Retail Investor Attention and Mutual Fund Performance: Evidence from EDGAR Log Files

Hao Ding*

April 4, 2025

Abstract

I develop a measure of retail investor attention to mutual funds, *Total Views*, by distinguishing between retail and sophisticated investors' access to fund shareholder reports (N-CSR) via EDGAR. *Total Views* positively predicts retail fund flows and performance, with a 0.27% rise in future flows and a 0.02% improvement in alpha. An equal-weighted high-minus-low portfolio based on abnormal *Total Views* yields positive returns. *Total Views* strengthen the flow–performance relationship. Investor attention on reports from outperforming funds helps attract additional inflows but does not cause more outflows for underperforming funds. Further analyses show that fund shareholder reports offer valuable, non-time-sensitive information throughout the year.

Keywords: Retail Investor Attention, EDGAR, Mutual Fund, Shareholder Reports

JEL Classification: G11, G12, G14

*Hao Ding is at Oxford-Man Institute of Quantitative Finance, University of Oxford, hao.ding@eng.ox.ac.uk.
I thank Constantinos Antoniou and Roman Kozhan for their guidance and support. I also thank John Thanassoulis, Tyler Shumway and conference participants at 2024 FMA Annual Meeting, 2025 Swiss Society for Financial Market Research (SGF) Conference for their valuable comments and feedback.

Investor attention has long been recognised as a significant factor in financial markets. The availability and accessibility of mutual fund information can shape investment decisions, drive fund flows, and impact future performance. However, measuring retail investor attention to mutual funds is challenging because these funds are normally associated with limited Google search volume and social media activity—metrics commonly used to gauge retail investor interest in the stock market. Developing an effective proxy for measuring retail investor attention to mutual funds is imperative. Detailed internet traffic logs from the U.S. Securities and Exchange Commission's (SEC) Electronic Data Gathering, Analysis and Retrieval system (EDGAR), the central repository for investment company filings (including semi-annual shareholder reports N-CSR forms for mutual funds), enable capturing investor attention by tracking filing views from individual investors.

In this paper, I introduce *Total Views*, a measure representing the total number of times each mutual fund shareholder report is accessed through EDGAR by retail investors. By leveraging this measure as a proxy for investor attention, I show that *Total Views* is positively related to future mutual fund flows and performance. Similar to how YouTubers are compensated based on the number of views their videos receive, mutual funds benefit from the attention paid to their disclosures. Understanding this relationship provides valuable insights into retail investor behaviour and the impact of information acquisition on investment decisions within the mutual fund industry.

The SEC's EDGAR system has been a focal point in numerous studies as a means of understanding investor behaviour through the lens of information acquisition. For example, studies like F.W. Li and Sun (2022) show that an abnormal number of IP addresses searching for firms' financial statements strongly predicts future stock returns and firm fundamentals, suggesting that investors' costly information acquisition activities reveal their private expectations of firm value. Drake, Roulstone, and Thornock (2015) demonstrate that investor attention, measured through the number of downloads of financial statements, correlates significantly with the speed at which information is incorporated into stock prices. Their research primarily focuses on 10-K filings and highlights the role of firm-specific events, such as earnings announcements, in driving investor attention to these reports. While investor attention proxies such as trading volume (Barber and Odean, 2007), business press coverage (Rees, Sharp, and Twedt, 2015), and Google searches (Da, Engelberg, and Gao, 2011) have shown efficacy in predicting stock price movements, EDGAR activity provides a unique perspective on financial information acquisition. Drake, Quinn, and Thornock (2017) extend the research by providing a demographic analysis of EDGAR usage, showing key determinants of how and where EDGAR resources are utilised. Previous literature illustrates that EDGAR traffic data is valuable in analysing information acquisition and investment behaviour.

Additionally, recent studies document EDGAR's widespread use among various financial market participants, including financial analysts (Gibbons, Iliev, and Kalodimos, 2021), the Federal Reserve (E.X. Li et al., 2023), hedge funds (Crane, Crotty, and Umar, 2022), and institutional investors (T. Chen, Dong, and Lin, 2020) (Drake, Johnson, et al., 2020). The majority of the literature on EDGAR traffic concentrates on firm-level disclosures and stock price movements in response to investor attention garnered through EDGAR downloads of filings such as 10-K forms. While the existing literature provides robust methodologies for measuring investor attention and linking it to firm-level outcomes, there is a noticeable gap when it comes to mutual funds. Despite the extensive research on EDGAR usage and investor attention, studies focusing on mutual fund disclosures are scarce.

Compared to public companies, mutual funds have different disclosure requirements, such as longer reporting intervals (semi-annually) and different delivery requirements under rule 30e-3 of the Investment Company Act of 1940. Mutual fund families have the option to transmit shareholder reports by making these reports accessible at a website address provided in a notice to investors. Additionally, traditional attention metrics like the Google Search Volume Index (SVI) (Da, Engelberg, and Gao, 2011) are less effective in the mutual fund context, as mutual funds are not as frequently searched online as publicly traded firms. It is very common for mutual fund companies to have a zero SVI for extended periods. Given the lack of investor attention metrics like Google SVI or trading volume, which are both not always applicable to every mutual fund[1], a new measure of investor attention is needed. This paper proposes the use of *Total Views*, representing the cumulative view count of mutual fund shareholder reports on EDGAR. This measure offers several advantages: unlike Google SVI, which shows minimal activity for mutual funds, *Total Views* directly measures the attention specific to mutual fund disclosures. It captures a wider array of investor interactions, including those who may not rely on search engines for their information needs.

The semi-annual shareholder reports (N-CSR, N-CSRS) are the primary textual filings for mutual funds, similar to annual and quarterly filings (10-K, 10-Q) for firms. The other regular fund filing, the mutual fund prospectus (N-1A), is filed only on an annual basis and is considerably shorter. Shareholder reports include updated information about the funds and are intended to help retail investors assess and monitor their fund investments on an ongoing basis. They are prepared separately for each fund or, if the fund has multiple share classes, for each

[1] Google SVI for ETFs is available (C.-C. Lee, M.-P. Chen, and C.-C. Lee, 2021)

share class of a fund. The reports cover information about expenses, performance, statistical tables, and graphical representations of the fund. They provide thorough information about the fund and serve as the primary information source for investors. Funds must file Form N-CSR with the SEC on EDGAR within 10 days of disseminating annual and semi-annual reports to shareholders. They must be distributed to fund investors in hard copy or digitally. Shareholder reports are available on EDGAR, fund company websites, Bloomberg, etc. However, viewing of reports from methods other than EDGAR cannot be measured. This paper examines the EDGAR log file datasets, which provide information on internet traffic logs for EDGAR filings through SEC.gov. Each access to the filing is recorded, making it possible to measure the views as a proxy for investor attention.

Mutual funds are a key investment product for US households. Retail investors (i.e., households) held the vast majority (88%) of the $22.1 trillion in US mutual fund net assets at year-end 2022, according to the 2023 Investment Company Fact Book. Moreover, more than half of US households owned mutual funds, according to the annual survey of mutual fund ownership. However, retail investors do not usually have access to information such as monthly holding positions and fund characteristics, as institutional investors do, because these require significant search costs. As a result, shareholder reports are often the first and sometimes only source of information for retail investors to make investment decisions. Institutional investors also do not necessarily need to go to EDGAR to download the reports, as these are also archived on the Bloomberg Terminal, which retail investors rarely have access to.

The EDGAR log data from 2003 to 2017 provide detailed logs about each viewing, including the IP address of the viewer, date and time of the viewing, and file name. This enables the calculation of view counts for each filing and allows the classification of bot crawlers and human viewers, as well as retail and sophisticated (institutional) investors. This paper focuses on views of web forms in HTML format, as they are in a user-friendly format for human investors using browsers to access filings on the SEC website. Loughran and McDonald (2017) suggest that the magnitude of non-robot investor requests was surprisingly low before 2012. The lack of annual report requests suggests that investors generally were not conducting fundamental research on stocks. However, Cong, Du, and Vasarhelyi (2018) show that investors access machine-readable formats at a higher rate than traditional formats like HTML or PDF after 2012, especially for small companies, suggesting that the usage of XBRL files enhances transparency and accessibility, thereby potentially reducing information asymmetry. G. Chen and Zhou (2019) also support this finding, showing that systematic requests for EDGAR filings increase significantly after the adoption of XBRL. The increase in systematic requests is more pronounced for firms with lower information accessibility and accounting comparability. As more machines and bots take advantage of EDGAR filings, it is crucial to be able to distinguish human viewings. Ryans (2017) takes a more nuanced approach by examining methodologies for counting human views in EDGAR log files, separating them from bot views to get a clearer picture of actual investor demand for SEC filings. His work underscores the importance of accurately measuring human activity on EDGAR to draw reliable conclusions about investor behaviour.

This objective of this paper is to relate mutual fund investors' attention to fund flows and performance. Based on the robot screening procedure of Drake, Roulstone, and Thornock (2015), I add a file extension-based filter to distinguish between retail and sophisticated investors. The rationale is that retail investors generally do not have access to bots, so eliminating views from bot-related IP addresses helps filter for retail investor behaviour. This method differs from that of Drake, Johnson, et al. (2020), as it does not require identifying IP address ownership.

Cross-sectional regressions of *Total Views* on fund flows indicate that total retail views positively affect fund flows by 0.27%; however, this effect loses statistical significance when institutional funds are included, confirming that *Total Views* represent attention exclusively from retail investors. This result aligns with findings in the stock market, where Chi and Shanthikumar (2018) show that retail investor trading—both buying and selling—is significantly related to EDGAR searches for 10-K and 10-Q filings, more so than to Google searches. Retail investors' trading direction is influenced by the accounting information they read in the filings. Chi and Shanthikumar (2018) also find that the positive relationship between retail trading and EDGAR searches is strongest for the most easily readable 10-K and 10-Q filings; therefore, I control for report readability in the regression model. Finally, they find that retail investor trading predicts higher returns on days with heavier EDGAR searches, indicating that more research leads to better profits. Similarly, I find that *Total Views* positively predict fund performance, with a regression coefficient of 0.02%, regardless of whether institutional funds are included.

After demonstrating that retail investor attention can predict future fund flows and performance, I explore the asset pricing implications of *Total Views* by constructing high-minus-low hedge portfolios based on abnormal *Total Views* deciles. This strategy proves profitable, yielding a 0.18% risk-adjusted return for retail funds, and 0.14% when including institutional funds. However, the profitability is significant only in equal-weighted portfolios, not in value-weighted portfolios, suggesting that the alpha is concentrated in smaller funds. Finally, I examine the impact of *Total Views* on the flow-performance relationship. I find that *Total Views* enhance the flow–performance relationship. Investor attention to reports from outperforming funds helps attract additional inflows but does not lead to more outflows for underperforming funds.

I use an interaction term with a dummy variable to demonstrate that both annual and semi-annual reports have similar predictive power for fund flows and performance, suggesting that retail investor attention to shareholder reports is equally valuable regardless of the type of report. Furthermore, the impact of *Total Views* on future fund performance and flows remains consistent throughout the year, with no significant difference between the first and second quarters post-filing. This indicates that the retail attention to mutual fund shareholder reports is not time-sensitive. The conclusion that retail investor attention leads to higher fund inflows is strengthened by robustness tests that account for non-linearity and use marketing (12b-1) fees as an alternative retail fund classification. Furthermore, Oster's sensitivity test confirms the findings are unlikely to be skewed by omitted variable bias.

This paper aims to fill the existing gap by examining the relationship between investor attention to mutual fund disclosures (as measured by *Total Views*) and subsequent fund flows and performance. By doing so, it contributes to the literature in two key ways: it sheds light on EDGAR mutual fund disclosure traffic, a relatively underexplored area in mutual fund investor attention studies, and introduces a novel measure of retail investor attention for mutual funds that offers more precise insights into investor behaviour and fund performance. The introduction of the *Total Views* measure enhances our understanding of how investors engage with mutual fund information and its impact on fund flows and performance.

The remainder of this paper is organised as follows. Section 1 introduces a dynamic portfolio choice model with ivestor attention. Section 2 defines the *Total Views* measure and develops the hypotheses. Section 3 describes the data, presents summary statistics. Section 4 tests the relationship between *Total Views* and future fund flows and performance. Section 5 presents additional analyses that compare the relationships across various groups. Section 5 concludes the paper.

1 A Dynamic Portfolio Choice Model with Investor Attention

1.1 Information Structure

We consider a continuous-time overlapping generations framework where each period t investors decide how to allocate their wealth across mutual funds. For simplicity, focus on a representative fund j in period t. The timeline within period t is as follows:

Public Signal (Past Returns): At the start of period t, investors observe fund j's publicly available past performance $r_{j,t-1}$ (e.g. net return in period $t-1$). This past return serves as a public signal about the fund's true quality. We model the fund's (unobserved) skill or true alpha α_j as a random variable that influences returns.

$$r_{j,t-1} = \alpha_j + \varepsilon_{j,t-1} \tag{1}$$

where $\varepsilon_{j,t-1}$ is a noise term with zero mean (e.g. $\varepsilon \sim N(0,\sigma^2)$). All investors observe $r_{j,t-1}$, which updates their beliefs about α_j.

Attention Choice (Costly Private Signal): After seeing $r_{j,t-1}$, each investor decides whether to acquire a costly private signal about fund j's true alpha. Specifically, investor i can obtain an informative signal by, for example, reading the fund's shareholder report on EDGAR. Acquiring this signal incurs a utility cost of c. If the investor pays c, they receive a private signal

$$s_{j,t} = \alpha_j + \eta_{i,j,t} \tag{2}$$

where $\eta_{i,j,t}$ is a noise term independent of $\varepsilon_{j,t-1}$ with variance $\mathrm{Var}(\eta_{i,j,t}) = \frac{1}{c}$. Thus, higher cost c corresponds to a more precise signal. Investors who choose not to pay c receive no additional information beyond the public $r_{j,t-1}$.

All investors share common prior beliefs about α_j (updated after observing $r_{j,t-1}$). Let the prior (after public signal) for α_j be normal with mean $\mu_{j,t-1}$ and precision (inverse variance) $\tau^{(\text{pub})}_{j,t-1}$. If investor i acquires the private signal $s_{j,t}$, they update their belief about α_j using Bayes' rule. Because both prior and signal are normal, the posterior belief about α_j remains normal with updated precision equal to the sum of prior and signal precisions. In particular, since the private signal has variance $1/c$ (precision c), the posterior precision for an attentive investor is:

$$\tau^{(\text{post})}_{j,t} = \tau^{(\text{pub})}_{j,t-1} + c \tag{3}$$

Equivalently, the posterior variance is $\dfrac{1}{\tau^{(\text{pub})}_{j,t-1} + c}$, which is lower than the prior variance, reflecting the value of the additional information. Investors who do not acquire $s_{j,t}$ retain the prior belief based on $r_{j,t-1}$ alone (precision $\tau^{(\text{pub})}_{j,t-1}$).

1.1.1 Portfolio Allocation

After updating beliefs, investors choose how much to allocate to fund j (and other assets) to maximise expected utility. We assume each investor has constant elasticity of substitution (CES) utility over final wealth contributions from each fund. Specifically, let $x_{i,j,t}$ be the amount investor i allocates to fund j in period t, which yields a random gross return $R_{j,t+1} = 1 + r_{j,t}$ next period. The investor's end-of-period wealth is $W_{i,t+1} = \sum_j x_{i,j,t} R_{j,t+1}$ (summing over all funds j and possibly a risk-free asset with return R_f). The utility function is given by a CES aggregator of wealth across funds (with elasticity of substitution θ and risk aversion parameter γ):

$$U_i = \mathbb{E}_t \left[\frac{1}{1-\gamma} \Big(\sum_j \omega_j \left(x_{i,j,t} R_{j,t+1} \right)^{\frac{\theta-1}{\theta}} \Big)^{\frac{\theta(1-\gamma)}{\theta-1}} \right] \tag{4}$$

where ω_j are preference weights (or 1 if symmetric). This formulation nests constant relative risk aversion (CRRA) when $\theta = 1$ (no substitution across assets, only risk aversion across states) and allows $\theta > 1$ for some substitution effect among funds. In words, the investor derives utility from a diversified portfolio of funds, with diminishing marginal utility for extremely concentrated bets. Higher expected returns (alphas) on fund j increase the optimal allocation $x_{i,j,t}$, but with decreasing marginal benefit due to risk aversion and the CES preferences.

For portfolio optimisation, investor i chooses $x_{i,j,t}$ optimally, given beliefs about α_j. In equilibrium, with normally distributed returns and a CRRA setup, the optimal allocation to fund j is increasing in the investor's posterior mean estimate of α_j and decreasing in return variance. For instance, if $\theta = 1$ (CRRA utility), the first-order condition equates expected marginal utility per dollar across funds. With a risk-free asset (return R_f), an investor will allocate to fund j if the expected excess return $E_t[r_{j,t}|\text{info}]$ is positive and sufficiently high. Intuitively, more precise information (higher posterior precision) allows the investor to invest more confidently in high-α funds. Conversely, without paying attention, an investor relies on the coarser public information and may invest less in fund j due to higher uncertainty about its alpha.

1.2 Equilibrium Attention Allocation

Denote by $A_{j,t}$ the equilibrium measure of investors paying attention to fund j in period t. In equilibrium, $A_{j,t}$ is determined endogenously as the fixed point of investors' optimal attention decisions. We assume a unit mass of investors, so $A_{j,t} \in [0, 1]$ can be interpreted as the fraction of investors who acquire the private signal for fund j. Investors decide to pay attention if the expected benefit (in terms of higher expected utility from better-informed portfolio choice) exceeds the cost c. Formally, let $\Delta U_{i,j,t}$ denote the increment in expected utility for investor i from acquiring the private signal about fund j (relative to remaining uninformed about j). An investor i will pay attention to fund j if and only if:

$$\Delta U_{i,j,t} = \mathbb{E}_t\big[U_i \mid \text{info } s_{j,t}\big] - \mathbb{E}_t\big[U_i \mid \text{no info on } j\big] \geq c \tag{5}$$

This is the attention choice condition. Equivalently, one can derive a marginal utility condition: the investor compares the expected marginal utility gain from improved portfolio allocation to the marginal cost c. For risk-averse investors, $\Delta U_{i,j,t}$ is increasing in the posterior variance reduction achieved by the signal, and in the dispersion between fund j's potential outcomes (since more uncertainty means more value from resolving that uncertainty).

In equilibrium, We determine $A_{j,t}$ such that investors are making optimal decisions given their expectations and no one has an incentive to deviate. If all investors are ex ante identical (with the same cost c and preferences), typically the attention choice will be corner solution (either all or none pay attention) unless indifference. To obtain an interior equilibrium $0 < A_{j,t} < 1$, We introduce a slight heterogeneity in attention costs or benefits. For example, suppose investors have heterogeneous cost c_i or slightly different prior beliefs. Then there will be a cut-off type c^* such that all investors with $c_i < c^*$ pay attention and those with $c_i > c^*$ do not. The equilibrium attention fraction $A_{j,t}$ is the measure of investors with $c_i \leq c^*$. At the margin, the investor with cost c^* is indifferent, so $\Delta U_{j,t}(c^*) = c^*$. This yields a fixed-point equation for $A_{j,t}$ because $\Delta U_{j,t}$ itself may depend on how many others are informed:

- **Diminishing Returns Feedback:** If more investors pay attention and identify fund j as high-α, more capital will flow into j, potentially diminishing its future returns (e.g. due to limited investment capacity or price impact). We capture this by assuming that if a fraction $A_{j,t}$ become informed and invest in j, the fund's effective alpha in the next period is a decreasing function of inflows. For instance, suppose fund j's net alpha in period t given inflow $Flow_{j,t}$ is $\tilde{\alpha}_{j,t} = \alpha_j - \kappa F_{j,t}$ for some $\kappa > 0$. If informed investors collectively contribute to a larger $Flow_{j,t}$ when $A_{j,t}$ is higher, then each informed investor's expected gain from information is lower when $A_{j,t}$ is large (since the alpha advantage erodes). Conversely, if few investors pay attention, an attentive investor can exploit j's alpha with less competition. This negative feedback ensures an interior equilibrium: $A_{j,t}$ adjusts so that the marginal benefit of attention equals cost c.

- **Equilibrium Fixed Point:** Let $B_{j,t}(A)$ denote the expected benefit ΔU for a representative investor from becoming informed given that a fraction A of investors are also informed (this function encapsulates the feedback effects above). In equilibrium, $A_{j,t}$ satisfies:

$$A^*_{j,t} = \mathbb{P}\{\Delta U_{i,j,t} \geq c\} = \mathbb{P}\{c_i \leq B_{j,t}(A_{j,t})\} \tag{6}$$

where the second expression integrates over the distribution of c_i. If c_i is uniformly distributed on $[0, \bar{c}]$, for example, then $A_{j,t} = \frac{B_{j,t}(A_{j,t})}{\bar{c}}$. The fixed-point (self-consistency) condition is $B_{j,t}(A_{j,t}) = c^*$ for the marginal type. Intuitively, $A_{j,t}$ is high when past returns $r_{j,t-1}$ are strong (yielding a high expected α_j and thus large $B_{j,t}$), but $A_{j,t}$ cannot grow arbitrarily large because as more investors become informed, the incremental value of being informed falls.

1.2.1 Impact on Fund Flows

Let $Flow_{j,t}$ denote the net flow into fund j during period t (e.g. percentage increase in assets under management due to investor trading). We posit that fund flows are increasing in the level of investor attention because attentive investors act on their information. Specifically, higher $A_{j,t}$ means more investors closely monitoring fund j and rebalancing their portfolios based on fund j's prospects:

- If fund j had strong past performance, many investors pay attention ($A_{j,t}$ high). Most of these attentive investors receive signals that (on average) confirm high α_j. They will increase their allocations to fund j, resulting in buy orders/inflows. In addition, uninformed investors also tend to invest somewhat in response to the public good performance. Thus, good past performance coupled with high attention yields large inflows.

- If fund j had poor past performance, few investors pay attention ($A_{j,t}$ low). Uninformed investors might withdraw some capital based purely on the poor public signal, but many investors may simply ignore the fund (inertia or lack of interest). Only a small fraction become informed; those who do will likely confirm low α_j and may reduce their holdings, but since $A_{j,t}$ is small, these outflows are limited. Furthermore, because We assume no short-selling of mutual funds, new investors cannot bet against the fund's poor performance — they can at most choose not to invest. Thus, poor performance with low attention leads to relatively modest outflows.

Combining these effects, the model produces an asymmetric flow–performance curve. Let $Flow_{j,t} = f(r_{j,t-1})$ represent the equilibrium mapping from last period's return to current net flow. Because $A_{j,t}$ amplifies the reaction to performance on the upside more than on the downside, $f(r)$ will be convex. In other words, the elasticity of flows to past returns is higher when returns are high than when returns are low. We can sketch this formally. Suppose baseline (uninformed) flow response to performance is $f_{\text{base}}(r)$ (which could be relatively linear or only mildly convex), and the incremental flow due to attention is an increasing function of both A and the information those attentive investors glean. We can write, in reduced form:

$$Flow_{j,t} = f_{\text{base}}(r_{j,t-1}) + g(A_{j,t}(r_{j,t-1}),\, r_{j,t-1}) \tag{7}$$

- $f_{\text{base}}(r_{j,t-1})$ captures the baseline (uninformed) reaction of investors to the fund's past performance $r_{j,t-1}$. This component typically generates moderate and symmetric responses.
- The additional component $g(A_{j,t}, r_{j,t-1})$ captures how investor attention positively amplifies investment flows. This is where the explicit positive relationship between attention and flows is represented:
$$\frac{\partial g(A_{j,t}, r_{j,t-1})}{\partial A_{j,t}} > 0 \tag{8}$$

This derivative explicitly states that, holding past returns constant, increased investor attention leads directly to larger inflows. Investors who pay attention ($A_{j,t}$ fraction) receive more precise private signals about the fund's true skill (α_j). With better information, these attentive investors are more confident in their assessment of the fund's quality and will invest more aggressively if their private signals are favourable. As a result, greater attention translates directly into increased aggregate buying activity, leading to positive net inflows. In contrast, when attention is low, flows depend primarily on uninformed investors who react less aggressively and more cautiously to public information alone.

The equilibrium attention allocation condition explicitly captures the positive feedback loop between attention and flows through the marginal benefit condition:

- Investors pay attention if the incremental expected utility gain from informed investing exceeds the attention cost c. Specifically, as more investors pay attention and invest informed, the expected returns from informed investing (net of risk) become clearer and thus incentivise further attention.
- Positive feedback loop: the reason investors initially pay attention is because higher attention, through better information, leads to higher expected returns and thus higher flows.

Thus, the core explicit link between investor attention $A_{j,t}$ and investment flows is the positively sloped partial derivative.

1.3 Attention and the Flow–Performance Relationship

A key implication of the model is an asymmetric flow–performance relationship driven by attention. We show that the attention allocation $A_{j,t}$ is increasing and convex in past returns $r_{j,t-1}$, and that this translates into stronger inflows after good performance than outflows after poor performance for fund j.

1.3.1 Attention as an Increasing Convex Function of Past Returns

Because higher $r_{j,t-1}$ raises investors' expected α_j, the incentive to acquire information grows with past performance. In other words, $a(r)$ satisfies $a'(r) > 0$: funds that performed better last period attract more attentive investors. Moreover, the marginal effect of performance on attention itself increases with r — i.e. $a''(r) > 0$, implying convexity. Intuitively, an improvement in performance from "good" to "great" triggers a larger jump in attention than an improvement from "poor" to "mediocre." This convexity arises because when performance is exceptionally good, investors expect a high α_j and thus a very large benefit to being informed (making it well worth paying cost c), whereas when performance is poor, many investors already suspect α_j is low and the potential benefit of additional information is minimal (so few pay attention). This leads to a convex flow–performance relationship.

1.3.2 Impact on Flow-performance Relationship

For good performance, $A_{j,t}$ is large and g is positive (many buy), whereas for poor performance $A_{j,t}$ is near zero and g may be slightly negative (some sell) but limited in magnitude. The first derivative $\frac{\partial F}{\partial r}$ is thus larger for high r than for low r, and the second derivative $\frac{\partial^2 F}{\partial r^2} > 0$ across much of the range, reflecting convexity. In practical terms, top-performing funds experience disproportionate inflows, while bottom-performing funds do not symmetrically experience such extreme outflows.

This asymmetry is consistent with observed investor behaviour. Investors enthusiastically chase past winners (high r), especially after doing further research (attention), but they are more hesitant to penalise past losers to the same extent. In our model, this is endogenously determined by the costly attention equilibrium: attention gravitates towards likely winners, fueling further inflows to those funds, whereas losers are largely neglected, so their outflows are muted. Additionally, since new investors cannot short a bad fund (they simply avoid it) and existing investors may exhibit inertia or behavioural biases (e.g. the disposition effect, holding onto losers), outflows in reaction to poor performance are limited.

In summary, the equilibrium $A_{j,t}$ serves as the mediating variable linking past performance to future flows. The model yields testable implications: (i) $A_{j,t}$ rises non-linearly with $r_{j,t-1}$, and (ii) the flow–performance slope is steeper for positive performance than for negative performance. These implications align with empirical evidence of a convex flow–performance relationship in mutual funds and highlight the role of investor attention in amplifying the success of winning funds while softening the punishment of underperforming funds.

In this model, the positive relationship between investor attention $A_{j,t}$ and investment flows is explicitly embedded in the equilibrium concept through the function $g(A_{j,t}, r_{j,t-1})$. Specifically, attention affects flows through the following mechanism:

1.4 Dynamic Evolution

A dynamic model unfolds over multiple periods, allowing for learning and path-dependence. A fund that genuinely has high skill α_j will, on average, deliver positive excess returns repeatedly. Each period of outperformance attracts a new wave of retail attention (EDGAR views) and fresh inflows. Thus, skilful funds experience a cumulative attention effect: growing awareness leads to growing assets under management. Importantly, because attention is limited, this process is gradual – not all investors recognise the manager's skill immediately. This contrasts with a frictionless rational market where investors would instantly allocate so much capital that future excess returns are driven to zero (Berk and Green (2004)). Instead, here, even a truly skilled fund may continue to earn positive risk-adjusted returns for a while because not everyone has caught on at once. In other words, limited attention slows down the flow of capital, allowing persistent outperformance in equilibrium for high-attention funds. This aligns with empirical findings that an abnormal attention surge predicts a modest but positive improvement in future fund alpha. On the flip side, a fund with poor true skill ($\alpha_j < 0$) will tend to underperform repeatedly, but if retail investors avoid paying attention to bad news, outflows will be sluggish. Some inattentive investors might remain in the fund despite subpar performance (perhaps hoping for mean-reversion or simply not noticing the underperformance), which can allow the fund's assets to decay slowly rather than experiencing an efficient run. Thus, underperforming funds do not lose assets as quickly as they would if all investors were fully attentive. Over time, extremely poor funds may eventually attrition (investors who do monitor will exit, and new investors are unlikely to enter), but the key friction is that the response is delayed and dampened. This captures asymmetric attention—investors disproportionately pay attention to out-performers and under react to under-performers.

2 Mechanism Explanation

In this framework, investor attention affects investment decisions and fund returns by mediating information flow. The economic mechanism can be summarised as follows: when a fund performs well, it grabs the limited attention of retail investors, leading more of them to acquire information (e.g. read the shareholder report) and subsequently invest, thereby increasing fund flows. This attention-driven inflow of capital is not mere noise; it is directed by informed belief-updating about the fund's skill. If the outperformance indeed reflected managerial skill (and not pure luck), these new inflows are "smart money (Zheng 1999)" that chase a fund with genuinely higher expected returns. The result is that funds which attract more attention continue to do well as they accumulate assets – a form of reinforcing feedback between past success, investor attention, and future success. Empirically, this mechanism is evidenced by the finding that abnormal EDGAR views positively predict both higher subsequent fund flows and a slight improvement in future risk-adjusted returns. The model provides intuition for this: high attention is a costly signal that investors are seeking out and processing favourable information about the fund, which they would only do if they anticipated superior future performance. In effect, attention-rich funds are those that investors (on average) have identified as likely winners, and thus these funds enjoy continued inflows and performance momentum. This is consistent with studies in other markets showing that retail attention proxies can predict returns – for example, increases in Google search volumes foreshadow higher stock prices. A rejection of the null hypothesis that attention doesn't matter (H1/H2 in the paper) confirms that indeed attention contains information about future flows and performance. Our model's mechanism formalises this insight: information acquisition (viewing reports) is valuable to investors and thus is undertaken more for funds where it can pay off – those likely to deliver good future returns.

Another key feature explained by the model is the asymmetric reaction to positive vs. negative performance. Retail investors tend to strongly chase recent winners but only weakly punish losers. In the framework, this asymmetry arises because of both rational and behavioural factors. Among funds in which investors are not currently invested, if a fund achieves top-decile returns, investors update their beliefs and conclude that it might be a hidden gem.

Consequently, they flock to evaluate the fund and invest if convinced, given the substantial potential upside of investing in a truly skilled fund. In contrast, if a fund performs poorly, many investors simply ignore it, as the inability to short sell provides little incentive to expend effort investigating an underperforming fund. For investors who are already invested, behavioural biases reinforce this pattern: they exhibit inattention to bad news and display disposition effects. Empirical evidence shows investors are less likely to even log in to view their portfolio after losses (Sicherman et. al 2016), which translates in our context to fewer EDGAR report views for underperforming funds. Consequently, past winners get a lot of eyes and new money, whereas past losers do not trigger proportional outflows. This explains why investor attention strengthens the flow–performance relationship primarily on the upside: when attention is high, a given positive return result leads to an even larger inflow than it otherwise would (attention magnifies the sensitivity to good performance), yet attention does little to exacerbate outflows for poor performance (since attention wanes when returns are negative).

The overall effect is a convex flow-performance curve an outcome long documented in mutual fund markets (Sirri and Tufano 1998). Investors buy winners much more than they sell losers, which attention-based model endogenously reproduces. Notably, this pattern can arise even if all investors are Bayesian and rational about updating, provided they face information-gathering costs and constraints. Thus, a behavioural-style equilibrium (performance chasing with inertia) is achieved from a friction that is arguably rational (limited attention) but captures a realistic cognitive limitation. In equilibrium, the gradual diffusion of information through costly attention also implies that skilled funds do not immediately lose their superior returns. Rather, high-attention funds can earn positive abnormal returns for some time because not all capital that could exploit their skill arrives at once. This offers a formal mechanism for why an attention-based long–short strategy (long funds with high retail attention, short those with low attention) might deliver positive alpha, as found in the data. By concentrating on funds that investors have identified (via costly attention) as likely winners, such a strategy piggybacks on the collective information of attentive investors. In sum, the model shows that investor attention is a conduit through which information influences fund investments and prices: it determines who learns what about which funds, thereby affecting capital allocation and subsequent performance outcomes. This mechanism is distinct from, but complementary to, classical drivers like past returns or fund characteristics – attention links those drivers to actual investor behaviour.

2.1 The *Total Views* Measure

The *Total Views* measure represents manual downloads from retail investors. To capture manual downloads, I exclude all IP addresses that use bots to download files from the EDGAR system on the same day. The raw EDGAR log files contain records for each filing download request. I first match the requests for mutual fund shareholder reports using the unique filing identifier 'accession'. These requests include access to both '.htm' and '.txt' extension files. Following Loughran and McDonald (2017), I mark IP addresses that access '.txt' files as bot requests. The HTM files represent web forms associated with most filings and are typically read by individuals, while TXT files contain the complete form, including all documents and HTML markup, and are typically used for machine reading of the documents.

After identifying the file types, I follow Drake, Roulstone, and Thornock (2015) and mark IP addresses as bot requests if they show more than 1,000 downloads in a single day or more than five downloads per minute. The idea is to identify bot patterns characterised by high-frequency and high-volume requests that are unlikely to be made by humans. Additionally, I mark all IP addresses with an empty user agent. A user agent is a string of information that a browser sends to a website's server when a connection is made, identifying the browser and its capabilities. An empty user agent strongly indicates that the traffic source does not want to be identified and is often associated with sophisticated investors, such as hedge funds, scraping EDGAR filings while attempting to conceal their identity.

Bot IPs are identified by:

1. Downloads of '.txt' file; or
2. Downloading more than 1,000 files in one day; or
3. Downloading more than 5 items per minute; or
4. Downloads with an empty user agent.

I exclude all requests associated with any IP address marked as a bot by at least one of these four rules. Then, I aggregate the filtered request records at the file-quarter level to obtain the total number of EDGAR requests for a single filing during a three-month period.[2] The Fisher–Pearson coefficient of skewness is 2.37, indicating that the distribution of the total number of EDGAR requests is heavily right-tailed. Following the procedure in Drake, Roulstone, and Thornock (2015), I take the natural logarithm of one plus the count to construct the *Total Views* measure.

[2]Details about filtering EDGAR logs are introduced in Section 2.1.

2.2 Hypothesis Development

Investor attention plays a pivotal role in financial markets, influencing investment decisions and asset prices. Traditional theories of information acquisition suggest that investors allocate their limited attention based on the perceived benefits of acquiring additional information (Grossman and Stiglitz, 1980) (Diamond and Verrecchia, 1981). These theories have been substantiated in various contexts, such as stock markets, where increased attention to corporate disclosures has been linked to price movements and trading behaviour (Da, Engelberg, and Gao, 2011) (Barber, Huang, and Odean, 2016). However, the impact of retail investor attention on mutual funds, particularly through the lens of EDGAR view data, remains underexplored. In the context of mutual funds, the decision to access shareholder reports through the EDGAR system can be interpreted as an indicator of retail investors' interest in the fund's future performance.

According to Ivkovic and Weisbenner (2009), mutual fund flows are significantly influenced by investor behaviour, particularly in response to past performance and fund characteristics. They suggest that investors' actions, such as buying or redeeming fund shares, are based on their expectations of future returns. This aligns with the notion that the effort to acquire information reflects investors' private expectations and, thus, could have predictive value for investment behaviour. If retail investors seek information with the intention of making informed investment decisions, a positive correlation between the viewing data and future fund flows would be expected.

Sicherman et al. (2015) investigate the role of investor attention by examining online account logins, demonstrating that investors selectively pay attention to their portfolios based on market conditions. They show that investors are less likely to log into their accounts after market declines, highlighting how attention is related to sentiment and not purely to trading opportunities. This concept is relevant for mutual funds, as retail investors may similarly avoid reviewing shareholder reports during periods of poor performance, leading to a potential disconnect between investor attention and investment decisions. Therefore, retail investor attention in the context of mutual funds could reflect either an informed decision-making process or a psychological response to market conditions. Kaniel and Parham (2017) find that retail investors tend to exhibit herding behaviour, where their attention and trading activities are influenced by previous trading patterns and information. The first hypothesis, building on these insights, proposes that:

H1: Retail investor attention is not associated with investment decisions.

This hypothesis posits that retail investor attention, as measured by the *Total Views* of mutual fund shareholder reports, does not significantly influence subsequent investment decisions, as reflected in future fund flows. A rejection of this null hypothesis would indicate that retail investor attention plays a role in guiding investment choices, suggesting that the information in shareholder reports is indeed being utilised by retail investors to make investment decisions.

In addition to influencing investment decisions, retail investor attention might also indicate expectations regarding a fund's future performance. Studies such as those by Hsieh, Chan, and Wang (2020) and Yung and Nafar (2017) have shown that proxies for retail investor attention, such as the Google Search Volume Index (SVI), are associated with future performance predictions in stock markets. If investor attention indeed reflects their performance expectations, we would expect a significant relationship between viewing data and future fund performance. This hypothesis is stated as follows:

H2: Retail investor attention is not associated with future fund performance.

This hypothesis suggests that there is no significant relationship between the *Total Views* of mutual fund shareholder reports and future fund performance. By analysing the relationship between viewing data and subsequent performance metrics, I aim to determine whether investor attention can predict future performance. A rejection of this null hypothesis would imply that retail investor attention is an informative signal for future fund performance, indicating that the information acquisition efforts of these retail investors are justified by subsequent performance outcomes. Testing these two hypotheses will provide insights into the role of retail investor attention in the mutual fund industry and its impact on fund flows and performance.

3 Data

3.1 EDGAR log dataset

To measure total human views, I obtained the EDGAR server logs from 1 January 2006 to 30 June 2017. These log files record each 'click', capturing a user's request to open a filing file. There are two EDGAR traffic log file datasets.[3] Only the first dataset, which covers from 1 January 2003 through 30 June 2017, provides detailed information about user IP addresses,

[3] The data is available for download at `https://www.sec.gov/data-research/sec-markets-data/edgar-log-file-data-sets`

Apache log agents, Apache access codes, and crawler identifiers. Thus, the dataset used in this paper spans up to 30 June 2017. Although the EDGAR database archives all mutual fund filings since 2003, it was only in 2006 that the SEC introduced the unique share class-level (Class ID) identifier. Mutual fund shareholder reports (N-CSR forms) are filed at the investment company level, and a company may use the same filing for several funds within the fund family. The Class ID is crucial for uniquely identifying individual funds. Therefore, I follow the most commonly used time span for the analysis of mutual fund shareholder reports, beginning from 1 January 2006. Since the report was first filed in Q1 2006, *Total Views* in the sample start from Q2 2006.

First, I use the Master Index of EDGAR Dissemination Feed[4] to select mutual fund shareholder report filing forms (N-CSR forms for annual reports and N-CSRS forms for semi-annual reports). Then, I merge the EDGAR traffic log file dataset with the master index of mutual fund shareholder reports using the unique filing identifier 'accession'. For each daily EDGAR log file, I perform the following procedure:[5]

1. Keep requests if $code = 200$. Keep only the successful delivery of the requested document by the EDGAR server.

2. Keep requests if $idx = 0$. Remove index page observations, as index pages serve as the initial page when viewers search for a fund, directing them to a link to the actual filing.

3. Keep requests if $crawler = 0$. Drop identified web crawlers, such as search engines like Google and the Internet Archive Wayback Machine.

4. Make list A of IP addresses for requests where the last three characters of $doc = 'txt'$. '.txt' files containing HTML markup are typically used for machine reading.

5. Make list B of high-volume bot requests consisting of IP addresses with more than 1,000 file downloads in a single day.

6. Make list C of high-frequency bot requests consisting of IP addresses with more than 5 downloads per minute.

7. Make list D of IP addresses with an empty user agent for requests where $noagent = 1$. Some of these IP addresses are associated with financial institutions such as hedge funds.[6]

[4]Data available at `https://www.sec.gov/Archives/edgar/full-index/`

[5]Data files and sample Python code are available from the author.

[6]IP address registration information is available from Historical Whois Data (WhoWas) by the American Registry for Internet Numbers (ARIN) at `https://www.arin.net/reference/research/whowas/`

8. Calculate the *Total Views* per filing as the number of log records for the filing where the IP address is not in list A, B, C, or D.[7]

9. *Total Views* is recorded for the following two quarters, as shareholder reports are filed every six months.

The file size of the filing, word count of the form, form readability (Coleman–Liau Index, Flesch Reading Ease Score), and sentiment (Loughran-McDonald negative and uncertainty word proportions) measures are gathered from the CRSP database. Filings containing fewer than 100 words are dropped to remove documents with potential technical issues. These features are merged with the EDGAR filing data using the 'accession' identifier, the unique SEC document accession number associated with the requested document.

3.2 Fund Characteristics

Mutual fund characteristics, including the quarterly reported *Expense Ratio* and *Turnover Ratio*, month-end *Total Net Assets*, *Net Monthly Returns*, and *Fund Age* (in months), are collected from the CRSP Survivor-Bias-Free database. They are merged with shareholder reports data first on 'crsp_cik_map', which matches 'CIK' and 'CRSP_FUNDNO', the unique identifier created by CRSP for each mutual fund. A 'CRSP_FUNDNO' is assigned to each fund or fund class combination so that class-specific information can be captured in the database. In situations where a fund may have multiple classes, a separate 'CRSP_FUNDNO' is assigned to each class. If a 'CIK' is mapped to multiple 'CRSP_FUNDNO', tickers in shareholder reports and mutual fund data are used to match the filing to the individual fund level. Index funds and ETFs are excluded according to the 'index_fund_flag' and 'et_flag' variables in CRSP, as is standard practice in mutual fund literature.

However, the EDGAR viewing data are only available at the filing file level, meaning viewing counts are calculated for each file and cannot be attributed to a specific fund class. Therefore, I aggregate the fund characteristics to the filing level by combining multiple classes (if any) into a single 'giant' fund following Jordan and Riley (2015), where the fund characteristics are weighted by Total Net Assets (TNA).[8]

[7] Only users who download fund shareholder reports using bots will be excluded. If a user employs bots to download other filings, such as firm-level 10-K or 10-Q forms, but does not use bots to download fund reports, the records associated with that user are retained in the dataset as manual views.

[8] Cao, Yang, and Zhang (2023) find that the returns of funds under the same CIK are highly correlated. I find similarly high correlations for fund characteristics within the same CIK.

3.3 Summary Statistics

Table 1 presents descriptive statistics for 101,662 filing-quarter observations from Q2 2006 to Q2 2017. There are, on average, 1,081 fund families (identified by CIK) in the sample. *Total Views* is winsorised at [0, 99%], while the remaining variables are winsorised at [1%, 99%]. The average number of views per filing is 186, ranging from 0 to 644 views. This is significantly lower than the average of 7,834 views for 10-K filings of firms (Drake, Roulstone, and Thornock, 2015). This difference is not surprising, as reflected by the Google Search Index, which often shows a score of zero for many mutual funds. Shareholder reports are lengthy, averaging more than 43,000 words, with an average file size of 3.62 MB. These reports are of moderate difficulty to read, consistent with the fact that around 39% of mutual fund retail investors hold a college or postgraduate degree (Institute, 2023). On average, the shareholder reports contain 1% negative and uncertainty words according to the Loughran-McDonald dictionary.

The fund families in the sample have an average net asset of $572 million, but the net asset of the median fund is $154 million, with an interquartile range of $52 million to $419 million. This suggests that there is a considerable breadth of both large and small funds in the sample. With respect to other control variables, the average fund in the sample has a quarterly turnover of 85.25 per cent, consistent with the fact that they are active funds, an expense ratio of 1.05 per cent, a net return of 0.47 per cent, and an age of 171 months. The fund performance metrics show that fund families in the sample receive average inflows of 3.63 per cent, with an average alpha of 0.27 per cent.

3.4 Correlations

Figure 1 intuitively shows the pairwise correlations using the Pearson correlation coefficient along with the corresponding p-values. I find that *Total Views* is positively correlated with the length and size of the filing (i.e., the Pearson correlation coefficients are 0.34 and 0.35, respectively, and both are statistically significant at the 1% level). Longer reports likely take more time to read, and viewers may not finish in one go, necessitating multiple accesses. Additionally, longer reports are harder to read, as noted by Loughran and McDonald (2014), which increases the likelihood of multiple accesses. Lastly, longer reports may contain more information, making it worthwhile for viewers to revisit them.

Moreover, report length is highly correlated with report size, with a Pearson correlation coefficient of 0.8, significant at the 1% level. Fund net total assets and age are also correlated. The expense ratio is positively correlated with the turnover ratio and negatively correlated with

Table 1: **Summary Statistics**

	Mean	Std	Min	Q1	Median	Q3	Max
Panel A: Shareholder report features							
$Total\ Views$	186	84	0	110	179	240	644
$Word\ Count\ (1,000)$	43.33	55.23	6.65	14.92	22.84	45.10	329.99
$File\ Size\ (MBs)$	3.6160	5.2767	0.0997	0.7929	1.6423	3.9843	30.8833
$Flesch\ Reading\ Ease$	32.8802	6.0292	15.7364	29.4475	32.7033	36.3387	50.1386
$Coleman - Liau$	16.7963	2.5360	14.1650	15.4822	16.0787	17.0084	29.2820
$LM\ Negative$	0.0106	0.0027	0.0052	0.0087	0.0104	0.0123	0.0185
$LM\ Unvertainty$	0.0099	0.0027	0.0047	0.008	0.0095	0.0115	0.0183
Panel B: Fund characteristics							
$Total\ Net\ Asset\ (mil\$)$	571.85	1285.98	1.77	51.86	154.31	419.39	8513.95
$Expense\ Ratio$	0.0105	0.0044	0.001	0.0077	0.0101	0.0130	0.0244
$Turnover\ Ratio$	0.8525	1.0183	0.0300	0.3015	0.5487	0.9757	6.3924
$Net\ Returns$	0.0047	0.0221	-0.0745	-0.0025	0.0061	0.0155	0.0618
$Age\ (months)$	171	118	11	86	151	229	673
Panel C: Fund flows and performance							
$Fund\ Flow\%$	3.6331	10.6222	-9.3145	-0.0359	0.0282	2.5147	66.7764
$Alpha\%$	0.2721	3.2193	-11.2546	-1.1264	0.1911	2.0461	8.5261

N = 101,662

Table 1 reports summary statistics for shareholder reports, fund characteristics, fund flows and performance. Statistics are based on quarterly data. Fund age is rounded up to months. Word counts are reported in thousands of words, file size in MBs, and total net assets in million dollars. *Total Views* is winsorised at [0, 99%], while the remaining variables are winsorised at [1%, 99%]. The sample period spans from Q2 2006 to Q2 2017, comprising 101,662 observations from 1,663 investment companies (CIKs). On average, there are 1,081 fund families each year.

net total assets.

Figure 1: **Correlation Matrix with P-value**

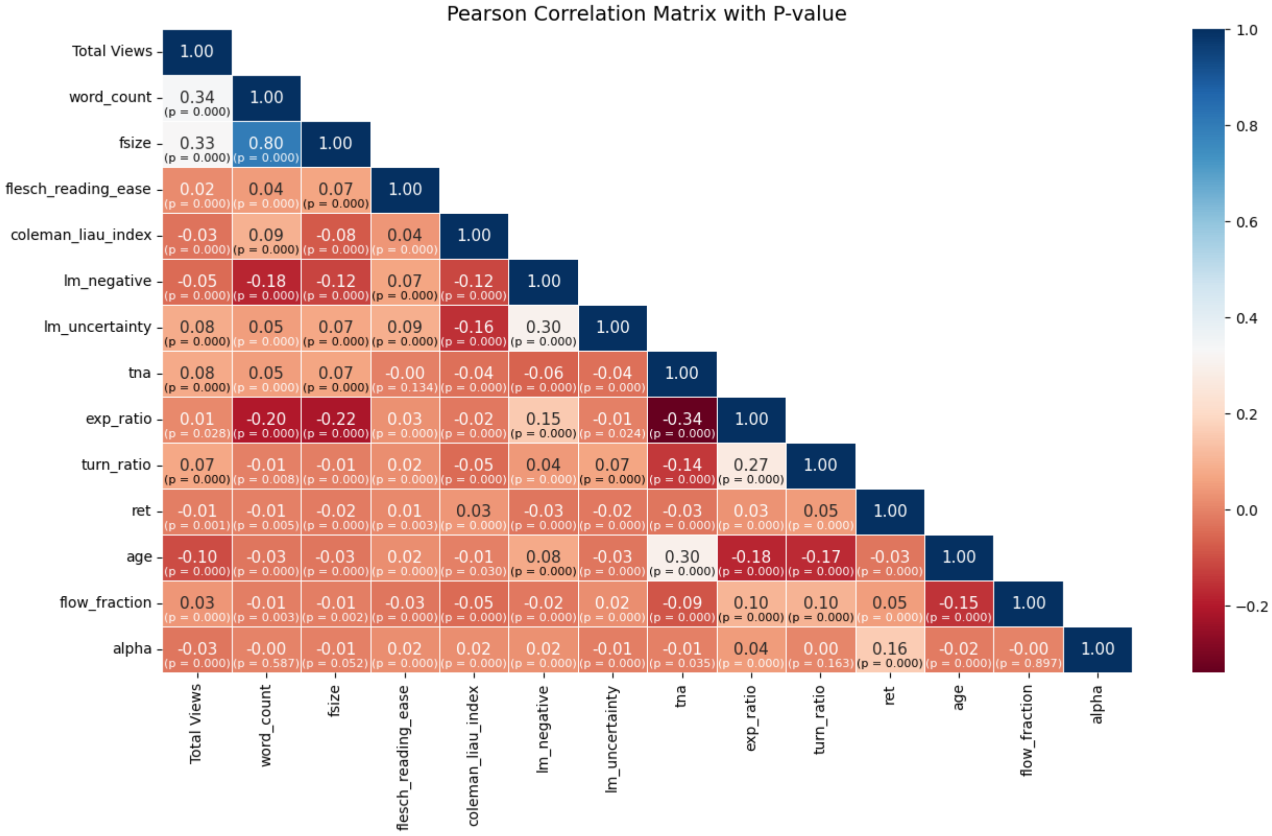

Figure 1 displays the pairwise Pearson correlation among the shareholder reports features, fund characteristics, fund flows and alpha. The blue (red) colour represents a positive (negative) correlation, with the darker shade indicating a higher correlation. The P-value is shown in brackets.

4 Main Results

4.1 Fund Flows and Performance

I first analyse the relationship between *Total Views* and investors' reactions—fund flows—and then investigate whether they consequently predict future fund returns. If investors value shareholder reports, they will react to the information conveyed by purchasing or selling the corresponding mutual fund, thus influencing fund flows. Attracting capital flows has always been an important goal for mutual funds.

Hypotheses H1 and H2 examine the ability of *Total Views* to influence fund flows and predict fund performance. I test these hypotheses in the following regression analysis, where fund flows and performance are regressed on *Total Views*. The regressions control for various factors, including shareholder report features such as word count, file size, readability and sentiment measures; fund characteristics such as expense ratio (logged), turnover ratio (logged), total net assets (logged), fund age (logged), and previous returns; quarter and time fixed effects. Additionally, control for previous return rank and squared previous return rank for flow% (Cremers and Petajisto, 2009); squared total net assets (logged), and previous inflows for alpha (Hillert, Niessen-Ruenzi, and Ruenzi, 2024). I cluster standard errors by fund to account for residual serial correlation. A statistically significant b_1 would reject Hypothesis 1 (2), suggesting that the *Total Views* of shareholder reports are related to fund flows (performance).

$$\begin{aligned} Flow\ (Alpha)\ _{i,t} = a &+ \mathbf{b_1 Total\ Views}_{i,t-1} + b_2 Word\ count_{i,t-1} + b_3 File\ size_{i,t-1} \\ &+ b_4 Flesch_{i,t-1} + b_5 Coleman - Liau_{i,t-1} + b_6 LM\ negative_{i,t-1} \\ &+ b_7 LM\ uncertainty_{i,t-1} + b_8 ln(TNA)_{i,t-1} + b_9 ln(Exp\ ratio)_{i,t-1} \\ &+ b_{10} ln(Turn\ ratio)_{i,t-1} + b_{11} RET_{i,t-1} + b_{11} ln(Age)_{i,t-1} \\ &+ b_d Controls_d + \sum b_m Quarter\ FEs_m + \sum b_n Time\ FEs_n + \epsilon_{i,t} \end{aligned} \tag{9}$$

I define fund flows following Kostovetsky and Warner (2020), use CRSP monthly data of assets under management (AUM) and net returns to calculate monthly fund flows. $Flow(\%)$ in month t is defined as:

$$Flow(\$) = AUM_t - AUM_{t-1} \times (1 + Net\ Monthly\ Return_t) \tag{10}$$

Fund flows are adjusted for mergers and then standardised by dividing them by the previous period's assets under management (AUM).

$$Flow\% = \frac{Flow(\$)}{AUM_{t-1}} \tag{11}$$

I evaluate fund performance using Alpha, the risk-adjusted excess return. Alpha is defined as the difference between the fund's excess return in quarter t and the return calculated by multiplying the factor realisations by the loading from the four-factor model (Carhart, 1997). It provides a clearer picture of a fund's performance as it accounts for systematic risk, offering investors insights into the fund's ability to outperform the market after adjusting for risk.

In Table 2, I present the results for retail funds only in columns (1) and (2) and for all funds in columns (3) and (4). Retail funds are identified using the 'retail_fund' indicator from CRSP. Fund family level classification is performed by taking the mode of the indicator dummy from the class level. There are 70,021 observations (69%) associated with retail funds, which have shorter reports, averaging 39,652 words in length compared to 43,332 words for the full sample. Additionally, retail funds have a lower average *Total Views* of 179, compared to 186 for the full sample. This is expected, as retail funds primarily target retail investors, who generally possess less financial knowledge compared to professional investors. Longer reports may be more difficult for them to read.

For retail funds, column (1) shows a strong positive association between *Total Views* and future fund flows, which is statistically significant at the 5% level. The coefficient on *Total Views* is 0.27%, implying that an additional 100 views can lead to a 1.29% increase in future fund flows. In column (2), there is similar evidence of a positive association between *Total Views* and future fund performance, statistically significant at the 10% level. The coefficient on *Total Views* is 0.02%, suggesting that an additional 100 views can result in a 0.09% increase in returns.

However, for all funds, column (3) provides no evidence that *Total Views* is predictive of future fund flows. Since views from user IPs associated with bot usage—primarily advanced institutional investors—are excluded, it is not surprising that *Total Views* from retail investors do not explain future flows when institutional funds are included. In column (4), I find that *Total Views* still predict future fund performance for all funds. With a coefficient of 0.02% and significance at the 5% level, *Total Views* positively predict fund returns. Although fund flows for institutional funds cannot be estimated by retail investor attention—as retail capital flows do not constitute the majority of institutional fund flows—retail investor attention remains 'smart' and can still serve as an indicator of future fund performance.

These results support the idea that $Total\ Views$, as a proxy for retail investor attention, are useful in predicting investment decisions (flows) and fund performance (alpha). This finding is consistent with Sirri and Tufano (1998), which show that fund flows are directly related to the media attention received by funds. Therefore, hypotheses H1 and H2 are rejected. The rest of the paper uses the retail funds sample unless otherwise specified.

4.2 The High-minus-Low Abnormal *Total Views* Portfolios

I next examine whether the positive association between retail investor attention and future fund performance holds in a portfolio setting that accounts for the portfolio's exposure to known asset pricing risk factors. I follow C.M. Lee and So (2017) and orthogonalise *Total Views* to a set of control variables because it is important to examine the incremental information content of

Table 2: **Fund Flows and Performance vs. Report Views**

	(1) Retail Funds	(2)	(3) All Funds	(4)
	Flow%	Alpha	Flow%	Alpha
Total Views	0.0027**	0.0002*	0.0009	0.0002**
	[2.099]	[1.693]	[0.798]	[1.978]
Word Count	0.0043	0.0001	0.0062*	0.0001
	[1.056]	[0.029]	[1.824]	[0.151]
File Size	-0.0421	0.0026	-0.0570	0.0028
	[-0.856]	[0.668]	[-1.585]	[0.994]
Flesch Reading Ease	-0.0432**	0.0012	-0.0247	0.0001
	[-2.298]	[0.634]	[-1.551]	[0.087]
$Coleman - Liau$	-0.1663***	-0.0012	-0.1565***	-0.0026
	[-3.761]	[-0.238]	[-3.966]	[-0.579]
LM negative	-75.2569	6.5393	-69.1673	9.5847
	[-0.975]	[0.660]	[-1.135]	[1.168]
LM uncertainty	28.5888	6.4044	26.2301	1.9838
	[0.504]	[1.007]	[0.490]	[0.387]
Total Net Asset (ln)	0.0289	0.0124	-0.1562	0.0373
	[0.189]	[0.225]	[-1.244]	[0.806]
Expense Ratio (ln)	0.5036	0.2799***	1.1458**	0.2167***
	[0.968]	[5.194]	[2.254]	[5.399]
Turnover Ratio (ln)	0.5716**	-0.0476	0.4266*	-0.0546**
	[2.520]	[-1.533]	[1.848]	[-2.063]
Net Returns	19.1249***	10.9971***	14.8997***	11.3978***
	[3.496]	[9.978]	[3.044]	[12.114]
Age (months) (ln)	-3.3549***	-0.0374	-2.3383***	-0.0486
	[-6.040]	[-0.900]	[-5.391]	[-1.474]
Return Rank	0.2591		1.1495**	
	[0.385]		[2.077]	
$Return\ Rank^2$	-0.0084		-0.1637*	
	[-0.072]		[-1.769]	
$(Total\ Net\ Asset\ (ln))^2$		0.0042		0.0007
		[0.863]		[0.164]
Inflow		-0.0043*		-0.0029
		[-1.796]		[-1.553]
Reporting Quarter FEs	Y	Y	Y	Y
Time FEs	Y	Y	Y	Y
Adjusted R^2	0.101	0.600	0.064	0.601
	N = 70,021		N = 101,662	

Table 2 presents the estimated coefficients and t-statistics (in brackets) from the OLS regression of fund flow% and alpha for retail funds only and for all funds. The regressions control for report features, fund-level characteristics, and fixed effects, with standard errors clustered by fund. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. The adjusted r-squared is reported.

information acquisition beyond that found in other contemporaneously observable information signals. To do this, I regress *Total Views* on all of the common control variables included in regression (9) and obtain the residual values, which are labelled as Abnormal *Total Views*. I rank them into deciles in each quarter. Then, I estimate the alpha for equal- and value-weighted portfolios using the four-factor model (Carhart, 1997). Finally, I compare the alpha estimates of the portfolios with the highest (decile 10) and lowest (decile 1) levels of Abnormal *Total Views*. The difference in the alpha between the extreme portfolios is the hedged return to a strategy that longs funds with high Abnormal *Total Views* and shorts funds with low Abnormal *Total Views*.

Table 3 presents the four-factor alphas for retail-only and all funds. In column (1), I find a significant return for the equal-weighted hedge portfolio based on Abnormal *Total Views* for retail funds. Specifically, the hedge return is 0.18%, significant at the 1% level. In contrast, in column (2), the return for the value-weighted hedge portfolio for retail funds is statistically indistinguishable from zero. The outperformance of the equal-weighted strategy over the value-weighted strategy suggests that the predictable returns to *Total Views* are concentrated in smaller funds. This finding is consistent with similar results in the stock market, as documented by Drake, Roulstone, and Thornock (2015), where investors expect larger returns from small stocks due to the greater costs associated with information acquisition and processing, further supporting the concept of 'disclosure processing costs' as discussed in Blankespoor, deHaan, and Marinovic (2020).

I observe similar findings for the Abnormal *Total Views* hedge strategy when institutional funds are included in the sample. In column (1), the equal-weighted high-minus-low strategy based on Abnormal *Total Views* for all funds yields a positive return of 0.14%, also significant at the 1% level. However, in column (2), the return for the value-weighted high-minus-low portfolio remains insignificant.

Overall, the results of the portfolio test are consistent with the cross-sectional regression findings, supporting the existence of the predictive ability of retail investor attention for future performance in both retail-only and all funds. This underscores the value of retail investor attention.

Table 3: **Four Factor adjusted Returns of Portfolios Formed Based on Abnormal *Total Views***

	(1)	(2)	(3)	(4)
	Retail Funds		All Funds	
	Equal-Weighted%	Value-Weighted%	Equal-Weighted%	Value-Weighted%
10 (Highest)	0.39	0.36	0.34	0.32
9	0.25	0.19	0.25	0.13
8	0.29	0.18	0.27	0.18
7	0.26	0.14	0.22	0.09
6	0.30	0.21	0.25	0.17
5	0.27	0.08	0.26	0.12
4	0.32	0.40	0.29	0.28
3	0.36	0.44	0.31	0.36
2	0.38	0.49	0.33	0.33
1 (Lowest)	0.20	0.27	0.20	0.19
High-Low	0.18***	0.09	0.14***	0.13
t-stat	[3.341]	[0.665]	[2.991]	[1.209]

Table 3 presents the equal- and value-weighted portfolio alphas using portfolio deciles of Abnormal *Total Views*. The portfolios are formed each quarter based on their level of Abnormal *Total Views*, calculated as the residual from the regression of *Total Views* on report features and fund characteristics. T-statistics are provided in brackets. *, **, *** indicate statistical significance of the difference in alpha between the highest and lowest quintile portfolios at the 10%, 5%, and 1% levels, respectively.

4.3 Flow-Performance Relationship in Retail Funds

As *Total Views* is associated with future fund flows and performance, I then test whether *Total Views* can affect the relationship between fund flows and performance. Previous studies show that fund flows typically positively respond to past performance (Jonathan and Richard, 2004): outperformance attracts inflows, and underperformance causes outflows. I regress fund flows on *Total Views*, as well as the interaction of $Total\ Views \times Net\ Returns$, to test the impact of retail investor attention on the flow-performance relationship. Use excess return (Jonathan and Richard, 2004) at $t-1$ as the past performance measure.[9] The independent variables include all control variables from the previous regression (9).

Table 4 presents the regression of fund flows on *Total Views* and the interaction of *Total Views* with past returns. In column (1), I find no evidence that *Total Views* affects the flow-performance relationship, as the coefficient of the interaction term is statistically insignificant.

[9]Fund flows at time t, *Total Views* at time $t-1$.

Adding the interaction term does not change the significant positive association between *Total Views* and future flows, with a coefficient of 0.28%, significant at the 5% level.

In column (2), I introduce a past return dummy to separate funds into outperforming and underperforming funds. The outperforming dummy equals one for funds with non-negative past returns, and zero otherwise. The underperforming dummy equals one for funds with negative returns, and zero otherwise. Decomposing past performance results in a lower coefficient for *Total Views* at 0.17%, significant at the 10% level. For non-negative returns, the coefficient of the interaction term with the positive dummy is 0.12% and is statistically significant at the 10% level, indicating that investor attention further enhances fund inflows for outperforming funds, in addition to the flows related to their original performance. However, when past returns are negative, the coefficient of the interaction term with the negative dummy is 0.04% and insignificant, showing that investor attention does not contribute to further outflows for underperforming funds.

As a result, retail attention affects future fund flows by helping outperforming funds to attract additional inflows, resulting in a 0.30% attention-driven flow, only marginally higher than the overall impact regardless of performance. Attention to good performance explains nearly half of the attention-driven flow, while attention to bad performance cannot explain anything. The asymmetry reflects the fact that retail investors can only go long on funds but cannot short-sell. Retail investors can buy any funds with better performance whether they already hold them or not but can only sell funds they hold. Reading reports from outperforming funds gives retail investors more confidence and supports their decision to make additional investments. Also, the disposition effect (Weber and Camerer, 1998) and mental accounting (Thaler, 1999) could contribute to the effect.

5 Additional Analyses

The results in Section 3 show that retail investor attention positively predicts fund flows and performance. The portfolio tests further demonstrate the asset pricing implications. In this section, I explore where the predictive power of *Total Views* is strongest, investigating whether it is driven by specific disclosure forms or the timing of access.

Table 4: **Flow-Performance Relationship in Retail Funds**

	(1)		(2)	
Variable	Flow%			
Total Views	0.0028**	[2.165]	0.0017*	[1.695]
Total Views $\times$ *Net Returns*	-0.0210	[-0.834]		
Total Views $\times$ *Outperforming(+)*			0.0012*	[1.757]
Total Views $\times$ *Underperforming(−)*			0.0004	[0.336]
Word Count	0.0043	[1.056]	0.0043	[1.053]
File Size	-0.0421	[-0.855]	-0.0422	[-0.857]
Flesch Reading Ease	-0.0433**	[-2.300]	-0.0432**	[-2.299]
Coleman − Liau	-0.1660***	[-3.753]	-0.1664***	[-3.767]
LM negative	-75.3092	[-0.975]	-75.3343	[-0.976]
LM uncertainty	28.5637	[0.503]	28.5691	[0.503]
Total Net Asset (ln)	0.0287	[0.188]	0.0301	[0.196]
Expense Ratio (ln)	0.5040	[0.969]	0.5105	[0.979]
Turnover Ratio (ln)	0.5713**	[2.518]	0.573**	[2.514]
Net Returns	15.322**	[2.070]	20.9492***	[2.789]
Age (months) (ln)	-3.3544***	[-6.037]	3.3568***	[-6.027]
Return Rank	0.2600	[0.387]	0.2566	[0.380]
Return Rank^2	-0.0086	[-0.074]	-0.0081	[-0.070]
Reporting QuarterFEs	Y		Y	
Time FEs	Y		Y	
Adjusted R^2	0.101		0.101	

N = 70,021

Table 4 presents the estimated coefficients and t-statistics (in brackets) from the OLS regression of fund flow% on *Total Views* with the interaction term *Total Views* $\times$ *Net Returns* / *Total Views* $\times$ *Net Returns Dummy*. The regression control for report features, fund-level characteristics and fixed effects, with standard errors clustered by fund. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. The adjusted r-squared is reported.

5.1 Annual vs. Semi-Annual Reports

Annual reports are viewed more frequently and are more comprehensive in terms of content compared to semi-annual reports. For example, the annual report of the Vanguard PRIMECAP Fund, filed on 30 September 2021, is 29 pages long and covers six sections: Your Fund's Performance at a Glance, Advisor's Report, About Your Fund's Expenses, Performance Summary, Financial Statements, and Trustees Approve Advisory Arrangement. In contrast, the semi-annual report filed on 31 March 2021 is only 17 pages long and includes just three sections: About Your Fund's Expenses, Financial Statements, and Liquidity Risk Management.

Annual shareholder reports (N-CSR forms) have a higher average of 188 views and a larger word count of 44,072, compared to 170 views and a word count of 34,757 for semi-annual reports (N-CSRS forms). The differences in views and length between the two groups are statistically significant at the 1% level, with t-statistics of 29.195 and 24.816, respectively. The significant difference in length and view counts between annual and semi-annual reports highlights the need to separately assess the impact of *Total Views* on fund flows and performance for each type of report.

Similar to Section 3.3, I create a form dummy for N-CSR annual reports and add the interaction of *Total Views* with the form dummy to regression (9). Table 5 presents the results of the regression with the interaction term. In columns (1) and (2), where fund flows are the dependent variable, I continue to find a positive coefficient for the main effect of *Total Views*, which is higher at 0.30% and statistically significant at the 5% level. This suggests that attention to semi-annual reports has an even stronger impact on fund flows, despite the form being shorter and receiving fewer views. However, the coefficient of the interaction term is not significant, indicating that the predictive power of retail investor attention on fund flows remains strong regardless of whether the attention is on annual or semi-annual reports.

In columns (3) and (4), I find similar evidence that *Total Views* positively predict future fund performance, with a coefficient of 0.03%, significant at the 5% level. The interaction term is not significant, indicating that the relationship between *Total Views* and alpha is not significantly affected by the form type. Unlike findings in the stock market, where investor response to annual (10-K) filings is stronger than to quarterly (10-Q) filings (Griffin, 2003) (F.W. Li and Sun, 2022), both annual and semi-annual shareholder reports are equally important in the attention-versus-decision and attention-versus-reward relationships for mutual funds.

One possible explanation is that mutual fund investments are typically long-term compared to stocks, and there are no major events like earnings announcements that dramatically drive price movements. While mutual fund annual reports provide a comprehensive overview of the entire year, semi-annual reports focus on a more specific and recent time period of six months. Consequently, the information in semi-annual reports may be more relevant and timely for investors, enhancing their decision-making process. As a result, both types of reports are equally valuable for retail investors.

Table 5: **Fund Flows and Performance vs. (Semi)Annual Report Views**

	(1)	(2)	(3)	(4)
	Flow%		Alpha	
Total Views	0.0030**	[2.209]	0.0003**	[1.962]
Total Views × *Annual Dummy*	-0.0005	[-0.809]	-0.0001	[-1.090]
Word Count	0.0045	[1.077]	0.0001	[0.118]
File Size	-0.0430	[-0.868]	0.0024	[0.620]
Flesch Reading Ease	-0.0441**	[-2.293]	0.0010	[0.535]
Coleman − Liau	-0.1669***	[-3.766]	-0.0014	[-0.264]
LM negative	-71.1439	[-0.893]	7.3852	[0.708]
LM uncertainty	26.1717	[0.447]	5.9052	[0.918]
Total Net Asset (ln)	0.0276	[0.180]	0.0124	[0.225]
Expense Ratio (ln)	0.4980	[0.957]	0.2787***	[5.140]
Turnover Ratio (ln)	0.5704**	[2.520]	-0.0478	[-1.538]
Net Returns	19.1241***	[3.495]	10.9978***	[9.978]
Age (months) (ln)	-3.3549***	[-6.039]	-0.0374	[-0.900]
Return Rank	0.2597	[0.387]		
Return Rank^2	-0.0085	[-0.073]		
(*Total Net Asset (ln)*)^2			0.0042	[0.856]
Inflow			-0.0043*	[-1.797]
Reporting Quarter FEs	Y		Y	
Time FEs	Y		Y	
Adjusted R^2	0.101		0.600	

N = 70,021

Table 5 presents the estimated coefficients and t-statistics (in brackets) from the OLS regression of fund flow% and alpha for annual and semi-annual reports views. The regression control for report features, fund-level characteristics and fixed effects, with standard errors clustered by fund. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. The adjusted r-squared is reported.

5.2 Views in the First and Second Quarter Post-Filing

Lastly, I examine whether the impact of *Total Views* on future fund flows and performance is time-sensitive. Drake, Roulstone, and Thornock (2015) show that views surge around earnings announcement days. A similar pattern is observed for mutual fund disclosures. The average number of views during the quarter immediately following the filing is significantly higher—199 views compared to 178 views during the second quarter—at the 1% significance level, with a t-statistic of 20.268. If the information in shareholder reports is time-sensitive, the impact of *Total Views* during the first quarter following the filing should be greater than during the second quarter, when the information is no longer fresh. To investigate this question, I add the interaction of *Total Views* with a quarter dummy variable which takes the value one for the first quarter immediately after the filing to regression (9) to investigate the question.

Table 6 presents the results of the regression analysis with the interaction term. In columns (1) and (2), where fund flows are the dependent variable, I continue to observe a larger positive coefficient for the main effect of *Total Views*, with a coefficient of 0.32%, statistically significant at the 5% level. However, the coefficient for the interaction term is not significant. This suggests that retail investor attention has a strong impact on fund flows regardless of timing, indicating that its predictive power remains consistent over time.

In columns (3) and (4), similar evidence is found, showing that *Total Views* positively predict future fund performance, with a coefficient of 0.03%, significant at the 10% level. The interaction term is also not significant, implying that the relationship between *Total Views* and alpha is not significantly influenced by the timing of access.

Overall, the results suggest that retail attention to shareholder reports is not time-sensitive. Unlike 10-K filings, mutual fund reports remain equally valuable throughout the year, as they contain less time-sensitive information. This can be attributed to the longer investment horizon of mutual funds compared to stocks.

5.3 Robustness Test: Non-linear Effects

In this section, I extend the baseline specification by adding the squared term of $Total\ Views$ to examine potential non-linearities in the relationship between retail investor attention and fund outcomes. The results indicate that while the coefficient on the linear term ($Total\ Views$) remains positive and is approximately three times larger—significant at the 1% level—the coefficient on the squared term is negative and statistically significant at the 5% level.

This pattern suggests that high (but not excessively high) levels of attention may be bene-

Table 6: **Fund Flows and Performance vs. Quarterly Views**

	(1)	(2)	(3)	(4)
	Flow%		Alpha	
Total Views	0.0032**	[2.018]	0.0003*	[1.692]
Total Views $\times$ *Time Dummy*	-0.0007	[-1.348]	-0.0001	[-0.727]
Word Count	0.0042	[1.018]	-0.0001	[-0.001]
File Size	-0.0425	[-0.860]	0.0025	[0.662]
Flesch Reading Ease	-0.0436**	[-2.313]	0.0012	[0.618]
$Coleman - Liau$	-0.1665***	[-3.763]	-0.0012	[-0.241]
LM negative	-75.3031	[-0.976]	6.5377	[0.660]
LM uncertainty	29.7520	[0.524]	6.4817	[1.019]
Total Net Asset (ln)	0.0268	[0.175]	0.0125	[0.227]
Expense Ratio (ln)	0.4940	[0.954]	0.2792***	[5.172]
Turnover Ratio (ln)	0.5709**	[2.514]	-0.0476	[-1.535]
Net Returns	19.1074***	[3.491]	10.9971***	[9.978]
Age (months) (ln)	-3.3503***	[-6.041]	-0.0371	[-0.893]
Return Rank	0.2568	[0.383]		
$Return\ Rank^2$	-0.0081	[-0.070]		
$(Total\ Net\ Asset\ (ln))^2$			0.0042	[0.856]
Inflow			-0.0043*	[-1.798]
Reporting Quarter FEs	Y		Y	
Time FEs	Y		Y	
Adjusted R^2	0.101		0.600	

N = 70,021

Table 6 presents the estimated coefficients and t-statistics (in brackets) from the OLS regression of fund flow% and alpha for views in the first and the second quarter after filing. The regression control for report features, fund-level characteristics and fixed effects, with standard errors clustered by fund. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. The adjusted r-squared is reported.

ficial for fund flows, whereas extremely high attention could begin to dampen them. For fund performance (alpha), however, neither the linear term nor the squared term of $Total\ Views$ is statistically significant, implying that this non-linear effect is specific to flows rather than to alpha.

5.4 Robustness Test: Alternative Retail Funds Classification

The term 12b-1 fee was introduced by SEC Rule 12b-1 of the 1940 Investment Company Act. These fees are charged by the fund to cover its marketing and distribution expenses. An alternative way to classify retail funds, therefore, is to group funds that charge an annual 12b-1 fee below 25 basis points as "no-load" funds (i.e., direct-sold), and those that charge at least 25 basis points as broker-sold. Broker-sold funds typically cater to a less sophisticated retail investors (Bergstresser, Chalmers, and Tufano, 2009).

Following this alternative classification (i.e., *actual_12b1* < 0.25% as retail), I identify a similar number of retail funds—68,219 in total—compared to 70,021 using the standard CRSP indicator. Despite the difference in sample size, the relationship between Total Views and Flow% remains strong. Specifically, the coefficient remains positive (approximately 0.60%) and statistically significant at the 1% level. However, retail investor attention no longer exhibits a statistically significant relationship with future fund performance (alpha).

5.5 Robustness Check: Omitted Variable Bias

To address potential omitted variable bias, I use Oster (2019) coefficient stability approach, quantifying how strongly an unobserved factor must correlate with both the independent variable ($Total\ Views$) and dependent variable (Flows%) to nullify the estimated effect.

First, I run two regressions:

- **Baseline Model**: Including only the predictors significant in the full model Table 2 column (1).
- **Full Model**: Baseline Model plus all controls, the same as Table 2 column (1).

Second I calculate the Oster's Delta (δ) that measures how strong omitted variables must be to eliminate the resulting impact of $Total\ Views$ on fund performance.

$$\delta = \frac{\beta_{full}}{\beta_{baseline} - \beta_{full}} \times \frac{R^2_{max} - R^2_{full}}{R^2_{full} - R^2_{baseline}} \tag{12}$$

Table 7: **Robustness Test: Non-linear Effects**

	(1)	(2)	(3)	(4)
	Flow%		Alpha	
Total Views	0.0090***	[3.476]	0.0005	[1.361]
Total Views^2	-0.0001**	[-2.510]	-0.0001	[-0.802]
Word Count	0.0046	[1.124]	0.0001	[0.066]
File Size	-0.0386	[-0.786]	0.0027	[0.707]
Flesch Reading Ease	-0.0432**	[-2.297]	0.0012	[0.634]
Coleman − Liau	-0.1650***	[-3.746]	-0.0012	[-0.226]
LM negative	-74.4128	[-0.966]	6.5758	[0.664]
LM uncertainty	30.9305	[0.547]	6.5125	[1.029]
Total Net Asset (ln)	0.0343	[0.224]	0.012	[0.219]
Expense Ratio (ln)	0.5053	[0.972]	0.2801***	[5.193]
Turnover Ratio (ln)	0.5714**	[2.521]	-0.0476	[-1.532]
Net Returns	19.1444***	[3.501]	10.9984***	[9.978]
Age (months) (ln)	-3.3603***	[-6.051]	-0.0376	[-0.905]
Return Rank	0.2515	[0.374]		
Return Rank^2	-0.0071	[-0.062]		
(Total Net Asset (ln))^2			0.0043	[0.874]
Inflow			-0.0043*	[-1.801]
Reporting Quarter FEs	Y		Y	
Time FEs	Y		Y	
Adjusted R^2	0.101		0.600	

N = 70,021

Table 7 presents the estimated coefficients and t-statistics (in brackets) from the OLS regression of fund flow% on *Total Views* and the squared term to account for nonlinearity. The regression control for report features, fund-level characteristics and fixed effects, with standard errors clustered by fund. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. The adjusted r-squared is reported.

Table 8: **Robustness Test: Alternative Retail Funds Classification**

	(1)	(2)	(3)	(4)
	Flow%		Alpha	
Total Views	0.0060***	[5.146]	0.0001	[1.074]
Word Count	-0.0050	[-1.196]	-0.0001	[-0.120]
File Size	0.0207	[0.378]	0.0031	[0.753]
Flesch Reading Ease	-0.0044	[-0.231]	0.0033	[1.632]
$Coleman - Liau$	-0.1487***	[-3.201]	0.0038	[0.734]
LM negative	-91.8454	[-1.215]	6.8115	[0.678]
LM uncertainty	50.7152	[0.847]	10.6946*	[1.874]
Total Net Asset (ln)	-0.1184	[-0.687]	0.0150	[0.347]
Expense Ratio (ln)	0.0638	[0.077]	0.3378***	[4.585]
Turnover Ratio (ln)	1.0846***	[2.803]	-0.0542**	[-2.383]
Net Returns	6.6311	[0.988]	9.9729***	[7.074]
Age (months) (ln)	-3.1475***	[-6.474]	-0.0093	[-0.263]
Return Rank	-0.7508	[-1.316]		
Return Rank^2	0.1347	[1.358]		
$(Total\ Net\ Asset\ (ln))^2$			0.0014	[0.318]
$Inflow$			-0.0004	[-0.239]
Reporting Quarter FEs	Y		Y	
Time FEs	Y		Y	
Adjusted R^2	0.107		0.597	

N = 68,219

Table 8 presents the estimated coefficients and t-statistics (in brackets) from the OLS regression of fund flow% on *Total Views* For retail funds classified according to 12-b1 fees. The regression control for report features, fund-level characteristics and fixed effects, with standard errors clustered by fund. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. The adjusted r-squared is reported.

R^2_{max} is set as $1.3 \times R^2_{full}$ or at most 1.

Applying Oster (2019) robustness test, I obtain a delta (δ) of approximately 6.39. This indicates that omitted variables would need to be more than six times as strongly correlated with both the treatment ($Total\ Views$) and our outcome (Flow%) as compared to the observed controls to completely negate the positive effect. Such a substantial δ strongly suggests that the findings are robust to omitted variable bias.

Similarly, applying Oster (2019) robustness test for alpha (fund performance), I obtained a negative delta ($\delta \approx -0.0514$). The negative δ from the sensitivity analysis indicates that the baseline estimates (without controls) may not be reliable. This finding highlights the importance of incorporating appropriate controls to properly identify the relationship between retail investor attention ($Total\ Views$) and mutual fund performance.

6 Conclusion

This study investigates the role of retail investor attention in predicting mutual fund flows and performance, using *Total Views*, the total number of mutual fund shareholder report views accessed through the SEC EDGAR system, as a measure of retial attention. The findings indicate that retail investor attention is positively correlated with both future fund flows and fund performance, suggesting that retail investors actively use the information contained in shareholder reports to make informed investment decisions. Specifically, for retail funds, an increase in *Total Views* is associated with a 0.27% increase in future fund flows and a 0.02% increase in fund performance (alpha). However, the predictive power of *Total Views* is not significant for institutional funds, highlighting that *Total Views* primarily capture the attention of retail investors.

The portfolio analysis shows that an equal-weighted high-minus-low strategy based on abnormal *Total Views* yields a significant abnormal return of 0.18% for retail funds. This effect is concentrated in smaller funds, as significant returns are found in equal-weighted portfolios but not in value-weighted portfolios. A similar result of 0.14% is observed when institutional funds are included in the sample, underscoring the asset pricing implications of retail investor attention in the mutual fund industry.

Retail investor attention also strengthen the flow-performance relationship. *Total Views* increase the impact of past good performance on future flows, with the coefficient of the interaction term between *Total Views* and positive past returns dummy being 0.12%. This suggests that retail investors are more likely to increase their investments in outperforming funds if they have already viewed the reports. Conversely, the coefficient for the interaction with negative past returns dummy is not significant.

Further analysis indicates that both annual and semi-annual reports have similar predictive power for fund flows and performance, suggesting that retail investor attention to shareholder reports is valuable regardless of the type of report. Moreover, the impact of *Total Views* on future fund performance and flows remains consistent throughout the year, with no significant difference between the first and second quarters post-filing. This implies that the information contained in mutual fund shareholder reports is not time-sensitive. Robustness tests addressing non-linearity and using marketing fees (12b-1) as an alternative retail-fund classification further validate that retail investor attention translates into higher fund inflows. In addition, Oster's sensitivity test confirms the findings are robust to omitted variable bias.

Despite these findings, the *Total Views* measure may also hav some limitations. It does not capture access to fund shareholder reports through investment company websites, data providers like Yahoo Finance, or platforms such as Bloomberg Terminal. These alternative sources of information are more likely to be used by either naive retail investors, who may not seek out detailed shareholder disclosures, or sophisticated professional investors with access to advanced financial databases such as Bloomberg Terminal, CRSP, or Morningstar. Naive retail investors are often viewed as uninformed, unskilled noise traders, while professional investors tend to be more informed and may use automated methods, such as bots, to access information. Consequently, *Total Views* capture attention that falls between these two extremes. Future research could address the attention from both uninformed and highly informed market participants to gain a more comprehensive understanding.

Overall, the results of this paper suggest that mutual fund shareholder reports provide valuable, non-time-sensitive information for retail investors. By introducing a novel measure of retail investor attention specific to mutual funds, this study demonstrates that attention to these reports can predict future fund flows and performance. The findings contribute to the literature and have practical implications for fund managers and policymakers, supporting initiatives such as the Tailored Shareholder Reports, which aim to improve the readability and accessibility of mutual fund disclosures, thereby potentially further benefiting retail investors.

References

Barber, B.M., Huang, X., and Odean, T., 2016. Which Factors Matter to Investors? Evidence from Mutual Fund Flows. *The review of financial studies* [Online], 29(10), pp.2600–2642. Available from: https://doi.org/10.1093/rfs/hhw054.

Barber, B.M. and Odean, T., 2007. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The review of financial studies* [Online], 21(2), pp.785–818. Available from: https://doi.org/10.1093/rfs/hhm079.

Bergstresser, D., Chalmers, J.M.R., and Tufano, P., 2009. Assessing the costs and benefits of brokers in the mutual fund industry. *The review of financial studies* [Online], 22(10), pp.4129–4156. Available from: https://doi.org/10.1093/rfs/hhp022.

Blankespoor, E., deHaan, E., and Marinovic, I., 2020. Disclosure processing costs, investors' information choice, and equity market outcomes: a review. *Journal of accounting and economics* [Online], 70(2), p.101344. Available from: https://doi.org/https://doi.org/10.1016/j.jacceco.2020.101344.

Cao, S.S., Yang, B., and Zhang, A.L., 2023. Managerial risk assessment and fund performance: evidence from textual disclosure. *Ssrn* [Online]. Available from: https://doi.org/http://dx.doi.org/10.2139/ssrn.4060307.

Carhart, M.M., 1997. On persistence in mutual fund performance. *The journal of finance* [Online], 52(1), pp.57–82. Available from: https://doi.org/https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.

Chen, G. and Zhou, J., 2019. XBRL Adoption and Systematic Information Acquisition via EDGAR. *Journal of information systems* [Online], 33(2), pp.23–43. Available from: https://doi.org/10.2308/isys-52140.

Chen, T., Dong, H., and Lin, C., 2020. Institutional shareholders and corporate social responsibility. *Journal of financial economics* [Online], 135(2), pp.483–504. Available from: https://doi.org/https://doi.org/10.1016/j.jfineco.2019.06.007.

Chi, S. and Shanthikumar, D.M., 2018. Do retail investors use sec filings? evidence from edgar search [Online]. Available at SSRN: https://ssrn.com/abstract=3281234 or http://dx.doi.org/10.2139/ssrn.3281234. Available from: https://doi.org/10.2139/ssrn.3281234.

Cong, Y., Du, H., and Vasarhelyi, M.A., 2018. Are XBRL Files Being Accessed? Evidence from the SEC EDGAR Log File Dataset. *Journal of information systems* [Online], 32(3), pp.23–29. Available from: https://doi.org/10.2308/isys-51885.

Crane, A., Crotty, K., and Umar, T., 2022. Hedge funds and public information acquisition. *Management science* [Online], 69(6), pp.3241–3262. Available from: https://doi.org/10.1287/mnsc.2022.4466.

Cremers, M. and Petajisto, A., 2009. How active is your fund manager? a new measure that predicts performance. *The review of financial studies* [Online], 22(9), pp.3329–3365. Available from: http://www.jstor.org/stable/40247664.

Da, Z., Engelberg, J., and Gao, P., 2011. In search of attention. *The journal of finance* [Online], 66(5), pp.1461–1499. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2011.01679.x. Available from: https://doi.org/https://doi.org/10.1111/j.1540-6261.2011.01679.x.

Diamond, D.W. and Verrecchia, R.E., 1981. Information aggregation in a noisy rational expectations economy. *Journal of financial economics* [Online], 9(3), pp.221–235. Available from: https://doi.org/https://doi.org/10.1016/0304-405X(81)90026-X.

Drake, M.S., Johnson, B.A., Roulstone, D.T., and Thornock, J.R., 2020. Is There Information Content in Information Acquisition? *The accounting review* [Online], 95(2), pp.113–139. Available from: https://doi.org/10.2308/accr-52498.

Drake, M.S., Quinn, P.J., and Thornock, J.R., 2017. Who Uses Financial Statements? A Demographic Analysis of Financial Statement Downloads from EDGAR. *Accounting horizons* [Online], 31(3), pp.55–68. eprint: https://publications.aaahq.org/accounting-horizons/article-pdf/31/3/55/16038/acch-51736.pdf. Available from: https://doi.org/10.2308/acch-51736.

Drake, M.S., Roulstone, D.T., and Thornock, J.R., 2015. The determinants and consequences of information acquisition via edgar. *Contemporary accounting research* [Online], 32(3), pp.1128–1161. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1911-3846.12119. Available from: https://doi.org/https://doi.org/10.1111/1911-3846.12119.

Gibbons, B., Iliev, P., and Kalodimos, J., 2021. Analyst information acquisition via edgar. *Management science* [Online], 67(2), pp.769–793. Available from: https://EconPapers.repec.org/RePEc:inm:ormnsc:v:67:y:2021:i:2:p:769-793.

Griffin, P.A., 2003. Got information? investor response to form 10-k and form 10-q edgar filings. *Review of accounting studies* [Online], 8(4), pp.433–460. Available from: https://doi.org/10.1023/A:1027351630866.

Grossman, S.J. and Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. *The american economic review* [Online], 70(3), pp.393–408. Available from: http://www.jstor.org/stable/1805228 [Accessed September 25, 2024].

Hillert, A., Niessen-Ruenzi, A., and Ruenzi, S., 2024. Mutual fund shareholder letters: flows, performance, and managerial behavior. *Management science* [Online], 0(0), null. Available from: https://doi.org/10.1287/mnsc.2021.03417.

Hsieh, S.-F., Chan, C.-Y., and Wang, M.-C., 2020. Retail investor attention and herding behavior. *Journal of empirical finance* [Online], 59, pp.109–132. Available from: https://doi.org/https://doi.org/10.1016/j.jempfin.2020.09.005.

Institute, I.C., 2023. Available from: https://www.ici.org/system/files/2023-05/2023-factbook.pdf.

Ivkovic, Z. and Weisbenner, S., 2009. Individual investor mutual fund flows. *Journal of financial economics* [Online], 92(2), pp.223–237. Available from: https://doi.org/https://doi.org/10.1016/j.jfineco.2008.05.003.

Jonathan, B.B. and Richard, C.G., 2004. Mutual fund flows and performance in rational markets. *Journal of political economy* [Online], 112(6), pp.1269–1295. Available from: http://www.jstor.org/stable/10.1086/424739 [Accessed April 19, 2024].

Jordan, B.D. and Riley, T.B., 2015. Volatility and mutual fund manager skill. *Journal of financial economics* [Online], 118(2), pp.289–298. Available from: https://doi.org/https://doi.org/10.1016/j.jfineco.2015.06.012.

Kaniel, R. and Parham, R., 2017. Wsj category kings – the impact of media attention on consumer and mutual fund investment decisions. *Journal of financial economics* [Online], 123(2), pp.337–356. Available from: https://doi.org/https://doi.org/10.1016/j.jfineco.2016.11.003.

Kostovetsky, L. and Warner, J.B., 2020. Measuring innovation and product differentiation: evidence from mutual funds. *The journal of finance* [Online], 75(2), pp.779–823. Available from: https://doi.org/https://doi.org/10.1111/jofi.12853.

Lee, C.M. and So, E.C., 2017. Uncovering expected returns: information in analyst coverage proxies. *Journal of financial economics* [Online], 124(2), pp.331–348. Available from: https://doi.org/https://doi.org/10.1016/j.jfineco.2017.01.007.

Lee, C.-C., Chen, M.-P., and Lee, C.-C., 2021. Investor attention, etf returns, and country-specific factors. *Research in international business and finance* [Online], 56, p.101386. Available from: https://doi.org/https://doi.org/10.1016/j.ribaf.2021.101386.

Li, E.X., Lind, G., Ramesh, K., and Shen, M., 2023. Externalities of Accounting Disclosures: Evidence from the Federal Reserve. *The accounting review* [Online], 98(5), pp.401–427. Available from: https://doi.org/10.2308/TAR-2022-0105.

Li, F.W. and Sun, C., 2022. Information acquisition and expected returns: evidence from edgar search traffic. *Journal of economic dynamics and control* [Online], 141. Markets and Economies with Information Frictions, p.104384. Available from: https://doi.org/https://doi.org/10.1016/j.jedc.2022.104384.

Loughran, T. and McDonald, B., 2014. Measuring readability in financial disclosures. *The journal of finance* [Online], 69(4), pp.1643–1671. Available from: https://doi.org/https://doi.org/10.1111/jofi.12162.

Loughran, T. and McDonald, B., 2017. The use of edgar filings by investors. *Journal of behavioral finance* [Online], 18(2), pp.231–248. eprint: https://doi.org/10.1080/15427560.2017.1308945. Available from: https://doi.org/10.1080/15427560.2017.1308945.

Oster, E., 2019. Unobservable selection and coefficient stability: theory and evidence. *Journal of business & economic statistics* [Online], 37(2), pp.187–204. Available from: https://doi.org/10.1080/07350015.2016.1227711.

Rees, L., Sharp, N., and Twedt, B., 2015. Who's heard on the street? determinants and consequences of financial analyst coverage in the business press. *Review of accounting studies* [Online], 20(1), pp.173–209. Available from: https://doi.org/10.1007/s11142-014-9298-3.

Ryans, J., 2017. Using the edgar log file data set [Online]. Available at SSRN: https://ssrn.com/abstract=2913612 or http://dx.doi.org/10.2139/ssrn.2913612. Available from: https://doi.org/10.2139/ssrn.2913612.

Sicherman, N., Loewenstein, G., Seppi, D.J., and Utkus, S.P., 2015. Financial Attention. *The review of financial studies* [Online], 29(4), pp.863–897. eprint: https://academic.oup.com/rfs/article-pdf/29/4/863/24451268/hhv073.pdf. Available from: https://doi.org/10.1093/rfs/hhv073.

Sirri, E.R. and Tufano, P., 1998. Costly search and mutual fund flows. *The journal of finance* [Online], 53(5), pp.1589–1622. Available from: https://doi.org/https://doi.org/10.1111/0022-1082.00066.

Thaler, R.H., 1999. Mental accounting matters. *Journal of behavioral decision making* [Online], 12(3), pp.183–206. Available from: https://doi.org/https://doi.org/10.1002/(SICI)1099-0771(199909)12:3<183::AID-BDM318>3.0.CO;2-F.

Weber, M. and Camerer, C.F., 1998. The disposition effect in securities trading: an experimental analysis. *Journal of economic behavior & organization* [Online], 33(2), pp.167–184. Available from: https://doi.org/https://doi.org/10.1016/S0167-2681(97)00089-9.

Yung, K. and Nafar, N., 2017. Investor attention and the expected returns of reits. *International review of economics & finance* [Online], 48, pp.423–439. Available from: https://doi.org/https://doi.org/10.1016/j.iref.2016.12.009.