Mendel University in Brno

MENDELU Working Papers in Business and Economics

105/2025

The Power of the Crowd: Retail Investors and the Cost of Capital

Stephen P. Ferris, Jan Hanousek Jr., Jan Hanousek, Jolana Stejskalová

MENDELU Working Papers in Business and Economics
Research Centre
Faculty of Business and Economics
Mendel University in Brno
Zemědělská 1, 613 00 Brno
Czech Republic
http://vyzc.pef.mendelu.cz/en
+420 545 132 605

Citation

Ferris, S. P., Hanousek, J. Jr., Hanousek, J., Stejskalová, J. (2025). The Power of the Crowd: Retail Investors and the Cost of Capital. *MENDELU Working Papers in Business and Economics* 105/2025. Mendel University in Brno.

Abstract

Ferris, S. P., Hanousek, J. Jr., Hanousek, J., Stejskalová, J.: **The Power of the Crowd: Retail Investors and the Cost of Capital**

Using a natural experiment based on technical improvements to Google Trends data, we are able to identify the attention of unsophisticated retail investors more clearly and disentangle its impact on equity trading. We find that this trading has a significant negative effect on a firm's implied cost of capital. Further, we discover that the attention of unsophisticated investors decreases liquidity. These adverse effects of trading by unsophisticated investors are more pronounced for smaller firms with lower institutional ownership. As a remedy, firms respond to the increased trading of unsophisticated investors by reducing the textual complexity of their financial statements.

Key words

retail investors, investor attention, cost of capital, trading, sophistication, earnings complexity

JEL: G12, G14

Contacts

Jan Hanousek, Professor of Finance at Mendel university in Brno, e-mail: jan.hanousek.uf@mendelu.cz.

Acknowledgements

We are grateful for the comments of Konstantin Sokolov, Shaddy Douidar, Huaixin Wang, conference participants at the European Financial Management Association, Slovak Economic Association Meeting, Financial Management Association, and seminar participants at the University of Memphis, University of Mississippi, and Mississippi State University.

"*Given the broad awareness and brand recognition of Reddit, including as a result of the popularity of r/wallstreetbets among retail investors, and the direct access by retail investors to broadly available trading platforms, the market price and trading volume of our Class A common stock could experience extreme volatility for reasons unrelated to our underlying business or macroeconomic or industry fundamentals, which could cause you to lose all or part of your investment if you are unable to sell your shares at or above the initial offering price*." Reddit Inc. IPO prospectus[1], page 66.

Introduction

Retail investors play a significant role in equity markets, with their behavior influencing a wide range of market dynamics. Early studies, such as Kumar and Lee (2006), Kaniel et al. (2012), and Barrot et al. (2016), emphasize the positive effects of retail participation, particularly in providing liquidity. More recent studies, however, suggest this view might be overly optimistic. Bradley et al. (2023) argue that retail investors can coordinate their trades and potentially push prices away from their fundamental values. Goldstein et al. (2013) theorize that retail investors can start and propagate trading frenzies, which amplify volatility and hinder price discovery. Chapkovski et al. (2023) contend that retail investors, motivated by emotions and gamified platforms, often make investment decisions unrelated to valuation fundamentals.

As unsophisticated investors increasingly participate in trading, often driven by market trends (Bradley et al., 2023), social media influences (Benetton et al., 2023), and emotional decision-making (Chapkovski et al., 2023), their impact can disrupt fundamental price discovery mechanisms, and ultimately affecting market stability.

This kind of disruption can substantially affect the firm's cost of capital. A firm's cost of capital is primarily determined by the perceived risk and liquidity of its equity. When driven by unsophisticated motives, retail investor behavior has the potential to decrease liquidity (Eaton et al., 2022) and, consequently, increase a firm's cost of capital. When retail investors engage in herding behavior or speculative trades, stock prices can deviate from their fundamental values. This introduces even more uncertainty and risk, forcing firms to compensate investors with higher returns (Brunnermeier, 2009). This has the effect of increasing a firm's cost of capital. By differentiating between the attention of sophisticated and

[1] Available here https://www.sec.gov/Archives/edgar/data/1713445/000162828024006294/reddits-1q423.htm

unsophisticated investors, we explore how varying levels of investor sophistication affect a firm's cost of equity capital.

Our empirical analysis yields several important findings. First, we find that while aggregate retail investor attention correlates with a lower implied cost of capital, attention from unsophisticated investors significantly increases the cost of capital. Stocks that attract higher levels of attention from unsophisticated investors experience a 10% higher implied cost of capital than the global historical average (Lee et al., 2021). We also show that stocks of smaller firms that receive attention from unsophisticated investors underperform the market by 233 basis points in the subsequent month. Conversely, larger firms outperform the market by 81 basis points, suggesting that rational arbitrageurs can offset this adverse effect (e.g., Shiller, 1984; Shleifer, 2000; Lee, 2001).

Second, our analysis reveals the effect of retail investor attention on market liquidity. While aggregate attention increases liquidity, attention from unsophisticated investors reduces liquidity by 8 %. This result is particularly prominent in smaller firms, where retail investors are more common. Unsophisticated investors are more likely to herd, and thus, they are less likely to make contrarian trades and provide market liquidity (Kaniel et al., 2008; Kelley and Tetlock, 2013). Consistent with Eaton et al. (2022), we find that unsophisticated investors could harm market liquidity, which is the channel most likely to affect a firm's cost of capital.

Lastly, our analysis shows that firms actively attempt to mitigate any adverse effect on their cost of capital from trading of these investors. We observe that firms significantly reduce the complexity of their financial statements. Notably, they reduce the amount of text in the statement, replacing it with more pictures and infographics and reducing the number of complex words. These changes in presentation, style, and content are likely due to the increased trading by less sophisticated retail investors. By reducing the complexity of their financial statements, firms can better communicate with these investors. Moreover, reduced earnings complexity leads to more unified forecasts by financial analysts (e.g., Li, 2008; Loughran and McDonald, 2014; Ertugrul et al., 2017; Bae et al., 2023). As a result, investors receive less ambiguous information regarding business operations and projections.

Our study makes three significant contributions to finance literature. First, we extend the existing research analyzing retail investors by comparing the effects of sophisticated and unsophisticated investors. Using technological improvements to Google Search Volume Intensity (SVI) data, we are able to separately measure the attention of unsophisticated and sophisticated investors. This separation allows us to quantify how retail investors of varying sophistication influence market behavior, a contribution that directly aligns with recent studies of trading by retail investors (Eaton et al., 2022).

The 2021 update to Google Trends is a natural experiment that allows us to more accurately determine the sophistication of traders in a firm's stock. We observe that investors identified by our methodology behave in line with the theoretical models of Ho and Stoll (1981) and Grossman and Miller (1988), as their attention appears persistent. As their theory suggests, the persistent order flow by less sophisticated investors creates a fluctuation in inventory value, which can adversely affect liquidity and the firm's cost of capital.

Second, our study examines the longer-term consequences of trading by unsophisticated investors. Like Eaton et al. (2002), we find that trading by unsophisticated investors has an adverse effect on liquidity. In contrast to Eaton et al. (2022), however, our study is more than an analysis of the immediate impacts of unsophisticated trading on market quality. We show that unsophisticated investor trading not only disrupts liquidity but also has enduring effects on the firm's cost of obtaining new capital. Further, we test and discover that the impact of reduced liquidity is sensitive to the firm's size. Our results help establish the long-term effects of unsophisticated investors on market behaviors, which is a new direction in the trading literature.

Our third contribution is analyzing how firms redesign their financial reporting in response to increased unsophisticated investor attention. Firms receiving such attention simplify their 10-Q and 10-K filings and reduce the complexity of their content. We find that firms use more infographics and pictures than text to convey information. Moreover, we discover that firms use fewer complex words in their statements. This indicates a reduction in the overall complexity of financial reporting. While these results align with previous studies on corporate disclosure strategies (Li, 2008; Loughran and McDonald, 2014; Ertugrul et al., 2017; Bae et al., 2023), they document a novel link between the behavior of retail investors and subsequent changes in firm disclosure. Moreover, these results outline how firms respond to unsophisticated investor attention and potentially mitigate the negative effects without further regulation.

We organize our study into six separate sections. In section two, we review the existing literature on the effects of retail investor trading on market dynamics, mainly focusing on equity performance. Section three explains our data collection and methodology, highlighting our use of Google Trends as a natural experiment. In section four, we present our empirical findings regarding the effects of retail investor attention on the cost of capital, liquidity, and earnings complexity. Section five contains various robustness checks, including placebo tests, to verify our identification strategy. Finally, in section six, we summarize our findings and discuss their implications for corporate governance, market regulation, and the role of retail investors in financial markets.

1 The Consequences of Equity Trading by Retail Investors

A firm's cost of equity capital is affected by its stock performance in the capital markets. Diamond and Verrecchia (1991) show that increased liquidity attracts large investors, which can reduce the firm's cost of capital. Easley and O'Hara (2004) find that information asymmetry significantly affects the cost of capital, as investors demand higher returns to hold stocks in the presence of greater information asymmetry. Similarly, Malkiel et al. (1997) suggest that the standard deviation of a firm's returns should be positively correlated with the firm's future returns.

Many of these channels influence the cost of capital and are also affected by retail investors. Cookson and Niessner (2020) find that investor disagreement has a significant effect on the trading volume and volatility of a firm's equity. Research by Hirshleifer and Teoh (2003), Hou et al. (2009), and Hirshleifer et al. (2011) reports that investor attention is necessary for price discovery. Further, as noted previously, retail investors can significantly affect liquidity (e.g., Kaniel et al., 2008; Kelley and Tetlock, 2013; Eaton et al., 2022).

The literature examining the effects of retail investor trading on a firm's cost of capital is very narrow. Cao et al. (2015) analyzed the impact of reputation on the cost of capital and found that changes in reputation are associated with subsequent changes in the firm's investor base. A strengthened reputation improves investor recognition, leading to enhanced risk sharing. Pastor et al. (2022) show that investor demand for green assets influences the cost of capital. However, explicit testing of the effect on a firm's equity cost of capital by retail investor trading has not yet been done.

Studies by Kaniel et al. (2008, 2011) and Kelley and Tetlock (2013) suggested that trading by retail investors can provide liquidity to other market participants; therefore, their participation is positive for the market. Barrot et al. (2016) further confirmed this liquidity contribution by retail investors, especially when conventional providers are constrained. But market activity, such as GameStop's short squeeze, demonstrates the adverse effect that retail investors can have on the market through their coordinated trading (Bradley et al., 2023). Further, retail investors can cause and propagate trading frenzies (Goldstein et al., 2013; Bradley et al., 2023), adversely affecting the market's liquidity.

Eaton et al. (2022) suggested that these conflicting results can be explained by the variance in sophistication among retail investors. Using outages of the Robinhood trading platform, they find that Robinhood investors worsen liquidity, increase volatility, and increase market order imbalances. Different levels of sophistication among investors have also been observed in studies analyzing social media (Farrell et al., 2022; Bradley et al., 2023) and are implicit in the theoretical modeling of Grossman and Stiglitz (1980).

Consequently, we hypothesize that unsophisticated retail investors negatively affect a firm's equity cost of capital. Consistent with Eaton et al. (2022), we expect that liquidity is the channel by which a firm's cost of capital is affected. Unsophisticated investors are more likely to be influenced by herding (Eaton et al., 2022; Chapkovski et al., 2023) and, thus, more likely to demand liquidity when traditional liquidity providers are restricted. [2] Conversely, sophisticated investors tend to be contrarian, implying they are more likely to be liquidity providers (Kaniel et al., 2008; Kelley and Tetlock, 2013).

It is doubtful, however, that retail investors can affect the entire market. Kumar and Lee (2006) note that even in the presence of systematic noise trading, which pushes prices away from value fundamentals, the activities of rational arbitrageurs can offset this behavior (e.g., Shiller, 1984; Shleifer, 2000; Lee, 2001). Not every stock, however, has a sufficient number of institutional investors to arbitrage the value departures due to noise trading. For instance, small firms have lower analyst coverage and fewer institutional investors, making arbitrage corrections less likely (O'Brien and Bhushan, 1990). Consequently, we contend that the adverse effects of trading due to unsophisticated investors will be more significant for smaller firms.

[2] An example where conventional liquidity providers are constrained is a trading frenzy. As Goldstein et al. (2013) show, a trading frenzy can occur when speculators place a large weight on information such as a rumor. During the frenzy, the frantic speculative trading leads to significant pressure on prices. This can cause market makers to perceive themselves as uninformed and thus decrease market liquidity (Green and Smart, 1999).

2 Data and the Measurement of Investor Sophistication

2.1 Google's Search Value Intensity

This study uses Google Trends data for the ticker searches of all publicly listed firms covering the sample period from 2004 to 2018. The SVI reported by Google ranges from 0 to 100, where 100 corresponds to the period of the highest search intensity for the given term. Google data, however, is not constant. Eichenauer et al. (2021) provided a detailed study of the consistency of Google search data. They observed that the frequency of searches and regional search data for the same period can vary significantly across collection dates. There are two reasons for this variance.

First, Google does not report the total number of searches. Instead, Google provides an index created from a random sampling of their data. This random sampling can cause significant deviations for small population samples, such as regional searches or higher-frequency data (e.g., daily). It does not, however, cause substantial deviations for large population samples.

The second reason for this variance is technological improvement in data collection. Over the years, Google search has undergone numerous changes. [3] These changes focus on improving the data and filtering out unrelated searches. Google continually improves its Search Volume Intensity (SVI) measure (Eichenauer et al., 2021) and applies these improvements backward to existing data. Thus, the time-series values are retroactively changed based on these technical enhancements designed by Google.

It is because Google retroactively re-estimates the time series of its SVI values after each technical improvement that generates the natural experiment we use in this study. Specifically, we use the SVI values for all publicly listed firms from 2004 to 2018 collected at two different points in time. The first sample, referred to as the "original," was collected in December 2019. These values are estimated before the last technical improvement[4] in 2021. Our second sample, referred to as "improved," is collected in December of 2022. [5] Thus, we have SVI data for a sample of firms at the same point in time, but calculated

[3] For example, Google has made over 5,000 improvements to their searches in 2021 alone (https://blog.google/products/search/danny-25-years-of-search/).

[4] This improvement in 2021 consisted of introducing a new feature in Trends called "Spikes," which highlighted sudden increases in search volume for a particular query. However, Google also notes that it changed its algorithm to filter out spam and irrelevant results more effectively.

[5] We collect the search data for the same period several times to test the robustness of our results. As noted by Eichenauer et al. (2021), we observe slight differences in the datasets. However, the differences are insignificant and close to zero for datasets downloaded close to each other. Our results remain statistically identical when using data collected at different times for the improved sample or when using an average, supporting the conclusion that

using two different methods. Again, we want to emphasize that these changes are applied retroactively. This means that any data collected since 2021 will consist of these improved SVI measures.

We plot the differences between the original and the improved SVI in Figure 1.

[Figure 1]

As an illustration, Figure 1 presents the SVI for Google's ticker (i.e., "GOOG." [6]) for both the original and improved samples. Our analysis shows that the original sample SVI tends to be higher than the improved SVI. [7] This result is consistent with (Google's) claim that its measurement of SVI is improved by excluding searches that Google considered unrelated.

Our starting sample consists of 4,518 unique firms over the 2010 to 2018 period. [8] Following Da et al. (2011), we collect the monthly SVI of the firm's tickers rather than its actual name. [9] Similarly, we remove firms whose tickers consist of only one or two characters (e.g., "C" for Citi group) as well as firms whose tickers have generic meanings (e.g., "DO" for Diamond Offshore Drilling). Overall, our final sample consists of 288,793 firm-month observations. We provide summary statistics for all variables in Appendix B.

technological improvements drive the differences in original and improved datasets. We offer a more detailed explanation in Section 5.

[6] The search is not case-sensitive. Moreover, Google Trends offers search data on both "GOOG" as a term and as a topic, with the latter also including searches such as "GOOG price." However, Google does not share all searches they include for any given topic, meaning that using "GOOG" as a topic might include searches that investors would not search. As a result, we focus on searching for tickers as a term consistent with past literature (e.g., Da et al., 2011).

[7] It is possible for the improved SVI to be larger than the original SVI, as is visible in the picture. This may be caused by a significant spike in the searches caused by unexpected major news concerning the company that was previously misclassified by the Google algorithm as unrelated. However, we observe that most of the time (for roughly 70% of the sample), the SVI of the original is larger than or equal to the SVI of the improved dataset.

[8] Google Trends data started in 2004, and we have data for searches from 2004 to 2018. However, our identification test use data only available since 2010, such as institutional investor attention and the algorithm of Boehmer et al. (2021). We, therefore, restrict our sample to 2010-2018 for our presented results. The rest of our results are robust to the sample reduction, and results using a total sample are available upon request.

[9] Da et al. (2011) further show that while the level of SVI can be used, they prefer the change in levels of SVI, using the past six months of data. However, base SVI is preferred in our study since our primary variable is the difference between SVI and SVI^*. Using change instead of base SVI could eliminate the technological improvement in SVI^* and thus reduce the power and effectiveness of our tests.

We contend that Google classified searches as unrelated or irrelevant because unsophisticated investors initiated them. In the following section, we justify using this measure as a proxy for unsophisticated investor attention.

2.2 Measuring Unsophisticated Trading

Google categorizes searches as either related (sophisticated) or unrelated (unsophisticated), similar to theoretical models used in literature (e.g., Grossman and Stiglitz, 1980). We can decompose the $SVI_{i,t}$ of stock *i* in month t into two components: (1) searches by sophisticated ($S_{i,t}$) investors, and (2) searches by unsophisticated investors ($U_{i,t}$): [10]

$$SVI_{i,t} = U_{i,t} + S_{i,t} \tag{1}$$

We assume that the improved SVI (i.e., SVI*) is lower or equal to SVI due to technological improvements. This occurs because of the removal of unrelated searches originating from unsophisticated investors. That is $U^*_{i,t} \leq U_{i,t}$ where $U^*_{i,t}$ represents the searches initiated by unsophisticated investors that remain in the SVI measure after the 2021 technical improvement implemented by Google. Consequently, $SVI^*_{i,t}$, the SVI calculated after the 2021 improvements can be expressed as:

$$SVI^*_{i,t} = U^*_{i,t} + S_{i,t}, \text{ where } U^*_{i,t} \leq U_{i,t} \tag{2}$$

Thus, the difference between SVI and SVI* is:

$$SVI_{i,t} - SVI^*_{i,t} = U_{i,t} - U^*_{i,t} + S_{i,t} - S_{i,t} \tag{3}$$

$$SVI_{i,t} - SVI^*_{i,t} = U_{i,t} - U^*_{i,t} \tag{4}$$

$$SVI_{i,t} - SVI^*_{i,t} = U^\dagger_{i,t} \tag{5}$$

We represent the difference between SVI and SVI* as $U^\dagger_{i,t}$, which proxies for the attention of unsophisticated investors.[11,12]

---

[10] Ding and Hou (2015) show that investors are more likely to use tickers to find information about a company, whereas consumers generally use the company's name. As a result, we assume that the majority of ticker searches will originate from investors.

[11] It is important to note that it is not the attention of all unsophisticated investors, only those identified by Google during the latest significant technological improvements. However, $U^\dagger_{i,t}$ should have very low type one error, meaning that attention captured should stem from unrelated searches, as any differences stemming from sampling should be minimal.

[12] Let us note that our specified equations hold even with negative SVI difference. This can occur during a situation, where Google incorrectly flagged some searches as unrelated in previous technological improvements. In this case,

Google does not disclose how they classify a search as relevant or unrelated. We do know, however, what information they use in their screening algorithm (e.g., Dotson et al., 2017; Ahmadi et al., 2023). Google tracks the user's reported demographics, search history, and what the user does after each search. Consequently, the searches that are most likely to be deleted by Google are those by individuals who typically do not follow investing news or regularly search for stock information or finance topics. These are likely to be first-time investors or individuals unfamiliar with the information channels of the capital markets.

The literature offers several theories explaining how unsophisticated investors can adversely affect the market. Adverse selection models, described by Glosten and Milgrom (1985) and Kyle (1985), argued that noise traders' transitory orders dilute private information's value. Alternatively, Ho and Stoll (1981) and Grossman and Miller (1988) developed inventory risk models, where persistent order flows by inexperienced investors create fluctuations in the value of inventory. By analyzing whether the difference in SVI is persistent, we can better understand the behavior of identified investors. We provide the results of this analysis in Appendix C.

We observe a significant persistence in SVI differences, suggesting that unsophisticated investors identified by our approach behave in a manner consistent with Ho and Stoll (1981) and Grossman and Miller (1988). Therefore, their attention can be driven by sentiment or induced by behavioral bias rather than transitory trading. Moreover, our results show that Google is not removing truly unrelated searches. If that were true, we should not expect persistence in the removed searches. We provide further evidence that the changes in SVI are due to technological improvements in Section 5.

2.3 Changes in SVI as a proxy of unsophisticated retail investors

In the previous section, we argued that technological improvements in Google Search data allow us to capture the attention of unsophisticated retail investors. We provide several tests in this section to further verify our identification method. First, it is essential to show that the SVI is related to retail volume and thus can be used to examine the effect of retail trading on the firm's equity cost of capital and share liquidity. To measure retail volume, we use the Barber et al. (2023) improvement on the Boehmer et al.

$SVI_{i,t} - SVI^*_{i,t} = U^{\dagger}_{i,t} + S^{\dagger}_{i,t}$, where $U^{\dagger}_{i,t} < S^{\dagger}_{i,t}$, and $S^{\dagger}_{i,t} < 0$. However, this should not affect our empirical results, as we expect there to be different signs for SVI difference and SVI improved. Our results are robust to the exclusions of observations with negative SVI difference, and they are available upon request.

(2021) algorithm, which itself is based on the method of Lee and Ready (1991).[13] Using this approach, we can approximate the monthly retail volume for each stock.

We construct two variables to undertake our empirical analysis. The first is *Retail Volume Scaled,* which is the monthly retail volume scaled by total share volume. The second is *Abnormal Retail Volume,* which is the percentage change in the ratio of retail volume to the average retail volume for the stock over the past three months. For control variables, we introduce *Size,* defined as the market value of equity; *Book to Market Ratio; Illiquidity* as defined by Amihud (2002); *Past Returns* described by Brennan et al. (2012); *Volatility* which is defined as the standard deviation of daily returns for the month; *Volume*, defined as the average share volume for the month, as well as firm and year fixed effects. Detailed definitions of these control variables are provided in Appendix A.

[Table 1]

The results of our analysis are presented in Table 1. Models 1 through 3 use *Retail Volume Scaled* as the dependent variable, while models 4 through 6 use *Abnormal Retail Volume*. We observe that both versions of SVI (original and improved) significantly predicts retail volume. This result is consistent with the conclusions of Ding and Hou (2015). Moreover, our *SVI Difference* is also a significant predictor of retail volume. These findings show that SVI is significantly and positively related to retail volume, thus supporting our approach to studying unsophisticated investors.

Next, we must verify that our primary variable, *SVI Difference*, captures the trading activity of less sophisticated investors. To test this hypothesis, we use the approach of Ben-Rephael et al. (2017), who proxy institutional investor attention by search activity on Bloomberg. They analyze the relation between institutional and retail investor attention, which they similarly proxy with Google search activity. Using their methodology, we examine the relationship between institutional attention and our unsophisticated retail investor trading activity measures.

[Table 2]

Table 2 contains the results of our analysis. Our dependent variable is average institutional attention. This is measured as the average abnormal institutional investor attention (AIAC), as defined by Ben-Rephael et al. (2017), for the given month and stock. We observe that *SVI Improved*, the retail investor attention,

[13] Barber et al. (2023) suggest that this improvement yields high and homogenous accuracy rates across all stocks. For a more detailed explanation of the algorithm, see Barber et al. (2023).

is positively related to institutional attention. Institutional attention leads to *SVI Improved,* which confirms the findings of Ben-Rephael et al. (2017). Conversely, we observe that *SVI Difference,* our proxy for unsophisticated retail investor attention, is negatively related to institutional investor attention. This result shows that unsophisticated retail investors target stocks that institutional investors avoid and vice versa. For various practical, fiduciary, and compliance reasons, it is reasonable to assume that institutional investors are sophisticated investors. Similar behavior or targeting identical stocks should be a measure of sophistication by retail investors. Therefore, the investors whose attention is captured by *SVI Difference* can be described as unsophisticated. These two results support our use of the *SVI Difference* to examine the effect of unsophisticated investors on the firm's equity cost of capital.

3 Empirical Findings

3.1 Implied Cost of Capital

We first examine the effect of unsophisticated retail investors on the firm's implied cost of capital (ICC). ICC is fundamental to several key corporate decisions, from determining the hurdle rate for investments to the composition of a firm's capital structure and subsequent profitability of corporate operations. To undertake our empirical testing, we follow Gebhardt et al. (2001) and the methodology of Hou et al. (2012). Thus, our approach uses regression-based forecasts rather than analysts' earnings. This method is preferable since analysts' earnings forecasts are less available for smaller firms. These are the same firms that attract greater attention from unsophisticated investors. This method has also been used in previous research by Donangelo (2014), İmrohoroğlu and Tüzel (2014) and Pastor et al. (2022).

Similar to Pastor et al. (2022), we use the Gebhardt et al. (2001) model to estimate the ICC since it provides the most precise expected returns estimates in the cross-section (Lee et al., 2021). Consistent with the methodology of Hou et al. (2012), we use regression-based forecasts based on historical data for each industry to obtain earnings-per-share forecasts. The implied cost of capital is the internal rate of return that equates the present value of future dividends to the current stock price.[14] Following this

[14] Hou et al. (2012) use ten years of historical data for the given industry to forecast earnings for each stock up to three years ahead. They then assume that the stock will revert to the industry average ROE from year 4 to year 12, with terminal perpetuity after year 12. They further assume that the dividend-payout ratio will remain constant for the entire forecasting horizon. A combination of bisection and Newton-Raphson algorithms solves the resulting equation. They discard observations where the discount rates have different solutions from the two algorithms. For a full description of the methodology, we refer to Lee et al. (2021) and the online appendix of Pastor et al. (2022).

methodology, we then calculate ICC for each stock and month. Since we observe the SVI for the entire month, we use the closing price of the given month when calculating the ICC.

[Table 3]

We report the results of our analysis in Table 3. We find that the estimated coefficients for *SVI Original* and *SVI Improved* are both significantly positive. These results suggest that the attention of investors leads to higher expected returns and, thus, a lower cost of capital. Conversely, the *SVI Difference* has a significantly negative effect, which implies lower expected returns. Our findings suggest that the negative effects caused by the trading activity of unsophisticated investors lead to lower expected performance. Therefore, equity trading by unsophisticated investors increases the equity cost of capital for the affected firms. We analyze the economic significance of our findings as well as test for causality in the following section.

3.1.1 Causality and Economic Significance

The previous section establishes a significant statistical effect on a firm's implied cost of capital (ICC) due to investor attention. However, because the SVI variables are normalized by their maximum value and we only observe searches deleted due to the latest technological improvement, economic interpretation can be difficult. Consequently, we use propensity score matching and the potential outcome framework of the Rubin Causal Model (Holland, 1986) to estimate the Average Treatment Effect on Treated (ATET). This model is based on two outcomes, one with and one without treatment:

$$y_{0,i} = \mu_0 + \varepsilon_{0i} \text{ and } y_{1,i} = \mu_1 + \varepsilon_{1,i} \tag{6}$$

Formally, the model can be written as $y_{T,i} = \mu_T + \varepsilon_{T,i}$, where subscript T=1 denotes the treatment, and T=0 represents the control group. We observe only one outcome for each firm *i*, either $y_{0,i}$ or $y_{1,i}$. The counterfactual outcomes must be estimated. In summary, the Rubin Causal Model helps us understand the effect of a treatment by comparing what happens to a group that receives the treatment to what would have happened if they had not. We estimate the Average Treatment Effect on the Treated (ATET), where $y_{0,i}$ is calculated using the nearest-neighbor approach with an extensive set of controls.[15]

[15] The treatment effect $E[y_{1i} - y_{0i}]$ is under random assignment equal to $\mu_1 - \mu_0$.

In this study, the outcome variable is the firm's implied cost of capital (ICC). The terms μ_1 and μ_2 are the indicators of whether there is a high level of attention from unsophisticated investors while controlling for various characteristics. We perform exact matching on year, month, and the Fama-French 48 industry classifications, while the approximate coordinate for matching is firm size.

We follow established procedures for constructing the control group (e.g., Rosenbaum and Rubin, 1985; Rubin, 2008) and evaluating the Average Treatment Effect on Treated (ATET). To define the treatment and control group, we use the distribution of the *SVI Difference*. Specifically, we assign monthly values above the 80th percentile[16] as the treatment group (i.e., attention driven by unsophisticated investors), while values below the 20th percentile serve as a control group. We report the results of our analysis in Panel A of Table 4.

[Table 4, Panel A]

We find that the treatment effect is highly significant, indicating a decrease of 11 basis points in the implied cost of capital following high levels of attention from unsophisticated investors. This effect is also highly economically significant. Lee et al. (2021) report that the historical average ICC equals 1.09%.[17] Therefore, firms located in the highest pentile of unsophisticated investor attention have approximately a 10% lower expected return than those in the lowest pentile.

The ATET approach allows us to establish causality and demonstrate the economic significance of our results. In Panel B of Table 4, we provide a balance plot for firm size, our approximate coordinate for matching. Overall, these findings show that our sample is well balanced and support our causal interpretation of unsophisticated retail trading's effect on the equity cost of capital.

3.1.2 Cost of Capital Using Realized Returns

The preceding section describes the effects of unsophisticated investors on the firm's implied cost of equity capital. Another approach is to employ realized returns to proxy for expected returns (e.g.,

[16] We omit the middle 60% to better isolate the effects of the attention of unsophisticated investors. This split is also chosen to offer a balancing of covariates used for matching. Results using other sample splits of the sample are consistent with these presented results.

[17] We observe an average ICC for our sample to be 1.36%, which would correspond to an increase of 8% following attention by unsophisticated investors. Our sample, however, does not contain all firms due to removals based on ticker length. Hence, our sample is smaller than that of Lee et al. (2021).

Armstrong et al. 2010).[18] By analyzing realized returns, we can quantify the effect such trading has on equity investors.

To undertake this analysis, we follow Armstrong et al. (2010). Specifically, we construct twenty-five (5x5) equal-weighted portfolios for each month based on a two-dimensional sort using investor attention and firm size. We then compute buy-and-hold returns for each portfolio one month ahead.

To construct our portfolios, firms are first ranked into pentiles based on unsophisticated retail investor attention (*SVI Difference*). Then, each of these five unsophisticated retail investors' attention portfolios is sorted into five size-based portfolios, resulting in twenty-five different portfolios. We then use the five-factor model of Fama and French (2015) to evaluate the performance of each portfolio. Armstrong et al. (2010) claim that employing this portfolio approach mitigates possible concerns regarding noise in future returns.

We use excess return as our measure of performance. Firms that suffer from poor performance are less able to attract buyers for their securities. This results in a decline in equity values and increases the firm's cost of capital. Thus, there is an inverse relation between the firm's performance and its equity cost of capital.

[Table 5]

We report the results of this portfolio analysis in Table 5. We present the results for the largest pentile of unsophisticated investor attention, which is the focus of our study.[19] We observe that the smallest firms earn significantly negative alphas following the high attention of unsophisticated investors. Specifically, these firms earn 230 basis points less. As the firm's size increases, however, the firm begins to gain positive and significant alphas. The largest firms earn 81 basis points. Consequently, the arbitrage portfolio, which shorts the smallest firms and purchases the largest firms in the pentile, gains 312 monthly basis points.[20]

[18] A rich literature compares the merits of future returns versus the implied cost of capital (e.g., Easton and Monahan, 2003; Guay et al., 2011; McInnis, 2010).

[19] Other pentiles are available upon request.

[20] Let us note that this approach does not consider rebalancing costs. Results in Section 4.2 show that high attention from unsophisticated investors leads to negative effects on liquidity. This means that this trading strategy has large transaction and rebalancing costs. This implies that actual returns following this strategy might be significantly smaller.

These results show that only the returns of small firms are adversely affected, with the shares of large firms unaffected. This is consistent with Kumar and Lee (2006), who contend that rational arbitrageurs offset the noise trading of unsophisticated investors. Nevertheless, for small firms, which are less traded by institutional investors, trading by unsophisticated investors adversely affects their performance. This has the effect of increasing the equity cost of capital for these smaller firms.

3.2 Market Liquidity Effects

To test the effect of retail investors on liquidity, we use the effective spread as a measure of liquidity (Chordia et al., 2001; Fang et al., 2009; Eaton et al., 2022).[21] Specifically, we use the dollar-weighted effective spread scaled by the midquote. Using the share-weighted effective spread yields identical results. Similar to Fang et al. (2009), we calculate the monthly effective spread by taking the average daily effective spread for the given month.[22] Since the effective spread does not follow a normal distribution, we use its natural logarithm transformation in our regression analysis. This measure of effective spread is negatively related to market liquidity, with larger positive values indicating worsening liquidity.

[Table 6]

We report the results of our analysis of liquidity in Table 6. While *SVI Original* and *SVI Improved* lead to increased liquidity, the *SVI Difference* significantly reduces liquidity. These findings suggest that unsophisticated retail investors do not improve liquidity but actually reduce it. This result is consistent with Eaton et al. (2022) and helps to explain the mixed evidence for the liquidity provision hypothesis established in the literature. While retail participation positively impacts liquidity, unsophisticated traders will likely reinforce any liquidity shortages rather than correct them (e.g., Goldstein et al., 2013). This negative effect can be caused by retail investors' effect on inventory costs (Baldauf et al., 2024) or by placing market orders during periods of higher volatility (Goldstein et al., 2013).

However, Kumar and Lee (2006) observed that institutional investors should act as rational arbitrageurs and offset any adverse impact on market liquidity from unsophisticated trading. Therefore, we contend that the effectiveness of rational arbitrageurs in the market will depend on the participation level of

---

[21] The Liquidity Provision Hypothesis (Glosten and Milgrom, 1985) discusses how market makers help provide liquidity, even in the presence of information asymmetry, where some traders have more information than others. Their model explains how the bid-ask spread compensates the market maker for the risk of trading with someone who might have superior information, thereby ensuring that liquidity is maintained in the market.

[22] We use a dollar-weighted effective spread scaled by midquote, calculated using the approach of Holden and Jacobsen (2014). Results are identical when using share-weighted effective spread instead.

institutional investors. For large firms with significantly more institutional investors, their retail volume will be dwarfed by the trading of institutions. We should, therefore, observe that unsophisticated retail trading fails to decrease liquidity for the largest stocks.

To test this conjecture, we interact the *SVI Difference* with size quartiles. The results of this analysis are presented in model (4). The smallest firms constitute the base category. We find that the smallest firms are adversely affected by unsophisticated retail investors and have significantly worse liquidity (Shleifer and Vishny, 1997). Larger firms, however, are either unaffected or have increased liquidity. These results are consistent with the findings reported by Shiller (1984).

To facilitate economic interpretation and establish causality, we employ the Average Treatment Effect on Treated (ATET) approach using propensity score matching, as in the preceding section.

[Table 7, Panel A]

We observe in Panel A of Table 7 that the treatment effect is highly significant, indicating an approximate 8% increase in the effective spread following high attention from unsophisticated traders. This result is economically substantial and helps explain the negative effect that retail investors have on the equity cost of capital. As Diamond and Verrecchia (1991) note, increased liquidity attracts large investors, thus reducing the firm's equity cost of capital. Since we observe that unsophisticated investors decrease the liquidity of a firm's equity, those firms become less attractive to large investors. This, in turn, is likely to increase the firm's equity cost of capital.

We report the balancing plot for firm size based on the log of total assets in Panel B. Overall; these results show that our sample is well-balanced and supports our methodology and causal interpretation.

3.3 Firm Response and Earnings Quality

Annual and quarterly reports filed by public firms (i.e., 10-K and 10-Q statements) are important channels by which a firm communicates valuation-relevant information to its stakeholders. Management might seek to offset the negative effects of unsophisticated investors by reducing the complexity of earnings reporting.

Intuitively, more complex statements might lead to greater information ambiguity which can make the stock more volatile (Ertugrul et al., 2017). This might occur due to the limited information-processing expertise of individual investors (Lawrence, 2013). The complexity of financial statements is also related to more dispersed analyst forecasts and unexpected earnings (Lehavy et al., 2011). Thus, reducing the textual complexity of quarterly earnings reporting should result in less disagreement among analysts. Retail investors searching online for information about a firm should then receive consistent analyst reports, with less variability in their recommendations. Since financial statements are the primary channel through which firms communicate with shareholders, reducing their complexity might be a response to the attention of unsophisticated investors.

Contemporary research uses several metrics to measure the complexity of such statements, such as the gross and net file size of the electronic version of the statement (Loughran and McDonald, 2014), and the ratio of complex words (Loughran and McDonald, 2023). Loughran and McDonald (2016) specify that net file size, which is a natural logarithm of the statement size containing only text, should be preferred to gross file size, which also contains figures, tables, and infographics. In this study, however, we include both. Since we are interested in how management responds to the attention of unsophisticated investors, we employ the percentage change in the gross and net file size from the previous quarter instead of the natural logarithm.

[Table 8]

We report the results of our analysis in Table 8. Since financial statements are only issued quarterly, we calculate the maximum SVI for a given quarter.[23] We chose the maximum SVI since the mean could cause a downward bias effect in measuring the effect. [24] We see that gross size is unaffected in models (1) and (2). We observe in models (3) and (4), however, that higher levels of unsophisticated investor attention lead to a decrease in the net file size of the statement. This suggests that the managers write less text in their financial statements. These results might indicate that managers replace text with graphics or other visuals. Both imply that managers try to reduce the complexity of their financial statements. Finally, we note in models (5) and (6) that quarterly earnings have a lower incidence of complex words. We measure complexity as the number of complex words scaled by the number of total words using the data and methodology of Loughran and McDonald (2023). These results indicate that managers reduce the amount of text and the complexity of the words they choose in response to unsophisticated investor attention.

[23] Given the fewer observations and the assumed seasonality for fiscal quarters, we employ a different set of fixed effects for the regression. Namely, we use year and fiscal quarter fixed effects, as reported in the regression.

[24] Maximum might also be preferred, as even only one month of unsophisticated investor attention can have pervasive effects. Using the mean, however, yields similar conclusions, albeit weaker estimates.

As in previous sections, we estimate the average treatment effect on the treated (ATET) approach using propensity score matching to facilitate economic interpretation. In this analysis, we match exactly in the fiscal quarter and year, to facilitate more robust matching. For approximate matching we use firm size.

[Table 9, Panel A]

We report our results in Panel A of Table 9. The gross file size does not appear to be significantly different after the attention of unsophisticated investors. The net file size, however, decreases by approximately 5% following the attention of unsophisticated investors. We observe that the median change from one quarter to the next corresponds to a 4.1% increase in sample size. This is consistent with Loughran and McDonald (2023) who reported increased statement complexity over our sample period. Our result is economically significant and suggests that management changes how they present guidance and performance results in their financial statements. Conversely, the complexity of statements decreases by 0.2%, with a sample median of 0.44%.

In aggregate, these results show that management responds to the attention of unsophisticated investors. Their primary focus appears to be reducing text length. Given that the gross file size does not change, there is a suggestion that text is being replaced with more easily processed graphics, figures, and other visuals. The likely reason for this change is the presence of less sophisticated retail investors.

While it is unclear whether unsophisticated investors are attracted by conditions caused by complex quarterly earnings statements, our results show that firms strategically change the style and manner of their financial reporting. The change is likely motivated by the attention of unsophisticated investors' attention and the firm's desire to mitigate any negative effects. Given the rich literature detailing the positive effects associated with a reduction in the complexity of financial reporting (e.g., Loughran and McDonald, 2014; Bae et al., 2023), it also suggests that firms have the tools available to mitigate the negative effects due to retail investors without the need for additional agency or federal regulation.

4 Placebo Tests

In this study, we contend that changes in SVI are due to technological improvements, consistent with Google's practice of improving its algorithms and applying any changes retroactively. Eichenauer et al. (2021) test the consistency of Google Trends data and find that small populations and higher-frequency data might suffer from a sample bias. While this bias should not be significant for large populations and the monthly data used in this study, we test for the possibility that sample bias rather than technological improvements accounts for our findings.

If our results are driven by sample bias, it implies that the *SVI Difference* has a purely random distribution, lacking any systematic component. Therefore, we use a placebo test to randomly assign unsophisticated investor attention to some firms. If this placebo test does not lead to a significant effect on the dependent variable, it rejects the hypothesis that the *SVI difference* is random.[25] Thus, the differences in SVI would be systematic, and the explanation would be the technological improvement by Google in their measurement algorithm. Such results would then verify our identification of unsophisticated retail investors.

We use two different placebo tests. In the first placebo test, denoted as Placebo Test 1, we randomly assign the low and high differences in searching algorithms for each firm and month. This process is done using a generator of pseudo random numbers from the uniform (0,1) distribution. The high and low probability is set at 20%, consistent with Section 4.

This approach, however, does not allow for any momentum in the attention of investors since the distribution is randomly generated for each month. Consequently, we employ a second placebo test, denoted as Placebo Test 2. This test assigns, with the same probability, high and low attention in two subsequent periods rather than generating every month individually. To ensure robustness in our results, we replicated the process 100 times and reported the mean estimates. We use the same approach as that in Section 4 to evaluate the placebo treatment. We report our results in Table 10.

[Table 10]

We focus our analysis primarily on the effects on the cost of capital and liquidity. Models (1) and (3) use the implied cost of capital as defined in Section 4.1. Models (2) and (4) use the dollar-weighted effective spread scaled by the midquote. We find for Placebo Test 1 and 2 that the treatment effect is insignificant. These results reject any concern that the SVI difference is randomly generated. These findings support

[25] An alternative test of whether the *SVI difference* is impacted by sample bias could be done by using SVI data downloaded at different times both prior and post the technological improvements. While we repeated our tests with SVI downloaded at later times, as well as the mean of all downloaded new SVI data, we only have one SVI data collected prior to the last technological advancements. Due to the nature of a placebo test and its robustness, it should be considered a superior approach. Nevertheless, we replicate the analysis using SVI downloaded at different times post the 2021 improvement. The results are very similar and lead to same conclusion. For reasons of reporting brevity, they are not presented here, but are available upon request.

our identification strategy and verify that the SVI differences are driven by technological improvements. This, in turn, allows us to use the difference to proxy for the attention of unsophisticated investors.

Summary and Discussion

Extensive literature establishes how retail investor participation can lead to short-term price effects in the equity market. However, there is much less research examining the long-term impact of such trading. We address this limitation by analyzing the effect retail investor attention, especially that of unsophisticated retail investors, has on the firm's cost of capital. Our findings show that unsophisticated retail investor attention significantly negatively affects a firm's implied cost of capital, primarily through diminished market liquidity. This occurs especially in smaller firms with lower institutional ownership. Notably, stocks receiving high attention from unsophisticated investors experience a 10% higher implied cost of capital than the historical average. The stocks of smaller firms underperform following such attention, while those of larger firms with greater institutional ownership tend to outperform. This indicates that institutional investors help mitigate the adverse effects of trading by unsophisticated investors. Our results are consistent with the modeling of unsophisticated and inexperienced investor behavior effect on inventory risk by Ho and Stoll (1981) and Grossman and Miller (1988).

We also find that firms adjust their financial reporting by simplifying their disclosures in response to increased attention from unsophisticated investors. This occurs by reducing the amount of text in the statement and using fewer complex words. Our results further suggest that management uses visuals such as graphics or figures to replace text in their financial disclosures. This suggests that firms are actively trying to mitigate the negative effects of unsophisticated trading by better communicating with their less skilled and knowledgeable shareholders. This finding establishes new connections between financial reporting, investor relation practices, and the trading of unsophisticated retail investors.

Given the surge in retail investor trading in the market and the debate about the effect of that participation on market dynamics, the issue of additional market regulation arises. While our results show the adverse effects that occur due to the involvement of unsophisticated investors, we also find that firms do respond to this trading. This suggests that the calls for further regulation might be an overreaction.

Overall, the results of our study deepen our understanding of the role retail investors, especially unsophisticated retail investors, play in the stock market. Given certain momentum in sentiment and attention, any effects brought on by retail investor participation can have more long-term effects, which impact even core corporate outcomes. Given the rise in retail investor participation and a proliferation of alternative information sources and gamified platforms, understanding how retail investor behavior shapes capital markets remains an important question. Understanding whether corporate governance mechanisms, not only changes in disclosures, mitigate the adverse effects of unsophisticated investors could help firms determine whether they can take a more strategic approach to managing their investor base.

References

Ahmadi, I., Abou Nabout, N., Skiera, B., Maleki, E., & Fladenhofer, J. (2023). Overwhelming targeting options: Selecting audience segments for online advertising. *International Journal of Research in Marketing*.

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

Armstrong Ch. S., Core J. E., Taylor D. J., Verrecchia R. E., 2010, When Does Information Asymmetry Affect the Cost of Capital? *Journal of Accounting Research* 49(1), 1-40.

Bae, J. W., Belo, F., Li, J., Lin, X., & Zhao, X. (2023). The opposing effects of complexity and information content on uncertainty dynamics: Evidence from 10-K filings. *Management Science*, *69*(10), 6313-6332.

Baldauf, M., Mollner, J., & Yueshen, B. Z. (2024). Siphoned apart: A portfolio perspective on order flow segmentation. *Journal of Financial Economics*, *154*, 103807.

Barber, B. M., Huang, X., Jorion, P., Odean, T., & Schwarz, C. (2023). A (sub) penny for your thoughts: Tracking retail investor activity in TAQ. *The Journal of Finance*.

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

Ben-Rephael, A., Da, Z., & Israelsen, R. D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *The Review of Financial Studies*, *30*(9), 3009-3047.

Benetton, M., Mullins, W., Niessner, M., Toczynski, J., 2023, Celebrity Influencers: This is not Financial Advice, Working paper, available here https://bpb-us-w1.wpmucdn.com/sites.usc.edu/dist/b/1001/files/2023/11/Niessner_CelebrityInfuencers.pdf

Boehmer E., Jones Ch. M., Zhang X., Zhang X., 2021, Tracking Retail Investor Activity, *The Journal of Finance* 76(5), 2249-2305.

Bradley, D., Hanousek Jr, J., Jame, R., & Xiao, Z. (2023). Place Your Bets? The Value of Investment Research on Reddit's Wallstreetbets. *The Review of Financial Studies*, hhad098.

Brennan M. J, Chordia T., Subrahmanyam A., Qing T., 2012, Sell-order liquidity and the cross-section of expected stock returns, *Journal of Financial Economics* 105(3), 523-541.

Brunnermeier, M. K. (2009). Deciphering the liquidity and credit crunch 2007–2008. *Journal of Economic perspectives*, *23*(1), 77-100.

Cao, Y., Myers, J. N., Myers, L. A., & Omer, T. C. (2015). Company reputation and the cost of equity capital. *Review of Accounting Studies*, *20*, 42-81.

Chapkovski P., Khapko M., Zoican M., 2023, Trading gamification and investor behavior. Management Science (forthcoming), Available at SSRN: https://ssrn.com/abstract=3971868 or http://dx.doi.org/10.2139/ssrn.3971868

Cookson, J. A., & Niessner, M. (2020). Why don't we agree? Evidence from a social network of investors. *The Journal of Finance*, *75*(1), 173-228.

Chordia, T., Roll, R., Subrahmanyam, A., 2001. Market liquidity and trading activity. *The Journal of Finance* 56, 501–530.

Da, Z., Engelberg, J., Gao, P., 2011, In search of attention, *Journal of Finance* 66 (5), 1461–1499.

Donangelo, A. (2014). Labor mobility: Implications for asset pricing. *The Journal of Finance*, *69*(3), 1321-1346.

Dotson, J. P., Fan, R. R., Feit, E. M., Oldham, J. D., & Yeh, Y. H. (2017). Brand attitudes and search engine queries. *Journal of Interactive Marketing*, *37*(1), 105-116.

Diamond D. W., Verrecchia R. E., 1991, Disclosure, Liquidity, and the Cost of Capital, *The Journal of Finance* 46(4), 1325-1359.

Ding R., Hou W., 2015, Retail investor attention and stock liquidity, *Journal of International Financial Markets, Institutions and Money* 37, 12-26.

Easley, D., & O'Hara, M. (2004). Information and the cost of capital. *The journal of finance*, *59*(4), 1553-1583.

Easton, P. D., & Monahan, S. J. (2003). An Evaluation of the Reliability of Accounting Based Measures of Expected Returns: A Measurement Error Perspective. *Accounting Review* 80 (2005): 501–38.

Eaton, G. W., Green, T. C., Roseman, B. S., & Wu, Y. (2022). Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics*, *146*(2), 502-528.

Eichenauer V. Z., Indergand R., Martínez I. Z., Sax Ch., 2021, Obtaining consistent time series from Google Trends, *Economic Inquiry* 60, 694-705.

Ertugrul, M., Lei, J., Qiu, J., & Wan, C. (2017). Annual report readability, tone ambiguity, and the cost of borrowing. *Journal of Financial and Quantitative Analysis*, *52*(2), 811-836.

Fama E. F., French K. R., 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116(1), 1-22.

Fang V. W., Noe T. H., Tice S., 2009, Stock market liquidity and firm value, *Journal of Financial Economics* 94, 150-169.

Farrell M., Green T. C., Jame R., Markov S., 2022, The democratization of investment research and the informativeness of retail investor trading, *Journal of Financial Economics* 145(2), 616-641.

Gebhardt, W. R., Lee, C. M., & Swaminathan, B. (2001). Toward an implied cost of capital. *Journal of accounting research*, *39*(1), 135-176.

Goldstein I., Ozdenoren E., Yuan K., 2013, Trading frenzies and their impact on real investment, *Journal of Financial Economics* 109, 566-582.

Glosten, L. and Milgrom, P. 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, Journal of Financial Economics 14(1), 71-100.

Green, J., Smart, S., 1999, Liquidity Provision and Noise Trading: Evidence from the "Investment Dartboard" Column, *The Journal of Finance* 54(5), 1885-1899.

Grossman S. J., Stiglitz J. E., 1980, On the Impossibility of Informationally Efficient Markets, *The American Economic Review* 70(3), 393-408.

Grossman, S. J., & Miller, M. H. (1988). Liquidity and market structure. *Journal of Finance*, *43*(3), 617-633.

Guay, W., Kothari, S. P., & Shu, S. (2011). Properties of implied cost of capital using analysts' forecasts. *Australian Journal of Management*, *36*(2), 125-149.

Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2011). Limited investor attention and stock market misreactions to accounting information. *The Review of Asset Pricing Studies*, *1*(1), 35-73.

Hirshleifer D., Teoh S. H., 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting Economics* 36(1-3), 337-386.

Ho, T., & Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial economics*, *9*(1), 47-73.

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

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945-960.

Hou, K., Van Dijk, M. A., & Zhang, Y. (2012). The implied cost of capital: A new approach. *Journal of Accounting and Economics*, *53*(3), 504-526.

Hou, K., Peng, L., Xiong, W., 2009, A tale of two anomalies: The implications of investor attention for price and earnings momentum, Working paper, Ohio State University and Princeton University.

İmrohoroğlu, A., & Tüzel, Ş. (2014). Firm-level productivity, risk, and return. *Management Science*, *60*(8), 2073-2090.

Kaniel R., Saar G., Titman S., 2008. Individual Investor Trading and Stock Returns, *The Journal of Finance* 68(1), 273-310.

Kaniel R., Liu S., Saar G., Titman S., 2012, Individual Investor Trading and Return Patterns around Earnings Announcements, *The Journal of Finance* 67, 639-680.

Kelley E. K., Tetlock P. C., 2013, How Wise Are Crowds? Insights from Retail Orders and Stock Returns, *The Journal of Finance* 68(3), 1229-1265.

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

Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of accounting and economics*, *56*(1), 130-147.

Lee C. M. C., Ready M. J., 1991, Inferring Trade Direction from Intraday Data, *The Journal of Finance* 46(2), 733-746.

Lee C. M. C., 2001, Market efficiency and accounting research: A discussion of 'capital market research in accounting' by S.P. Kothari, *Journal of Accounting and Economics* 31(1-3), 233–253.

Lee, C. M., So, E. C., & Wang, C. C. (2021). Evaluating firm-level expected-return proxies: implications for estimating treatment effects. *The Review of Financial Studies*, *34*(4), 1907-1951.

Lehavy, R., Li, F., & Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The accounting review*, *86*(3), 1087-1115.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and economics*, *45*(2-3), 221-247.

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *the Journal of Finance*, *69*(4), 1643-1671.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187-1230.

Loughran, T., & McDonald, B. (2023). Measuring firm complexity. *Journal of Financial and Quantitative Analysis*, 1-28.

Malkiel, B. G., & Xu, Y. (1997). Risk and return revisited. *Journal of Portfolio Management*, *23*(3), 9.

McInnis, J. (2010). Earnings smoothness, average returns, and implied cost of equity capital. *The Accounting Review*, *85*(1), 315-341.

O'Brien, P. C., & Bhushan, R. (1990). Analyst following and institutional ownership. *Journal of Accounting Research*, 28, 55-76.

Pastor, Ľuboš, Robert F. Stambaugh, and Lucian A. Taylor. "Dissecting green returns." *Journal of Financial Economics* 146.2 (2022): 403-424.

Rosenbaum, P. R., and D. B. Rubin. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* 39: 33–38. https://doi.org/10.2307/2683903.

Rubin, D. B. 2008. For objective causal inference, design trumps analysis. *Annals of Applied Statistics* 2: 808–840. https://doi.org/10.1214/08-AOAS187.

Shiller, Robert J., 1984, Stock prices and social dynamics, *Brookings Papers on Economic Activity* 2, 457–510.

Shleifer A., 2000, Inefficient Markets – An Introduction to Behavioral Finance, *Oxford University Press*, Oxford, UK.

Shleifer A., Vishny R. W., 1997, The Limits of Arbitrage, *The Journal of Finance* 52(1), 35-55.

Figure 1: Search Volume Index (SVI) of Google Ticker Symbol

This figure plots the Search Volume Index (SVI) values obtained from the sample collected before technological improvement in 2021, *SVI original*, and the SVI values obtained after the technological improvement, *SVI improved*. We plot the SVI data for the ticker "GOOG", which corresponds to Google. The series ended in 2015, since Google then went through restructuring and Alphabet Inc. became the parent company.

We observe that *SVI original* is almost always higher than SVI improved, which is consistent with Google removing searches during technological improvements. However, it is possible for SVI-improved to be larger than SVI-original. Google explains that such a situation is possible, especially during surges of activity, when they may mistakenly classify more trades as unrelated than the actual number. However, situations where the difference between SVI original and SVI improved is significantly negative only constitute a small fraction of the sample. We observe two major spikes in the figure. The first spike was in 2006, which is likely the result of Google acquiring YouTube, as well as Google introducing several new features, such as Google Translate and Google News. The second spike, which was later corrected, was in 2013 when Google was declared to not have violated antitrust or anticompetition statutes by FTC. The reason for the possible correction was Google incorrectly classifying searches as unrelated, which instead were users searching up the impact of this news on the stock.

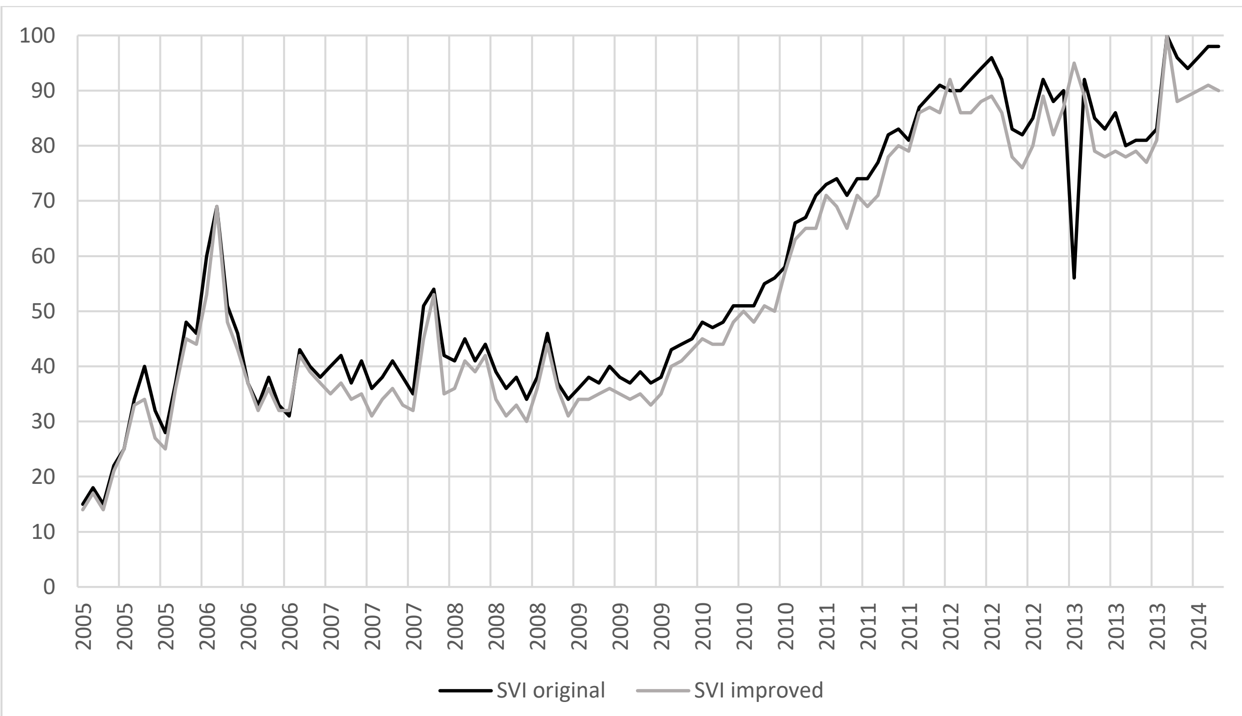

Table 1: Investor Attention and Retail Volume

This table reports the effects of investor attention on retail volume. To capture retail volume, we use the Barber et al. (2023) improvement on the Boehmer et al. (2021) algorithm, which uses the Lee and Ready (1991) quote midpoint signing method. Columns 1 to 3 employ as a dependent variable retail volume scaled by the total volume for the month. Columns 4 to 6 employ the abnormal retail volume as a dependent variable, which is the percentage change of retail volume to the average retail volume for the stock. The average retail volume is calculated by taking the average retail volume for stock over the past three months. Control variables for every regression include Size, Book to market, Past profitability, Amihud's illiquidity, Volatility, Volume and year and firm dummies. Standard errors are clustered at the firm level to control for unobserved time-invariant firm-level heterogeneity. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

	Retail volume scaled			**Abnormal retail volume**		
Variables	**Model (1)**	**Model (2)**	**Model (3)**	**Model (4)**	**Model (5)**	**Model (6)**
SVI original	0.0170***			0.0039***		
	(0.0015)			(0.0008)		
SVI improved		0.0182***			0.0039***	
		(0.0018)			(0.0011)	
SVI difference			0.0082***			0.0025***
			(0.0015)			(0.0006)
Size	-1.7226***	-1.7142***	-1.6752***	0.0567**	0.0563**	0.0644**
	(0.0772)	(0.0777)	(0.0788)	(0.0271)	(0.0275)	(0.0277)
Book to market	0.4420***	0.4419***	0.4561***	-0.1821***	-0.1907***	-0.1878***
	(0.0633)	(0.0680)	(0.0689)	(0.0322)	(0.0342)	(0.0336)
R_{m-1}	-0.1010*	-0.0997*	-0.0743	0.2625***	0.2637***	0.2691***
	(0.0560)	(0.0561)	(0.0563)	(0.0311)	(0.0312)	(0.0313)
$R_{[m-3,m-2]}$	0.9138***	0.9187***	0.9462***	0.1307***	0.1301***	0.1358***
	(0.0690)	(0.0692)	(0.0694)	(0.0317)	(0.0317)	(0.0312)
$R_{[m-6,m-4]}$	0.7447***	0.7415***	0.7491***	-0.0411**	-0.0431**	-0.0416**
	(0.0659)	(0.0659)	(0.0664)	(0.0174)	(0.0176)	(0.0173)
$R_{[m-12,m-7]}$	0.1460***	0.1456***	0.1481***	-0.0676***	-0.0684***	-0.0679***
	(0.0305)	(0.0305)	(0.0307)	(0.0107)	(0.0108)	(0.0107)
Illiquidity	-0.0014	-0.0014	-0.0014	-0.0030***	-0.0030***	-0.0030***
	(0.0014)	(0.0014)	(0.0014)	(0.0010)	(0.0010)	(0.0010)
Volatility	0.2178***	0.2217***	0.2366***	0.4027***	0.4038***	0.4069***
	(0.0153)	(0.0155)	(0.0160)	(0.0363)	(0.0362)	(0.0368)
Volume	0.0000*	0.0000*	0.0000*	0.0000*	0.0000**	0.0000**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	13.8032***	13.7127***	13.9727***	-1.6248***	-1.6081***	-1.5512***
	(0.5445)	(0.5524)	(0.5593)	(0.2583)	(0.2633)	(0.2593)
Firm fixed effects	YES	YES	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES	YES	YES
R^2	0.7010	0.6957	0.6943	0.1221	0.1221	0.1216

Number of observations	26,4876	263,768	263,768	264,838	263,730	263,730

Table 2: Institutional Investor Attention

This table reports the relationship between institutional investor attention and retail investor attention. The dependent variable is the institutional investor attention, measured using the AIAC defined by Ben-Rephael et al. (2017). This variable is defined by converting Bloomberg categorical scores of search activity into continuous values. We use the average AIAC for the given stock and month. For each explanatory variable, t-1 denotes a one-month lag, and t+1 denotes a one-month lead. Every regression includes year and industry fixed effects. Standard errors are clustered at the firm level to control for unobserved time-invariant firm-level heterogeneity. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

Variables	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
SVI improved t+1	0.0011***	0.0014***	0.0009***			
	(0.0002)	(0.0002)	(0.0001)			
SVI improved	0.0028***	0.0020***	0.0024***			
	(0.0002)	(0.0002)	(0.0002)			
SVI improved t-1	-0.0014***	-0.0028***	-0.0030***			
	(0.0002)	(0.0002)	(0.0002)			
SVI difference t+1				-0.0020***	-0.0004***	-0.0002
				(0.0002)	(0.0001)	(0.0001)
SVI difference				-0.0019***	-0.0005***	-0.0003**
				(0.0002)	(0.0001)	(0.0001)
SVI difference t-1				-0.0023***	-0.0008***	-0.0006***
				(0.0002)	(0.0001)	(0.0001)
Institutional attention t-1		0.7435***	0.5106***		0.7441***	0.5109***
		(0.0075)	(0.0052)		(0.0074)	(0.0052)
Institutional attention t-2			0.3168***			0.3168***
			(0.0043)			(0.0043)
Constant	-0.2191***	-0.1013***	-0.0930***	-0.1044***	-0.0729***	-0.0752***
	(0.0436)	(0.0111)	(0.0075)	(0.0392)	(0.0101)	(0.0069)
Industry fixed effects	YES	YES	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES	YES	YES
R2	0.1203	0.6082	0.6512	0.1143	0.6062	0.6493
Number of observations	168,113	166,529	164,379	168,113	166,529	164,379

Table 3: The Effect of Investor Attention on the Implied Cost of Capital

This table reports the effects of investor attention on the implied cost of capital (ICC). The dependent variable was calculated using the classic framework of Gebhardt et al. (2001), following the methodology of Hou et al. (2012), which uses regression-based forecasts instead of analysts' earnings. The implied cost of capital is then the internal rate of return that equates the present value of future dividends to the current stock price. The dependent variable is constructed at a monthly and firm level. Control variables for every regression include size, book to market, past profitability, Amihud's illiquidity, volatility, volume, year, and firm dummies. We report robust standard errors to control for unobserved time-invariant firm-level heterogeneity. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

Variables	**Model (1)**	**Model (2)**	**Model (3)**
SVI original	0.0005**		
	(0.0002)		
SVI improved		0.0008***	
		(0.0002)	
SVI difference			-0.0024***
			(0.0006)
Size	-0.0267***	-0.0283***	-0.0265***
	(0.0053)	(0.0054)	(0.0053)
Book to market	1.0008***	1.0000***	1.0014***
	(0.0339)	(0.0339)	(0.0338)
R_{m-1}	-0.2223*	-0.2229*	-0.2217*
	(0.1176)	(0.1176)	(0.1176)
$R_{[m-3,m-2]}$	-0.2485***	-0.2486***	-0.2463***
	(0.0839)	(0.0839)	(0.0839)
$R_{[m-6,m-4]}$	-0.1230**	-0.1227**	-0.1218**
	(0.0486)	(0.0486)	(0.0486)
$R_{[m-12,m-7]}$	-0.0721***	-0.0719***	-0.0714***
	(0.0222)	(0.0222)	(0.0222)
Illiquidity	-0.0130*	-0.0130*	-0.0130*
	(0.0078)	(0.0078)	(0.0079)
Volatility	0.0282***	0.0280***	0.0292***
	(0.0094)	(0.0094)	(0.0094)
Volume	0.0000***	0.0000***	0.0000***
	(0.0000)	(0.0000)	(0.0000)
Constant	1.5892***	1.5885***	1.6100***
	(0.1673)	(0.1673)	(0.1676)
Industry fixed effects	YES	YES	YES
Year fixed effects	YES	YES	YES
R^2	0.0519	0.0520	0.0520
Number of observations	210,626	210,626	210,626

Table 4: Investor Attention and the Implied Cost of Capital, ATET Approach

This table reports the Average Treatment Effect on Treated (ATET), measuring the impact of the increased attention of unsophisticated investors on the implied cost of capital. We use the distribution of the SVI difference, where we assign monthly values above the 80th percentile as the treatment group (attention driven by the unsophisticated investors), and values below the 20th percentile serve as the control group. We require exact matching on year, month, and Fama-French 48 industry classification, while the approximate coordinate for matching is the firm size.
In each column, we report the ATET conducted as an effect of the unsophisticated investors' high attention. The standard errors of the ATET (in parentheses) are computed with the robust option (at least two suitable matches for each treated). Below is the balance summary of the mean difference and variance ratio between the corresponding treated and control groups. Panel B shows the density plot outlining matching quality. ***, **, and * denote statistical significance on 1, 5, and 10% significance levels.

Panel A – Results of ATET analysis

	ICC
	(1)
High attention unsophisticated (ATET)	-0.111***
(std. error)	(0.028)
p-value	<0.001
Number of treated	41,298
Number of observations	47,268
Balance summary	
mean difference (size)	-0.003
variance ratio (size)	1.045

Panel B – Density plot of matching quality

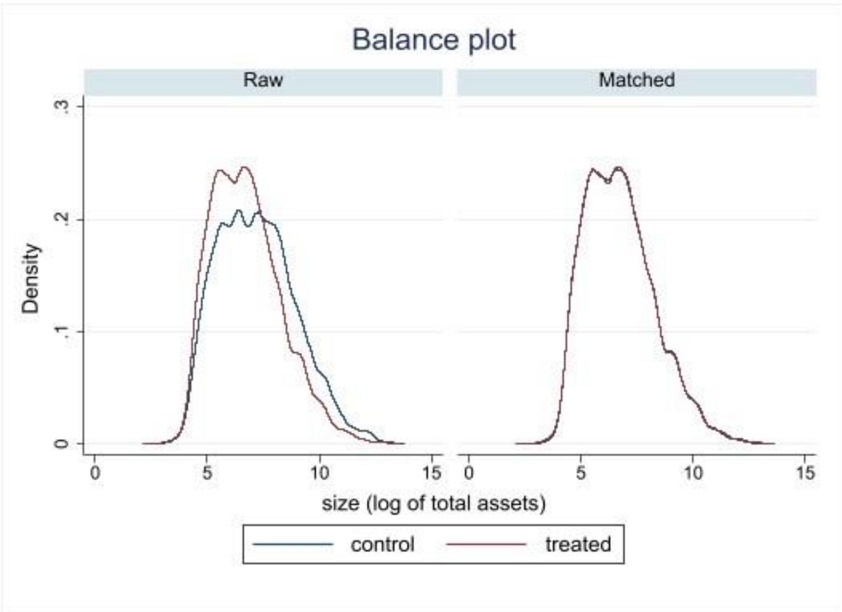

Table 5: Unsophisticated Investor Attention and the Cost of Equity Capital

This table reports the effect of unsophisticated retail investors' attention on the cost of equity capital. We form 25 equal-weighted portfolios for each month based on two-dimensional dependent sorts and compute one-month ahead buy-and-hold returns for each portfolio. Firms are first ranked into quintiles based on unsophisticated retail investor attention, which is the difference between the original SVI value and the improved SVI value. Then, within each number of unsophisticated retail investors' attention, they are sorted into five portfolios based on size, defined as the natural logarithm of the market value of equity. We also include an arbitrage portfolio created by buying the largest firms and selling the smallest ones within the given pentile of unsophisticated investor attention. The portfolio is rebalanced every month from January 2010 until December 2018. We use the Fama-French 5-factor model:

$$R_m^P - R_m^F = \beta_0 + \beta_1(R_m^M - R_m^F) + \beta_2 SMB_m + \beta_3 HML_m + \beta_4 RMW_m + \beta_5 CMA_m + \epsilon_m$$

where R_m^P is the monthly return of a particular portfolio, R_m^F is the one-month Treasury bill rate, and R_m^M is the value-weighted market return. ***, **, and * denote statistical significance at 1%, 5%, and 10% levels, respectively. We report the highest attention pentile results, other pentiles are available upon request.

Variables	Q1 (Smallest size)	Q2	Q3	Q4	Q5 (Largest size)	Arbitrage portfolio (smallest - largest)
$R_m^M - R_m^F$	0.916***	0.988***	0.978***	1.039***	0.998***	0.082
	(0.073)	(0.049)	(0.037)	(0.030)	(0.028)	(0.086)
SMB	0.747***	1.021***	1.137***	0.861***	0.449***	-0.297**
	(0.112)	(0.076)	(0.058)	(0.046)	(0.043)	(0.133)
HML	0.060	0.195**	0.086	0.017	-0.110**	-0.168
	(0.144)	(0.097)	(0.074)	(0.059)	(0.055)	(0.172)
RMW	-0.487***	-0.165	-0.209**	-0.148**	-0.163**	0.327
	(0.176)	(0.119)	(0.091)	(0.072)	(0.067)	(0.210)
CMA	0.174	-0.108	-0.128	-0.143	-0.059	-0.234
	(0.214)	(0.144)	(0.110)	(0.088)	(0.081)	(0.255)
Constant	-2.336***	0.418**	0.673***	0.834***	0.811***	3.120***
	(0.249)	(0.168)	(0.129)	(0.102)	(0.095)	(0.297)
R^2	0.797	0.913	0.950	0.963	0.955	0.157
N (obs)	108	108	108	108	108	108

Table 6: The Effect of Investor Attention on the Effective Spread

This table reports the effects of investor attention on the effective spread. The dependent variable is the natural logarithm of the dollar-weighted effective spread scaled by midquote. The dependent variable is negatively related to liquidity. Results are identical when using Share-weighted effective spread. Control variables for every regression include size, book to market, past profitability, Amihud's illiquidity, volatility, volume, year, and firm dummies. Standard errors are clustered at the firm level to control for unobserved time-invariant firm-level heterogeneity. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

Variables	Model (1)	Model (2)	Model (3)	Model (4)
SVI original	-0.0019***			
	(0.0003)			
SVI improved		-0.0019***		
		(0.0002)		
SVI difference			0.0012***	0.0079***
			(0.0004)	(0.0008)
SVI difference*Q2 size				-0.0091***
				(0.0010)
SVI difference*Q3 size				-0.0116***
				(0.0010)
SVI difference*Q4 size				-0.0062***
				(0.0013)
Size	-0.5407***	-0.5394***	-0.5454***	-0.5383***
	(0.0064)	(0.0065)	(0.0065)	(0.0065)
Book to market	-0.0063	-0.0029	-0.0066	-0.0095
	(0.0111)	(0.0117)	(0.0120)	(0.0117)
R_{m-1}	-0.1890***	-0.1889***	-0.1909***	-0.1895***
	(0.0068)	(0.0068)	(0.0069)	(0.0070)
$R_{[m-3,m-2]}$	-0.1902***	-0.1907***	-0.1910***	-0.1832***
	(0.0087)	(0.0088)	(0.0088)	(0.0087)
$R_{[m-6,m-4]}$	-0.0036	-0.0038	-0.0025	0.0015
	(0.0084)	(0.0084)	(0.0085)	(0.0083)
$R_{[m-12,m-7]}$	0.0116	0.0114	0.0117	0.0128*
	(0.0076)	(0.0075)	(0.0075)	(0.0078)
Illiquidity	0.0004**	0.0004**	0.0004**	0.0036**
	(0.0002)	(0.0002)	(0.0002)	(0.0015)
Constant	-1.9956***	-2.0069***	-2.0553***	-2.1215***
	(0.2099)	(0.2116)	(0.2050)	(0.1994)
Industry fixed effects	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES
R^2	0.7706	0.7717	0.7698	0.7724
Number of observations	267,583	266,475	266,475	265,899

Table 7: The Effect of Unsophisticated Investors' Attention on the Effective Spread, ATET Approach

This table reports the Average Treatment Effect on Treated (ATET), measuring the impact of the increased attention of unsophisticated investors on the effective spread. We use the distribution of the SVI difference, where we assign monthly values above the 80th percentile as the treatment group (attention driven by the unsophisticated investors) and values below the 20th percentile serve a control group. We require exact matching on year, month, and Fama-French 48 industry classification, while the approximate coordinate for matching is the firm size. The outcome variable is the Dollar-weighted effective spread scaled by midquote. Using Share-weighted effective spread scaled by midquote yields to identical results. In each column, we report the ATET conducted as an effect of the high attention of unsophisticated investors. The standard errors of the ATET (in parentheses) are computed with the robust option (at least two suitable matches for each treated). Below is the balance summary of the mean difference and variance ratio between the corresponding treated and control groups. Panel B shows the density plot outlining matching quality. ***, **, and * denote statistical significance on 1, 5, and 10% significance levels.

Panel A – Results of ATET analysis

	Mean effective spread
	(1)
High attention unsophisticated (ATET)	0.146***
(std. error)	(0.007)
p-value	<0.001
Number of treated	55,706
Number of observations	116,686
Balance summary	
mean difference (size)	-0.002
variance ratio (size)	1.031

Panel B – Density plot of matching quality

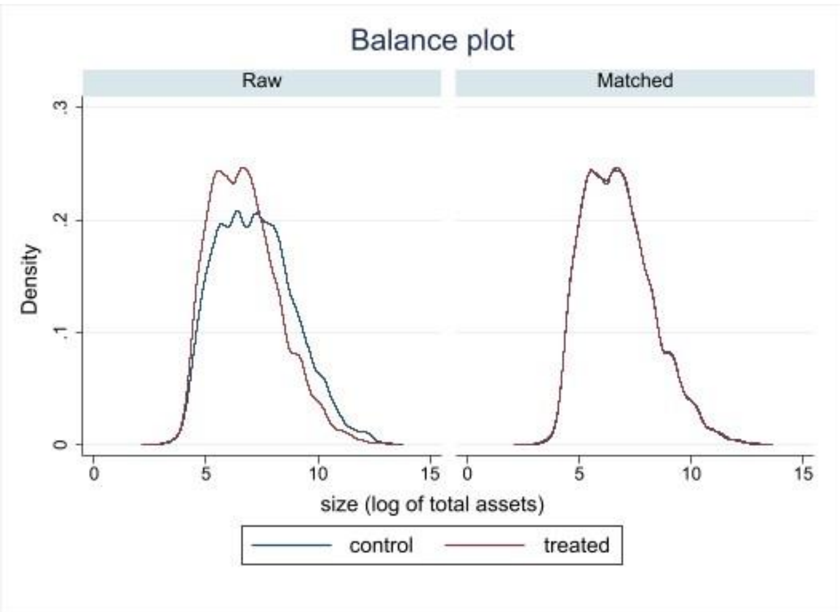

Table 8: The Complexity of Earnings Reporting and Investor Attention

This table reports the effects of investor attention on earnings transparency and complexity. We use three measures to capture earnings complexity: 1) the % change in the gross file size of the financial statement pdf file from one earnings to the next; 2) the % change in the net file size of the financial statement pdf file from one earnings to the next; 3) the ratio of complex words in the given earnings to total words, as defined by Loughran and McDonald (2024), reported in %. The gross file size includes pictures and infographics, while the net file size only includes text. The data is in quarterly format, where we compute the max SVI for the quarter preceding the quarterly earnings.
Control variables for every regression include Size, Book to market, past profitability, Amihud's illiquidity, Volatility, Volume, and year and fiscal quarter dummies. Standard errors are clustered at the firm level to control for unobserved time-invariant firm-level heterogeneity. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

	10-K/10-Q File size Gross % change		**10-K/10-Q File size Net % change**		**10-K/10-Q Complexity**	
Variables	**Model (1)**	**Model (2)**	**Model (3)**	**Model (4)**	**Model (5)**	**Model (6)**
SVI improved (3m max)	-0.027		0.075***		0.000***	
	(0.038)		(0.013)		(0.000)	
SVI difference (3m max)		0.281		-0.062*		-0.001***
		(0.310)		(0.036)		(0.000)
Size	-5.513***	-5.292***	-0.697***	-0.418*	0.060***	0.061***
	(1.427)	(1.572)	(0.245)	(0.238)	(0.001)	(0.001)
Book to market	3.152*	3.226*	-1.370**	-1.216*	0.080***	0.080***
	(1.770)	(1.823)	(0.654)	(0.651)	(0.004)	(0.004)
R_{m-1}	-9.327	-9.295	4.418	4.451	0.023**	0.023**
	(10.062)	(10.038)	(4.603)	(4.601)	(0.011)	(0.011)
$R_{[m-3,m-2]}$	-22.363***	-22.722***	-1.151	-1.022	-0.021***	-0.020***
	(5.539)	(5.602)	(3.227)	(3.225)	(0.007)	(0.007)
$R_{[m-6,m-4]}$	3.765	3.521	2.083	2.066	-0.011**	-0.010**
	(4.270)	(4.374)	(2.186)	(2.189)	(0.004)	(0.004)
$R_{[m-12,m-7]}$	14.174***	14.214***	0.554	0.507	-0.003	-0.003
	(5.037)	(5.052)	(1.018)	(1.018)	(0.002)	(0.002)
Illiquidity	-0.003	-0.001	-0.003	-0.006	-0.000*	-0.000*
	(0.036)	(0.036)	(0.020)	(0.020)	(0.000)	(0.000)
Volatility	0.019	-0.012	0.876***	0.955***	-0.000	0.000
	(1.245)	(1.276)	(0.234)	(0.234)	(0.001)	(0.001)
Volume	-0.000	-0.000	-0.000***	-0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	72.490***	68.346***	-27.748***	-26.408***	0.036***	0.047***
	(13.6140)	(11.2055)	(5.6678)	(5.6731)	(0.0129)	(0.0129)
Fiscal quarter fixed effects	YES	YES	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES	YES	YES
R^2	0.020	0.020	0.307	0.307	0.170	0.170
Number of observations	70,465	70,465	70,465	70,465	71,108	71,108

Table 9: The Complexity of Earnings Reporting and Investor Attention, ATET Approach

This table reports the Average Treatment Effect on Treated (ATET), measuring the impact of the increased attention of unsophisticated investors on various indicators of firm complexity. We first calculate the max SVI difference for the quarter preceding the quarterly earnings and assign monthly values above the 80th percentile as the treatment group (attention driven by the unsophisticated investors), and values below the 20th percentile serve a control group. We require exact matching on year and fiscal quarter, while the approximate coordinate for matching is the firm size. We use three outcome variables to capture earnings complexity: 1) the % change in the gross file size of the financial statement pdf file from one earnings to the next; 2) the % change in the net file size of the financial statement pdf file from one earnings to the next; 3) the ratio of complex words in the given earnings to total words, as defined by Loughran and McDonald (2024), reported in %. In each column, we report the ATET conducted as an effect of the high attention of unsophisticated investors. The standard errors of the ATET (in parentheses) are computed with the robust option (at least two suitable matches for each treated). Below is the balance summary of the mean difference and variance ratio between the corresponding treated and control groups. Panel B shows the density plot outlining matching quality. ***, **, and * denote statistical significance on 1, 5, and 10% significance levels.

Panel A – Results of ATET analysis

	10-K/10-Q File size Gross % change	**10-K/10-Q File size Net % change**	**10-K/10-Q Complexity**
	(1)	(2)	(3)
High attention unsophisticated (ATET)	6.575	-4.962***	-0.021***
(std. error)	(4.434)	(1.511)	(0.003)
p-value	0.138	0.001	<0.001
Number of treated	15,270	15,270	15,270
Number of observations	33,104	33,104	33,104
Balance summary			
mean difference (size)	-0.001	-0.001	-0.001
variance ratio (size)	1.004	1.004	1.004

Panel B – Density plot of matching quality

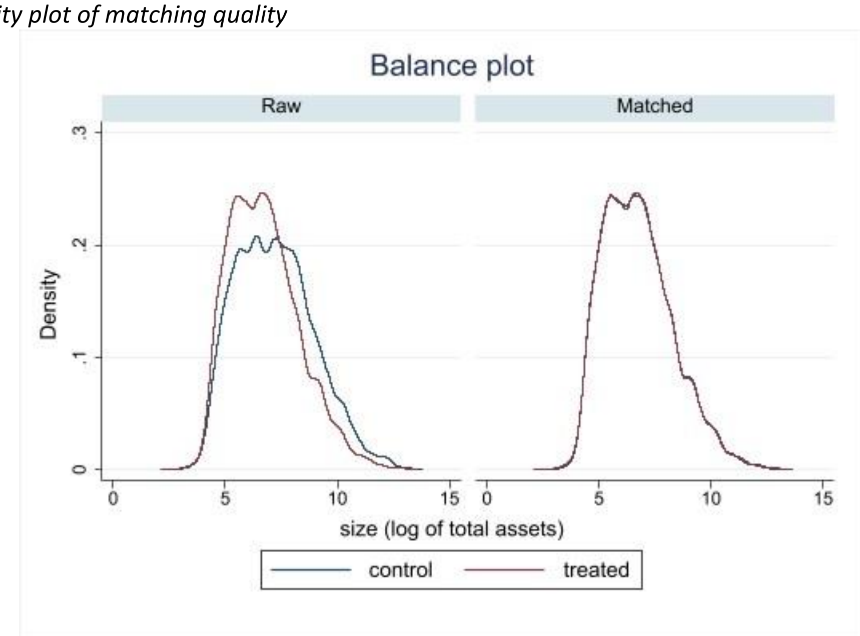

Table 10: Robustness Analysis, Placebo Tests

This table reports the Average Treatment Effect on Treated (ATET), measuring the impact of the randomly assigned increased search by unsophisticated investors. We use a generator of the pseudo-random numbers from the uniform (0,1) distribution. The Placebo test #1 randomly assigned low and high differences in searching algorithms (linked with unsophisticated investors) for each firm and month. The high and low probability was set to 20%, consistently with the ATET analysis. Before we conducted similar matching procedures as in Table 4 and Table 9, we again excluded about 60% of firms in the middle. Placebo test #2 assigns (with the same probability) the high and low in two subsequent periods. We use the same set of covariates to evaluate the placebo treatment and require the exact matching as in Table 4 and Table 7 (year, month, and Fama-French 48 industry classification); we also consider FF 12 industry classification for robustness.
The first two columns, (1)-(2), repeat the analysis with placebo assignments for the cost of capital; the last two columns, (3)-(4), correspond to the variable dollar-weighted effective spread scaled by misquote.
In each column, we report the placebo test conducted as the mean effect on treated (ATET), the difference between similar firms in high and low (placebo unsophisticated vs. the rest). We conducted 100 replications. The standard errors of the mean ATET (in parentheses) are computed with the robust option (at least two suitable matches for each treated). The p-value uses the normal approximation for the t-ratio representing mean ATE over the standard error of the mean ATET. Details of each simulation's results are available in the Internet appendix.
The balance summary provides the range of matched sample differences between mean and variance ratios for treated and control samples for all 100 simulations. Means differences close to 0 and variance ratios close to 1 indicated a very good fit for matched samples.

	Implied Cost of Capital		**Dollar-weighted Effective Spread (/midpoint)**	
	(1), Test 1	(2), Test 2	(3), Test 1	(4), Test 2
Test value (ATET)	0.0062	0.0090	-0.0022	-0.0032
(std. error)	(0.0224)	(0.0271)	(0.0069)	(0.0083)
p-value	0.390	0.370	0.627	0.651
Number of simulations	100	100	100	100
Placebo probability high/low	0.20	0.20	0.20	0.20
Balance summary				
Min mean difference (size)	0.007	0.010	0.005	0.007
Max mean difference (size)	0.012	0.017	0.008	0.013
Mean variance ratio (size)	1.065	1.083	1.053	1.073
Max variance ratio (size)	1.081	1.118	1.066	1.093

Appendix A: Variable Descriptions

Variable	*Description*
Dependent Variables	
Abnormal retail volume	To capture retail volume, we use the Barber et al. (2023) improvement on the Boehmer et al. (2021) algorithm, which uses the Lee and Ready (1991) quote midpoint signing method. Abnormal retail volume is the percentage change of retail volume to the average retail volume for the stock. The average retail volume is calculated by taking the average retail volume for the stock over the past three months. Data source: TAQ
Dollar-weighted effective spread scaled by midquote	The natural logarithm of the dollar-weighted effective spread scaled by the midquote was calculated using Holden and Jacobsen's (2014) methodology. The effective spread is calculated at daily frequency and then we take the average during the given calendar month. Source: TAQ
Implied cost of capital (ICC)	Implied cost of capital (ICC) calculated using approach of Hou et al. (2012) using the residual-income valuation model of Gebhardt et al. (2001): $P_{i,t} = B_{i,t} + \sum_{\tau=1}^{\infty} \frac{E_t[EPS_{i,t+\tau}] - r_e E_t[B_{i,t+\tau}]}{(1+r_e)^{\tau}}$, where $P_{i,t}$ is the current stock price, $EPS_{i,t+\tau}$ is the forecast of earnings per share in year $t+\tau$, and $B_{i,t+\tau}$ is the book value per share. The ICC, defined as r_e, is specific to firm *i* and time *t*. It is calculated at monthly frequency, where the stock price is equal to closing price for the given month. For a full description of the methodology, we refer to Lee et al. (2021) and online appendix of Pastor et al. (2022).
Retail volume scaled	To capture retail volume, we use the Barber et al. (2023) improvement on the Boehmer et al. (2021) algorithm, which uses the Lee and Ready (1991) quote midpoint signing method. We then define Retail volume scaled as the retail volume for the month scaled by the total volume for the month. Data source: TAQ
10-K/10-Q File size Gross % change	Calculated as a percentage change of Gross File Size from previous quarter to the current quarter. Reported in percentage. Source: Loughran and McDonald, https://sraf.nd.edu/sec-edgar-data/lm_10x_summaries/
10-K/10-Q File size Net % change	Calculated as a percentage change of Net File Size from previous quarter to the current quarter. Reported in percentage. Source: Loughran and McDonald, https://sraf.nd.edu/sec-edgar-data/lm_10x_summaries/
10-K/10-Q Complexity	Calculated as a ratio of complex words to total number of words in the statement. Reported in percentage. Source: Loughran and McDonald, https://sraf.nd.edu/sec-edgar-data/lm_10x_summaries/
Measures of Investor Attention	
SVI original	Search Value Index available through Google Trends. Index has a range from 0 to 100 and is scaled by the maximum value in the series. SVI original was collected in December 2019.

SVI improved	Search Value Index available through Google Trends. Index has a range from 0 to 100 and is scaled by the maximum value in the series. SVI improved was collected in December 2022.
SVI difference	Difference between SVI original and SVI improved
Firm Control Variables	
Book-to-market ratio	The book-to-market ratio is defined as book equity divided by market equity. Data sources: CRSP and Compustat.
Firm size	Firm size is the natural logarithm of the market value of equity. Data sources: CRSP and Compustat.
Illiquidity	Illiquidity is the sum of the absolute values of daily returns divided by the daily volume for the year, multiplied by 10^6. Defined by Amihud (2002). Data sources: CRSP and Compustat.
Past profitability	The group of variables $\mathrm{R}_{\mathrm{m}-1}$, $\mathrm{R}_{[\mathrm{m}-3,\mathrm{m}-2]}$, $\mathrm{R}_{[\mathrm{m}-6,\mathrm{m}-4]}$, and $\mathrm{R}_{[\mathrm{m}-12,\mathrm{m}-6]}$, which stand for returns over the last month, months 3 to 2, 6 to 4, and 12 to 6, respectively. Defined by Brennan et al. (2012). Data sources: CRSP and Compustat.
Volatility	Standard deviation of excess daily returns for the given month. Source: CRSP
Volume	Total share volume. Source: CRSP

Appendix B: Summary Statistics

	N	Mean	SD	P25	Median	P75
Measures of investors' attention						
SVI original	288,793	41.261	25.124	19.000	39.000	61.000
SVI improved	287,636	37.818	26.474	14.000	35.000	59.000
SVI difference	287,636	3.488	9.789	-1.000	1.000	6.000
Dependent variables						
Retail volume scaled	285,634	5.994	4.983	2.884	4.263	7.171
Abnormal retail volume	285,456	0.133	1.807	-0.294	-0.064	0.260
Implied cost of capital	218,882	1.363	1.021	0.400	1.400	2.000
Dollar-weighted scaled by midquote	288,725	-3.341	0.980	-4.120	-3.474	-2.754
10-K/10-Q File size Gross % change	73,249	25.617	80.595	-15.751	4.727	37.328
10-K/10-Q File size Net % change	73,249	22.665	84.075	-20.350	4.075	26.727
10-K/10-Q Complexity	74,869	0.492	0.236	0.314	0.442	0.627
Firm Characteristics						
Size	284,308	7.035	1.692	5.717	6.876	8.129
Book to market	274,091	0.670	19.765	0.278	0.511	0.817
Illiquidity	288,723	0.396	40.756	0.000	0.002	0.015
Volatility	288,688	2.347	1.796	1.354	1.943	2.857
R_{m-1}	288,709	1.012	0.127	0.952	1.009	1.066
$R_{[m-3,m-2]}$	288,536	1.025	0.179	0.937	1.019	1.101
$R_{[m-6,m-4]}$	287,373	1.041	0.223	0.930	1.030	1.135
$R_{[m-12,m-6]}$	282,125	1.111	0.459	0.922	1.068	1.228

Appendix C: Characterization of Unsophisticated Investors

This table reports the tests of persistence of unsophisticated investor attention. The dependent variable is the SVI difference, while the main variables of interest are lagged observations of SVI difference by one and two months, respectively. We report results with increasing fixed effects. Standard errors are clustered at the firm level to control for unobserved time-invariant firm-level heterogeneity. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

Variables	Model (1)	Model (2)	Model (3)	Model (4)
SVI difference (t-1)	0.3959***	0.3954***	0.1847***	0.1838***
	(0.0038)	(0.0038)	(0.0055)	(0.0055)
SVI difference (t-2)	0.3499***	0.3494***	0.1462***	0.1453***
	(0.0036)	(0.0036)	(0.0052)	(0.0051)
Constant	0.8852***	0.8491***	2.3165***	2.3519***
	(0.0245)	(0.0377)	(0.0341)	(0.0692)
Firm fixed effects	NO	YES	NO	YES
Year fixed effects	NO	NO	YES	YES
R^2	0.4505	0.4508	0.5370	0.5374
Number of observations	284,262	284,262	284,262	284,262