Mutual Fund Shorts and Endogenous Research Intensity: Theory and Evidence

Boone Bowles and Adam V. Reed*

June 2025

Abstract

Using granular data linking mutual fund holdings with their research behavior, we show that mutual fund managers acquire 5% more information about their shorts than their longs while shorts also earn higher abnormal returns. We reconcile these facts with a model of endogenous information acquisition that embeds short-side frictions. Extending the model with sequential-signals predicts—and the data confirm—that the highest-alpha shorts, paradoxically, are the least researched. Though this challenges conventional wisdom that more research leads to better returns, the intuition is straightforward: managers concentrate research on borderline cases, but extensive research is unnecessary for clear winners.

JEL classification: D83, G14, G23

Keywords: mutual funds, information economics, short selling.

*Mays Business School, Texas A&M University (boone.bowles@tamu.edu), Kenan-Flagler Business School, University of North Carolina (adam_reed@unc.edu). The authors thank Dora Horstman, Yunzhi Hu, Matthew Ringgenberg, Jesse Davis, and conference and seminar participants at Texas A&M University, University of Cologne - Center for Financial Research, the Southwestern Finance Association Annual Meeting (2024), the Midwest Finance Association Annual Meeting (2024), the Eastern Finance Association Annual Meeting (2024), and the 16th Annual Paris Hedge Fund Conference (2025). All errors are our own. We (the authors) have no conflicts of interest to disclose.

1 Introduction

Short selling plays a critical role in market efficiency while also posing distinct frictions and risks. Unlike buying, shorting is impeded by locate fees, borrowing constraints, and potentially unlimited downside risks. These frictions have inspired extensive literature examining whether and how short selling impacts price discovery and asset pricing.[1] However, relatively little is known about how short-selling constraints shape institutional investors' decisions around information acquisition.

A particularly important class of institutional investors in this context is active mutual fund managers, who control a large share of public equity capital and often operate under strict portfolio constraints, including limitations on short selling. Despite their importance in capital markets, how mutual funds approach short selling—how they choose which stocks to short, how much information they gather beforehand, and what returns those decisions produce—has not been widely researched. These open questions make mutual funds a natural setting for studying how information acquisition interacts with short-selling behavior.

This paper asks how active mutual fund managers learn when short selling is costly and how that learning shapes their trades and payoffs. Empirically, we accomplish this by linking two complementary data sets: (i) position-level holdings for U.S. equity mutual fund families, and (ii) the time-stamped request logs from the SEC's EDGAR filing system. Merging the two lets us observe—at the fund family-stock-quarter level—the exact public filings a manager downloads *before* a position appears in their portfolio. This linkage gives us an unusually granular view of the information-production process inside a major class of institutional investors.

Data in hand, we document two significant facts. First, before a new position is estab-

[1]E.g., Jones and Lamont (2002), Boehmer et al. (2008), and Engelberg et al. (2018).

lished, mutual fund managers ramp up their EDGAR requests significantly in the quarters prior, retrieving roughly 5% *more* EDGAR filings for stocks they will short than for those they will hold long—a gap that persists after controlling for a host of confounding factors. Second, these very shorts subsequently deliver larger absolute risk-adjusted returns, averaging 125 basis points (bps) per quarter versus 30 bps for comparable new long positions. Together, these facts raise a natural question: what makes it optimal for managers to devote the most research effort to trades (e.g., shorts) that end up delivering the best returns?

To explain these patterns we introduce a tractable model, nested in the Grossman and Stiglitz (1980) framework, that puts a single twist on the classic setup: short positions carry extra, side-specific costs. A mutual fund manager chooses how much noisy public information to buy before deciding whether to take a fixed-size long or short position. The usual convex cost of precision applies to both longs and shorts, and both also face a fixed monitoring charge, but only shorting requires an additional locate fee and a size-dependent surcharge that captures fee-volatility and recall risk. These frictions raise the precision threshold the manager must clear before it is worthwhile to short.

Echoing the Grossman and Stiglitz (1980) intuition, the manager collects information until the expected benefit of additional information exactly matches its marginal cost. Because the threshold is higher for shorts, the optimal precision—and hence observable research effort—is greater on the short side. At the same time, shorts are executed only when the signal implies a larger mispricing, mechanically leading to higher absolute abnormal returns. Taken together, the model accounts for Fact 1 (more research for shorts than longs) and Fact 2 (larger absolute returns for shorts than longs).

Importantly, this model is transparent enough to yield additional testable predictions when we extend it to let managers acquire information sequentially, following the logic formalized by Banerjee and Breon-Drish (2020). In this setting the manager samples signals until the expected mispricing—and posterior precision—is large enough to justify a short. When the very first signal is extreme, the position looks like a *clear winner* and little additional research is required; when early signals are more equivocal, the manager purchases further precision and the trade becomes a *borderline short.* The framework therefore predicts an inverse relation inside the short portfolio: positions preceded by more requests should realize smaller absolute abnormal returns, whereas longs—unburdened by the short-side cost threshold—should display no such pattern.

Putting this prediction to the data confirms the clear-winner-versus-borderline-short logic. Among shorts, a one-standard-deviation *increase* in filing requests corresponds with roughly a 40-basis-point *decline* in absolute returns. For long positions, by comparison, the same increase in requests has no consistently discernible effect. This asymmetric slope—more specifically the inverse slope among shorts—is precisely what the sequential-sampling model implies, yet would be difficult to anticipate without the formal model and can be detected only because our data link mutual fund EDGAR downloads to mutual fund positions. Together, the evidence offers a clean, out-of-sample validation of the model's most nuanced economic insight.

Furthermore, complementary evidence underscores the broader economic logic of informed trading. Mutual fund families that engage more intensively in information acquisition realize consistently higher benchmark-adjusted returns. This finding strongly suggests that mutual fund managers' overall information-gathering efforts are productive, reinforcing the rational choice interpretation of information acquisition. Therefore, the negative slope observed among short positions is not evidence of inefficient information acquisition. Instead, it highlights that greater research effort is rationally allocated to less obvious opportunities, which still contribute positively to overall performance. Additionally, we validate the economic intuition of our model by demonstrating that an independent measure of ex ante mispricing reliably predicts future stock returns—specifically, stocks identified as overpriced indeed exhibit low subsequent returns. This predictive validity provides further external confirmation of the "clear winner" logic driving managers' differential information strategies.

Our analysis makes several important contributions to the literature on short selling, information acquisition, and mutual fund behavior. First, we provide direct, position-level evidence linking information acquisition to short selling, addressing a critical gap left by earlier work that relies on aggregate proxies. Unlike studies utilizing aggregate search data (Da et al. (2011), Ben-Rephael et al. (2017)) or aggregated short-interest measures (Seneca (1967), Hong and Stein (2003), Asquith et al. (2005), Boehmer et al. (2008)), our data uniquely reveal mutual fund manager actions, allowing us to better observe the relationship between information gathering and subsequent portfolio decisions. Our focus on longs versus shorts within mutual fund families differs fundamentally from early evidence that aggregates short positions across investors.

Second, our findings deepen the understanding of short-selling constraints by pinpointing their influence directly at the information-acquisition stage, uncovering a crucial but under-explored channel through which short-selling frictions affect market outcomes. Prior studies such as Engelberg et al. (2018) and An et al. (2021) have documented the broad effects of borrowing constraints and institutional limits; however, our work distinctively identifies the endogenous information acquisition process, shaped by marginal costs and benefits, as the key mechanism underlying these constraints' impact on institutional investor behavior and asset prices.

Third, we contribute to a small but growing literature on mutual fund shorting behavior.

Building on Almazan et al. (2004), who document widespread self-imposed prohibitions, and Chen et al. (2013), who provide an early look at mutual fund shorting skill, we show that 48% of the mutual fund families in our study engage in short selling. We further demonstrate that those short positions are preceded by more information acquisition and deliver higher abnormal returns than corresponding long positions within the same portfolios. Our findings complement recent work by An et al. (2021), who examine the growth and underperformance of long-short mutual funds, by showing that shorting activity is more common and more skillfully executed within the broader mutual fund universe than previously assumed.

Fourth, we add new evidence to the empirical literature on institutional information acquisition. Our position-level approach complements recent EDGAR studies of hedge funds and other institutions (Chen et al. (2020a), Chen et al. (2020b), Crane et al. (2022), Gibbons et al. (2021), Gibbons (2023)) by linking downloads to both sides of the same mutual fund portfolios. Also, to our knowledge, we are the first to document a negative link between pre-trade information acquisition and subsequent abnormal returns for shorts—a pattern predicted by our model and unseen in prior work and a finding that overturns the conventional expectation that more research should lead to better performance and reveals a previously hidden aspect of how skilled investors allocate their attention across investment opportunities.

Fifth, we extend the theoretical literature by explicitly incorporating sequential information acquisition into the classical Grossman-Stiglitz (1980) framework. By doing so, our model bridges insights from sequential search theory (Banerjee and Breon-Drish (2020)) and earlier foundational work on costly information acquisition (Hellwig (1980), Verrecchia (1982)). This theoretical advancement yields tractable and novel predictions regarding managerial behavior under costly short selling, predictions we empirically validate using our mutual fund setting.

The remainder of the paper proceeds as follows. Section 2 details the data sources and construction of key variables. Section 3 presents our two core empirical facts about the short-versus-long differences in information acquisition and returns. Section 4 introduces and develops our theoretical framework, deriving predictions consistent with our empirical findings. Section 5 tests the model's novel prediction regarding the within-short relationship between information acquisition and returns, and explores additional implications about holding durations. Section 6 provides additional tests and Section 7 concludes. A full description of the data and a detailed theoretical derivation are provided in the Online Appendix.

2 Data

This study draws on two primary data sources. First, we obtain quarterly mutual fund holdings from the Center for Research in Security Prices Survivor-Bias-Free U.S. Mutual Fund Database (*CRSP Mutual Funds*). For each mutual fund family i, stock j, and quarter t we record the number of shares reported (nbr_shares_{ijt}). A long position,($Long_{ijt}$) equals one when $nbr\_shares_{ijt} > 0$, while a short position, ($Short_{ijt}$) is one when $nbr\_shares_{ijt} < 0$. We also flag the first quarter in which a family initiates a position as $NewLong_{ijt}$ or $NewShort_{ijt}$, respectively.

Second, we measure information acquisition with the U.S. SEC's EDGAR Log Files, which record every electronic request for a public filing. Each observation details the filing requested, the date and time of the request, and the requester's IP address.[2] Following

[2]For example, the filing request is for the 2013 annual report (10-K) for IBM. The request was made at 10:14 am on March 1, 2014. The request originated from the IP address 123.123.123.*abc*.

the unmasking procedure detailed in Appendix A.1, we link IP addresses to mutual fund families. For every i–j–t we then count the number of filings family i requests about firm j during quarter t and label this $Requests_{ijt}$. Our baseline measure of mutual fund information acquisition is $MFIA_{ijt} = ln(1 + Requests_{ijt})$. To capture pre-trade information acquisition we also compute $Requests_{ijt,t-1}$, the two-quarter sum of requests, and use the same transformation.[3]

We construct our sample beginning with all common stocks and all CRSP mutual fund families between 2010 and 2017 (30 quarters). We retain an i-j pair only if mutual fund family i reports a non-zero position in stock j at least once during the sample window. We further require that (i) the family takes *at least one* short position during the window, and (ii) its average quarterly EDGAR activity is ≥ 5 requests. Observations in which a family simultaneously holds both long *and* short positions in the same stock are dropped.

The final panel contains 4,937,880 family-stock-quarter observations. Roughly 37% indicate a long position, whereas 2% indicate a short position—half of which occur without a contemporaneous long. Aggregated to the family-quarter level, the median family submits 1,219 filing requests per quarter—the mean is 7,122—and at least one request appears in 89% of family-quarters. Our inclusion criteria reduce the fund count from the CRSP universe of 115 families to 55 that actively short. Summary statistics are reported in Table 1, and Figure 1 plots the time series of aggregate long, short, new-long, and new-short positions. Overall, the data afford a novel view of both sides of mutual fund portfolios and of the information-gathering efforts that precede them.

[Table 1 about here.]

[3] Additional stock-level variables and fund characteristics come from CRSP, Compustat, Thomson Reuters 13F, RavenPack, and the mispricing scores of Stambaugh and Yuan (2017). A complete variable dictionary appears in Appendix A.2.

[Figure 1 about here.]

3 Two Empirical Facts

This section establishes two stylized facts that motivate the remainder of the analysis. First, mutual fund families devote more information-gathering effort to their short positions than to their long positions. Second, those short positions, on average, subsequently earn better (i.e., more negative) risk-adjusted returns than the corresponding long positions within the same portfolios.

3.1 Fact 1: Research Intensity and Position Type

We proxy for research intensity with number of EDGAR filings downloaded by mutual fund family i for company j during quarter t. To test whether information acquisition differs across position types we estimate

$$MFIA_{ijt} = \beta_1 Long_{ijt} + \beta_2 Short_{ijt} + \gamma_i + u_j + \omega_t + \varepsilon_{ijt}, \tag{1}$$

where $MFIA$ (defined earlier) includes either one quarter of requests or two quarters and the model includes mutual fund family, stock, and quarter fixed effects.

The coefficient estimates for *Long* and *Short* capture the intensity of information acquisition for long and short positions relative to non-holdings, within a mutual fund family. To test whether managers differentially acquire information for shorts and longs we test the linear combination $\beta_2 - \beta_1 = 0$. The results from estimating Equation 1 are shown in Panel A of Table 2.

[Table 2 about here.]

Two patterns stand out. First, the positive and statistically significant estimates for both *Long* and *Short* indicate that managers gather more information about the stocks they hold—on either side of their portfolios—compared to the stocks they do not hold. This is true when information acquisition and positions are measured simultaneously or when considering lagged information acquisition. In terms of economic significance, mutual fund managers make approximately 10% more requests for their long positions compared to stocks they do not hold, and they make roughly 15% more requests for their short positions.

Second, with respect to the difference between longs and shorts, mutual fund managers acquire roughly 5% more information about their shorts than about their longs, and this difference is statistically significant and consistent over various model specifications.

One concern is that the differential could be driven by large, legacy positions that require little incremental analysis. Panel B addresses this possibility by re-estimating Equation 1 with $NewLong$ and $NewShort$—indicator variables for positions initiated during quarter t. Here, the differential widens to as much as 7%, and the effect is strongest when information is aggregated over the current and previous quarter. Thus, in the months leading up to a new position, managers ramp up research on prospective shorts but not on prospective longs. Overall, the results in Table 2 establish our first empirical fact.

Fact 1: *Mutual fund families allocate disproportionately more research effort to the positions they hold short.*

3.2 Fact 2: Position Type and Returns

Next we ask whether shorts earn better returns than longs within mutual fund families. In other words, we compare the long side of an investor's portfolio with the short side of that same portfolio. We estimate

$$Return_{ijt} = \alpha + \beta Short_{ijt} + \lambda Z_{jt} + \gamma_i + \varepsilon_{ijt}, \tag{2}$$

where $Return_{ijt}$ is either the excess return of stock j during quarter t ($ExRet$) or the stock's characteristic-adjusted return ($DGTW$).[4] Vector Z includes the market return and three risk factors: SMB, HML, and UMD.[5] The model also includes a fund family fixed effect.

We estimate the model over three subsamples to compare the differences between longs and shorts. In the first two subsamples (only longs or only shorts) the intercept α captures the abnormal return. The third subsample includes both long and short positions, thus α captures the abnormal return for longs while $Short$ measures the additional return to shorts and $Short + \alpha$ is the total abnormal return for shorts. Table 3 shows the results.

[Table 3 about here.]

The evidence for longs is mixed: abnormal returns using the risk-factor model average 29 bps per quarter, but characteristic-adjusted returns average only 8 bps and are statistically insignificant. Shorts are different: across all specifications they deliver sizable negative abnormal returns—between 118 and 132 bps per quarter—implying that shorts outperform longs by more than 100 bps. Formally, the absolute long-short gap of 117 bps (Column 7) is significant at the 1% level.[6]

To address the confounding effects of timing, we also estimate Equation 2 on new positions only and consider both *contemporaneous* returns and *next quarter* returns. Table 4 reports

[4]Daniel et al. (1997).

[5]Fama and French (1993) and Carhart (1997).

[6]Since negative returns are earned by short positions, the absolute return differential is defined as the total return earned by shorts less the total return earned by longs. Based on Equation 2, this is $-(Short{+}2\alpha)$.

the results and shows a clear pattern when timing is accounted for (Columns 4-7): new shorts deliver gains of up to 145 bps per quarter while new longs do not earn large abnormal returns. Indeed, the long-short gap widens to almost 200 bps in this setting and remains statistically significant.

[Table 4 about here.]

The histograms of characteristic-adjusted returns in Figure 2 visualizes and reinforce these findings: long-position returns cluster tightly around zero, whereas short-position returns exhibit a pronounced left tail. Together, the results in Table 3, Table 4, and Figure 2 establish our second empirical fact.

Fact 2: *Mutual funds' short positions consistently outperform their long positions.*[7]

[Figure 2 about here.]

4 Model

Our empirical analysis establishes two new facts. First, mutual fund families undertake more information acquisition before initiating short positions than before comparable long positions. Second, those shorts subsequently earn substantially larger abnormal returns than the longs. In this section we develop a theoretical framework that contains these two facts while clarifying why managers concentrate their research effort on short positions that also generate the higher average payoffs. Accordingly, we introduce a concise analytical framework—a variant of the Grossman and Stiglitz (1980) costly-information model augmented with

[7]This outperformance—over 100 bps per quarter ($\approx$ 4% per year)—remains economically meaningful even after accounting for plausible lending fees: 91% of stocks have lending fees under 1% per year (value-weighted the mean $\approx$ 0.25%), and excluding the 5% most expensive issues yields a mean fee of only 85 bps (D'avolio (2002); Blocher and Whaley (2015)).

short-selling frictions—that rationalizes both empirical findings. Then, extending the baseline framework to incorporate sequential signal acquisition, the model generates additional cross-sectional predictions, which we test empirically in the sections that follow.

4.1 Baseline Model

The model has two periods $t \in \{0, 1\}$ and a risky asset with equally likely binary fundamental values $V \in \{\nu_H, \nu_L\}$: the *value gap* is $\Delta \equiv \nu_H - \nu_L > 0$. Competitive noise traders pin down the pre-trade price at the unconditional mean, $P_0 = \frac{\nu_H+\nu_L}{2}$. Before trading, a risk-neutral investor may purchase a costly binary noisy signal ($x \in \{\nu_H, \nu_L\}$) that coincides with the true state with probability $P(x = V) = \frac{1}{2} + \frac{\pi}{2}$, where $\pi \in [0, 1]$ is the chosen precision. Precision is costly: $c(\pi) = \kappa\pi^2$, $\kappa > 0$. After observing the signal, the investor selects position side $s \in \{L \text{ (long)}, S \text{ (short)}, 0\}$ and position size q.

Opening any position incurs fixed monitoring costs $F > 0$, leading the investor to trade a minimal optimal block $q_{\min} > 0$ or abstain.[8] Long positions incur no additional costs ($C_L = F$) while short positions incur a locate fee $\phi > 0$ and a borrow surcharge $\eta q_{\min} > 0$. ($C_S = F + \phi + \eta q_{\min}$.)

At $t = 0$ the investor chooses precision π, observes the private signal x, updates beliefs via Bayes' rule, and chooses a trade side (s). At $t = 1$ the fundamental V is revealed and the investors payoff is $\Pi \equiv q_{min}(V - P_0) - C_s - c(\pi)$.

Equilibrium is characterized by a single optimal research intensity, π^*; side-specific precision thresholds $\{\pi_L^{thr}, \pi_S^{thr}\}$ where $\pi_s^{thr} = \frac{2C_s}{q_{\min}\Delta}$; and an initial price P_0, such that (i) π^* maximizes expected profit ex ante; (ii) the investor can trade side $s \in \{L, S\}$ only if $\pi^* \geq \pi_s^{thr}$; and (iii) the uninformed market makers break even: $P_0 = \frac{(\nu_H+\nu_L)}{2}$. This model yields three

[8] See Appendix A.3.1 for further details on the minimum tranche.

propositions:[9]

Proposition 1. *Shorts require higher precision thresholds than longs:* $\pi_S^{thr} > \pi_L^{thr}$.

Intuition. Extra short-side fees ($C_S > C_L$) and the minimum-tranche requirement $q_{\min} > 0$ jointly raise the break-even hurdle. Put differently, a short bet must be backed by a more accurate signal before it is worthwhile.

Proposition 2. *Conditional on a trade, the average precision for shorts is larger than for longs:* $E[\pi^* \mid s = S] > E[\pi^* \mid s = L]$.

Intuition. Shorts only occur in the high-precision equilibrium, whereas long positions are observed in both the high-precision equilibrium ($\pi^* \geq \pi_S^{thr}$) and the long-only equilibrium ($\pi^* < \pi_S^{thr}$).

Proposition 3. *Conditional on a trade, the average absolute abnormal return is larger for shorts than for longs:* $E\big[\,|\alpha|\,\big|\, s = S\big] \ > \ E\big[\,|\alpha|\,\big|\, s = L\big]$.

Intuition. A short is executed only if (i) precision satisfies $\pi^* \geq \pi_S^{thr}$ and (ii) the posterior value gap meets the stricter *value-gap threshold*: $\tau_s = \frac{C_s}{q_{\min}}$, with $\tau_S > \tau_L$. Both forces imply a larger perceived mispricing—and hence a larger realized abnormal return—than for a long.

Together Propositions 1 and 2 deliver the identical ordering we observe in our first empirical fact: mutual fund managers gather more information before short positions than before comparable long positions. Proposition 3 mirrors our second empirical fact by implying that, once a trade is made, shorts should earn systematically larger absolute abnormal returns than longs.

[9] Appendix A.3.1 provides proofs and closed-form expressions.

4.2 Extended Model: Sequential Signals

The baseline model explains why mutual fund managers research shorts more intensively and why shorts yield larger absolute abnormal returns. These findings—with the underlying theory and intuition—invite a finer question: within a mutual fund's short book (or long book), do trades requiring more research earn systematically different returns than those requiring less? At first glance the baseline logic implies a positive link between research effort and returns on either side of the book. But anecdotally, within the set of shorts some may be *clear winners* that seem overpriced after minimal analysis, whereas other trades, *borderline positions*, require extensive research.

To examine this within-side relation we extend the baseline model by allowing for *sequential signals*.[10] All primitives remain as in the baseline model except that now the investor acquires information sequentially (i.e., one filing at a time) rather than in a single initial signal. Each signal is an independent observation of the fundamental with additive Gaussian noise: $x_j = V + \varepsilon_j, \;\; \varepsilon_j \sim \mathcal{N}(0, \sigma^2)$. Each draw has precision $\pi_0 = \frac{1}{\sigma^2}$ and costs $\kappa \pi_0^2$. By conjugacy, after N draws the total precision is $N\pi_0$ and the posterior mean is $m_N = \frac{1}{N} \sum_{j=1}^{N} x_j$.

After each draw the investor compares the *posterior dollar mispricing* $|m_N|$ with the side-specific threshold, $\tau_s = \frac{C_s}{q_{\min}}$. The investor can then (i) trade by taking side $s \in \{L, S\}$ at block size $q_{\min}$ if $|m_N| \geq \tau_s$; (ii) continue research by paying $\kappa \pi_0^2$ for another draw if the option value of more information is positive; or (iii) stop research and hold no position if the option value of another draw is negative.[11]

Equilibrium is defined by precision thresholds $\{\tau_L, \tau_S\}$ and initial price P_0 such that (i) the stopping rule described above is optimal given τ_s and P_0; (ii) each threshold τ_s satisfies

[10] See Banerjee and Breon-Drish (2020).
[11] Appendix A.3.2 for further details.

$\tau_s = \frac{C_s}{q_{\min}}$, so the investor is indifferent at $|m_N| = \tau_s$; and (iii) uninformed market makers break even in expectation by setting $P_0 = \frac{(\nu_H + \nu_L)}{2}$.[12] This extended model delivers an important novel prediction:

Proposition 4. *Expected absolute abnormal returns of executed shorts decrease in the number of signals acquired.*

Intuition. More information draws reduce the posterior's variance such that, in expectation, the posterior crosses the threshold τ_S by a smaller margin. Consequently, on average, requiring more draws yields smaller posterior dollar mispricing and, hence, smaller average absolute returns.

Proposition 4 provides a sharp, novel prediction distinguishing our sequential-sampling model from simpler alternatives: a systematic negative relation between information acquisition and subsequent absolute returns *within short positions.* Intuitively, shorts requiring extensive research are precisely those where investors acquire just enough conviction to trade, yielding smaller abnormal returns relative to *clear winners* identified after minimal analysis. We test this prediction empirically using our unique dataset in the next section.

In contrast, the model predicts a much weaker relation for longs. While longs share certain trading costs (F, $q_{\min}$), they do not face short-side frictions (e.g., locate fees, borrowing costs). Consequently, the trading threshold τ_L is lower, making it easier to clear even with weaker initial signals. Thus, longs requiring multiple signals before execution do not differ systematically in realized mispricing compared to those identified quickly, implying a flatter relationship between information acquisition and returns among longs.

[12] Appendix A.3.2 shows that these three conditions jointly deliver a unique equilibrium and derives closed-form expressions for the investor's continuation values and the option value of an additional signal. All proofs are also shown in the appendix.

5 Empirical Test of Proposition 4

In this section, we assess the key prediction of the sequential-sampling model: conditional on executing a short, when mutual fund managers acquire more public filings they should, on average, realize smaller absolute abnormal returns. We proceed in two steps. First, we visualize the raw relation between research intensity ($Requests$) and subsequent abnormal returns for longs versus shorts—and for *new* longs and shorts. Second, we estimate a panel regression that formally tests whether the slope between information acquisition and absolute returns is indeed negative for shorts but flat for longs.

Figure 3 plots abnormal returns against contemporaneous filing requests separately for all long positions (Panel A) and all short positions (Panel B). For longs, the scatter cloud shows no discernible trend, and a fitted line is essentially flat. By contrast, the fitted line slopes upward for shorts, indicating that positions accompanied by more EDGAR requests tend to earn less negative returns (i.e., worse returns). Since more information should improve performance under a naive "more is better" view, this positive slope among shorts is perhaps counterintuitive; yet it is consistent with the model's prediction.

[Figure 3 about here.]

To ensure this pattern is not driven by stale holdings, Figure 4 repeats the scatterplots focusing solely on new long and new short positions. The inverse relation for shorts is more prominent in this setting, whereas the slope for new longs again hovers near zero.results.

[Figure 4 about here.]

To formally test these patterns we estimate

$$Return_{ijt} = \alpha + \beta_1 MFIA_{ijt} + \delta Short_{ijt} + \beta_2 Short_{ijt} X MFIA_{ijt} + \lambda Z_{jt} + \gamma_i + u_j + \epsilon_{ijt}, \quad (3)$$

which is similar to Equation 2 but with the inclusion of $MFIA$ and the interaction of $Short$ and $MFIA$. The model also includes mutual fund family and stock fixed effects.

The estimate for $MFIA$ tests the relation between information acquisition and returns for long positions. The linear combination of $MFIA$ and the interaction term ($ShortXMFIA$) tests the relation for shorts. Also, the intercept in this model (α) estimates the abnormal return to long positions since *Long* is the omitted category (the table replaces α with *Long* for interpretability). Table 5 shows the results.

[Table 5 about here.]

The estimate for $MFIA$ is positive and significant when testing returns and contemporaneous information acquisition (Columns 2–3), but is negative or statistically insignificant when including lagged information acquisition (Columns 4–5). Taken together, these results suggests a statistically indistinguishable (or flat) relation between research and returns among longs—as predicted by the model.

Among shorts, however, the evidence is clear that more research reduces absolute returns. Column 2 of Table 5 reports an estimate of 0.26 shorts, which means that doubling research for a given short *reduces* its absolute quarterly abnormal return by 26 bps. This negative relation holds whether research and returns are measured contemporaneous or with a lag, and regardless of whether we use the four-factor model or characteristic adjustments. Further, we obtain similar results focusing on new shorts and new longs (Table 6). In this setting there is some evidence that information acquisition and returns are positively related for new longs, but for new shorts the relation is still negative and still large. The total effect of 0.54 in Column 3 suggests that doubling research for a new short decreases the abnormal return by 54 bps per quarter.

[Table 6 about here.]

Overall, these findings confirm Proposition 4: conditional on a short being executed, more research activity predicts smaller absolute abnormal returns.

6 Alternative Explanation and Additional Evidence

The inverse research-return slope for executed shorts is intuitive following the sequential-sampling model, but there is a competing interpretation for the empirical finding. Perhaps managers viewing extra EDGAR filings simply bog themselves down in *public* information that holds little incremental value, so the inverse relation reflects wasted effort rather than the clear-winner-versus-borderline logic of the model. Fortunately, we can test this alternative empirically by examining whether heavier research correlates with poorer overall performance at the family level. We can also examine whether managers genuinely scale back research for potential shorts that are "clear winners," as the sequential-signal framework predicts.

6.1 Is public-document research wasteful?

Could the inverse relation between research and absolute returns among executed shorts reflect wasted effort? If so, fund families that generally do more public-document research should underperform. We test this by aggregating our sample to family-quarter observations and using the following regression model to explain family-level benchmark-adjusted returns:

$$Return_{it} = \alpha + \beta MFIA_{it} + \lambda X_{it} + \omega_t + \epsilon_{it}, \quad (4)$$

where $Return_{it}$ is the average benchmark-adjusted return across the actively managed funds of family i during quarter t and X_{it} includes various controls.[13] The variable of interest, $MFIA_{it}$, measures total requests for all stocks by mutual fund family i during quarter t.[14]

Table 7 shows the results using contemporaneous and lagged measures of information acquisition, and finds that the research-return slope is *positive*: fund families that do more EDGAR research have better returns. Indeed, the estimates suggest that if a family doubled its research efforts it would improve next-quarter's benchmark alpha by roughly 1.4 bp—small in percentage terms but $7 million for a median $50 billion family. Thus, public-document research is productive on average, contradicting the "wasteful research" interpretation.

[Table 7 about here.]

6.2 Evidence with Clear-Winners

The evidence in Section 5 shows that for executed short positions, more research is related to lower absolute returns. These findings support the intuition in behind Proposition 4: that shorts requiring extensive research are *borderline shorts* whereas those requiring less research are *clear winners*. This section provides further support of Proposition 4 by showing that stocks that are clear winner shorts *ex ante* are the same stocks that require less attention from managers.

First, we show that overpriced stocks—identified *ex ante* using indicators based on the mispricing measure of Stambaugh and Yuan (2017)—earn low returns. Table 8 highlights

[13] Control variabels include family size ($TNAM$), total fund flows over the last year ($Netflow12$), average portfolio concentration (HHI), average expense ratios ($ExpenseRatio$), and average portfolio turnover ($Turnover$).

[14] $MFIA_{it}$ aggregates $MFIA_{ijt}$ across all stocks for family i during quarter t.

this result, showing the results from testing the following regression model:

$$Return_{jt} = \alpha + \beta_1 Underpriced_{jt-1} + \beta_2 Overpriced_{jt-1} + \lambda Z_{jt} + \epsilon_{jt}, \tag{5}$$

where $Return_{jt}$ is either the excess return of stock j during quarter t ($ExRet$) or the characteristic-adjusted return ($DGTW$).[15]

[Table 8 about here.]

Second, we combine the *ex ante* overpricing indicator with an indicator for low uncertainty: an overpriced stock with a low level of uncertainty is a clear winner with respect to shorts.[16] Then for new short positions we estimate the following model:[17]

$$\begin{aligned} MFIA_{ijt} = {} & \beta_1 LowDisagree_{jt} + \beta_2 Overpriced_{jt} \\ & + \beta_3 LowDisagreeXOverpriced_{jt} + \gamma_i + u_j + \omega_t + \epsilon_{ijt}. \end{aligned} \tag{6}$$

For new short positions, the interaction of *LowDisagree* and *Overpriced* indicates a clear winner since the stock has low uncertainty and the overpricing foreshadows negative future returns.[18] The model always includes mutual fund family and quarter fixed effects and includes stock fixed effects where noted. Table 9 shows the results.

[Table 9 about here.]

[15] The variables of interest are *Underpriced* and *Overpriced*, which are indicators based on the mispricing measure of Stambaugh and Yuan (2017).

[16] To proxy for low uncertainty we define *LowDisagree* as equal to one for each stock-quarter observation with *Disagreement* in the lowest quintile.

[17] We estimate the analogous model for new long positions.

[18] Combining *LowDisagree* with *Underpriced* indicates clear winners in the case of longs.

In the case of new long positions, there is a weak relation or no relation between managers' research and clear winners. This result supports Proposition 4 and findings from Tables 5 and 6.

For new short positions, where *LowDisagreeXOverpriced* represents the clear winners, Proposition 4 is also supported: managers do less research for clear winners. Column 5 shows that mutual fund managers acquire 8% less information about overpriced stocks with low disagreement compared to other stocks. When including stock-fixed effects, managers acquire up to 12% less information when stocks are overpriced and have low disagreement. Taken together, the evidence favours the sequential-sampling logic of Proposition 4: managers do less research when the first signals already point to an obvious mispricing..

7 Conclusion

This paper links the holdings of U.S. equity mutual funds with observations of their research behavior from the EDGAR Log Files and, in doing so, observes how portfolio managers allocate research effort for both long and short positions. Two empirical facts stand out: (i) funds allocate 5% more research to stocks they will short than to those they will buy, and (ii) those shorts earn absolute abnormal returns more than 100 bps per quarter higher than comparable longs.

Then, using a parsimonious endogenous information model with short-specific frictions we reconcile these facts while generating a novel insight about research, executed trades, and performance. The model predicts that short positions preceded by more research should deliver *smaller* absolute alpha. The data confirm this inverse slope between research and returns for shorts—and a flat one for longs—validating the model's clear-winner-versus-borderline-trade logic.

By showing that information gathering—not just trade execution—is a primary margin on which short-selling frictions operate, we sharpen the classic narrative about costly-arbitrage. We also demonstrate that public filings are a reliable source of alpha when investors optimally ration attention and research efforts. Further research in this area should seek to better understand the factors that drive information acquisition decisions for other investor types and for other types of trades. Investigating the impact of information acquisition on other aspects of portfolio management and performance could also provide valuable insights for investors and financial professionals.

References

Almazan, A., Brown, K. C., Carlson, M., and Chapman, D. A. (2004). Why constrain your mutual fund manager? *Journal of Financial Economics*, 73(2):289–321.

An, L., Huang, S., Lou, D., and Shi, J. (2021). An anatomy of long-short equity funds. *Available at SSRN 3813790*.

Asquith, P., Pathak, P. A., and Ritter, J. R. (2005). Short interest, institutional ownership, and stock returns. *Journal of Financial Economics*, 78(2):243–276.

Banerjee, S. and Breon-Drish, B. (2020). Strategic trading and unobservable information acquisition. *Journal of Financial Economics*, 138(2):458–482.

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

Blocher, J. and Whaley, R. E. (2015). Passive investing: The role of securities lending. *Working Paper Series*.

Boehmer, E., Jones, C. M., and Zhang, X. (2008). Which shorts are informed? *The Journal of Finance*, 63(2):491–527.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.

Chen, H., Cohen, L., Gurun, U., Lou, D., and Malloy, C. (2020a). Iq from ip: Simplifying search in portfolio choice. *Journal of Financial Economics*, 138(1):118–137.

Chen, H., Desai, H., and Krishnamurthy, S. (2013). A first look at mutual funds that use short sales. *Journal of Financial and Quantitative Analysis*, 48(3):761–787.

Chen, Y., Kelly, B., and Wu, W. (2020b). Sophisticated investors and market efficiency: Evidence from a natural experiment. *Journal of Financial Economics*, 138(2):316–341.

Crane, A., Crotty, K., and Umar, T. (2022). Hedge funds and public information acquisition. *Management Science*.

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

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of finance*, 52(3):1035–1058.

D'avolio, G. (2002). The market for borrowing stock. *Journal of Financial Economics*, 66(2-3):271–306.

Engelberg, J. E., Reed, A. V., and Ringgenberg, M. C. (2018). Short-selling risk. *The Journal of Finance*, 73(2):755–786.

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

Gibbons, B. (2023). Public market information and venture capital investment. *Journal of Financial and Quantitative Analysis*, 58(2):746–776.

Gibbons, B., Iliev, P., and Kalodimos, J. (2021). Analyst information acquisition via edgar. *Management Science*, 67(2):769–793.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408.

Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of economic theory*, 22(3):477–498.

Hong, H. and Stein, J. C. (2003). Differences of opinion, short-sales constraints, and market crashes. *The Review of Financial Studies*, 16(2):487–525.

Jones, C. M. and Lamont, O. A. (2002). Short-sale constraints and stock returns. *Journal of Financial Economics*, 66(2-3):207–239.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.

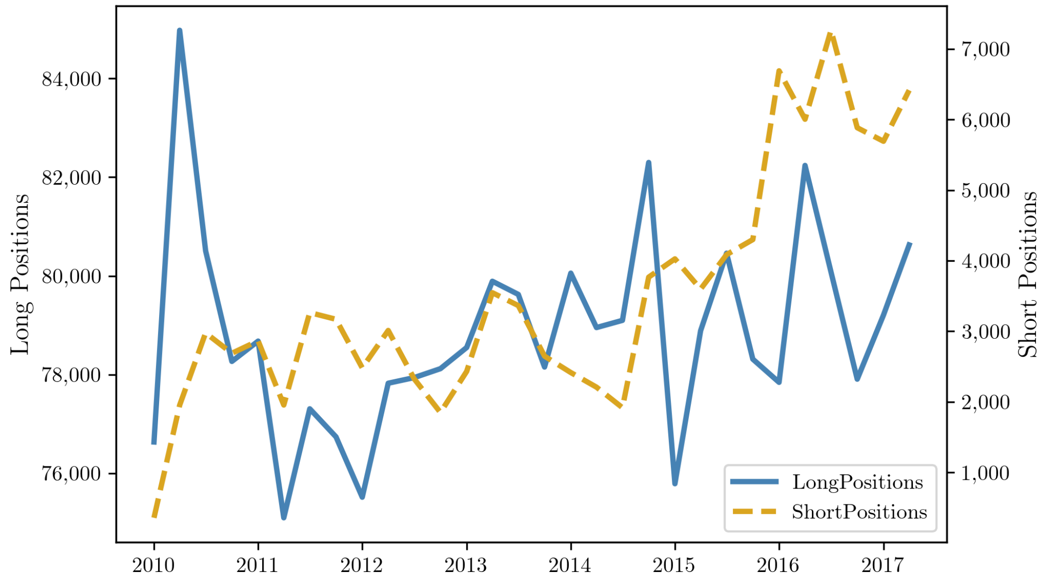
Lorden, G. (1970). On Excess Over the Boundary. *The Annals of Mathematical Statistics*, 41(2):520 – 527.

Seneca, J. J. (1967). Short interest: bearish or bullish? *The Journal of Finance*, 22(1):67–70.

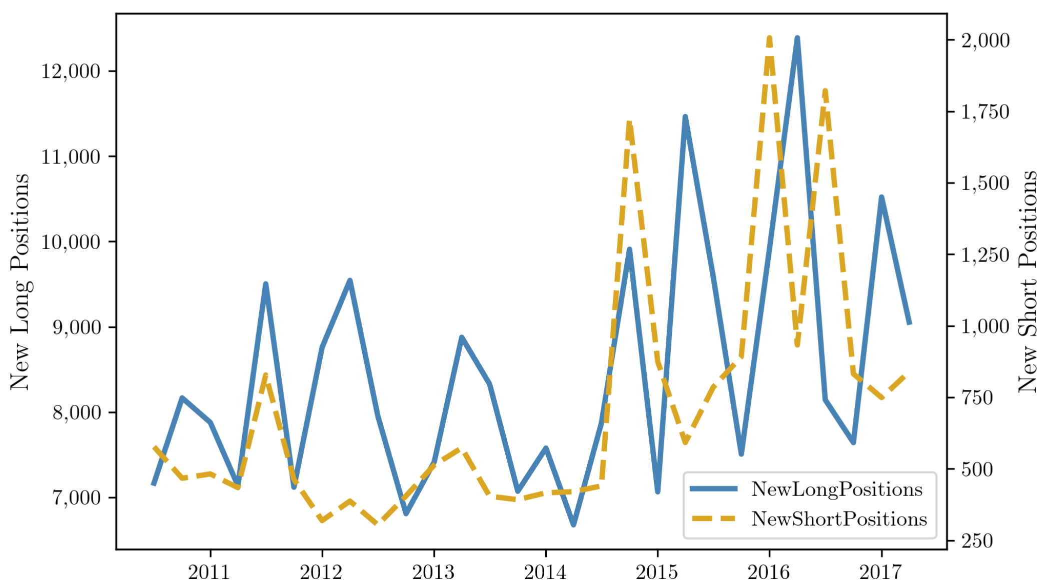
Sirri, E. R. and Tufano, P. (1998). Costly search and mutual fund flows. *The journal of finance*, 53(5):1589–1622.

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

Verrecchia, R. E. (1982). Information acquisition in a noisy rational expectations economy. *Econometrica: Journal of the Econometric Society*, pages 1415–1430.



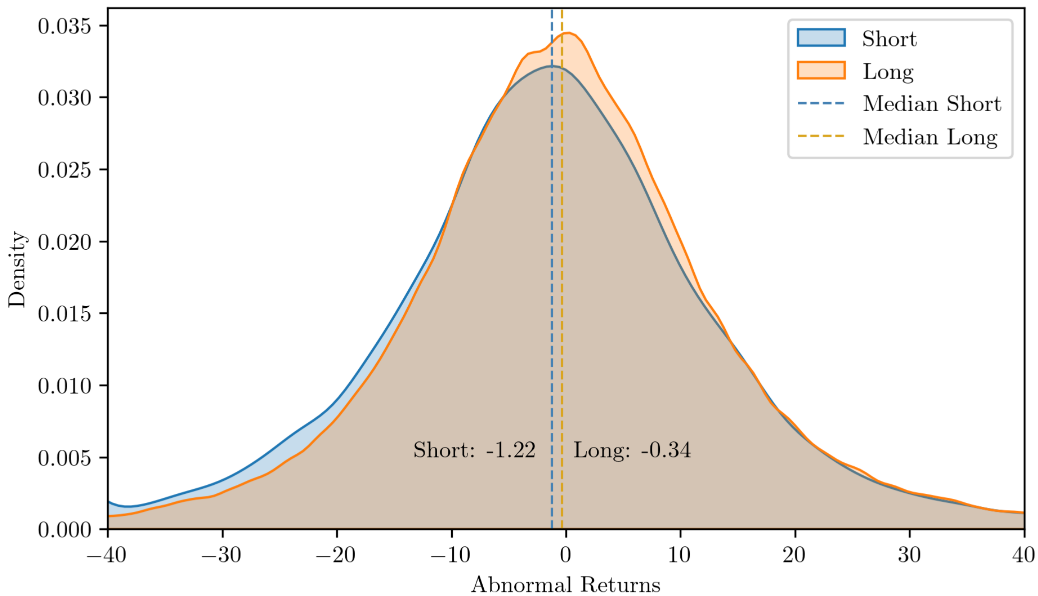
(A) All Long and Short Positions



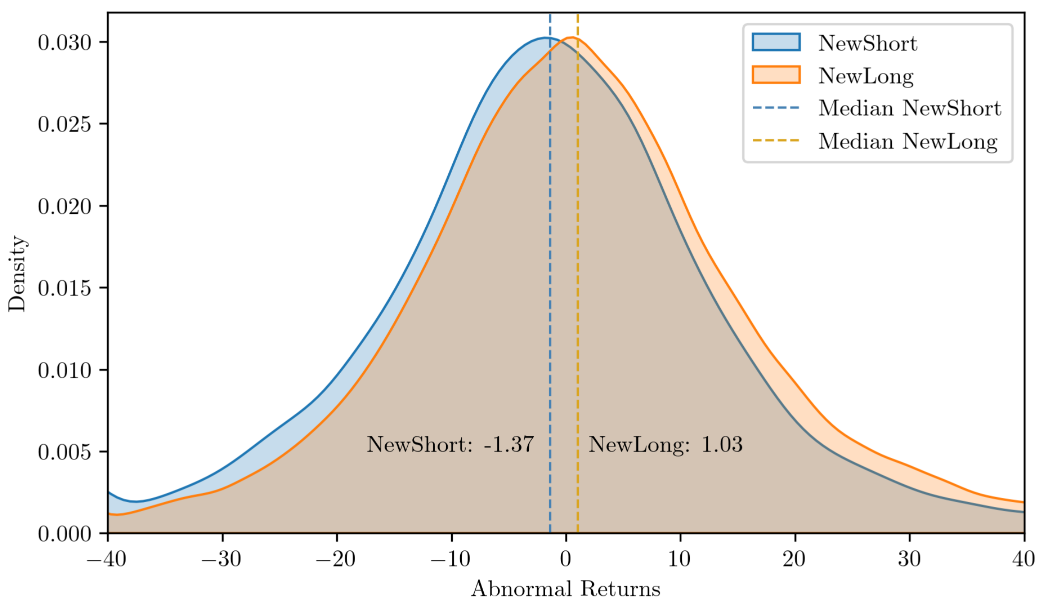
(B) New Long and New Short Positions

Figure 1: Time Series of Long and Short Positions

These time series plots show the changes in positions held by mutual funds over time. Panel A shows all long and short positions while Panel B shows the changes in new long and new short positions.



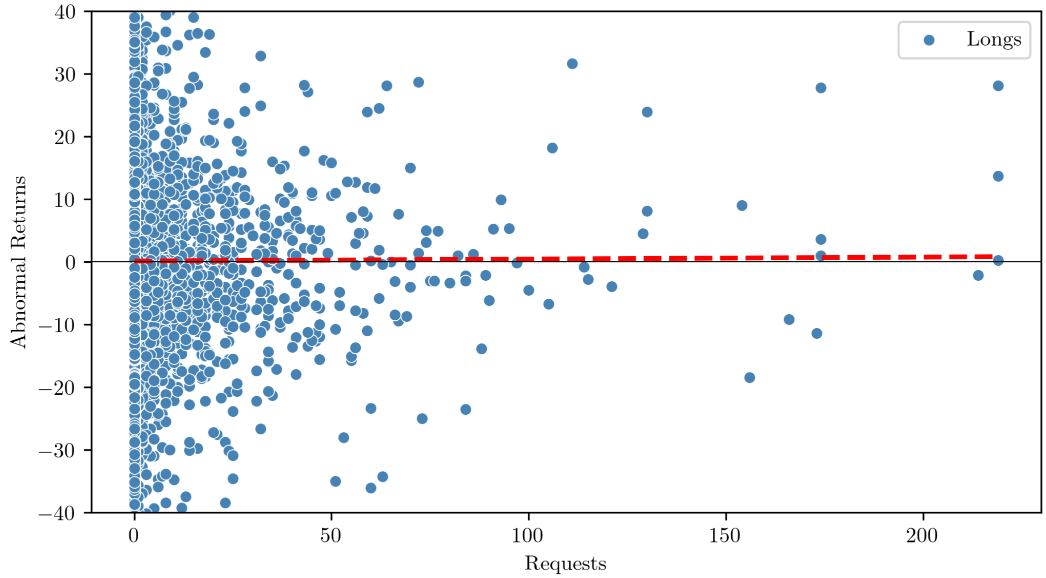
(A) All Long and Short Positions



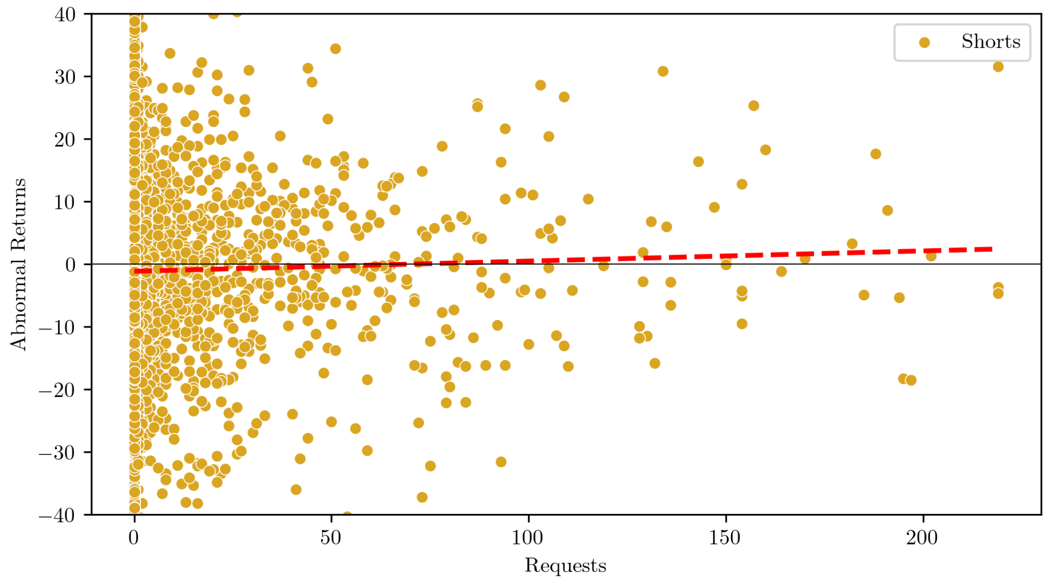
(B) New Long and New Short Positions

Figure 2: Histograms of Returns by Position

This figure shows the histogram of abnormal returns, as measured using equally-weighted DGTW returns, for long and short positions (Panel A) and for new long and new short positions (Panel B).

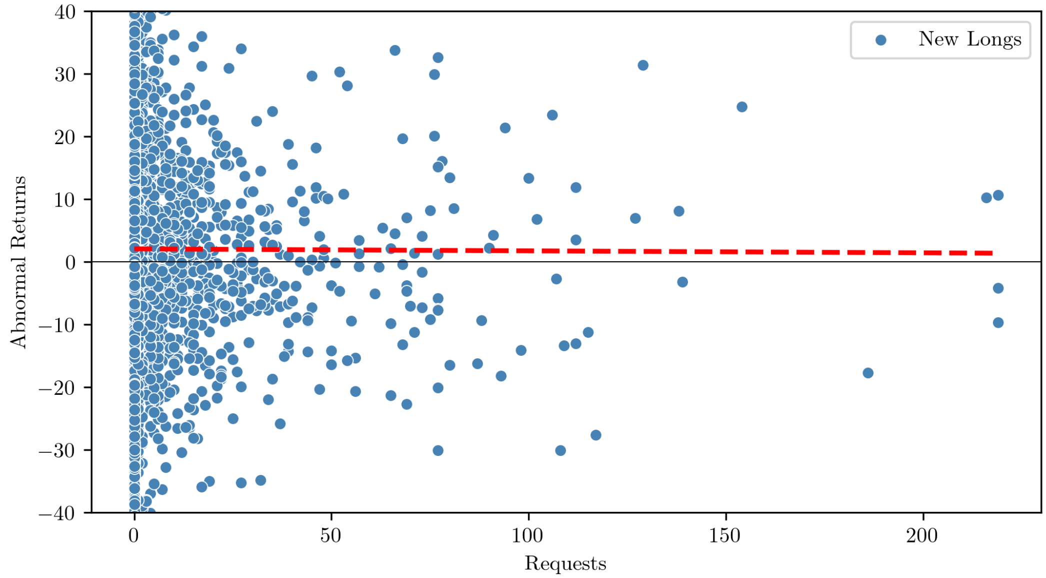


(A) All Long Positions

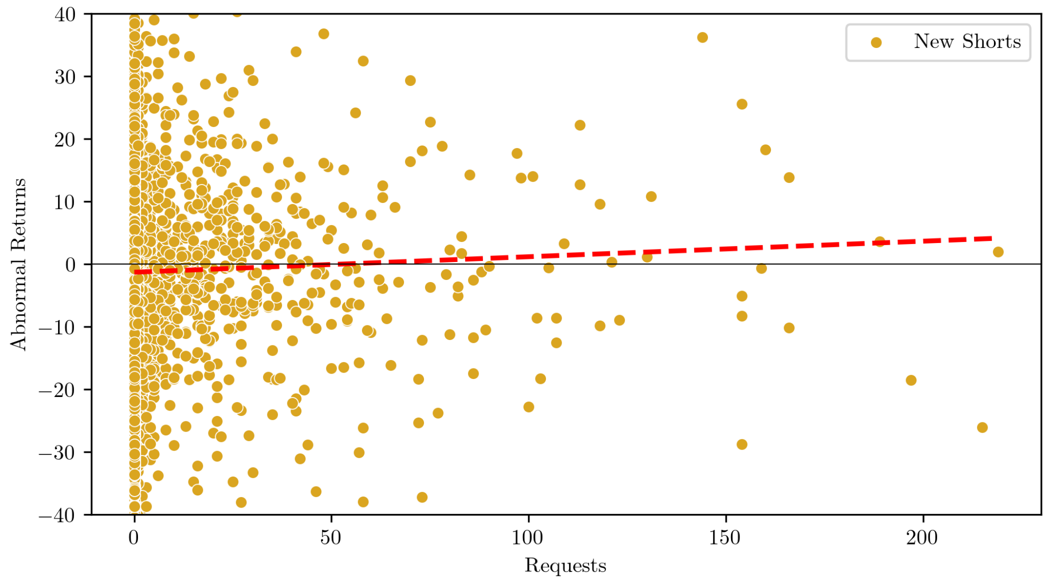


(B) All Short Positions

Figure 3: Scatter Plot of Requests and Abnormal Returns
This figure shows scatter plots of abnormal returns, as measured using equally-weighted DGTW returns, with requests for both long positions (Panel A) and short positions (Panel B).



(A) New Long Positions



(B) New Short Positions

Figure 4: Scatter Plot of Requests and Abnormal Returns: New Positions
This figure shows scatter plots of abnormal returns, as measured using equally-weighted DGTW returns, with requests for both new long positions (Panel A) and new short positions (Panel B).

Table 1: Summary Statistics

This table provides summary statistics for our sample. The subscripts for each variable refer to mutual fund family i, stock j, and quarter t. The entire panel includes $4,937,880$ observations and four variables are summarized over the entire panel. This table also provides summary statistics after collapsing the panel at the mutual fund family-by-quarter level (subscripts it) for which there are approximately $1,500$ observations, and after collapsing the panel at the stock-by-quarter level (subscripts jt) for which there are over $100,000$ observations for most of the variables. The period covers 30 quarters, from January 2010 through June 2017, and 115 different mutual fund families, of which 55 have at least one short position in our sample.

(1)	(2)	(3)	(4)	(5)
Variable	Mean	Std. Dev.	Median	N
Requests_{ijt}	1.92	21.93	0	4,937,880
$\mathbb{1}\{\text{Requests}_{ijt} \geq 1\}$	0.13	0.34	0	4,937,880
Long_{ijt}	0.37	0.48	0	4,937,880
Short_{ijt}	0.02	0.14	0	4,937,880
NewLong_{ijt}	0.04	0.20	0	4,937,880
NewShort_{ijt}	0.01	0.09	0	4,937,880
Requests_{it}	7,122	19,278	1,219	1,650
$\mathbb{1}\{\text{Requests}_{it} \geq 1\}$	0.89	0.32	1	1,650
TNAM_{it} (*millions*)	147,063	341,110	47,510	1,343
Funds_{it}	69	61	55	1,343
Stocks_{it}	185	177	111	1,179
Turnover_{it}	60	42	54	1,333
HHI_{it}	81	82	50	1,179
Expense Ratio_{it}	1.2	.33	1.2	1,333
Netflow12_{it}	0.002	0.018	0.001	1,336
MCAP_{jt} (*millions*)	4,486	18,555	591	163,526
Volume_{jt} (*thousands*)	1,210	5,608	234	163,526
ShortInt_{jt}	0.05	0.06	0.03	103,422
Mispricing_{jt}	50	13	50	75,907
Disagreement_{jt}	0.56	1.60	0.14	115,566
IdioVol_{jt}	0.02	0.02	0.02	102,946
News_{jt}	197	876	71	95,257
InstOwn_{jt} (*millions*)	98	344	24	79,075

Table 2: Information Acquisition and Positions

This table shows the results from estimating Equation 1. The dependent variable, $MFIA_{ijt}$, is either one quarter of requests ($Requests_{ijt}$) or two quarters of requests ($Requests_{ijt,t-1}$). In Panel A, the independent variables indicate whether family i has a long or short position in stock j in quarter t. In Panel B, the independent variables indicate whether family i has a *new* long or *new* short position in stock j in quarter t. In Columns 3 and 6, $MFIA$ is lagged one quarter while in Columns 4 and 7, $MFIA$ is lagged two quarters. The model includes family, stock, and quarter fixed effects. Standard errors are clustered by family-stock and quarter and are shown in parentheses. Indicators ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A. Long or Short Positions.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	MFIA = Requests_{ijt}			MFIA = $(\text{Requests}_{ijt,t-1})$		
	MFIA	L1.MFIA	L2.MFIA	MFIA	L1.MFIA	L2.MFIA
Long	0.11***	0.09***	0.08***	0.14***	0.12***	0.11***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Short	0.16***	0.15***	0.13***	0.20***	0.18***	0.16***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
$H_0 : Short - Long = 0$	0.05**	0.05***	0.05**	0.05**	0.05**	0.05**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Obs.	4,880,267	4,715,840	4,552,324	4,715,840	4,552,324	4,389,536
R^2	0.37	0.37	0.38	0.42	0.42	0.43
Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel B. New Long or New Short Positions.

	MFIA = Requests_{ijt}			MFIA = $(\text{Requests}_{ijt,t-1})$		
	MFIA	L1.MFIA	L2.MFIA	MFIA	L1.MFIA	L2.MFIA
NewLong	0.06***	0.01	-0.01	0.06***	0.00	-0.01
	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
NewShort	0.08***	0.06**	0.05**	0.09***	0.07**	0.05*
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
$H_0 : NewShort - NewLong$	0.02	0.05**	0.05**	0.03	0.07**	0.06**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Obs.	4,880,267	4,715,840	4,552,324	4,715,840	4,552,324	4,389,536
R^2	0.37	0.37	0.38	0.42	0.42	0.43
Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 3: Long and Short Positions and Returns

This table shows the results from estimating Equation 2. In the even-numbered columns, the dependent variable is the excess return of stock j during quarter t ($ExRet$) and the independent variables include quarterly risk-factors from the four-factor model (Fama and French (1993) and Carhart (1997)). In the odd-numbered columns, the dependent variable is the characteristics-adjusted return ($DGTW$). Columns 2 through 5 report α as the average abnormal return of the stocks in the sample. In Columns 6 and 7, α measures the abnormal return to stocks in long positions while $Short$ measures the additional return to stocks held in short positions. Thus, $Short + \alpha$ measures the total return to stocks in short positions. In addition to estimating over both long and short positions, the table also reports estimates when restricting the sample to only long positions and short positions. The model includes a mutual fund family fixed effect. Standard errors are clustered by stock and quarter and are shown in parentheses. Indicators ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Long Only		Short Only		All Positions	
	ExRet	DGTW	ExRet	DGTW	ExRet	DGTW
MktReturn	1.01***	0.03**	1.02***	0.05**	1.01***	0.03**
	(0.02)	(0.01)	(0.05)	(0.02)	(0.02)	(0.01)
SMB	0.54***		0.53***		0.54***	
	(0.04)		(0.07)		(0.04)	
HML	0.04		-0.10*		0.03	
	(0.03)		(0.06)		(0.03)	
UMD	-0.01		-0.20***		-0.02	
	(0.03)		(0.05)		(0.03)	
α	0.29*	0.08	-1.29***	-1.18***	0.29**	0.08
	(0.14)	(0.10)	(0.25)	(0.14)	(0.14)	(0.10)
Short					-1.61***	-1.33***
					(0.26)	(0.23)
$H_0 : Short + \alpha = 0$					-1.32***	-1.25***
					(0.31)	(0.17)
Obs.	1,721,856	1,317,857	47,118	35,367	1,768,976	1,353,224
R^2	0.19	0.00	0.16	0.00	0.19	0.00
Family FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 4: New Long and New Short Positions and Returns

This table shows the results from estimating Equation 2 on new positions. In the even-numbered columns, the dependent variable is the excess return of stock j during quarter t ($ExRet$) and the independent variables include quarterly risk-factors from the four-factor model (Fama and French (1993) and Carhart (1997)). In the odd-numbered columns, the dependent variable is the characteristics-adjusted return ($DGTW$). In the results below, α measures the abnormal return to stocks in new long positions while $NewShort$ measures the additional return to stocks held in new short positions. Thus, $NewShort + \alpha$ measures the total return to stocks in new short positions. Estimates are reported when using contemporaneous new positions, new positions from last quarter, and new positions from two quarters ago. The model includes a mutual fund family fixed effect. Standard errors are clustered by stock and quarter and are shown in parentheses. Indicators ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All Positions					
	Contemporaneous		New Position in $t-1$		New Position in $t-2$	
	ExRet	DGTW	ExRet	DGTW	ExRet	DGTW
MktReturn	1.03***	0.08***	1.05***	0.05***	1.06***	0.06**
	(0.02)	(0.02)	(0.04)	(0.02)	(0.04)	(0.02)
SMB	0.68***		0.60***		0.60***	
	(0.05)		(0.06)		(0.07)	
HML	0.05		-0.05		-0.05	
	(0.05)		(0.06)		(0.07)	
UMD	-0.00		-0.04		-0.03	
	(0.04)		(0.04)		(0.03)	
α	2.26***	2.08***	-0.52**	-0.44***	-0.81***	-0.60***
	(0.17)	(0.20)	(0.22)	(0.12)	(0.25)	(0.17)
NewShort	-3.64***	-3.37***	-0.93***	-0.69***	-0.20	-0.37*
	(0.35)	(0.36)	(0.30)	(0.25)	(0.22)	(0.21)
$H_0 : NewShort + \alpha = 0$	-1.38***	-1.29***	-1.45***	-1.13***	-1.01**	-0.96***
	(0.42)	(0.28)	(0.35)	(0.26)	(0.27)	(0.23)
Obs.	197,116	139,044	198,931	134,938	187,356	128,111
R^2	0.19	0.01	0.15	0.00	0.16	0.00
Family FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 5: Long and Short Positions, Returns, and Information Acquisition

This table shows the results from estimating Equation 3. In the even-numbered columns, the dependent variable is the excess return of stock j during quarter t (*ExRet*) and the independent variables include quarterly risk-factors from the four-factor model (Fama and French (1993) and Carhart (1997)). In the odd-numbered columns, the dependent variable is the characteristics-adjusted return ($DGTW$). In the results below, *Long* measures the abnormal return to stocks in long positions while $Short$ measures the additional return to stocks held in short positions. Further, $MFIA$ measures the effect of information acquisition on stocks held in long positions while the interaction term, $ShortXMFIA$, measures the additional effect of information acquisition on stocks in short positions. Thus, $ShortXMFIA+MFIA$ measures the total effect of information acquisition on the returns of stocks in short positions. The model includes mutual fund family and stock fixed effects. Standard errors are clustered by stock and quarter and are shown in parentheses. Indicators ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

(1)	(2)	(3)	(4)	(5)
	All Positions			
	MFIA = Requests_{ijt}		MFIA = $\text{Requests}_{ijt,t-1}$	
	ExRet	DGTW	ExRet	DGTW
MktReturn	1.01***	0.03**	1.01***	0.04***
	(0.02)	(0.01)	(0.02)	(0.01)
SMB	0.54***		0.53***	
	(0.04)		(0.04)	
HML	0.03		0.03	
	(0.04)		(0.04)	
UMD	-0.01		-0.01	
	(0.03)		(0.03)	
Long	0.24	0.03	0.27*	0.08
	(0.15)	(0.08)	(0.15)	(0.09)
Short	-1.28***	-1.09***	-1.29***	-1.12***
	(0.17)	(0.13)	(0.16)	(0.14)
MFIA	0.06*	0.08***	-0.04*	-0.01
	(0.03)	(0.03)	(0.02)	(0.02)
ShortXMFIA	0.20**	0.22**	0.18***	0.21***
	(0.08)	(0.08)	(0.06)	(0.07)
$H_0 : ShortXMFIA + MFIA = 0$	0.26***	0.30***	0.14**	0.19***
	(0.09)	(0.08)	(0.07)	(0.07)
Obs.	1,768,734	1,353,132	1,710,949	1,306,197
R^2	0.24	0.06	0.24	0.06
Family FE	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes

Table 6: New Long and New Short Positions, Returns, and Information Acquisition

This table shows the results from estimating Equation 3 on new positions. In the even-numbered columns, the dependent variable is the excess return of stock j during quarter t ($ExRet$) and the independent variables include quarterly risk-factors from the four-factor model (Fama and French (1993) and Carhart (1997)). In the odd-numbered columns, the dependent variable is the characteristics-adjusted return ($DGTW$). In the results below, $NewLong$ measures the abnormal return to stocks in long positions while $NewShort$ measures the additional return to stocks held in short positions. Further, $MFIA$ measures the effect of information acquisition on stocks held in long positions while the interaction term, $NewShortXMFIA$, measures the additional effect of information acquisition on stocks in short positions. Thus, $NewShortXMFIA + MFIA$ measures the total effect of information acquisition on the returns of stocks in short positions. The model includes mutual fund family and stock fixed effects. Standard errors are clustered by stock and quarter and are shown in parentheses. Indicators ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

(1)	(2)	(3)	(4)	(5)
	All Positions			
	MFIA = Requests_{ijt}		MFIA = $\text{Requests}_{ijt,t-1}$	
	ExRet	DGTW	ExRet	DGTW
MktReturn	1.04***	0.08***	1.04***	0.08***
	(0.02)	(0.02)	(0.02)	(0.02)
SMB	0.68***		0.68***	
	(0.05)		(0.05)	
HML	0.06		0.06	
	(0.06)		(0.06)	
UMD	-0.00		-0.00	
	(0.04)		(0.04)	
NewLong	2.13***	1.92***	2.18***	1.97***
	(0.16)	(0.16)	(0.17)	(0.17)
NewShort	-3.13***	-2.93***	-3.15***	-2.97***
	(0.33)	(0.31)	(0.33)	(0.32)
MFIA	0.32***	0.36***	0.10	0.14*
	(0.10)	(0.10)	(0.09)	(0.08)
NewShortXMFIA	0.12	0.18	0.12	0.20
	(0.17)	(0.15)	(0.14)	(0.12)
$H_0 : NewShortXMFIA + MFIA = 0$	0.45***	0.54***	0.22	0.34***
	(0.17)	(0.14)	(0.15)	(0.13)
Obs.	196,404	138,732	196,404	138,732
R^2	0.26	0.10	0.26	0.10
Family FE	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes

Table 7: Information Acquisition and Benchmark-Adjusted Returns

This table shows results from estimating Equation 4 on the quarterly sample of mutual fund family returns. The dependent variable is the average benchmark-adjusted return across the funds of family i during quarter t. The independent variable of interest, $MFIA_{it}$ is the number of requests made by family i during quarter t. The table uses $MFIA$ from the same quarter as when the returns are earned and in the previous four quarters. The other independent variables are described in Appendix A.2. The model includes quarter fixed effects. Standard errors are clustered by family and quarter and are shown in parentheses. Indicators ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

(1)	(2)	(3)	(4)	(5)	(6)
	L0.MFIA	L1.MFIA	L2.MFIA	L3.MFIA	L4.MFIA
	Return	Return	Return	Return	Return
MFIA	0.016*	0.014*	0.013*	0.011*	0.008
	(0.008)	(0.008)	(0.007)	(0.006)	(0.005)
TNAM	0.003	0.004	0.004	0.004	0.005
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
NetFlow12	-0.259	-0.227	-0.188	-0.138	-0.071
	(1.150)	(1.121)	(1.124)	(1.121)	(1.144)
HHI	-0.071***	-0.070***	-0.070***	-0.069***	-0.068***
	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)
ExpenseRatio	-0.090***	-0.089***	-0.089***	-0.088***	-0.087***
	(0.025)	(0.024)	(0.025)	(0.025)	(0.026)
Turnover	0.041	0.041	0.041	0.041	0.042
	(0.034)	(0.034)	(0.034)	(0.034)	(0.035)
Obs.	1,169	1,169	1,169	1,169	1,169
R^2	0.09	0.09	0.09	0.09	0.09
Quarter FE	Yes	Yes	Yes	Yes	Yes

Table 8: Mispricing and Returns

This table shows the results from estimating Equation 5 on the quarterly sample of stock returns. In Columns 2 through 4, the dependent variable is the excess return of stock j during quarter t ($ExRet$) and the independent variables (suppressed in the output) include quarterly risk-factors from the four-factor model (Fama and French (1993) and Carhart (1997)). In Columns 5 through 7, the dependent variable is the characteristics-adjusted return ($DGTW$) from (Daniel et al. (1997)). The indicator variable $L1.Underpriced$ equals one if stock j's mispricing score was in the bottom quintile in quarter $t-1$. The indicator variable $L1.Overpriced$ equals one if stock j's mispricing score was in the top quintile in quarter $t-1$. Standard errors are clustered by stock and quarter and are shown in parentheses. Indicators ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	ExRet	ExRet	ExRet	DGTW	DGTW	DGTW
L1.Underpriced	0.22	0.56		0.34	0.61**	
	(0.34)	(0.41)		(0.24)	(0.27)	
L1.Overpriced	-1.34**		-1.39**	-1.10**		-1.19**
	(0.60)		(0.63)	(0.43)		(0.44)
Obs.	75,218	75,218	75,218	72,648	72,648	72,648
R^2	0.18	0.18	0.18	0.00	0.00	0.00

Table 9: Information Acquisition and Clear Winners

This table shows results from estimating Equation 6 on new positions. The dependent variable, $MFIA$, is the number of requests from family i about stock j in quarter t. The independent variables $LowDisagree$, $Underpriced$, $Overpriced$, and their interactions identify whether stocks are clear winners in a given quarter. The model always includes mutual fund family and quarter fixed effects and includes stock fixed effects where noted. Standard errors are clustered by family-stock and quarter and are shown in parentheses. Indicators ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	New Long Only			New Short Only		
	MFIA	MFIA	MFIA	MFIA	MFIA	MFIA
LowDisagreeXUnderpriced	-0.04***	0.00	0.00			
	(0.01)	(0.01)	(0.01)			
LowDisagree			0.00			-0.00
			(0.01)			(0.03)
Underpriced			0.00			
			(0.01)			
Overpriced						-0.00
						(0.02)
LowDisagreeXOverpriced				-0.08**	-0.12**	-0.11*
				(0.04)	(0.05)	(0.06)
Obs.	110,887	110,770	110,770	13,600	13,125	13,125
R^2	0.45	0.48	0.48	0.63	0.66	0.66
Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	No	Yes	Yes	No	Yes	Yes

Online Appendix

This appendix provides additional details to supplement the main text. There are two sections of this appendix. Appendix A.1 contains details about the EDGAR Log Files and the process of unmasking mutual fund IP addresses. Appendix A.2 contains a table detailing the variables used in this paper. Appendix A.3 contains model details and extended proofs.

A.1 Unmasking IP Addresses in EDGAR Log Files

The EDGAR Log Files contain billions of observations of "requests" or "requests to view a filing." Each observation details the filing requested (accession number), the date and time of the request, and the requester (the IP address making the electronic request). A snapshot of the raw EDGAR Log Files is shown below:

IP Address	Date	Time	CIK	Accession Number
38.97.91.ecg	20170531	09:47:33	051143	000104746917001061
38.65.241.fhf	20170531	11:07:28	274191	000002741917000008
67.199.249.igg	20170624	12:27:02	320193	000032019317000009
216.223.41.aah	20170624	16:12:55	831259	000083125917000016

Given our focus in this paper on the requester, linking the masked IP addresses to identifiable investors (e.g., mutual fund families) is pivotal to our study. To unmask the IP addresses, we first notice the fourth octet in the examples above.[1] In place of the actual digits of the requesting IP address, the fourth octet is reported as a set of three letters. However, organizations typically register blocks of IP addresses, with the most common block fixing the first three octets and containing all 256 versions of the fourth octet.[2] In other words, only the first three octets are necessary to identify the organization that has registered that block of IP addresses.

Using this insight, we searched historical IP address registration records from 2010 through 2017 to identify the blocks of IP addresses registered to investment firms.[3] Then, using this hand-collected mapping between investment firms and IP addresses, we unmask

[1]For IP addresses, an *octet* is a group of eight bits, or the one to three digit numbers (from 0 to 255) separated by periods in the examples above.

[2]For example, all 256 IP addresses beginning with 38.97.91 will be registered to the same organization.

[3]IP registration records were acquired from MaxMind, https://www.maxmind.com/en/home.

the requesters in the EDGAR Log Files. As a result, the snapshot of raw data from above has been transformed into the following.

Investment Firm	Date	Time	Ticker	Filing
Abrams Capital	20170531	09:47:33	IBM	10-K for 2016
Harbor Capital	20170531	11:07:28	TGT	10-K for 2016
Crabel Capital	20170624	12:27:02	AAPL	10-Q for Q2 2017
Ronin Capital	20170624	16:12:55	FCX	Earnings for Q2 2017

Furthermore, the three letters used to mask the fourth octet is static, not dynamic. This means, for example, that def replaces the digits 146 for every instance of 146. This allows us to identify unique IP addresses. In other words, though an unmasked mutual fund may make 50 requests one day, we can observe how many different IP addresses made those requests.

Finally, we have adjusted the data to remove likely bots. As mentioned, the raw EDGAR Log Files contain billions of requests with many thousands of requests per day coming from single IP addresses. It is unlikely that these thousands of requests per day represent a human actually clicking on documents in EDGAR. It is much more likely that they represent computer programs (bots) downloading large quantities of data at a time. Given these IP addresses do not fit with the spirit of our research, we remove them from the data. The removal process is as follows: we remove IP addresses that either (i) make over $1,000$ requests in a day or (ii) make requests for over 100 different CIKs (i.e., firms).

A.2 Variable Details

Variable	**Description**
Information Acquisition (MFIA)	Measured at mutual fund family-stock-quarter (ijt) level.
Requests_{ijt}	The sum of the number of requests made by mutual fund family i for filings about stock j in quarter t. The sum is winsorized at the 99th percentile and scaled by adding one and taking the natural log. This data is primarily derived from the EDGAR Log Files.
$\text{Requests}_{ijt,t-1}$	The sum of the number of requests made by mutual fund family i for filings about stock j in quarter t and in quarter $t-1$. The sum is winsorized at the 99th percentile and scaled by adding one and taking the natural log. This data is primarily derived from the EDGAR Log Files.
Position Variables	Measured at mutual fund family-stock-quarter (ijt) level.
Long_{ijt}	Indicator equal to 1 if any of the mutual funds within mutual fund family i hold shares of stock j in a long position at the end of quarter t. We exclude all observations where mutual fund family i holds long and short positions of stock j during the same quarter. This data is derived from the CRSP Mutual Funds database.
Short_{ijt}	Indicator equal to 1 if any of the mutual funds within mutual fund family i hold shares of stock j in a short position at the end of quarter t. We exclude all observations where mutual fund family i holds long and short positions of stock j during the same quarter. This data is derived from the CRSP Mutual Funds database.
NewLong_{ijt}	Indicator equal to 1 if $Long_{ijt} = 1$ and $Long_{ijt-1} = 0$.
NewShort_{ijt}	Indicator equal to 1 if $Short_{ijt} = 1$ and $Short_{ijt-1} = 0$.
Stock Characteristics	Measured at stock-quarter (jt) level.
ShortInt_{jt}	The number of shares held short divided by the number of shares outstanding. This ratio is winsorized at the 99th percentile and scaled by adding one and taking the natural log. This data is derived from the Compustat database.
IdioVol_{jt}	The standard deviation of daily abnormal returns for stock j during quarter t, winsorized at the 99th percentile. Abnormal returns are estimated using the five-factor model with momentum. This data is derived from the CRSP database as well as Kenneth French's website.
Disagreement_{jt}	The standard deviation of analyst expectations from the IBES database. The standard deviation is winsorized at the 99th percentile and scaled by adding one and taking the natural log.
News_{jt}	The total number of news articles and press releases for stock j during quarter t based on the RavenPack database. The sum of news articles is winsorized at the 99th percentile and scaled by adding one and taking the natural log.
MCAP_{jt}	The total market capitalization of firm j calculated as the share price multiplied by shares outstanding. The product is scaled by taking the natural log.
InstOwn_{jt}	The number of shares owned by institutions which file 13F reports, winsorized at the 99th percentile and scaled by adding one and taking the natural log. This data is derived from the Thomson Reuters 13F database.
Volume_{jt}	Share volume traded during quarter t, winsorized at the 99th percentile and scaled by taking the natural log. This data is derived from CRSP.
Mispricing_{jt}	Average mispricing score for quarter t. The score is from Stambaugh and Yuan (2017). For interpretation, we divide the score by 100 and subtract 0.50 to center on zero.

Underpriced_{jt}	Indicator equal to 1 if Mispricing_{jt} is in the bottom quintile of all stocks in quarter t.
Overpriced_{jt}	Indicator equal to 1 if Mispricing_{jt} is in the top quintile of all stocks in quarter t.
ExRet_{jt}	The excess return of stock j during quarter t. The excess return is calculated as the raw return less the risk-free rate. The excess return is multiplied by 100 for interpretability and is winsorized at the 1st and 99th percentiles.
DGTW_{jt}	The abnormal return of stock j during quarter t calculated as the difference between stock j's raw return less the appropriate equally-weighted DGTW benchmark return (Daniel et al. (1997)). The abnormal return is multiplied by 100 for interpretability and is winsorized at the 1st and 99th percentiles. This return is referred to as the "characteristic-adjusted return."
LowDisagree_{jt}	Indicator equal to 1 if Disagreement_{jt} is in the bottom quintile during quarter t.
$\text{LowDisagreeXUnderpriced}_{jt}$	Indicator equal to 1 if Underpriced_{jt} is 1 and stock j has low disagreement. Low disagreement is defined as Disagreement_{jt} in the bottom quintile during quarter t.
$\text{LowDisagreeXOverpriced}_{jt}$	Indicator equal to 1 if Overpriced_{jt} is 1 and stock j has low disagreement. Low disagreement is defined as Disagreement_{jt} in the bottom quintile during quarter t.

Family Characteristics	Measured at the family-quarter (it) level.
TNAM_{it}	Total net assets under management. Scaled by taking the natural log.
HHI_{it}	Herfindahl-Hirschman Index to measure portfolio concentration. Calculated as the sum of the squared portfolio share of each holding within a portfolio. Winsorized at the 99th percentile and divided by 100 for interpretability.
$\text{Expense Ratio}_{it}$	Total annual expenses and fees divided by year-end TNA. Winsorized at the 99th percentile.
Turnover_{it}	Minimum of aggregate purchases and sales of securities divided by average TNA over the calendar year. Winsorized at the 99th percentile and scaled using the natural logarithm.
Netflow12_{it}	The net growth in fund assets beyond reinvested dividends (Sirri and Tufano (1998)) over the past one year. Winsorized at the 1st and 99th percentiles.
Return_{it}	Measures the risk-adjusted returns for a given fund compounded over a six-month period. One risk-adjustment is to use "benchmark adjusted returns," where appropriate benchmark return is subtracted from the raw return of the fund. The second risk-adjustment is to account for the four-factor (FF4) model.

Macro Variables	Measured at the quarter (t) level.
MkReturn_{t}	The quarterly return on the market, from Kenneth French's database.
HML_{t}	The quarterly high-minus-low factor, from Kenneth French's database.
SMB_{t}	The quarterly small-minus-big factor, from Kenneth French's database.
UMD_{t}	The quarterly momentum factor, from Kenneth French's database.

A.3 Model Details and Proofs

This section provides derivations underlying the theoretical framework developed in Section 4 of the main text. We proceed in three parts. Section A.3.1 presents the baseline one-shot information-acquisition model, proves equilibrium existence and uniqueness, and establishes Propositions 1–3. Section A.3.2 extends the environment to sequential-signal acquisition, derives the optimal stopping rule, and proves Proposition 4. Section A.3.3 provides details on the minimum tranche size ($q_{\min}$) assumption.

Throughout the appendix we retain the notation and parameterization introduced in the main text unless explicitly noted otherwise.

A.3.1 Baseline Model

Equilibrium Existence and Uniqueness

The risky asset has a binary fundamental value ($V \in \{\nu_H, \nu_L\}$) and $\Delta \equiv \nu_H - \nu_L > 0$ is the value gap. A risk-neutral investor may observe a signal of value by choosing precision $\pi \in [0, 1]$ and paying cost $c(\pi) = \kappa\pi^2$, $\kappa > 0$. Conditional on V, the signal $x \in \{\nu_H, \nu_L\}$ satisfies $P(x = V) = \frac{1}{2} + \frac{\pi}{2}$. The posterior mean therefore lies $\frac{\pi\Delta}{2}$ above or below the pre-trade price ($P_0 = \frac{\nu_H + \nu_L}{2}$), which is pinned down by competitive noise traders at the unconditional mean. After observing the signal, the investor selects position side $s \in \{L \text{ (long)}, S \text{ (short)}, 0\}$ and position size q.

Opening any position incurs fixed monitoring costs $F > 0$, and investors abstain from trading or trade a minimal block $q_{\min} > 0$. Trading costs for long positions total $C_L = F$ while for short positions $C_S = F + \phi + \eta q_{\min}$ where both $\phi > 0$ and $\eta q_{\min} > 0$.

Lemma A1. *For any chosen precision π, optimal trade decisions are:*

$$\textit{Go Long iff } x = \nu_H \textit{ and } \frac{\pi\Delta}{2} q_{\min} \geq C_L, \qquad \textit{Go Short iff } x = \nu_L \textit{ and } \frac{\pi\Delta}{2} q_{\min} \geq C_S.$$

Proof. Conditional expected profit from a long after a high signal is $\frac{\pi\Delta}{2} q_{\min} - C_L$. The investor trades only when this is non-negative; the symmetric argument gives the short cut-off. ■

Define the precision hurdles for long and short trades:

$$\pi_L^{thr} \equiv \frac{2C_L}{\Delta q_{\min}}, \qquad \pi_S^{thr} \equiv \frac{2C_S}{\Delta q_{\min}} > \pi_L^{thr}, \tag{A1}$$

and let the investor's *ex-ante* expected profit from optimally trading after observing x be:

$$\Gamma(\pi) \equiv \frac{1}{2}\Big[q_{\min}\frac{\pi\Delta}{2} - C_L\Big]_+ + \frac{1}{2}\Big[q_{\min}\frac{\pi\Delta}{2} - C_S\Big]_+ - c(\pi),$$

where $[z]_+ \equiv \max\{z, 0\}$. There are three relevant ranges for π:

- For $\pi \leq \pi_L^{thr}$: $\Gamma(\pi) = -c(\pi) < 0$.
- For $\pi \in (\pi_L^{thr}, \pi_S^{thr}]$: only long trades can be taken, so $\Gamma(\pi) = \frac{1}{2}(q_{\min}\frac{\pi\Delta}{2} - C_L) - c(\pi)$.
- For $\pi > \pi_S^{thr}$: both long and short trades can be taken, so $\Gamma(\pi) \equiv \frac{1}{2}(q_{\min}\frac{\pi\Delta}{2} - C_L) + \frac{1}{2}(q_{\min}\frac{\pi\Delta}{2} - C_S) - c(\pi)$.

Because $c(\pi)$ is strictly convex and the linear benefit pieces are continuous with kinks at both π_L^{thr} and π_S^{thr}, $\Gamma(\pi)$ is strictly concave on each region and continuous everywhere, so it attains a unique global maximizer $\pi^* \equiv \arg\max_{\pi \geq 0} \Gamma(\pi)$. If the maximal expected profit is non-positive, $\Gamma(\pi^*) \leq 0$, the investor acquires no information and never trades; otherwise she chooses the interior optimum $\pi^* > 0$ and follows the cut-off rule in Lemma A1. This proves **existence** and **uniqueness** of the equilibrium. ■

Proof of Proposition 1

Proposition 1 states that the precision threshold is higher for shorts than for longs: $\pi_S^{thr} > \pi_L^{thr}$.

Proof. Precision thresholds for longs and shorts are defined in Equation A1. By construction, $C_S = C_L + \phi + \eta q_{\min} > C_L$ where the locate fee and borrow surcharge are both greater than zero ($\phi, \eta q_{\min} > 0$). Therefore, $\pi_S^{thr} > \pi_L^{thr}$. ■

Proof of Proposition 2

Proposition 2 states that the average precision (or research intensity) for observed shorts is higher than for observed longs.

Proof. Recall the two precision thresholds ($\pi_L^{thr} = \frac{2C_L}{\Delta q_{\min}}$ and $\pi_S^{thr} = \frac{2C_S}{\Delta q_{\min}} > \pi_L^{thr}$), let the gross trading benefit be defined as $B(\pi) = \frac{1}{2}\left[q_{\min}\frac{\pi\Delta}{2} - C_L\right]_+ + \frac{1}{2}\left[q_{\min}\frac{\pi\Delta}{2} - C_S\right]_+$, and note that the slope of this benefit (the marginal benefit curve), with respect to chosen precision π, is a step-function with three flat regions:

- $B'(\pi) = 0$ for $\pi \leq \pi_L^{thr}$;
- $B'(\pi) = b_L = \frac{q_{\min}\Delta}{4}$ on $(\pi_L^{thr},\ \pi_S^{thr}]$ (only longs can cover fixed costs);
- $B'(\pi) = b_S = 2b_L = \frac{q_{\min}\Delta}{2}$ when $\pi > \pi_S^{thr}$ (both longs and shorts can cover fixed costs).

Also note that the investor's net objective, the ex-ante expected profit, is $\Gamma(\pi) = B(\pi) - \kappa\pi^2$ and the marginal cost is $2\kappa\pi$.

Matching the marginal-benefit step-function to the upward sloping marginal cost curve gives two candidate interior optima:

$$\pi^L(\kappa) = \frac{q_{\min}\Delta}{8\kappa}, \quad \pi^S(\kappa) = \frac{q_{\min}\Delta}{4\kappa};$$

which are feasible only if they lie on their intended segments. Define

$$\kappa_{high} = \frac{(q_{\min}\Delta)^2}{16C_L}, \quad \kappa_{low} = \frac{(q_{\min}\Delta)^2}{8C_S}$$

and note that $\kappa_{low} < \kappa_{high}$ whenever $C_S > 2C_L$. Thus, whenever short-selling costs are significantly higher than the costs of long positions, the investor's optimal precision falls into exactly one of the following cost bands.

- **High-cost band**. When $\kappa > \kappa_{high}$ the marginal cost curve is always higher than the marginal-benefit step function and, thus, the optimal precision is $\pi^* = 0$ and neither long or short trades are executed.[4]

- **Moderate-cost band**. When $\kappa_{low} < \kappa < \kappa_{high}$ the marginal cost curve intersects the marginal-benefits step-function only in the long-only region, thus, the optimal precision is $\pi^* = \pi^L(\kappa) \in (\pi_L^{thr}, \pi_S^{thr})$ and only long trades are executed.

- **Low-cost band**. When $\kappa < \kappa_{low}$ the marginal cost curve intersects the marginal-benefits step-function in the short region (or is always below b_S), thus, the optimal precision is $\pi^* = \pi^S(\kappa) \geq \pi_S^{thr}$ and both longs and shorts are executed.

[4]At $\kappa = \kappa_{high}$ the investor is indifferent between $\pi = 0$ and π_L^{thr}; we take $\pi^* = 0$ without loss of generality.

Shorts are only observed in the low-cost band and are always backed by high precision ($\pi^* \geq \pi_S^{thr}$) while longs are observed in both the low-cost band and the moderate-cost band, thus some longs are backed by lower precision ($\pi_L^{thr} < \pi^* < \pi_S^{thr}$). Hence

$$E[\pi^* \mid s = S] \geq E[\pi^* \mid s = L],$$

with strict inequality when $C_S > 2C_L$ (so the moderate-cost band has positive width). ■

Proof of Proposition 3

Proposition 3 states that the average absolute abnormal return for observed shorts is higher than for observed longs.

Proof. After observing the binary signal $x \in \{\nu_H, \nu_L\}$ the investor's posterior mean is $E[V|x] = P(V = \nu_H|x)\nu_H + P(V = \nu_L|x)\nu_L$, so the perceived deviation from price is always $|E[V|x] - P_0| = \frac{\pi\Delta}{2}$. This is the size of the "gap" the investor believes exists between the fundamental value and the market price.

A trade of side $s \in \{L, S\}$ is executed only if the expected gross gain on the minimum block covers the fixed cost C_s; Lemma 1 therefore implies the return hurdle $|E[V|x] - P_0| \geq \frac{C_s}{q_{\min}} = \tau_s$. Because $\tau_S > \tau_L$, a short is opened only when the perceived mispricing is strictly larger than the minimum mispricing that triggers a long.

Shorts appear only in the low-cost band identified in Proposition 2; there the chosen precision satisfies $\pi^* \geq \pi_S^{thr}$, hence the mispricing at entry obeys $|E[V|x] - P_0|_{s=S} \geq \frac{\pi_S^{thr}\Delta}{2}$. Longs are observed (i) in the same low-cost band when $\pi^* \geq \pi_S^{thr}$, and (ii) in the moderate-cost band when $\pi_L^{thr} < \pi^* < \pi_S^{thr}$. In the second scenario, the mispricing lies strictly below $\frac{\pi_S^{thr}\Delta}{2}$. Hence the distribution of mispricings for longs is a mixture of large mispricings (from the low-cost band) and strictly smaller ones (from the moderate-cost band), whereas the short-side distribution contains only the large mispricings.

The realized abnormal return on a unit position is $\alpha = V - P_0$. Conditional on the investor's entry rule, its *absolute* expectation is monotone in the perceived mispricing. Because a correct-side payoff of $+\frac{\Delta}{2}$ occurs with probability $\frac{1+\pi}{2}$ while a wrong-side payoff $-\frac{\Delta}{2}$ occurs with probability $\frac{1-\pi}{2}$, larger perceived mispricing at entry translates into larger average $|\alpha|$. Therefore, $E[|\alpha| \big| s = S] \geq E[|\alpha| \big| s = L]$, with strict inequality when $C_S > 2C_L$ (so the moderate-cost band has positive width). ■

A.3.2 Extended Model: Sequential Signals

Equilibrium Existence and Uniqueness

Let the investor observe sequential Gaussian signals $x_j = V + \varepsilon_j,\ \varepsilon_j \sim \mathcal{N}(0, \sigma^2)$, each costing $\kappa\pi_0^2$ where $\pi_0 = \frac{1}{\sigma^2}$. After N draws the posterior mean and precision are $m_N = \frac{1}{N}\sum_{j=1}^{N} x_j$, and $\pi_N = N\pi_0$. Let $V_N(m)$ be the investor's *continuation value* after N draws and posterior mean m. At that node the investor chooses to either (i) **trade:** take side s on the minimum block $q_{\min}$ if $|m| \geq \tau_s$, earning $q_{\min}(|m| - \tau_s)$; (ii) **draw:** pay the cost $c = \kappa\pi_0^2$ and receive $E[V_{N+1}(m')]$; or (iii) **stop:** do nothing and receive value 0. Hence,

$$V_N(m) = \max\Big\{0,\ q_{\min}(|m| - \tau_L),\ q_{\min}(|m| - \tau_S),\ E[V_{N+1}(m')] - c\Big\}. \tag{A2}$$

Because the Gaussian update is linear and costs are constant, $V_N(\cdot)$ is even and strictly increasing in $|m|$ once $|m| \geq \tau_L$. This implies thresholds $0 < \hat{m}_N^L \leq \hat{m}_N^S$ such that the unique optimal action is **draw** if $|m| < \hat{m}_N^L$ and $E[V_{N+1}(m')] - c > 0$; **quit** if $|m| < \hat{m}_N^L$ and $E[V_{N+1}(m')] - c \leq 0$; **long** if $\hat{m}_N^L \leq m < \hat{m}_N^S$; and **short** if $|m| \geq \hat{m}_N^S$. Posterior variance is monotone decreasing in N, so the short boundary $h_N := \hat{m}_N^S$ is weakly increasing. Suppose

$h_1 > \tau_S$. When $\tau_S = |m_1| < h_1$ the investor would strictly improve by shorting now rather than paying c to continue, contradicting optimality. Hence, $h_1 = \tau_S$ and, by monotonicity, $h_N = \tau_S$ for every N; the long boundary must equal τ_L by the same argument. Therefore, the optimal stopping rule is: (i) continue sampling while $|m_N| < \tau_L$ and the option value of another draw exceeds its cost; (ii) quit with no position if $|m_N| < \tau_L$ but the option value of another draw has fallen to zero; (iii) trade a long block as soon as $\tau_L \leq |m_N| < \tau_S$; and (iv) trade a short block as soon as $|m_N| \geq \tau_S$.

For this rule the investor is indifferent at the thresholds by definition of $\tau_s = \frac{C_s}{q_{\min}}$, and market makers break even at $P_0 = \frac{\nu_H + \nu_L}{2}$. Thus (τ_L, τ_S, P_0) constitutes an equilibrium. Now, assume a distinct pair $(\tilde{\tau}_L, \tilde{\tau}_S) \neq (\tau_L, \tau_S)$ also supports an equilibrium. If $\tilde{\tau}_S < \tau_S$, lowering the short boundary lets the investor save at least one signal cost while never reducing trade pay-offs, strictly increasing expected profit—a contradiction. If $\tilde{\tau}_S > \tau_S$ or the long boundary moves, a symmetric contradiction arises. Hence, the boundary pair (τ_L, τ_S) is unique. ■

Proof of Proposition 4

Proposition 4 states that the average absolute abnormal return for observed shorts decreases in the number of signals acquired.

Proof. Let the posterior dollar gap be $g_N \equiv |m_N - P_0|$ and its conditional standard deviation be $\sigma/\sqrt{N}$. A short block is traded the first time the gap reaches the fixed hurdle $g_N \geq \tau_S$ where $\tau_s = \frac{C_s}{q_{\min}}$. Denote this draw count by T_S.

Conditional on trading after exactly N signals ($T_S = N$), the gap must lie in $[\tau_S, \infty)$, with a truncated-normal tail. A standard formula gives the conditional mean of the overshoot:

$$E[g_N - \tau_S \mid T_S = N] = \left(\frac{\sigma}{\sqrt{N}}\right)\frac{\varphi(a_N)}{\Phi(-a_N)}, \quad a_N \equiv \frac{\tau_S\sqrt{N}}{\sigma}, \qquad \text{(A3)}$$

where φ and Φ are the standard-normal pdf and cdf.

Because $\Phi(-a_N)$ grows and $\varphi(a_N)$ shrinks with a_n, the fraction is bounded above by $\frac{1}{\sqrt{2\pi}}$. Thus, $E[g_N \mid T_S = N] \leq \tau_S + \frac{\sigma}{\sqrt{2\pi N}}$ and the right side is strictly decreasing in N.

At settlement the one-block abnormal return is $\alpha = V - P_0$. Conditional on executing the short after precisely N filings, the posterior mean is m_N and the residual uncertainty in V is $\mathcal{N}(m_N, \frac{\sigma^2}{N})$. Therefore,[5]

$$E[|\alpha| \mid T_S = N] = \tau_S + \frac{\sigma}{\sqrt{2\pi N}} + \frac{\sigma\sqrt{2}}{\sqrt{\pi N}}, \tag{A4}$$

is the sum of three positive terms whose second and third pieces shrink at the rate $\frac{1}{\sqrt{N}}$. Hence, the whole expectation falls as N rises.

The details above focused on shorts. Long positions satisfy the same analytical bound as shorts (Equation A4) but with the lower hurdle $\tau_L < \tau_S$. The second and third terms of Equation A4—the overshoot of the posterior mean and the residual valuation noise—are identical on both sides and shrink like $\frac{1}{\sqrt{N}}$. What is different between longs and shorts is the baseline:

- **Shorts** must first clear the high dollar hurdle τ_S that reflects extra locate and borrow fees. Because τ_S materially exceeds τ_L, each additional signal cuts the *total* expected $|\alpha|$ by a perceptible amount, producing the downward slope highlighted in Proposition 4.

- **Longs** face the lower hurdle τ_L. Relative to the smaller baseline, the same $\frac{1}{\sqrt{N}}$ decline in the overshoot-and-noise terms is proportionally tiny, so the model predicts an almost

[5] See Lorden (1970).

flat relation between information acquired and subsequent absolute returns on the long side. ■

A.3.3 Minimum Tranche Size

Operational Practice. Large U.S. mutual-fund families rarely initiate a new position with one share. Opening a position—long or short—triggers portfolio-compliance checks, internal risk sign-off, back-office trade-ticket generation, and downstream monitoring of corporate events. Therefore, desks often impose an internal ticket floor: long orders are usually entered in round lots whose dollar value exceeds a preset minimum (often \$25–\$50 thousand for mid-cap or large-cap stocks), and prime brokers supply stock-loan locates in the same round-lot multiples on the short side. The model captures these institutional frictions with a fixed monitoring cost F and a non-divisible block $q_{\min}$. The surcharge $\eta q_{\min}$ for shorts mirrors the per-share borrow fee applied to that same lot.

Modelling Choice. Treating $q_{\min}$ as given is a simplification rather than a claim that managers could never trade different block sizes. In richer microstructure models investors do optimize quantity jointly with information (e.g., Kyle (1985)-type settings). Here, letting the analyst pick an arbitrarily small q would merely convert the fixed desk cost F into a per-share cost F/q, forcing us to append an additional first-order condition without altering the economics of information choice. Keeping $q_{\min}$ fixed therefore improves tractability while staying close to common desk practice.

Robustness. If one set $q_{\min} \to 0$ while keeping a strictly positive fixed cost F, the investor could open an infinitesimal stake but still incur the full monitoring expense—a degenerate corner that would require adding risk aversion or per-share execution costs to restore an interior optimum. Such extensions complicate notation yet leave the paper's qualitative insights intact: (i) higher short-side costs still raise the break-even precision hurdle; (ii) the κ-band logic that delivers Propositions 1–3 is unchanged; and (iii) in the sequential model, each additional draw would still shrink the posterior variance by $1/\sqrt{N}$ and hence reduce the expected overshoot term, preserving Proposition 4. We therefore retain the stylised—but industry-consistent—assumption of a fixed, non-optimized tranche size for analytical clarity.