The Retail Habitat

Toomas Laarits and Marco Sammon

First version: May 9, 2022
This version: October 31, 2022

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. As an additional consequence, the well-known earnings announcer risk premium is limited to low retail stocks only. Further, high-retail stocks are more sensitive to retail order flow and are especially expensive to trade around earnings announcements. Overall, the findings document a new dimension of investor heterogeneity and suggest the comparative advantage of retail in holding hard-to-value stocks.

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. The set of stocks that experience 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. As the figure shows, 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 a lot of attention to the determinants of retail trading in the cross-section.

In light of the heterogeneity in retail trading intensity, a natural question is whether high and low retail stocks perform differently in terms of returns. In the bottom panel of Figure 1 we plot the returns of a long-short portfolio that goes long the quintile of stocks with the highest share of retail-initiated trades the prior month, and goes short the quintile of stocks with the lowest share of retail-initiated trades. In addition to a broad outperformance since 2020, this retail-tilted portfolio shows intriguing performance at other times: it tends to not lose as much value in market downturns such as during the Global Financial Crisis, though at the cost of not gaining much during the subsequent recovery. In other words, there is substantial heterogeneity in the intensity of retail trading, and the stocks retail prefers to trade have a performance distinct from the aggregate market portfolio.

Prior work has documented 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. Based on this work, it would seem that stocks with a high retail presence could be fertile grounds for smart money to reap profits from exploiting the poor trading habits of retail investors.

In this paper we document new facts on the characteristics of retail-heavy stocks that can reconcile these seemingly contradictory facets of retail trading. We argue that retail traders make up a large part of volume in shares that hard to value: firms that have longer-duration cashflows, have high shares of intangible capital and high valuation uncertainty. Or, equivalently, institutional investors tend to shun trading in such stocks. Our evidence suggests that institutions' desire to avoid hard to value stocks makes it difficult for them to bet against retail.

Consistent with high retail stocks being hard to value, we show that these stocks have larger earnings news volatility and most of it comes from the firm-specific component of earnings news. Further, consistent with current cashflows being less relevant for hard to value firms, we show that high retail stocks' prices are less sensitive to standardized unexpected earnings (SUE). We also document a large role of retail investors around earnings announcements: high retail stocks have substantial abnormal retail trading before and on announcements, and the returns move in the same direction as retail flow. Finally, we document that the earnings announcer risk premium, the average return earned in response to scheduled earnings news, is positive only for low retail stocks.

We start by documenting new facts pertaining to the distribution of retail trading in the cross-section. Retail traders as a group tend to concentrate on a subset of stocks. The concentration of retail-initiated volume is more heavy-tailed than that of overall trading volume, and this concentration has only increased in the recent past. Moreover, retail trading intensity is persistent: stocks with high shares of retail trades continue to see a large share of retail trades in the subsequent year, suggesting that retail trading intensity is a function of persistent underlying characteristics.

Motivated by the empirical regularity of high retail trading intensity in hard-to-value stocks, we turn our attention to earnings announcements. We argue that because current earnings should be less 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, retail stocks' prices respond about half as much. This effect is almost unchanged after controlling for a litany of characteristics known to be correlated with retail activity (e.g., past returns, liquidity, and fundamentals as discussed in Kumar and Lee (2006), Balasubramaniam et al. (2021) and Luo et al. (2021)) and holds at almost every point along the firm size distribution.

We also demonstrate a direct impact of retail trades on return dynamics around earnings announcements. Broadly, retail-heavy stocks see large abnormal trading around earnings announcements. More specifically, in the pre-announcement period, this is true relative to total trading volume, but not relative to shares outstanding. This suggests that institutional investors tend to avoid trading before earnings announcements. On the earnings announcement day itself, retail trading is high both relative to volume and shares outstanding. Consistent with these two pieces of evidence, we find that prices tend to move in the direction of net retail order flow only for high retail stocks.

A natural question is why more institutions don't bet against retail. One explanation is that institutions are especially sensitive to trading costs (Di Maggio et al. (2022)) and high retail stocks are especially expensive to trade around earnings announcements. One explanation for this is that assets which are harder to value often have larger adverse selection costs (Krinsky and Lee (1996)).

Consistent with this logic, we find that high retail stocks have high bid-ask spreads in a 5-day window around earnings announcements, relative to the historical stock-level average. The result is even stronger when using unconditional bid-ask spreads, as high retail stocks are generally more expensive to trade. This result is robust to controlling for the extreme nature of news for high-retail stocks and a host of firm-level controls and fixed effects.

Finally, we show that high retail stocks have substantially lower earnings day returns than the average stock, consistent with earnings news representing a less relevant signal regarding these stocks' value. This result holds both on the stock-announcement level, or by constructing calendar-time portfolios that go long all announcing firms. As discussed in Savor and Wilson (2016), a potential explanation for the earnings announcer premium is that announcing firms provide information about non-announcing firms. This is unlikely to apply to high retail stocks, however, as their earnings surprises are mostly composed of idiosyncratic news. Consistent with this, the announcer premium is decreasing in the degree of retail trading activity.

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 strand 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. More broadly, Gabaix et al. (2022) use novel holdings data of U.S. households and study their rebalancing across asset classes.

Perhaps surprisingly, the literature on retail investors has not devoted a lot of 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 cross-sectional heterogeneity in retail trading intensity, and this can be explained by a metric which is not obvious from looking only at accounting numbers, past returns and factor betas. Our results here add to this literature by suggesting that difficult-to-value stocks attract particular retail attention or, equivalently, repel institutional 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.

Another way to view cross-sectional heterogeneity in retail trading intensity is that is that there are some stocks or events which non-retail investors (i.e., 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)). In addition, Di Maggio et al. (2021) show that some institutional investors tend to reduce their exposure to stocks before earnings announcements. Their argument is that extreme earnings-day returns can reflect poorly on investment managers' stock picking skill and lead to outflows.

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 i.e., why some firms have systematically larger absolute earnings surprises and earnings-day returns.[1] 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)). This implies the following additional testable prediction.

Prediction 1A: *High retail stocks should have more volatile earnings-day returns and earnings news.*

Prediction 1A may not apply equally to all types of volatility. For example, Di Maggio et al. (2021) show that institutions further reduce their exposure to earnings announcements when the news is likely to have a large idiosyncratic component. This implies the following refinement of prediction 1A.

Prediction 1B: *High retail stocks should have more idiosyncratic earnings-day stock price volatility and their earnings surprises should be driven by the idiosyncratic component of earnings news.*

Building on prediction 1, if high retail stocks are hard to value, when news about current cashflows arrives, it may have a relatively smaller affect 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

[1] In the online appendix, we show 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. See also e.g., Foster (1977).

not as relevant for total present value. Alternatively, in hard to value stocks, investors may focus on different pieces of the news, leading to more disagreement and ultimately to under-reaction (Hong and Stein (2007)). 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 1C: *High retail stocks should respond relatively less to earnings news.*

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 may be because they are aware of such adverse selection risk, while retail investors are not.[2] One way to measure adverse selection is through transaction costs, which implies the following prediction:

Prediction 1D: *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 a 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 word *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.

[2]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)).

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 bigger relative to a fixed quantity e.g., shares outstanding.

Further, because institutional investors are net sellers ahead of earnings announcements, there should be net buying by retail. The retail buying may be the results 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. 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 inflows.*

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.

Prediction 2B: *Gross retail activity should 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 leaning against retail trades. 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 to underperformance 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 1D above, however, high retail stocks may be hard to trade, and therefore institutions may be unwilling to provide liquidity in this instance.[3] Finally, there may be frictions which prevent institutions from trading the type of stocks that retail investors favor Beber 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 2C: *For high retail stocks, prices should move in the direction of net retail order 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 1B is true, high retail stocks' earnings news will have a relatively larger idiosyncratic component, which is less useful for valuing non-announcing firms. In either case, we would expect high retail stocks to have a smaller or non-existent announcer premium.[4] This leads to the following testable prediction:

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

3 Data

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

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

where $\text{RBuy}_{i,t}$ and $\text{RSell}_{i,t}$ are the number of shares in retail-initiated buy and sell trades, respec-

[3] If prediction 1D holds empirically, 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.

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

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

Retail trades are identified using the algorithm proposed in Boehmer et al. (2021) that relies on the regulation of U.S. security markets requiring price improvement for retail-initiated trades that are internalized. Note that $\text{RSVOL}_{i,t}$ 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 small stocks. All that matters for most of our findings, however, is that the *ordinal* ranking of stocks on gross retail activity is correct. We construct this measure using the TAQ millisecond data from 2007-2021.[5]

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 Wharton Research Data Services (WRDS) provided linking table.

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

For our analysis of how retail investors respond to news, we focus on earnings announcements. To this end, we need to identify 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

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

use the first trading day after the release date of quarterly earnings (RDQ) in Compustat if it is at least one day before the date identified using IBES (Livnat and Mendenhall (2006)). We use the mapping file from WRDS to link IBES data to CRSP.[6]

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. Finally, we show that although high and low retail stocks have similar average returns, retail-heavy stocks perform differently during economic crises.

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. Recent work in Boehmer et al. (2021) employs the regulatory structure of U.S. equity markets to identify retail-initiated trades. In the 2007 to 2021 sample, marketable retail orders identified by this 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 $\mathrm{RSVOL}_{i,t}$, is large relative to its unconditional mean.

In Panel A of Table 1 we show the moments of $\mathrm{RSVOL}_{i,t}$ in five equal-weighted portfolios constructed on the value of $\mathrm{RSVOL}_{i,\tau}$ 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 share volume originate from retail traders. Table 1 also shows the heavy tail in retail trading: the share of retail-initiated volume doubles going from the third to fourth quintile, and doubles again going from the fourth to fifth quintile.

The retail sort is quite 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 65% probability of remaining in the top quintile 12 months in the future. These same stocks have an over 88% 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

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

the size of the stock at time $t = -12$. Again we see substantial persistence in portfolio assignments over time. Among small stocks with the highest share of retail trading, over 70% are in the top two quintiles 12 months later. Among large stocks 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.

Returning to Table 1, there is a significant size effect contributing to this gap in retail share of trading; 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 $\text{RSVOL}_{i,\tau}$ 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 $5\times$ 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 $4\times$.

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 turnover tends to be larger among stock with high prior month retail share of trading. The monthly turnover ranges from 17.3% to 21.7% 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 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 $10\times$ 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. To better illustrate the concentration of retail trading, we compare the cumulative share of trading volume across stocks sorted from low to high volume, and the cumulative share of retail-initiated trading volume across stocks sorted from low to high retail volume. Specifically, 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 2 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.

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 Panel A of Table 3, we present fundamentals across the $\text{RSVOL}_{i,\tau}$ and size sort. We find that high retail stocks are smaller, younger, have higher book-to-market ratios and tend to have low or negative earnings yields.[7] Additionally, high retail stocks tend to have somewhat smaller market betas. Consistent with the findings in Luo et al. (2021), retail investors trade stocks tilted away from momentum.

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 CAPM beta are driven entirely by firm size differences. Indeed, within size quintiles high retail stocks tend to be high CAPM beta stocks. 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. This is accounted for by stocks like Tesla,

[7] 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).

Apple and Amazon.

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 tends to trade stocks that are hard to value. One dimension of difficulty to value is the duration of cash-flows 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 knowledge capital, more organization capital and more book intangible capital from mergers.[8] Further, high retail have more valuable patents, relative to their total market value.[9] Also consistent with high retail stocks being harder to value, they have higher mispricing scores (Stambaugh and Yuan, 2017) and more valuation uncertainty (Golubov and Konstantinidi, 2021). Some of these patterns, like those on organization capital, seem to mostly reflect a size effect, while others hold within firm size buckets.

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 $\mathrm{RSVOL}_{i,\tau}$, and these conditional averages show that retail heavy stocks tend to have higher volatility, even controlling for a size effect. In the second and third column we report averages of quote-based and trade-based intraday volatility, computed by averaging the squared 1-second returns each day.[10] These measures are also elevated for high retail stocks, though the difference mostly reflects a size effect. The remaining four columns summarize measures of liquidity. λ_1 and λ_2 stand for Kyle's lambda, estimated with and without an intercept.[11] Both measures of illiquidity are higher for high retail stocks, and this gap survives controlling for size. Espread and Rspread stand for the percent effective and realized spread, respectively.[12] Both are higher among high retail stocks, but

[8]These are measured as capitalized R&D, capitalized SG&A and "OffBS" from the WRDS Peters and Taylor Total Q dataset Peters and Taylor (2017)).

[9]We obtain the market value of patents 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.

[10]These 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.

[11]Motivated by Kyle (1985), these are computed by estimating a regression of log price changes on volume. For more details, see the *WRDS Intraday Indicators Formula Note*.

[12]Specifically, following Holden and Jacobsen (2014), the percent effective spread for any trade k is defined as: Percent Effective $\text{Spread}_k = (2D_k(P_k - M_k))/M_k$ 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

controlling for size this relationship only continues to hold for large stocks. Broadly, this evidence is consistent with prediction 1D that retail stocks should be relatively more expensive to trade.

4.3 Stock Returns across Retail Portfolios

Given differences in the characteristics between high and low retail stocks, a natural question is whether these two groups of firms have systematically different average returns. To quantify this, we again sort stocks into quintiles based on the previous month's retail trading intensity. Then we construct a portfolio which goes long the top quintile and short the bottom quintile, which we call the long-short (L-S) retail strategy. In the lower panel of Figure 1, the dark blue line plots the cumulative returns to this strategy when the portfolios are equal weighted, while the light blue line shows the returns to this strategy when the portfolios are value weighted. For comparison, the red line plots the cumulative return on the market factor from Ken French's website.

While the average returns to the value-weighted L-S retail strategy are not significantly different from the value-weighted market (t=0.32), the L-S retail strategy seems to earn different returns in crises. For example, the L-S retail strategy did not decline as much during the financial crisis and did well in 2020, essentially avoiding the decline associated with the onset of the COVID-19 pandemic. This effect, however, was strongest in the equal weighted L-S strategy, which puts more emphasis on some of the small retail heavy stocks e.g., Game Stop and AMC. This is consistent with the results in Greenwood et al. (2022), who find that the retail long-short did well on arrival of stimulus checks in the COVID era. That the retail long-short has a distinct performance from the market portfolio again underlines the specific nature of the stocks retail trades focus on,

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

as Percent Effective $\text{Spread}_k = \left(2D_k(P_k - M_{k+5})\right)/M_k$ where M_{k+5} is the midpoint 5 minutes after trade k. The realized spread is designed to capture how far the midpoint moves after trade k occurs.

5.1 Distribution of Standardized Unexpected Earnings

As a test of prediction 1A, 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.[13] Especially among small stocks, the retail distribution has a significant number of observations in both the upper and lower tails. While less stark, this pattern also holds among large stocks.

To visualize the differences in the distribution of earnings surprises among high and low retail stocks, each quarter, we sort firms into 5 buckets based on their realized SUE. Table 6 shows the count of all earnings announcements in our sample within each of these SUE buckets crossed with our retail intensity buckets. The 5th row shows that the top retail quintile has a significant tilt toward extreme events, with the majority of the distribution in the top or bottom SUE bucket. While this in part reflects a size effect, it also holds conditional on size. In the second and third panel of Table 6 we first condition on size, and then carry out the same calculation within size buckets. As the table shows, controlling for size, the high retail share stocks see more earnings events in the extreme SUE quintiles.

To better understand the differences implied by Table 6, in the top left panel of Figure 3, we plot the distribution of earnings-day returns for high and low retail stocks, defined as stocks in the top and bottom quintile of retail activity, formed over the month before the earnings announcement. We further zoom in on firms in the bottom 20% of market capitalization (formed each quarter) i.e., small stocks. The distribution of returns for small high retail stocks is more skewed to the left and has heavier tails.[14] The fact that high retail stocks have low earnings-day returns is not a new phenomenon, as one might expect given the poor fundamental performance of "meme stocks" in 2020 and beyond. The bottom left panel of Figure 3 shows this pattern is mirrored for stocks in the top 20% of market capitalization.

[13] 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.

[14] In all 4 panels in Figure 3 we Winsorize the variables of interest at the 5% and 95% level by quarter, which is why there are large bars at either end of the histograms.

To further document these differences, Panel A of Table 7 shows summary statistics for SUE among high and low retail stocks. High retail stocks have both more negative SUE on average as well as more volatile SUE on average, meaning, a higher standard deviation. Note that the median SUE is roughly zero across the retail sort. The bottom two parts of Panel A repeat the analysis within size quintiles. Again, the standard deviation of SUE is larger in the retail-heavy quintile. Collectively, the evidence in Figure 3 and Tables 6 and 7 are consistent with prediction 1A: high retail stocks both have more volatile earnings news and more volatile earnings-day stock returns.[15]

To further understand why the earnings of these stocks are so hard to predict, Panel B of Table 7 replicates panel A, 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). Specifically, we regress firm-level SUE on market-wide (value-weighted) SUE and SIC-2 industry-wide (value-weighted) SUE in five year rolling windows. The systematic component of earnings is the predicted value from this regression in the last year of the five year rolling window, while the idiosyncratic component is the residual. Panel B shows that the increased volatility of SUE is essentially all driven by the idiosyncratic component of SUE, suggesting that the larger SUE volatility relates to information is specific to these firms, rather than larger sensitivity to economy-wide news. This is consistent with prediction 1B i.e., that the larger fundamental volatility of high retail firms is coming from the idiosyncratic component of earnings news.

5.2 Return Sensitivity to Earnings Surprises

Prediction 2 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

$$r^i_{t,t+n} = \alpha + \beta \text{SUE}_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t}, \tag{3}$$

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 month $t-1$ returns (Kumar and Lee (2006), Balasubramaniam et al. (2021), Bali et al. (2021), Luo

[15] One alternative explanation for why high retail stocks have larger earnings surprises is that such stocks have lower analyst coverage on average, which leads to less accurate forecasts. This, however, is inconsistent with evidence in Martineau and Zoican (2019) who show that high retail stocks tend to receive above average coverage. In unreported results we find that, conditional on market capitalization, high retail stocks tend to receive above average analyst coverage.

[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.

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, β is the earnings response coefficient. We are interested in how this varies across retail portfolios, so we interact $\text{SUE}_{i,t}$ with dummy variables for each quintile of retail trading intensity in the month before the earnings announcement. The omitted group is the middle bucket of retail activity. Table 8 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. The difference in coefficients on $\text{SUE} \times Q5$ and $\text{SUE} \times Q1$ is over .6, compared to an unconditional effect of just over 1. In the second set of three columns we control for a litany of firm characteristics listed in the above paragraph. The weaker sensitivity of high retail stocks to earnings surprises is left virtually unchanged.

A potential concern with the results in Table 8 is that high retail stocks don't respond less to news, just 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.

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 8: retail investors select into small stocks, 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 9 we re-estimate the regression 3 but include size dummies that are interacted with SUE. As Table 9 shows, high retail share stocks are less responsive to earnings news across the size distribution, and this difference is statistically significant at the 5% level for all but the smallest size portfolios.

5.3 Retail Trading around Earnings Announcements

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

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

In Figure 4 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. The bottom panels show the same results but cumulate the daily data. The results in Figure 4 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.

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

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

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 $\tau - 1$, 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 $\text{Intensity}_{i,t}$ at the stock level over the previous 252 trading days before $t = -5$.[17] Table 10 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. The bottom panel shows that this is actually 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 essentially zero. This suggests that, consistent with prediction 2A, non-retail are trading less in the pre-earnings period rather than retail trading more. Second, consistent with prediction 2B, 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.

Next, we turn to the question of whether high retail stocks are more expensive to trade around

[17] 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.

earnings announcements. Given the results in Table 6, however, we 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)). In addition, given the results in Table 5, we need to account for the higher average level of trading costs in high retail stocks. To address both these concerns, we estimate the following regression:

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

where DM Effective Spread$_{i,t}$ i.e., the demeaned effective spread, is 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 11 shows that, consistent with prediction 1D, 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.

Having shown that high retail stocks have both an especially high retail trading intensity and especially high trading costs around earnings announcements, we re-visit our results on the responsiveness of high retail stocks to earnings news. To this end, we estimate a modified version of Equation 3:

$$r^i_{t,t+n} = \alpha + \beta \text{SUE}_{i,t} + \beta_1 1_{i \in Q1_{\tau-1}} + \beta_2 1_{i \in Q2_{\tau-1}} + \beta_4 1_{i \in Q4_{\tau-1}} + \beta_5 1_{i \in Q5_{\tau-1}} + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \tag{6}$$

where $1_{i \in Q1_{\tau-1}}$ are indicators for quintiles of gross or net retail flows, formed over the 22 trading days before the earnings announcements. Table 12 shows that what matters is gross retail flows, as high and low net flows look similar to high gross flows. In other words, 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.

Finally, we turn to prediction 2C i.e., whether or not prices are more likely to move with retail investors in high retail stocks. So far, we have been sorting firms into quintiles based on gross retail activity i.e., retail buys plus retail sells. To test prediction 2c, we construct a measure of retail order imbalance as

$$\text{mroibvol}_{i,t} = \frac{\text{RBuy}_{i,t} - \text{RSell}_{i,t}}{\text{Volume}_{i,t}} \tag{7}$$

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.

Table 13 contains the results of a regression of *mroibvol* on dummy variables for past retail trading intensity quintiles. In the columns labeled -1, *mroibvol* is measured the week before the focal week, while in the columns labeled 0, *mroibvol* is measured the same week and in the columns labeled 1 it is measured in the next week. 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 shows that, for the average stock, returns move in the opposite direction of retail order imbalance i.e., other investors tend to trade against retail. But, the sixth row shows that, for high retail stocks, returns move in the direction of *mroibvol*. Further, this effect is roughly $2\times$ as strong right around earnings announcements (i.e., t=0 and t=1). This is consistent with prediction 2C, which states that for high retail stocks, retail order flow and prices should move in the same direction.

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 a the aggregate announcer portfolio has alpha with respect to the buy-and-hold portfolio.

In Table 14 we document average returns around earnings announcements 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 10 day announcement window, containing five trading days prior to the announcement, and five trading days starting with the day the earnings news could have been first traded. 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.

Overall, Table 14 shows that high retail stocks see lower announcement time returns, consistent with prediction 3. The first column shows average returns over a two-day window across the size and retail sort. The coefficients on Q5 are negative and substantial: high retail stocks see returns that are -1.5% to -.25% lower , going from small to large stocks. Note too the intercept across the five panels: it is positive and ranges from .55% to .25% going from small to large stocks, indicating the presence of an unconditional earnings announcer risk premium in this sample. The second and third columns break this return into pre- and post-announcement components. The majority of the return gap across the retail sort stems from the large negative (and statistically significant) incremental return on announcement, meaning, the column denoted “Post.” This is driven by the negative returns around earnings announcements for high retail stocks documented in Figure 3 and Table 7.

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.

The analysis in Table 14 is done on the firm-announcement level. However, firm 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. 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 15. 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 14: 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 14. 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.

Overall, the results in Tables 14 and 15 are consistent with prediction 3: high retail stocks have low earnings day returns.

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 new fact, 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 then 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. In these stocks, prices tend to move with retail order imbalance, suggesting that frictions may prevent institutions from betting against retail.

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, however, 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

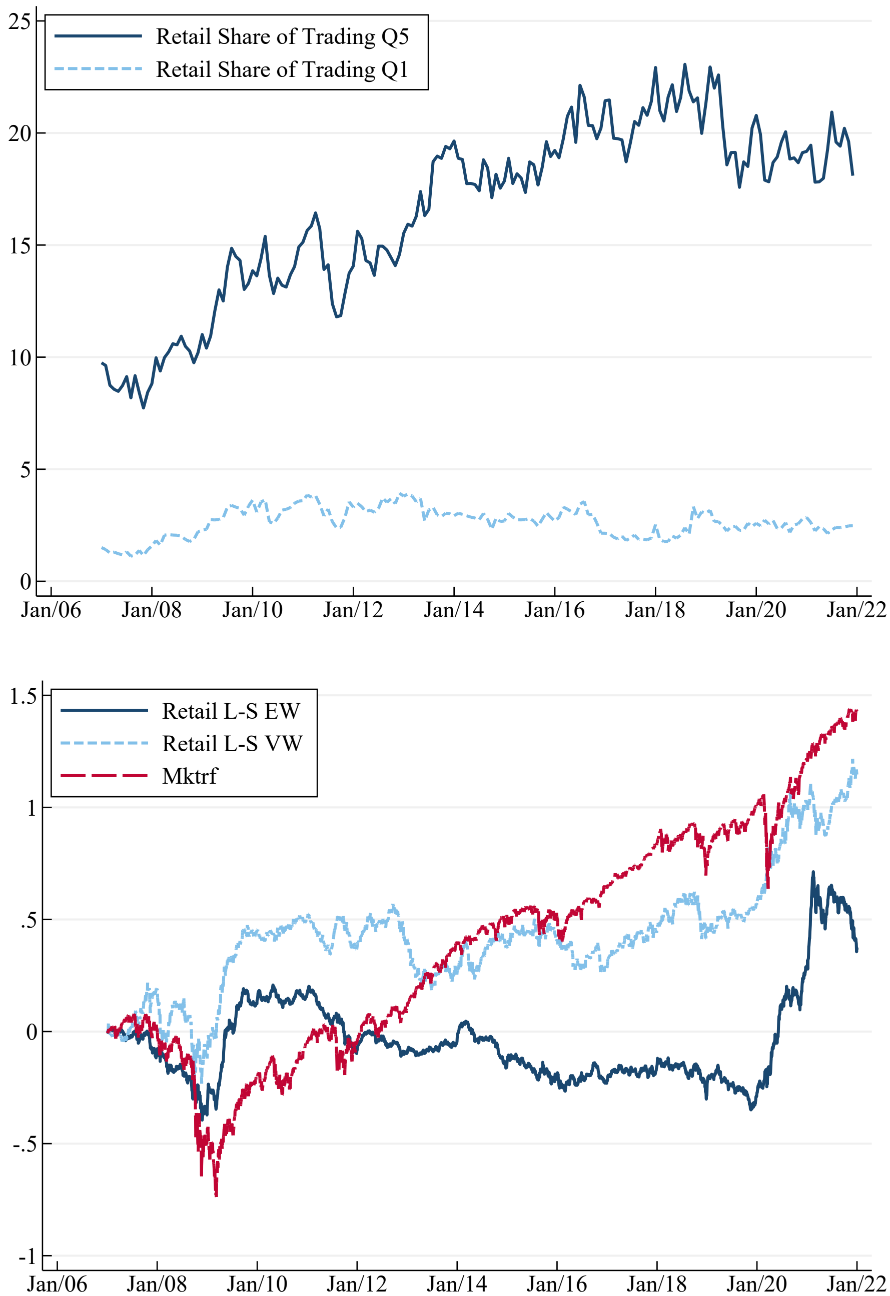

Figure 1: Retail share of trading volume. Top panel: Average retail share of trading volume in the top and bottom quintile sorted on previous month's retail trading intensity. Bottom panel: Cumulative log returns of equal weighted (EW) and value weighted (VW) retail intensity long-short portfolios.

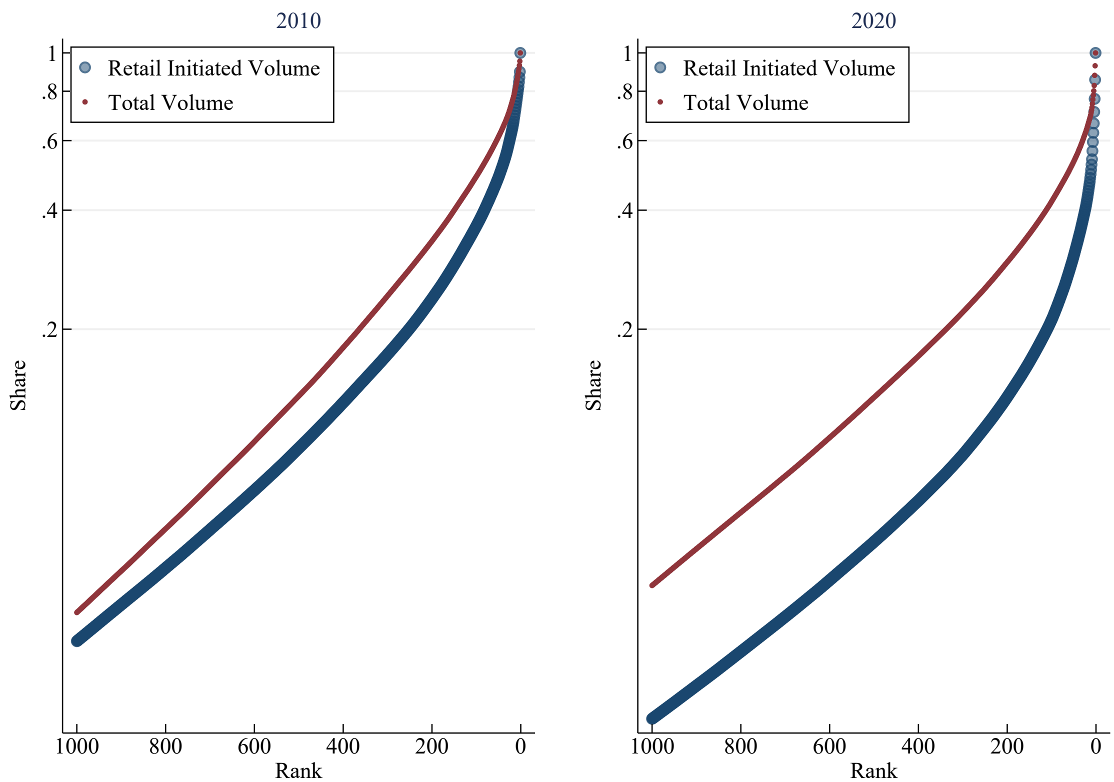

Figure 2: Cumulative Volume Share in 2010 and 2020. Each year, stocks are ranked on their retail-initiated turnover, and total turnover. Each data point represents the cumulative share of volume for stocks below the indicated rank.

Figure 3: Distribution of earnings returns and standardized unexpected earnings. Left two panels: distribution of earnings-day returns. Right two panels: distribution of standardized unexpected earnings. Top two panels: firms in the bottom 20% of market capitalization. Bottom two panels: firms in the top 20% of market capitalization. Returns and standardized unexpected earnings are Winsorized at the 5% and 95% level.

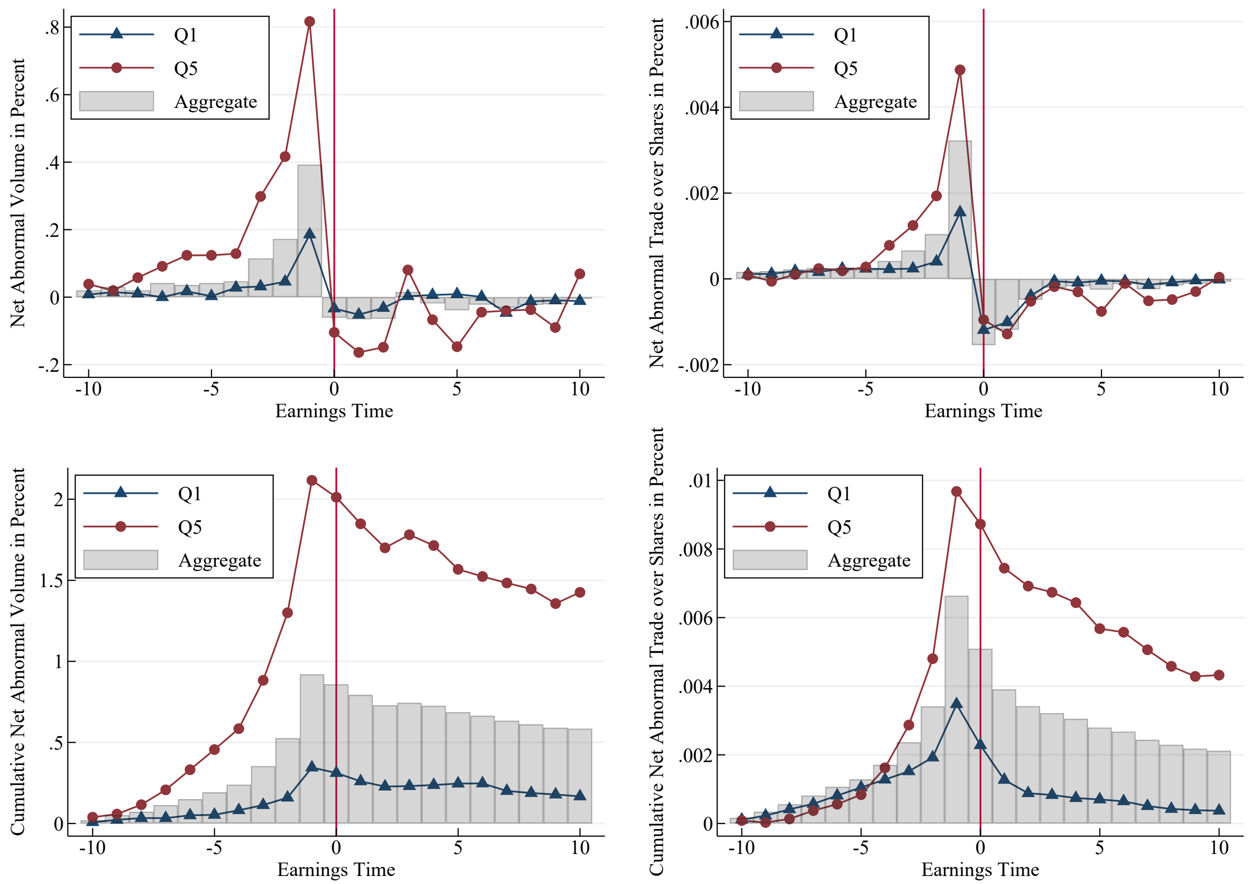

Figure 4: Abnormal Trading Volume around Earnings Announcements. Daily abnormal net retail volume and abnormal net retail share of volume, in percent units. Q1 represents the bottom quintile of retail intensity, while Q5 represents the top quintile. We subtract out the unconditional means in respective 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	0.79	0.91	2.14	2.97	107,989
	2	3.30	0.82	1.98	3.40	4.27	107,906
	3	4.87	1.18	3.02	4.98	6.29	107,866
	4	8.14	2.25	5.06	8.02	11.16	107,893
	5	18.02	7.02	10.27	16.68	27.55	107,839
1	1	4.65	3.13	0.74	4.24	9.13	21,679
1	5	26.23	7.70	16.15	26.48	35.63	21,524
5	1	2.14	0.68	1.09	2.25	2.92	21,639
5	5	8.40	3.60	4.24	7.78	12.71	21,501

Panel B. Monthly turnover							
Size	Port.	mean	sd	p10	p50	p90	count
	1	17.30	14.11	4.30	14.19	32.63	107,989
	2	18.12	14.22	5.09	14.60	34.74	107,906
	3	18.25	16.40	3.51	13.79	37.92	107,866
	4	19.45	22.12	1.98	11.88	46.45	107,893
	5	21.69	32.77	1.29	8.06	61.34	107,839
1	1	5.93	9.81	0.59	2.75	14.09	21,679
1	5	18.56	32.45	1.10	5.90	50.85	21,524
5	1	20.37	12.70	9.66	17.34	34.00	21,639
5	5	31.53	28.61	8.20	21.54	68.10	21,501

Panel C. Retail-initiated monthly turnover							
Size	Port.	mean	sd	p10	p50	p90	count
	1	0.31	0.34	0.05	0.24	0.64	107,989
	2	0.52	0.46	0.14	0.41	1.03	107,906
	3	0.78	0.79	0.14	0.57	1.66	107,866
	4	1.47	1.95	0.14	0.80	3.51	107,893
	5	3.44	5.27	0.19	1.24	9.95	107,839
1	1	0.30	0.61	0.01	0.10	0.76	21,679
1	5	3.84	5.82	0.25	1.37	12.12	21,524
5	1	0.38	0.26	0.14	0.32	0.67	21,639
5	5	2.69	3.73	0.49	1.39	6.02	21,501

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

Panel A.

	Retail Portfolio at $t = 0$				
$t = -12$	1	2	3	4	5
1	46.09	26.05	13.61	8.47	5.78
2	28.00	37.05	23.84	8.82	2.29
3	12.13	25.30	35.14	21.65	5.78
4	6.68	9.95	23.61	39.78	19.98
5	2.40	2.93	6.40	23.12	65.15

Panel B. Small stocks only.

	Retail Portfolio at $t = 0$				
$t = -12$	1	2	3	4	5
1	37.82	24.80	17.08	11.80	8.51
2	22.88	24.89	21.70	17.53	13.02
3	13.83	19.49	23.26	22.65	20.77
4	7.98	14.14	20.78	26.72	30.38
5	4.99	8.85	14.97	25.87	45.33

Panel C. Large stocks only.

	Retail Portfolio at $t = 0$				
$t = -12$	1	2	3	4	5
1	52.84	27.58	13.40	4.64	1.55
2	27.25	33.95	24.79	11.05	2.96
3	12.37	25.09	32.01	23.42	7.11
4	3.69	10.69	24.04	38.55	23.02
5	0.83	2.54	6.09	23.75	66.80

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	β_{CAPM}	β_{SMB}	β_{HML}	β_{UMD}
	1	1,762	3,932	21.61	43.45	11.55	0.62	0.04	1.03	0.61	0.26	-0.04
	2	1,807	6,255	22.72	45.62	13.19	0.58	0.04	1.09	0.63	0.19	-0.04
	3	984	9,750	22.51	39.66	13.80	0.62	0.02	1.10	0.71	0.15	-0.05
	4	347	10,147	19.67	26.39	12.25	0.76	-0.02	1.08	0.80	0.10	-0.09
	5	82	2,898	16.26	9.47	5.55	0.97	-0.13	0.88	0.67	-0.02	-0.08
1	1	66	73	15.38	10.31	-3.42	1.10	-0.06	0.47	0.38	0.11	-0.06
1	2	63	69	16.96	7.77	-3.35	1.09	-0.10	0.56	0.45	0.08	-0.08
1	3	55	61	17.12	6.18	-3.86	1.12	-0.15	0.60	0.47	0.03	-0.09
1	4	46	52	16.85	5.00	-4.68	1.09	-0.18	0.63	0.46	-0.02	-0.07
1	5	35	42	17.01	4.18	-3.49	1.07	-0.18	0.64	0.40	-0.03	-0.03
5	1	7,997	10,878	27.78	71.03	16.19	0.49	0.05	1.02	0.15	0.10	-0.01
5	2	9,515	14,241	30.05	73.25	16.54	0.46	0.06	1.03	0.13	0.07	0.00
5	3	11,101	19,390	31.51	75.79	16.96	0.45	0.06	1.03	0.12	0.03	-0.00
5	4	16,193	36,279	35.44	76.48	17.74	0.45	0.05	1.04	0.08	-0.02	-0.02
5	5	21,654	66,793	33.46	78.23	23.41	0.47	0.04	1.16	0.14	-0.11	-0.03

Table 3: Fundamentals in five retail share of trading sorted portfolios. The top 5 rows are based on five equal-weighted portfolios sorted on the retail share of trading volume in the previous month. The bottom 10 rows are based on twenty five equal-weighted portfolios first 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; β_{CAPM}, β_{SMB}, β_{HML} and β_{UMD} are 4-factor betas computed over the previous 252 trading days.

Valuation									
Size	Port.	CF	K_{Know}	K_{Org}	K_{OffBS}	K_{Int}	Mispricing	PAT	VU
	1	-1.619	0.06	0.22	0.28	0.54	48.13	1.64	0.73
	2	-1.174	0.07	0.22	0.29	0.54	48.41	2.32	0.73
	3	-0.188	0.09	0.27	0.37	0.61	49.92	2.85	0.76
	4	1.251	0.16	0.39	0.56	0.84	52.33	3.09	0.83
	5	1.421	0.30	0.55	0.91	1.20	54.10	2.23	0.93
1	1	-2.237	0.16	0.60	0.79	1.07	54.11	0.64	0.94
1	2	-1.778	0.24	0.72	1.01	1.34	53.28	0.91	0.94
1	3	-1.583	0.30	0.76	1.14	1.48	52.28	1.19	0.96
1	4	-0.825	0.37	0.73	1.18	1.51	51.57	1.26	0.96
1	5	-0.386	0.37	0.69	1.14	1.43	52.48	1.25	0.95
5	1	-4.017	0.04	0.13	0.16	0.43	45.00	2.67	0.65
5	2	-4.434	0.04	0.13	0.17	0.41	44.29	3.52	0.63
5	3	-4.245	0.05	0.13	0.18	0.40	44.49	4.42	0.62
5	4	-4.424	0.06	0.13	0.19	0.40	44.84	5.97	0.61
5	5	-2.582	0.07	0.12	0.19	0.40	47.83	9.16	0.64

Table 4: Valuation across five retail share of trading sorted portfolios. The top 5 rows are based on five equal-weighted portfolios sorted on the retail share of trading volume in the previous month. The bottom 10 rows are based on twenty five equal-weighted portfolios first sorted on market cap, then on the retail share of trading volume, both in the previous month. CF is cashflow duration, computed using the method in Gormsen and Lazarus (2021); K_{Know}, K_{Org}, K_{OffBS} and K_{Int} are from Peters and Taylor (2017); Mispricing is the mispricing score from Stambaugh and Yuan (2017); PAT is the real market value of patents over the past five years (data obtained from Kogan et al. (2017)) divided by market capitalization; VU is valuation uncertainty from Golubov and Konstantinidi (2021).

Size	Port.	SD	Ivol t	Ivol q	λ1	λ2	Espread	Rspread
	1	1.94	1.53	0.14	-0.01	0.94	0.44	0.29
	2	2.06	0.39	0.05	1.21	1.34	0.30	0.17
	3	2.35	0.71	0.09	1.30	2.06	0.44	0.25
	4	2.93	1.61	0.19	1.91	3.45	0.79	0.46
	5	4.05	3.56	0.48	6.38	8.72	1.56	0.89
1	1	2.95	9.03	1.10	-8.47	-0.95	2.83	2.10
1	2	3.30	6.83	0.76	-5.82	2.79	2.41	1.64
1	3	3.85	3.98	0.76	3.77	6.26	2.32	1.46
1	4	4.35	4.50	0.71	6.09	10.72	2.25	1.37
1	5	4.17	7.27	0.86	9.86	13.41	2.26	1.36
5	1	1.57	0.02	0.00	0.49	0.48	0.06	0.02
5	2	1.62	0.02	0.00	0.52	0.51	0.06	0.02
5	3	1.69	0.01	0.00	0.54	0.53	0.06	0.02
5	4	1.77	0.03	0.00	0.55	0.54	0.08	0.04
5	5	2.11	0.01	0.01	0.69	0.68	0.07	0.03

Table 5: Liquidity in five retail share of trading sorted portfolios. The top 5 rows are based on five equal-weighted portfolios sorted on the retail share of trading volume in the previous month. The bottom 10 rows are based on twenty five equal-weighted portfolios first 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 and Ivol q are intraday volatility computed from trades and quotes; λ_1 and λ_2 are Kyle's lambda, estimated with and without an intercept; Espread and Rspread are the effective and realized spread, computed using the methodology in Holden and Jacobsen (2014).

Number of firm-earnings events.						
Size	Port.	Low SUE	2	3	4	High SUE
	1	4,318	8,536	8,452	7,965	4,555
	2	4,399	8,220	8,362	8,079	5,045
	3	5,417	7,236	6,495	6,950	5,860
	4	7,447	5,548	3,603	5,010	7,297
	5	8,540	3,105	683	2,116	7,363
1	1	499	652	926	742	457
1	2	676	698	821	826	688
1	3	759	755	684	692	754
1	4	770	674	526	627	781
1	5	662	584	408	478	683
5	1	1,505	1,425	1,443	1,441	1,378
5	2	1,311	1,532	1,457	1,528	1,347
5	3	1,317	1,443	1,511	1,432	1,390
5	4	1,376	1,409	1,424	1,368	1,398
5	5	1,535	1,235	1,209	1,274	1,530

Table 6: Count of standardized earnings announcements across retail- and size-sorted portfolios. The top 5 rows are based on five equal-weighted portfolios sorted on the retail share of trading volume in the previous month. The bottom 10 rows are based on twenty five equal-weighted portfolios first sorted on market cap, then on the retail share of trading volume, both in the previous month. Quarterly earnings announcements from 2007 to 2021.

Panel A. Standardized SUE.

Size	Port.	Mean	SD	P10	P25	P50	P75	P90
	1	0.01	0.98	-0.33	-0.05	0.05	0.20	0.48
	2	0.03	1.04	-0.35	-0.04	0.06	0.21	0.53
	3	-0.02	1.43	-0.59	-0.08	0.06	0.25	0.70
	4	-0.19	2.25	-1.38	-0.26	0.05	0.38	1.15
	5	-0.55	3.45	-4.10	-1.00	0.00	0.75	2.50
1	1	-0.59	3.15	-3.56	-0.96	0.00	0.59	1.82
1	5	-0.91	4.36	-7.61	-1.96	0.00	1.28	3.92
5	1	0.06	0.48	-0.14	0.00	0.05	0.15	0.32
5	5	-0.00	1.16	-0.25	-0.01	0.05	0.17	0.43

Panel B. Idiosyncratic component of Standardized SUE.

Size	Port.	Mean	SD	P10	P25	P50	P75	P90
	1	0.36	1.22	-0.19	0.06	0.22	0.51	0.97
	2	0.38	1.26	-0.18	0.07	0.24	0.55	1.06
	3	0.34	1.56	-0.37	0.05	0.24	0.58	1.25
	4	0.18	2.27	-1.07	-0.06	0.25	0.70	1.78
	5	-0.14	3.33	-3.63	-0.67	0.22	1.05	3.05
1	1	-0.36	3.09	-3.43	-0.83	0.10	0.83	2.21
1	5	-0.52	4.14	-6.92	-1.74	0.09	1.49	4.25
5	1	0.41	0.96	-0.03	0.09	0.22	0.47	0.83
5	5	0.36	1.31	-0.08	0.09	0.23	0.51	0.97

Table 7: Summary statistics of earnings announcement surprises and returns across retail- and size-sorted portfolios. The top 5 rows of each panel are based on five 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 five equal-weighted portfolios first sorted on market cap, then on the retail share of trading volume, both in the previous month. The idiosyncratic component of SUE is computed following the methodology in Glosten et al. (2021). Quarterly earnings announcements from 2007 to 2021.

	Standardized Unexpected Earnings					
	(0, 0)	(0, 2)	(0, 4)	(0, 0)	(0, 2)	(0, 4)
SUE	1.049***	1.195***	1.232***	1.091***	1.243***	1.305***
	(16.20)	(13.39)	(13.01)	(15.22)	(12.97)	(12.84)
SUE x Q1	0.203**	0.150	0.160	0.253**	0.207	0.213*
	(2.83)	(1.47)	(1.64)	(3.28)	(1.91)	(2.00)
SUE x Q2	0.301***	0.270*	0.239*	0.342***	0.326**	0.270*
	(3.75)	(2.50)	(1.98)	(3.83)	(2.85)	(2.23)
SUE x Q4	-0.251***	-0.243**	-0.211*	-0.262***	-0.240**	-0.217*
	(-4.17)	(-2.87)	(-2.49)	(-4.00)	(-2.64)	(-2.38)
SUE x Q5	-0.478***	-0.511***	-0.510***	-0.425***	-0.437***	-0.451***
	(-7.38)	(-5.82)	(-5.65)	(-6.03)	(-4.80)	(-4.77)
Controls				Yes	Yes	Yes
N	150329	150483	150499	139435	139585	139599
R^2	0.05	0.05	0.05	0.05	0.05	0.05

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. 0 refers to the first day announcement information is tradeable during normal market hours. SUE is standardized unexpected earnings, defined 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 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.

	Cumulative Market-Adjusted Return					
	(0, 0)	(0, 2)	(0, 4)	(0, 0)	(0, 2)	(0, 4)
Size 1 x SUE	0.537***	0.671***	0.720***	0.588***	0.755***	0.816***
	(15.76)	(14.21)	(11.39)	(15.54)	(15.08)	(12.36)
Size 1 x SUE x Q5	0.0315	-0.0308	-0.0374	0.0901	0.00893	-0.0243
	(0.58)	(-0.45)	(-0.43)	(1.46)	(0.12)	(-0.25)
Size 2 x SUE	0.980***	1.166***	1.226***	1.019***	1.226***	1.292***
	(12.27)	(10.65)	(9.87)	(12.33)	(10.91)	(10.20)
Size 2 x SUE x Q5	-0.242**	-0.304*	-0.271	-0.205*	-0.284*	-0.235
	(-2.61)	(-2.59)	(-1.94)	(-2.03)	(-2.27)	(-1.60)
Size 3 x SUE	1.408***	1.521***	1.546***	1.470***	1.603***	1.656***
	(10.49)	(10.19)	(8.44)	(10.03)	(9.84)	(8.34)
Size 3 x SUE x Q5	-0.528***	-0.462**	-0.448*	-0.488**	-0.414*	-0.399
	(-3.75)	(-2.70)	(-2.20)	(-3.21)	(-2.18)	(-1.70)
Size 4 x SUE	1.668***	1.692***	1.815***	1.696***	1.809***	1.921***
	(9.83)	(7.93)	(8.25)	(9.43)	(8.18)	(8.37)
Size 4 x SUE x Q5	-0.809***	-0.731**	-0.773**	-0.786***	-0.836***	-0.836**
	(-4.59)	(-3.06)	(-2.96)	(-4.19)	(-3.45)	(-3.16)
Size 5 x SUE	1.468***	1.884***	2.150***	1.521***	1.849***	2.283***
	(4.05)	(4.40)	(5.40)	(3.76)	(3.95)	(5.04)
Size 5 x SUE x Q5	-0.714	-0.958	-1.199**	-0.725	-0.901	-1.296**
	(-1.63)	(-1.93)	(-2.87)	(-1.47)	(-1.65)	(-2.65)

Table 9: 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, Qk is an indicator variable for whether stock i was in retail intensity quntile k at the end of the month before the earnings announcement and Size j is an indicator for whether firm i was in size quintile j at the end of the month before the earnings announcement. SUE and returns winsorized at the 1st and 99th percentile. Quarterly earnings announcements from 2007 to 2021.

	Demeaned Retail as % of Trading Volume (percentage points)					
Timing:	-5 to -1	-3 to -1	-1	0	0 to 2	0 to 4
Q1	-0.212***	-0.201***	-0.248***	-0.320***	-0.239***	-0.194***
	(0.037)	(0.040)	(0.053)	(0.053)	(0.040)	(0.037)
Q2	-0.112***	-0.114***	-0.114***	-0.166***	-0.135***	-0.109***
	(0.026)	(0.029)	(0.039)	(0.040)	(0.029)	(0.026)
Q4	0.250***	0.274***	0.287***	0.278***	0.252***	0.219***
	(0.041)	(0.044)	(0.057)	(0.056)	(0.040)	(0.037)
Q5	1.576***	1.497***	1.429***	0.489***	0.685***	0.770***
	(0.094)	(0.101)	(0.119)	(0.108)	(0.093)	(0.088)
Observations	801,738	480,200	160,073	160,361	480,345	799,805
R-squared	0.025	0.025	0.029	0.029	0.022	0.02

	Demeaned Retail as % of Shares Outstanding (basis points)					
Timing:	-5 to -1	-3 to -1	-1	0	0 to 2	0 to 4
Q1	0.381***	0.241***	-0.201**	-4.558***	-2.126***	-1.300***
	(0.087)	(0.086)	(0.096)	(0.406)	(0.196)	(0.133)
Q2	0.288***	0.202***	-0.024	-2.342***	-1.111***	-0.661***
	(0.063)	(0.063)	(0.078)	(0.328)	(0.161)	(0.112)
Q4	-0.694***	-0.536***	-0.0905	4.198***	2.054***	1.281***
	(0.130)	(0.132)	(0.143)	(0.473)	(0.241)	(0.170)
Q5	-0.986*	-0.658	0.406	6.928***	3.879***	2.536***
	(0.502)	(0.488)	(0.531)	(1.039)	(0.702)	(0.615)
Observations	801,738	480,200	160,073	160,361	480,345	799,804
R-squared	0.047	0.046	0.047	0.084	0.048	0.037

Table 10: Retail activity and trading intensity around earnings announcements. Cross-sectional regression where left-hand-side variables are measure of retail trading intensity around earnings announcement. In the top panel, retail trading intensity is defined as (retail buys + retail sells)/(retail buys + retail sells + non-retail buys and sells) while in the bottom panel, retail trading intensity is defined as (retail buys + retail sells)/(shares outstanding). In all columns, we subtract the mean of these quantities computed over the previous 252 trading days. Qk is an indicator variable for whether stock i was in retail intensity quntile k at the end of the month before the earnings announcement. Time fixed effects are for year-month. Control variables include nominal price, returns from month $t-12$ to $t-2$, time since listing, market capitalization, book-to-market, gross profit margin, book long-term leverage, MAX (lottery demand) and month $t-1$ returns. Additional controls include betas on the three Fama-French factors as well as the momentum factor, idiosyncratic volatility and total volatility, all computed over the past 12 months. Standard errors are double clustered at the permno/year-month level.

	Demeaned Effective Spread					
	-5 to -1	-3 to -1	-1	0	0 to 2	0 to 4
Q1	0.00463	0.00411	0.00265	-0.0004	0.00281	0.00487
	(0.008)	(0.008)	(0.008)	(0.010)	(0.008)	(0.008)
Q2	-0.0069	-0.0069	-0.0115***	-0.0058	-0.0064	-0.0059
	(0.005)	(0.005)	(0.004)	(0.006)	(0.004)	(0.004)
Q4	0.0295***	0.0295***	0.0277**	0.0336***	0.0292***	0.0299***
	(0.010)	(0.011)	(0.011)	(0.009)	(0.009)	(0.009)
Q5	0.114***	0.115***	0.108***	0.130***	0.123***	0.120***
	(0.019)	(0.020)	(0.021)	(0.020)	(0.019)	(0.019)
Low SUE	0.0772***	0.0775***	0.0822***	0.106***	0.0933***	0.0926***
	(0.008)	(0.007)	(0.008)	(0.010)	(0.009)	(0.009)
2	0.00438	0.00579	0.00579	0.00912*	0.00468	0.00395
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.003)
4	0.00324	0.00146	-0.0027	0.00513	0.00121	3.6E-05
	(0.004)	(0.003)	(0.004)	(0.005)	(0.004)	(0.004)
High SUE	0.0688***	0.0698***	0.0653***	0.0586***	0.0540***	0.0511***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)
Constant	-7.578***	-7.370***	-7.571***	-8.428***	-9.478***	-9.830***
	(2.086)	(2.299)	(2.346)	(2.265)	(2.386)	(2.260)
Observations	141,195	141,149	140,814	140,915	141,145	141,183
R-squared	0.655	0.621	0.539	0.53	0.626	0.656
Time FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES

Table 11: Retail activity and demeaned trading costs around earnings announcements. Left-hand-side variables are average demeaned effective spread computed over various windows around earnings announcements. Demeaned effective spread is effective bid-ask spread minus average effective spread over the calendar month before the earnings announcement. 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 Quintiles of SUE are formed each quarter. 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.

	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 12: 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.

	All weeks			Announcement Weeks		
	-1	0	1	-1	0	1
Mroibvol	-0.850***	-0.139	0.192***	-0.859***	-1.151***	0.222
	(-13.92)	(-1.97)	(4.95)	(-6.23)	(-4.27)	(1.74)
Mroibvol x Q1	0.402***	0.191**	-0.0429	0.334*	1.121***	-0.106
	(8.53)	(3.27)	(-1.11)	(2.21)	(4.17)	(-0.60)
Mroibvol x Q2	0.155***	0.0713	0.0507	-0.0553	0.273	0.164
	(3.42)	(1.33)	(1.11)	(-0.31)	(1.10)	(0.86)
Mroibvol x Q4	-0.0506	0.182*	0.110*	-0.451*	0.254	0.153
	(-1.08)	(2.28)	(2.25)	(-2.41)	(0.86)	(0.80)
Mroibvol x Q5	0.0198	1.304***	0.194***	-0.159	2.633***	0.246
	(0.29)	(15.31)	(3.75)	(-0.76)	(7.53)	(1.39)
Constant	0.190***	0.218***	0.221***	0.180***	0.461***	0.307***
	(11.92)	(14.97)	(13.64)	(5.27)	(8.27)	(10.57)
Observations	2269182	2283259	2269773	173666	174349	173691
R^2	0.002	0.001	0.001	0.007	0.004	0.005

Table 13: 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. 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 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	-0.432	-0.289*	-0.0901	-0.497	-0.547**	0.0924
	(-1.33)	(-2.23)	(-0.34)	(-1.25)	(-2.74)	(0.30)
Size 1 x Q5	-1.533***	-0.133	-1.276***	-1.814***	0.552	-2.150***
	(-4.42)	(-0.76)	(-4.67)	(-3.77)	(1.94)	(-5.80)
Constant	0.550**	0.422***	0.124	0.964***	0.647***	0.306
	(2.81)	(5.03)	(0.74)	(3.64)	(5.49)	(1.41)
Size 2 x Q1	0.162	-0.105	0.257	0.0704	-0.125	0.179
	(0.90)	(-1.70)	(1.58)	(0.30)	(-1.22)	(0.88)
Size 2 x Q5	-1.067***	0.0367	-1.060***	-1.021**	0.169	-1.161***
	(-4.52)	(0.46)	(-5.17)	(-2.80)	(1.13)	(-3.88)
Constant	0.411**	0.181***	0.239*	0.623***	0.306***	0.329*
	(3.19)	(4.36)	(2.05)	(3.47)	(4.28)	(2.08)
Size 3 x Q1	0.273	-0.0756	0.348*	0.274	-0.167	0.452*
	(1.86)	(-1.58)	(2.36)	(1.49)	(-1.94)	(2.51)
Size 3 x Q5	-0.684**	0.0133	-0.695***	-0.514	0.180	-0.685**
	(-3.13)	(0.17)	(-3.42)	(-1.73)	(1.18)	(-2.86)
Constant	0.224*	0.0781*	0.154	0.315*	0.151*	0.163
	(2.17)	(2.10)	(1.54)	(2.38)	(2.39)	(1.38)
Size 4 x Q1	-0.0729	0.0126	-0.0895	0.0615	0.0144	0.0296
	(-0.62)	(0.34)	(-0.80)	(0.40)	(0.18)	(0.24)
Size 4 x Q5	-0.573**	0.114	-0.684***	-0.476*	0.136	-0.603**
	(-3.33)	(1.83)	(-4.32)	(-2.20)	(1.17)	(-3.30)
Constant	0.351***	0.0244	0.332***	0.418***	0.128*	0.306**
	(3.92)	(0.96)	(3.76)	(3.76)	(2.44)	(3.08)
Size 5 x Q1	0.145	0.0214	0.126	0.274**	0.0969	0.185*
	(1.60)	(0.60)	(1.48)	(2.60)	(1.77)	(2.06)
Size 5 x Q5	-0.241*	0.0857	-0.325**	-0.263	0.138	-0.409**
	(-2.07)	(1.60)	(-3.32)	(-1.63)	(1.71)	(-3.01)
Constant	0.256***	0.0960***	0.165**	0.396***	0.207***	0.198**
	(3.90)	(3.35)	(2.79)	(5.60)	(6.00)	(3.07)

Table 14: Cumulative returns around earnings announcements. Column headers refer to first and last days in return window. 0 is the first trading day on which announcement information is tradeable. Qk is an indicator variable for whether stock i was in retail intensity quintile k at the end of the month before the earnings announcement and Size j is an indicator for whether firm i was in size quintile j at the end of the month before the earnings announcement. Five panels sorted on size. Monthly fixed effects. Standard errors clustered by firm and month. Earnings announcements in 2007-2021.

		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	1.45	4.01	2.56	1.64	1.08	-0.55	1.47	4.78	3.31
1	2	1.50	3.75	2.25	1.69	3.55	1.86	1.60	4.09	2.49
1	3	1.41	2.27	0.87	1.41	5.01	3.60	1.62	-0.36	-1.98
1	4	1.63	1.70	0.07	1.34	6.23	4.88	1.94	-1.82	-3.76
1	5	1.93	3.47	1.54	1.35	9.14	7.79	2.35	-4.17	-6.52
2	1	0.54	2.27	1.72	0.78	1.34	0.56	0.59	3.34	2.74
2	2	0.81	3.25	2.44	0.95	2.86	1.90	0.90	2.92	2.03
2	3	1.16	2.47	1.31	1.22	2.40	1.19	1.23	2.81	1.58
2	4	1.36	1.25	-0.11	1.34	3.59	2.25	1.47	-0.27	-1.74
2	5	1.05	1.50	0.45	0.73	6.18	5.45	1.29	-2.33	-3.62
3	1	0.64	1.45	0.80	0.86	-0.04	-0.90	0.63	2.41	1.78
3	2	0.99	1.68	0.69	1.08	0.97	-0.11	0.96	2.73	1.78
3	3	0.98	0.54	-0.44	1.04	2.09	1.05	1.01	-0.26	-1.27
3	4	1.17	1.12	-0.04	1.17	2.20	1.03	1.19	0.62	-0.57
3	5	1.28	-0.39	-1.68	1.04	2.90	1.86	1.29	-2.47	-3.75
4	1	0.75	1.38	0.62	0.87	0.60	-0.27	0.76	2.00	1.24
4	2	0.87	0.06	-0.81	0.88	1.16	0.28	0.88	-0.66	-1.53
4	3	0.97	1.41	0.44	1.02	1.17	0.15	0.98	0.95	-0.03
4	4	1.09	2.40	1.31	1.15	0.93	-0.22	1.12	3.06	1.94
4	5	1.12	-0.54	-1.66	0.96	2.33	1.37	1.17	-1.83	-3.00
5	1	0.94	1.12	0.18	1.00	0.99	-0.01	0.98	1.54	0.56
5	2	0.88	1.12	0.24	0.89	0.87	-0.02	0.90	0.99	0.09
5	3	1.01	0.41	-0.60	1.02	0.65	-0.36	1.02	0.15	-0.87
5	4	1.04	1.34	0.30	1.05	1.11	0.05	1.06	0.29	-0.77
5	5	1.06	-0.11	-1.16	0.91	1.55	0.64	1.10	-1.72	-2.83

Table 15: 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.

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.

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 overconfidence. *Available at SSRN 3445660*.

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

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

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

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 finance*, 56(1):1–43.

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

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

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *The journal of finance*, 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 financial 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 finance*, 64(2):709–749.

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

Di Maggio, M., 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*.

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

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

Engelberg, J. E. and Parsons, C. A. (2011). The causal impact of media in financial 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 financial 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 efficiency 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., Laarits, T., and Wurgler, J. (2022). Stock market stimulus. 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 effects in financial markets. *The Journal of Finance*, 73(4):1567–1613.

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

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

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

Hong, H. and Stein, J. C. (2007). Disagreement and the stock market. *Journal of Economic perspectives*, 21(2):109–128.

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

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

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.

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

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

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

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

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

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

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

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: 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 effect. *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 finance*, 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, overconfidence and category learning. *Journal of Financial Economics*, 80(3):563–602.

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

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

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

Stambaugh, R. F. and Yuan, Y. (2017). Mispricing factors. *The review of financial studies*, 30(4):1270–1315.

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

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