE on average as well as a higher standard deviation of SUE on average. Again, the standard deviation of SUE is larger in the retail-heavy quintile. Collectively, the evidence in Table 6 is consistent with prediction 1A: high retail stocks both have more volatile earnings news and more volatile earnings-day stock returns.[15]

[14] 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.

[15] One alternative explanation for why high retail stocks have larger earnings surprises is that such stocks have

To further establish why the earnings of these stocks are so hard to predict, columns 5 and 6 of Table 6 replicate columns 3 and 4, but for the idiosyncratic component of earnings surprises. To decompose earnings news into idiosyncratic and systematic components, we follow the method in Glosten et al. (2021), regressing 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. Columns 5 and 6 show that the volatility of SUE is essentially all driven by the idiosyncratic component of SUE, suggesting that the larger SUE volatility relates to information that is specific to these firms, rather than larger sensitivity to economy-wide news. This 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.

Column 7 Table 6 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. High retail stocks tend to see a larger dispersion in analyst forecast errors, consistent with prediction 1A and more broadly with the notion that these are harder to value stocks. This result survives controlling for size. In the 8th column, we report the mean number of analysts covering each stock. This column shows that the number of analysts is similar for high and low retail stocks in the bottom of the size distribution and higher for high retail stocks in the top of the size distribution. This implies that the result in column 7 is not an artifact of fewer analysts providing estimates for high retail stocks. That firm-quarter level forecast errors are larger for high retail stocks is again consistent with the idea that retail traders select into hard to value stocks.

5.2 Return Sensitivity to Earnings Surprises

Prediction 1B builds on the logic that, by nature of being harder to value, high retail stocks likely respond less to earnings news than low retail stocks. To quantify this, we follow Kothari and Sloan (1992) and estimate earnings response regressions of the form

$$r^{i}_{t;t+n} = \quad + \quad \mathrm{SUE}_{i;t} + \quad X_{i;t} + \quad {}_{t} + \quad {}_{i} + \quad {}_{i;t}; \tag{3}$$

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.[16] 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

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, its increasing in the retail intensity.

[16]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.

month t 1 returns (Kumar and Lee (2006), Balasubramaniam et al. (2021), 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 3, 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 7 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 $Q5$ and SUE $Q1$ 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.

A potential concern with the results in Table 7 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.[17]

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 7: 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 8 we re-estimate the regression 3 but include dummy variables for quintiles of firm size, as well as their interaction with SUE. As Table 8 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

[17] In Appendix Table A1, we replicate Table 7, 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 in and where retail investors are rushing out before earnings announcement.

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 and 2A argue that retail investor trading activity should be especially high around earnings announcements, and that retail investors should be net buyers ahead of earnings announcements. To evaluate these hypothesises, in Figure 3 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 belonging to the top and bottom retail quintile around earnings announcements. 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 3 are consistent with predictions 2 and 2A. 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.

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:

$$\text{Retail Intensity}_{i,t} = \alpha + \beta_1 1_{i \in Q1_{t-1}} + \beta_2 1_{i \in Q2_{t-1}} + \beta_4 1_{i \in Q4_{t-1}} + \beta_5 1_{i \in Q5_{t-1}} + \gamma X_{i,t} + \delta_t + \eta_i + \epsilon_{i,t}, \quad (4)$$

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 $t-1$, where the middle quintile is the omitted group. $X_{i,t}$ are the same controls as Equation 3. 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 $t = -5$.[18] Table 9 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.[19] This suggests that, consistent with prediction 2A, institutional investors are trading less in the pre-earnings period rather than retail trading more. Second, and also consistent with prediction 2A, 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.

Next, we turn to the question of whether high retail stocks are more expensive to trade around earnings announcements. Prediction 1C argues that high retail stocks should not just be more expensive to trade on average, but they should be especially hard to trade around earnings announcements. Given the results in Table 5, however, we need to account for the higher average level of trading costs in high retail stocks. In addition, given the results in Table 6, we also need to account for the nature of the 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:

$$\begin{aligned}\text{DM Effective Spread}_{i,t} = {} & \alpha + \beta_1 1_{i \in Q1_{t-1}} + \beta_2 1_{i \in Q2_{t-1}} + \beta_4 1_{i \in Q4_{t-1}} + \beta_5 1_{i \in Q5_{t-1}} + \\ & \gamma_1 1_{i \in Q1 \text{SUE}_t} + \gamma_2 1_{i \in Q2 \text{SUE}_t} + \gamma_4 1_{i \in Q4 \text{SUE}_t} + \gamma_5 1_{i \in Q5 \text{SUE}_t} + \theta X_{i,t} + \delta_t + \eta_i + \epsilon_{i,t}\end{aligned} \quad (5)$$

[18] 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.

[19] 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 $t = 0$.

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 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 10 shows that, consistent with prediction 1C, 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.[20]

Prediction 2B argues that prices are more likely to move with retail investors' *net* demand in high retail stocks than low retail stocks. So far, we have been sorting firms into quintiles based on gross retail activity i.e., retail buys plus retail sells. So, to test prediction 2B, we construct a measure of retail order imbalance as

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

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

To test whether returns tend to move with or against net retail order flow, we estimate the following regression:

$$r_{i,t+\tau} = a + b\,mroibvol_{i,t} + \sum_{k=1}^{5} b_k 1_{i \in Qk_t} + \sum_{k=1}^{5} c_k 1_{i \in Qk_t} \times mroibvol_{i,t} + \gamma X_{i,t} + \delta_t + \delta_i + \varepsilon_{i,t} \tag{7}$$

where t is time in weeks. In words, the left-hand-side of Equation 7 are returns in the weeks relative to the "focal week" i.e., $\tau = 0$ and $1_{i \in Qk_t}$ are indicators for quintiles of gross or net retail flows, formed the previous month.

Table 11 contains the results. 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. If other investors tend to trade with retail, we would expect a positive relationship between $mroibvol$ and returns, while if other investors tend to trade against retail we would expect a negative coefficient.

The first row of the second column (indicated "0") shows that for the average stock, returns and mroibvol in week $\tau = 0$ move in the opposite direction, meaning retail investors are contrarian. The

[20]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 (Greenwood and Sammon, 2022).

fifth column restricts the sample to stock-weeks that happen to contain an earnings announcement and finds a much stronger negative relationship. The first row of columns indicated with “-1” shows that returns in week $= 1$ tend to have a negative correlation with mroibvol in week 0: retail traders are contrarian with respect to past returns. The positive coefficients in the first row in columns indicated “1” mean that mroibvol in week $= 0$ tend to have a positive correlation with returns in the subsequent week. In other words, the columns “1” show that past retail flow predicts future returns, consistent with Kaniel et al. (2008).

The remaining four rows show how these relationships differ across the retail sort. In particular, the columns indicated “0” in the fifth row, denoted Q5, show that high retail stock returns tend to move in the same direction as contemporaneous retail order flow, and particularly so in announcement weeks.

This result suggests that, for high retail stocks, returns move in the same direction as *mroibvol*, in contrast to the average stock. Further, comparing column “0” for all weeks to column “0” for announcement weeks shows that this effect is roughly 2 as strong right around earnings announcements. Overall, the results are consistent with prediction 2B, which states that for high retail stocks, retail order flow and prices should move in the same direction.

The columns indicated “-1” have a positive coefficient in the row corresponding to Q5, meaning retail investors tend to be especially contrarian relative to prior week’s returns in high retail stocks. Finally, the predictability of returns from current week mroibvol is stronger among the high retail share stocks, as shown in the columns “1”.

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 12, 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 12 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 12 shows that high retail stocks see lower announcement time returns, consistent with prediction 3. 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 6 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.

The analysis in Table 12 is done on the firm-announcement level. One concern with these results is that announcements tend to cluster in specific periods of the calendar year and these returns might not represent returns to an investor seeking exposure to the announcer risk premium. To better capture the announcement premium separate from a timing effect we follow Savor and Wilson (2016) and construct announcer and non-announcer portfolios. In Appendix Table A2, we show that this portfolio approach yields similar results to the firm-by-announcement level analysis.

In sum, the results in Tables 12 and A2 are consistent with prediction 3: high retail stocks do do not earn the earnings-announcer premium.

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. In these stocks, prices tend to move with retail order imbalance, suggesting that institutional investors choose not to lean against retail flow.

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 the findings document a new dimension of investor heterogeneity. Retail investors have a comparative advantage, relative to institutional investors, in holding and trading hard-to-value stocks. Future work can explore the specific motives and frictions that lead the different groups of investors to select stocks along this dimension.

8 Figures

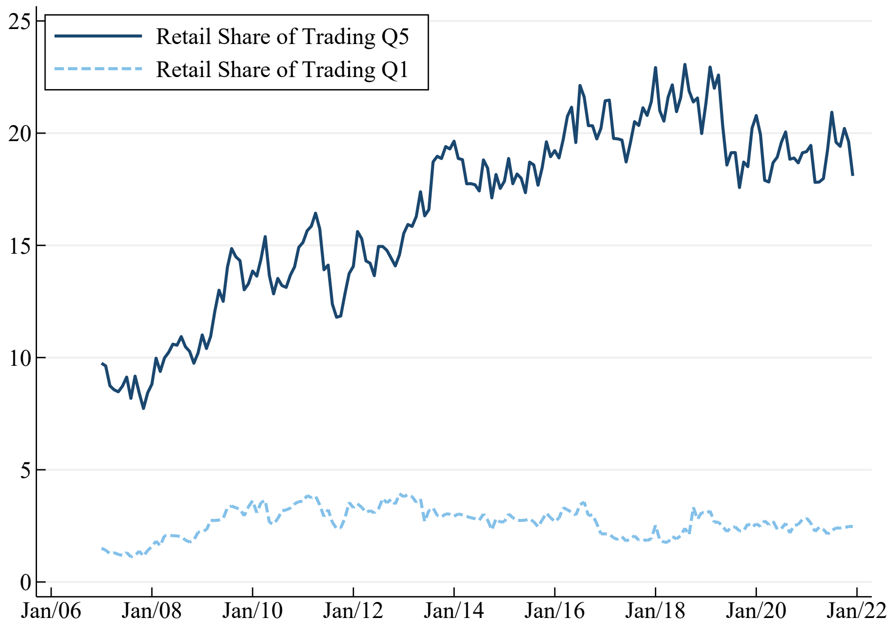

Figure 1: Retail share of trading volume. Average retail share of trading volume in the top and bottom quintile sorted on previous month's retail trading intensity.

Figure 2: Volume share of stocks with most trading volume. Retail volume share of most stocks with most retail-initiated trades. In each quarter we sum up the total dollar trading volume, and the total retail-initiated trading volume. The three lines in both graphs show the share of trading volume due to the 10, 50, and 100 most traded stocks in the quarter.

Figure 3: Abnormal Trading Volume around Earnings Announcements. Daily abnormal net retail volume and abnormal net retail share of volume, 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 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

Panel A. Retail share of trading volume

Size	Port.	mean	sd	p10	p50	p90	count
	1	2.03	0.73	0.97	2.11	2.92	92,581
	2	3.16	0.77	1.94	3.24	4.06	92,518
	3	4.45	1.03	2.80	4.56	5.66	92,502
	4	6.96	1.77	4.43	6.95	9.22	92,518
	5	15.29	6.24	8.83	13.88	24.06	92,439
1	1	3.86	2.00	1.29	3.79	6.56	18,585
1	5	23.22	6.56	14.51	23.38	31.30	18,449
5	1	2.21	0.69	1.15	2.32	2.99	18,550
5	5	8.54	3.65	4.36	7.92	12.91	18,415

Panel B. Monthly turnover

Size	Port.	mean	sd	p10	p50	p90	count
	1	18.35	14.16	6.10	15.02	33.61	92,581
	2	18.65	13.97	6.23	15.08	34.96	92,518
	3	19.17	15.79	5.10	14.89	37.92	92,502
	4	20.97	20.49	3.77	14.57	45.93	92,518
	5	26.27	32.86	2.38	13.24	69.25	92,439
1	1	8.52	11.05	1.35	5.13	18.64	18,585
1	5	23.24	34.55	1.91	9.21	65.26	18,449
5	1	19.98	12.07	9.72	17.08	33.16	18,550
5	5	30.88	28.08	8.24	20.93	66.43	18,415

Panel C. Retail-initiated monthly turnover

Size	Port.	mean	sd	p10	p50	p90	count
	1	0.33	0.34	0.08	0.25	0.64	92,581
	2	0.51	0.43	0.16	0.40	0.98	92,518
	3	0.75	0.71	0.19	0.56	1.52	92,502
	4	1.33	1.49	0.22	0.85	2.94	92,518
	5	3.66	5.03	0.29	1.66	9.80	92,439
1	1	0.33	0.68	0.02	0.16	0.75	18,585
1	5	4.45	6.04	0.35	1.82	14.37	18,449
5	1	0.38	0.26	0.15	0.32	0.68	18,550
5	5	2.67	3.71	0.50	1.37	5.98	18,415

Table 1: Trading in ve retail share of trading sorted portfolios. The top 5 rows of each panel are based on ve equal-weighted portfolios sorted on the retail share of trading volume in the previous month. The bottom 4 rows of each panel are based on twenty ve equal-weighted portfolios rst sorted on market capitalization, then on the retail share of trading volume, both in the previous month.

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 2: 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.

Size	Port.	Median Cap	Cap	Age	Prc	Past R	B/M	E/P	$_{CAPM}$	100-Inst.
	1	1,914	4,060	21.80	44.38	11.93	0.60	0.04	1.07	19.76
	2	2,036	6,233	22.83	46.68	13.30	0.56	0.04	1.11	20.65
	3	1,358	9,310	22.63	42.59	14.18	0.59	0.03	1.14	25.50
	4	596	13,122	20.74	32.52	13.59	0.67	-0.00	1.17	33.96
	5	152	5,565	14.62	12.96	5.84	0.82	-0.14	1.10	57.25
1	1	142	144	13.79	12.00	-1.48	0.94	-0.03	0.78	51.20
1	2	126	133	14.44	9.26	-2.73	0.94	-0.07	0.89	51.47
1	3	107	117	14.37	7.06	-4.25	0.97	-0.13	0.91	56.83
1	4	82	96	13.76	5.22	-6.52	0.97	-0.20	0.91	65.62
1	5	58	70	12.49	3.72	-10.64	0.90	-0.24	0.86	77.36
5	1	9,512	12,764	29.21	74.73	16.29	0.48	0.06	1.01	17.34
5	2	11,512	16,397	31.14	76.61	16.84	0.45	0.06	1.02	18.43
5	3	13,477	22,554	32.92	79.04	16.94	0.45	0.06	1.02	20.06
5	4	19,993	42,120	36.90	79.14	17.13	0.45	0.05	1.03	23.73
5	5	26,542	74,446	34.64	81.69	22.85	0.47	0.04	1.15	29.91

Table 3: Fundamentals in ve retail share of trading sorted portfolios. The top 5 rows are based on ve equal-weighted portfolios sorted on the retail share of trading volume in the previous month. The bottom 10 rows are based on twenty ve equal-weighted portfolios rst sorted on market cap, then on the retail share of trading volume, both in the previous month. Median cap is the median market capitalization while Cap is the mean; 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; $_{CAPM}$ is the market betacomputed over the previous 252 trading days; Inst. is institutional ownership from From 13F data, 100-Inst. is therefore another proxy for retail ownership.

Valuation								
Size	Port.	CF	K_{Int}	K_{Know}	K_{Org}	PAT	VU	Mispricing
	1	-0.19	-0.18	-0.24	-0.15	-0.14	-0.27	-0.16
	2	-0.15	-0.18	-0.21	-0.16	-0.04	-0.25	-0.14
	3	-0.05	-0.13	-0.14	-0.09	0.03	-0.16	-0.03
	4	0.11	0.05	0.02	0.07	0.12	0.09	0.13
	5	0.37	0.44	0.58	0.33	0.03	0.61	0.43
1	1	0.05	0.22	0.07	0.37	-0.28	0.59	0.33
1	2	0.14	0.47	0.29	0.54	-0.22	0.65	0.34
1	3	0.18	0.67	0.53	0.67	-0.16	0.69	0.36
1	4	0.21	0.79	0.84	0.68	-0.10	0.82	0.37
1	5	0.29	0.81	1.10	0.53	-0.12	0.86	0.48
5	1	-0.51	-0.30	-0.29	-0.33	0.05	-0.63	-0.41
5	2	-0.53	-0.31	-0.27	-0.32	0.20	-0.71	-0.47
5	3	-0.52	-0.31	-0.26	-0.31	0.35	-0.75	-0.46
5	4	-0.54	-0.32	-0.23	-0.31	0.61	-0.80	-0.44
5	5	-0.35	-0.32	-0.19	-0.33	1.01	-0.70	-0.20

Table 4: Valuation across ve retail share of trading sorted portfolios. The top 5 rows are based on ve equal-weighted portfolios sorted on the retail share of trading volume in the previous month. The bottom 10 rows are based on twenty ve equal-weighted portfolios rst sorted on market cap, then on the retail share of trading volume, both in the previous month. CF is cash ow duration, computed after Gormsen and Lazarus (2021); 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 ve 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).

Size	Port.	SD	Ivol t	2	Espread	Rspread
	1	1.90	0.43	1.21	0.25	0.12
	2	2.01	0.22	1.43	0.20	0.09
	3	2.26	0.23	2.11	0.28	0.12
	4	2.73	0.69	3.49	0.43	0.19
	5	4.00	1.19	8.77	0.92	0.48
1	1	2.64	3.08	2.56	1.41	0.88
1	2	3.07	2.04	5.98	1.20	0.63
1	3	3.57	2.21	8.20	1.23	0.65
1	4	4.23	2.79	11.05	1.36	0.78
1	5	4.58	1.75	14.30	1.48	0.80
5	1	1.54	0.02	0.44	0.06	0.02
5	2	1.60	0.02	0.48	0.05	0.02
5	3	1.68	0.01	0.49	0.06	0.02
5	4	1.73	0.03	0.50	0.07	0.03
5	5	2.05	0.01	0.61	0.07	0.03

Table 5: Liquidity in ve retail share of trading sorted portfolios. The top 5 rows are based on ve equal-weighted portfolios sorted on the retail share of trading volume in the previous month. The bottom 10 rows are based on twenty ve equal-weighted portfolios rst sorted on market cap, then on the retail share of trading volume, both in the previous month. SD is the standard deviation of daily stock returns in a given month; Ivol t is intraday volatility computed from trades; $_2$ is Kyle's lambda, estimated with an intercept; Espread and Rspread are the e ective and realized spread, computed using the methodology in Holden and Jacobsen (2014).

		Ann. Return		SUE		Idio. SUE		Analysts		
Size	Port.	Mean	SD	Mean	SD	Mean	SD	Disp.	Number	N
	1	0.35	6.49	0.01	1.02	0.38	1.83	0.16	8.94	31,478
	2	0.30	6.93	0.03	1.05	0.37	1.99	0.18	9.83	30,706
	3	0.21	7.47	-0.02	1.43	0.30	2.82	0.26	9.46	30,095
	4	0.06	8.28	-0.13	2.09	0.07	4.70	0.47	8.25	29,577
	5	-0.65	8.97	-0.49	3.42	-0.53	7.48	1.29	4.93	29,202
1	1	0.14	7.07	-0.35	2.58	-0.10	4.41	0.58	2.73	6,122
1	2	0.29	8.17	-0.46	3.07	-0.42	6.32	0.89	2.97	5,897
1	3	-0.06	8.69	-0.63	3.66	-0.86	8.68	1.33	3.03	5,948
1	4	-0.51	8.73	-0.68	3.96	-0.86	8.64	1.84	2.78	5,852
1	5	-1.37	8.78	-0.80	4.35	-1.23	10.09	2.20	2.26	5,740
5	1	0.19	5.01	0.06	0.48	0.43	1.30	0.10	14.55	6,211
5	2	0.17	5.31	0.07	0.41	0.44	1.04	0.09	15.69	6,241
5	3	0.14	5.52	0.07	0.56	0.42	1.16	0.11	16.49	6,114
5	4	0.10	5.52	0.05	0.71	0.40	1.08	0.13	17.69	6,001
5	5	-0.12	6.47	0.00	1.18	0.26	3.51	0.23	18.99	5,819

Table 6: Announcement-Level Summary Statistics. The top 5 rows are based on ve equal-weighted portfolios sorted on the retail share of trading volume in the previous month. The bottom 10 are based on twenty ve equal-weighted portfolios rst sorted on market cap, then on the retail share of trading volume, both in the previous month. Quarterly earnings announcements from 2007 to 2021.

	Standardized Unexpected Earnings					
	(0, 0)	(0, 2)	(0, 4)	(0, 0)	(0, 2)	(0, 4)
SUE	1.056	1.151	1.151	1.087	1.182	1.205
	(14.79)	(11.30)	(11.62)	(14.27)	(11.06)	(11.47)
SUE x Q1	0.158	0.152	0.204	0.210	0.211	0.262
	(1.93)	(1.26)	(1.83)	(2.38)	(1.68)	(2.20)
SUE x Q2	0.339	0.355	0.417	0.389	0.433	0.471
	(4.07)	(3.00)	(3.32)	(4.23)	(3.50)	(3.77)
SUE x Q4	-0.181	-0.139	-0.0793	-0.183	-0.121	-0.0566
	(-2.57)	(-1.39)	(-0.82)	(-2.53)	(-1.14)	(-0.54)
SUE x Q5	-0.474	-0.470	-0.434	-0.420	-0.394	-0.374
	(-6.85)	(-4.81)	(-4.61)	(-5.87)	(-3.99)	(-3.86)
Controls				Yes	Yes	Yes
N	148934	149100	149115	138195	138357	138370
R^2	0.09	0.09	0.09	0.10	0.10	0.10

Table 7: 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 rst day announcement information is tradeable during normal market hours. SUE is standardized unexpected earnings, de ned in Equation 2 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 pro t 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)
Observations	30359	30363	30363	28780	28784	28784

Table 8: 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. 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 2, Q_k 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.

Timing:	(-5,-1)	(-3,-1)	-1	0	(0,2)	(0,4)
	Demeaned Retail as % of Trading Volume (percentage points)					
Q1	-0.177***	-0.174***	-0.188***	-0.302***	-0.202***	-0.164***
	(0.032)	(0.033)	(0.042)	(0.048)	(0.036)	(0.034)
Q2	-0.0771***	-0.0894***	-0.0572*	-0.159***	-0.0997***	-0.0869***
	(0.021)	(0.023)	(0.030)	(0.037)	(0.026)	(0.022)
Q4	0.185***	0.191***	0.210***	0.211***	0.192***	0.163***
	(0.035)	(0.037)	(0.046)	(0.046)	(0.033)	(0.033)
Q5	1.282***	1.241***	1.338***	0.454***	0.566***	0.610***
	(0.085)	(0.090)	(0.101)	(0.100)	(0.086)	(0.081)
Observations	139,884	139,804	139,517	139,543	139,721	139,735
R-squared	0.102	0.079	0.049	0.053	0.068	0.08
Time FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES

Timing:	(-5,-1)	(-3,-1)	-1	0	(0,2)	(0,4)
	Demeaned Retail as % of Shares Outstanding (basis points)					
Q1	0.296***	0.149*	-0.237**	-4.499***	-2.097***	-1.296***
	(0.080)	(0.080)	(0.093)	(0.396)	(0.193)	(0.130)
Q2	0.204***	0.109*	-0.0649	-2.371***	-1.157***	-0.726***
	(0.054)	(0.056)	(0.075)	(0.316)	(0.156)	(0.107)
Q4	-0.502***	-0.370***	0.105	4.442***	2.207***	1.419***
	(0.114)	(0.117)	(0.133)	(0.470)	(0.241)	(0.170)
Q5	-0.707*	-0.348	1.037**	10.38***	5.443***	3.684***
	(0.412)	(0.394)	(0.432)	(1.060)	(0.653)	(0.524)
Observations	139,884	139,804	139,517	139,543	139,721	139,735
R-squared	0.07	0.06	0.051	0.098	0.083	0.074
Time FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES

Table 9: 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 de ned as (retail buys + retail sells)/(retail buys + retail sells + non-retail buys and sells) while in the bottom panel, retail trading intensity is de ned as (retail buys + retail sells)/(shares outstanding). In all columns, we subtract the mean of these quantities computed over the previous 252 trading days. kQ 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 xed e ects 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 pro t 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.

	Demeaned E ective Spread (basis points)					
	(-5,-1)	(-3,-1)	-1	0	(0,2)	(0,4)
Low Retail	-0.41	-0.388	-0.314	-0.845**	-0.631*	-0.388
	(0.265)	(0.284)	(0.357)	(0.413)	(0.335)	(0.254)
2	-0.085	0.0298	-0.218	0.0286	-0.133	-0.0637
	(0.171)	(0.181)	(0.253)	(0.334)	(0.245)	(0.156)
4	1.059**	1.254**	1.323**	1.365**	1.025***	1.019***
	(0.481)	(0.579)	(0.649)	(0.574)	(0.339)	(0.298)
High Retail	1.914**	1.961**	1.8	3.768***	3.339***	3.232***
	(0.786)	(0.953)	(1.249)	(1.158)	(0.994)	(0.871)
Observations	139,051	139,006	138,697	138,798	139,002	139,038
R-squared	0.101	0.096	0.09	0.103	0.103	0.114
Time FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES

Table 10: Retail activity and demeaned trading costs around earnings announcements. Left-hand-side variables are average demeaned e ective spread computed over various windows around earnings announcements. Demeaned e ective spread is e ective bid-ask spread minus average e ective spread over the calendar month before the earnings announcement. Q_k is an indicator variable for whether stock i was in retail intensity quintile k at the end of the month before the earnings announcement Quintiles of SUE are formed each quarter. Time xed e ects 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 pro t 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.

	All weeks			Announcement Weeks		
	-1	0	1	-1	0	1
Mroibvol	-1.015	-0.229	0.175	-1.091	-1.862	0.200
	(-14.17)	(-2.83)	(3.93)	(-7.34)	(-5.17)	(1.19)
Mroibvol x Q1	0.515	0.165	-0.0500	0.596	1.518	-0.113
	(9.23)	(2.66)	(-1.09)	(3.74)	(4.63)	(-0.55)
Mroibvol x Q2	0.210	0.0610	0.100	-0.0295	0.836	0.243
	(3.98)	(0.99)	(1.80)	(-0.15)	(2.50)	(1.07)
Mroibvol x Q4	-0.236	-0.0623	0.188	-0.620	-0.443	0.00380
	(-4.24)	(-0.99)	(2.87)	(-2.89)	(-1.15)	(0.01)
Mroibvol x Q5	-0.369	1.133	0.295	-0.785	1.672	0.329
	(-4.87)	(11.59)	(3.87)	(-3.05)	(3.60)	(1.18)
Constant	0.171	0.200	0.207	0.160	0.452	0.267
	(10.75)	(12.97)	(12.20)	(4.30)	(7.80)	(7.96)
Observations	1964869	1973772	1965183	150358	150896	150433
R^2	0.002	0.001	0.001	0.007	0.006	0.006

Table 11: Mroibvol 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. Q_k is an indicator variable for whether stock i was in retail intensity quintile k at the end of the month before the earnings announcement Dependent variable is return in excess of the equal-weighted market return.

	(-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 12: Cumulative returns around earnings announcements . Column headers refer to rst and last days in return window. 0 is the rst 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 Sizej is an indicator for whether rm i was in size quintile j at the end of the month before the earnings announcement. Five panels sorted on size. Monthly xed e ects. Standard errors clustered by rm and month. Earnings announcements in 2007-2021.

10 References

References

Balasubramaniam, V., Campbell, J. Y., Ramadorai, T., and Ranish, B. (2021). Who owns what? a factor model for direct stock holding. Technical report, National Bureau of Economic Research.

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 overcon dence in nancial 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. Available at SSRN 4202874.

Barber, B. M., Huang, X., Ko, K. J., and Odean, T. (2020). Leveraging overcon dence. 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 rst? The Review of nancial studies, 15(2):455{488.

Barber, B. M. and Odean, T. (2008). All that glitters: The e ect of attention and news on the buying behavior of individual and institutional investors. The review of nancial studies, 21(2):785{818.

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.

Bhattacharya, U., Daouk, H., and Welker, M. (2003). The world price of earnings opacity. The accounting review, 78(3):641{678.

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

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 nance, 56(1):1{43.

Cohen, L. and Lou, D. (2012). Complicated rms. Journal of nancial economics, 104(2):383{400.

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

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. The journal of nance, 66(5):1461{1499.

Daniel, K. and Moskowitz, T. J. (2016). Momentum crashes. Journal of Financial economics, 122(2):221{247.

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

DellaVigna, S. and Pollet, J. M. (2009). Investor inattention and friday earnings announcements. The journal of nance, 64(2):709{749.

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

Di Maggio, M., Egan, M., and Franzoni, F. (2022). The value of intermediation in the stock market. Journal of Financial Economics, 145(2):208{233.

Di Maggio, M., Franzoni, F. A., Kogan, S., and Xing, R. (2021). Avoiding volatility: Institutional trading before scheduled news. Available at SSRN 3786363.

Diether, K. B., Malloy, C. J., and Scherbina, A. (2002). Di erences 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 e ect 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.

Engelberg, J. E. and Parsons, C. A. (2011). The causal impact of media in nancial markets. the Journal of Finance, 66(1):67{97.

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

Foster, G. (1977). Quarterly accounting data: Time-series properties and predictive-ability results. Accounting Review, pages 1{21.

Frazzini, A. and Lamont, O. A. (2007). The earnings announcement premium and trading volume. Available at SSRN 986940.

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 e ciency of underlying securities. Management Science, 67(1):22{47.

Golubov, A. and Konstantinidi, T. (2021). Valuation uncertainty. Rotman School of Management Working Paper, (3850807).

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

Greenwood, R. and Sammon, M. C. (2022). The disappearing index e ect. Technical report, National Bureau of Economic Research.

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. In Proceedings of Paris December 2021 Finance Meeting EUROFIDAI-ESSEC.

Hartzmark, S. M. and Shue, K. (2018). A tough act to follow: Contrast e ects in nancial 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.

Hou, K. and Loh, R. K. (2016). Have we solved the idiosyncratic volatility puzzle? Journal of Financial Economics, 121(1):167{194.

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 e ciency. The Journal of nance, 48(1):65{91.

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 nance, 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 Sto man, N. (2017). Technological innovation, resource allocation, and growth. The Quarterly Journal of Economics, 132(2):665{712.

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 coe cients. 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: Journal of the Econometric Society, 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.

Liu, H., Peng, L., and Tang, Y. (2019). Retail attention, institutional attention. Baruch College Zicklin School of Business Research Paper, (2019-05):03.

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 e ect. Available at SSRN 3544949.

Martineau, C. and Mondria, J. (2022). News selection and asset pricing implications. Available at SSRN 4194851.

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

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 nance, 53(5):1775{1798.

Odean, T. (1999). Do investors trade too much? American economic review, 89(5):1279{1298.

Peng, L. and Xiong, W. (2006). Investor attention, overcon dence 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 nancial studies, 30(4):1270{1315.

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

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 Concentration of Retail Trading

As an alternative way to illustrate the concentration of retail trading relative to Figure 2, we compare the cumulative share of dollar trading volume across stocks sorted from low to high volume, and the cumulative share of retail-initiated dollar trading volume across stocks sorted from low to high retail volume. Speci cally, each day, for each stock, we compute the total dollars of trading volume from retail as the sum of retail buys and retail sales multiplied by the closing price. We also construct an analogue for total volume, multiplying the number of shares traded in TAQ by the closing price. Next, we add this up across all days in a given year, and rank stocks from the lowest to the highest dollar volume (with 1 denoting the stock with the most dollar volume). Finally, we compute the share of dollar volume attributed to each stock and then cumulate this share from the lowest to the highest ranked stock.

Figure A1 plots these cumulative shares for retail trades and total trades for two years: 2010 and 2020. In both years, the retail volume (large blue dots) sits below the total volume (small red dots), and has a steeper slope approaching the top ranked stock. This pattern, that retail trading is more concentrated than overall trading, holds for every year in our sample. Further, we see that in 2020, the gap between the retail and total lines widened, and the slope near the top rank stocks became even steeper. This suggests that retail became more concentrated during the pandemic, with the top 10 retail stocks making up about 40% of the total retail dollar volume in 2020.

A.2 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 modi ed version of Equation 3:

$$r^{i}_{t;t+n} = \quad + \quad SUE_{i;t} + {}_{1}1_{i2Q1 \quad 1} + {}_{2}1_{i2Q2 \quad 1} + {}_{4}1_{i2Q4 \quad 1} + {}_{5}1_{i2Q5 \quad 1} + X_{i;t} + {}_{t} + {}_{i} + {}_{i;t} \quad \text{(A1)}$$

where $1_{i2Qk \quad 1}$ are indicators for quintiles of gross or net retail ows, formed over the 22 trading days before the earnings announcements. In Table A1, 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 7, 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 ows ahead of the earnings announcement instead of gross ows. The coe cients on the \Low Flow" (i.e., most retail selling) and "High Flow"

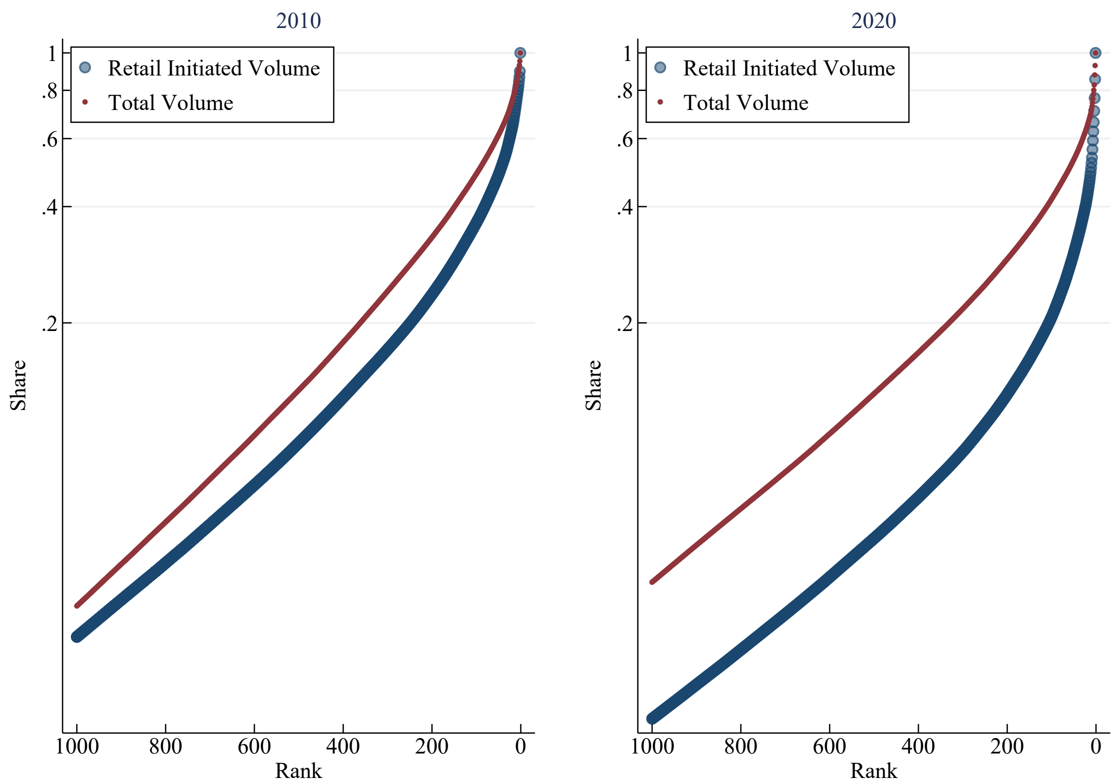

Figure A1: Cumulative Volume Share in 2010 and 2020. Each year, stocks are ranked on their retail-initiated turnover, and total turnover. Each data point represents the cumulative share of volume for stocks below the indicated rank.

(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.

	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 A1: 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.

A.3 Calendar Time Earnings Announcement Portfolio

In section 6 we show that high retail stocks earn lower returns in response to earnings announcement news.

To separate the earnings announcer premium from a timing effect, we follow Savor and Wilson (2016) and construct announcer and non-announcer portfolios. These portfolios go long stocks that are in their announcement window, and short all others stocks. The analysis is done within each of the size- and retail-sorted portfolios.

We report the results in Table A2. In the first set of three columns we report the average monthly returns from strategies that invest in announcing firms in a six trading day window around the earnings event. The second column, denoted "Ann.", shows the equal-weighted excess return of firms that are currently in their announcement window (and holds the risk-free asset if no such firms exist). The pattern of returns reflects that documented in Table 12: high retail stocks across the size spectrum tend to see lower returns on announcement. The pattern among non-announcing firms is weaker, and so the announcement long-short, denoted "Gap" again shows a decreasing pattern across the retail sort.

The second and third set of columns repeats the analysis but separates out pre- and post-announcement returns. The results are again consistent with the firm-level analysis in Table 12. The portfolio investing in high retail announcers in the pre- period does well, but that good performance is negated by the poor post-announcement behavior. Specifically, in Column 6, "Gap" tends to become more positive going from low to high retail, suggesting high retail stocks tend to have the highest pre-announcement returns. On the other hand, in Column 9, "Gap" tends to become more negative going from low to high retail, implying high retail stocks have the lowest post-announcement returns.

		Event window (-3, 2)			Event window (-3, -1)			Event window (0, 2)			
Size	Port.	N.-Ann.	Ann.	Gap	N.-Ann.	Ann.	Gap	N.-Ann.	Ann.	Gap	
1	1	0.95	0.32	-0.63	1.06	1.14	0.09	0.99	0.49	-0.49	
1	2	1.21	3.22	2.01	1.41	2.75	1.35	1.30	2.89	1.58	
1	3	1.20	3.20	2.00	1.27	5.02	3.75	1.39	2.36	0.97	
1	4	1.31	1.57	0.26	1.13	6.38	5.25	1.62	-2.18	-3.79	**
1	5	1.37	1.19	-0.18	0.55	8.98	8.43	1.79	-6.51	-8.29	***
2	1	0.60	2.30	1.71	0.84	0.51	-0.33	0.61	2.47	1.86	
2	2	0.89	4.13	3.24	1.08	2.04	0.96	0.97	4.57	3.61	**
2	3	0.87	2.01	1.14	1.00	2.59	1.58	0.91	1.52	0.61	
2	4	1.31	1.35	0.04	1.25	3.39	2.14	1.38	-1.14	-2.52	*
2	5	1.10	0.31	-0.79	0.84	3.93	3.09	1.25	-2.06	-3.31	*
3	1	0.67	1.41	0.74	0.85	0.50	-0.35	0.67	2.04	1.38	
3	2	0.99	1.30	0.32	1.05	0.77	-0.28	0.97	0.60	-0.37	
3	3	1.01	1.42	0.40	1.15	1.24	0.09	1.02	1.90	0.88	
3	4	1.14	1.50	0.37	1.11	2.69	1.58	1.16	0.46	-0.70	
3	5	1.41	1.46	0.05	1.31	2.53	1.22	1.43	-0.27	-1.70	
4	1	0.74	1.60	0.86	0.86	0.69	-0.17	0.77	2.08	1.31	
4	2	0.86	1.52	0.66	0.93	1.02	0.09	0.89	1.22	0.33	
4	3	0.96	0.95	-0.01	0.99	0.91	-0.08	0.96	0.53	-0.43	
4	4	1.11	1.77	0.66	1.13	0.34	-0.79	1.13	1.99	0.86	
4	5	1.12	-0.58	-1.70	0.93	2.25	1.32	1.16	-1.86	-3.02	**
5	1	1.01	1.05	0.04	1.03	0.98	-0.05	1.03	1.03	-0.00	
5	2	0.94	1.05	0.11	0.94	0.91	-0.03	0.96	0.78	-0.18	
5	3	0.99	0.97	-0.02	1.01	0.78	-0.23	1.00	0.41	-0.59	
5	4	1.02	0.78	-0.25	1.05	0.82	-0.23	1.05	0.24	-0.81	
5	5	1.07	-0.22	-1.29	0.93	1.52	0.59	1.12	-1.72	-2.83	***

Table A2: Announcer risk premium. Monthly returns of long-short strategies that are long announcing firms, short non-announcing firms within the indicated size-retail share portfolio during the indicated time period. Monthly data from 2007 to 2021. *, **, and *** indicate statistical significance of the returns on the last column at the 10%, 5%, and 1% levels, respectively.