

Do Hedge Funds Profit from Public Information?

Alan Crane

*Jones Graduate School of Business
Rice University, Houston, TX 77005, U.S.A.*

Kevin Crotty

*Jones Graduate School of Business
Rice University, Houston, TX 77005, U.S.A.*

Tarik Umar

*Jones Graduate School of Business
Rice University, Houston, TX 77005, U.S.A.*

Abstract

We examine whether hedge funds profit from public information. Using unique data on hedge funds' use of publicly-available SEC filings, we show that funds accessing filings in a month exhibit 1.5% higher annualized abnormal returns than non-users. Above-median users earn even higher returns. The effect is not due to fund-type differences. Performance declines with file complexity and increases with file uncertainty and competing hedge-fund views. Information processors (robotic downloaders and financial statement analysis specialists) exhibit weaker usage-return relations. Our results are less consistent with profitability resulting from information processing and more consistent with funds using public information to complement private signals.

Email addresses: Alan.D.Crane@rice.edu (Alan Crane), Kevin.P.Crotty@rice.edu (Kevin Crotty), Tarik.Umar@rice.edu (Tarik Umar)

1. Introduction

Do hedge funds profit from widely disseminated public information? Given that hedge funds are among the most sophisticated investors, the answer may be yes. Theory suggests that investors with complementary private signals or an information processing advantage can profit from public information (Kim and Verrecchia, 1994, 1997; Grossman and Stiglitz, 1980). On the other hand, the light regulation of hedge funds provides them flexibility to employ a wide variety of investment strategies (e.g., Fung and Hsieh, 2001; Stulz, 2007; Sun et al., 2012). These alternatives may be more attractive than strategies utilizing public information, leading sophisticated investors to respond less to public signals, consistent with evidence from mutual funds (Kacperczyk and Seru, 2007). More efficient markets with respect to public disclosures may also diminish the value of public information for hedge funds (Fama, 1970, 1991). While a large literature is dedicated to understanding both the performance and investment activities of hedge funds (for a survey treatment, see Agarwal et al., 2015), relatively little is known about whether hedge funds use and profit from public information, and if so, why. In this paper, we examine these questions in the context of information that is the epitome of public—mandated financial reports available to all market participants.

A necessary condition for hedge funds to profit from public information is that they acquire it in some way. However, this is difficult to test as we generally do not observe hedge funds' information sets. We overcome this challenge by using unique data that allows us to observe a subset of the information hedge funds acquire. This information is the very definition of public—SEC filings available at no cost to anyone with an internet connection. We compile a database of hedge funds' acquisition of financial disclosures from the SEC's EDGAR server. By mapping hedge fund internet protocol (IP) addresses to those accessing financial filings at the SEC, we are able to identify public information acquisition by hedge funds such as Renaissance Technologies, PanAgora, and AQR.

We document substantial variation in the use of public information by hedge funds both

across funds and within fund. While the median fund-month download amount is only 4 filings, the mean is 672. Moreover, there is variation in the type of information accessed. For an average fund-month, financial statements comprise about a third of the total downloads and unscheduled material disclosures (8-Ks) account for another 20% of filings accessed. Hedge funds also access other filings, such as insider trade disclosures (Form 4) and holdings of other institutions (13F). Finally, there is substantial time-series variation in the use of SEC filings. Conditional on use, the 90th percentile of use rises from about 50 filings per month in 2003 to almost 1,000 per month in 2017.

We then test whether the variation in public information access is related to performance. Hedge funds that access at least one filing have higher abnormal returns in the next month compared to funds that do not access filings. The result is statistically significant and economically large, representing a difference in abnormal returns of about 1.5% per year. More intensive information acquisition is also associated with higher subsequent abnormal returns, with the above-median users generating 2%-per-year higher returns than non-users. These results are consistent with hedge funds profiting from public information.¹

The fact that public information acquisition relates to performance is surprising. SEC filings are the quintessential “public” information, and therefore, usage of such information should not be profitable. However, we do not observe how the acquired information is actually used. The profitable use of public information suggests that either markets are grossly inefficient with respect to publicly-available data (which is unlikely), that there exist costs associated with processing public information, or that hedge funds also possess private information that is more valuable when used in conjunction with public information. In all of these cases, public information in the filings is responsible for performance differences. On the other hand, usage of public filings may merely be correlated with hedge fund types

¹It is worth noting that we do not observe information acquisition of public filings from other information intermediaries such as Bloomberg. If all funds obtain public information but do so from different sources, then we should observe no return differential as a function of EDGAR usage. That is, unobserved usage should bias *against* our findings. This is true unless there is a selection effect in which better funds use the SEC website rather than other sources. We use within-fund analysis to rule out such selection effects below.

that outperform.² To rule out such selection concerns, we examine within-fund variation in usage. Even within funds, performance is better subsequent to periods when those funds access more. This suggests that the relation we observe is more than just a selection effect where SEC filing usage proxies for fund type.

To investigate how funds profit from public information, we examine cross-sectional differences in performance as a function of the characteristics of the information accessed. Variation in performance as a function of the characteristics of the public information can help us disentangle whether the profitability we observe is due to costly information processing and/or complementary private information. Moreover, if performance varies as a function of the type of information, this also helps rule out the possibility that our results are driven by time-varying omitted fund characteristics. Performance does vary as a function of cross-sectional differences in filing characteristics, even within fund.

First, performance varies as a function of filing size. Specifically, when funds use larger filings, they do worse on average (approximately 75 basis points per year for a one-standard deviation larger file). Loughran and McDonald (2014) find that file size proxies for filing complexity, so our results suggest that funds do not perform better when analyzing complex information. This is evidence against the notion that the profitability of public information for hedge funds stems from an information processing channel. Second, performance is higher when other hedge funds also view the same filing in the same month. The number of other hedge fund views is generally quite small; therefore, we interpret this result as consistent with a small number of funds receiving correlated information. Again, this result is less consistent with an information processing advantage. We also find that viewing filings with uncertain language results in higher performance. Hedge funds could be better at resolving this uncertainty through either better processing ability or through complementary private information. Finally, funds perform better when they use filings of firms they have

²Kacperczyk and Seru (2007) find the opposite to be true in mutual funds. They find that better mutual funds rely less on public information as measured by analyst recommendations.

previously tracked. This result is consistent with the finding in Chen, Cohen, Gurun, Lou, and Malloy (2017) that institutions do better on trades around insider disclosures when they have previously tracked the firm. They argue that tracking firms is likely to be related to having private information.

To further understand the channel by which public information usage is profitable, we examine a subset of users that are ex-ante likely to have an information processing advantage. Specifically, we identify hedge funds that systematically scrape the SEC website. These “scrapers” earn 1.5% higher annualized abnormal returns than non-scrapers, but they do not generally exhibit a return-usage relation. We argue that these “scrapers” are more likely to profit from the processing of public information rather than from access to complementary private information. We then examine the performance of this subset of users as a function of the filing characteristics. Interestingly, we see the opposite results compared to the overall sample for several of the information characteristics. First, these “scrapers” perform better when files are longer (more complex). Second, they do better when they *have not* tracked the firm, which is consistent with scrapers focusing on information processing as opposed to building relationships with management teams. It also reinforces the view in the literature that tracking proxies for potential access to private information. The “scraper” results help validate our interpretation of the cross-sectional results in the full sample. Overall, hedge funds appear to profit from public information predominantly because of complementary private signals.

In a related test, we examine another subset of users – financial statement analysis specialists – who are also more likely to profit from advantages at processing public information rather than access to complementary private information. Specialists are funds for which both (1) the proportion of their total EDGAR usage that is due to 10-K/Q filings and (2) the total number of 10-K/Q filings they access (scaled by months in the sample) are above the cross-sectional medians for these measures. We examine the performance of the specialists and compare their performance to the non-specialists, or generalists. We do not find

that usage by specialists of 10-K/Q filings is significantly related to differences in subsequent performance. This result is less consistent with hedge funds developing an advantage at processing a certain filing type. Interestingly, the non-specialists outperform when they acquire one or more 10-K/Q filings, with no effect of intensity of usage on performance. Overall, these results are more consistent with hedge funds using public information to complement private signals.

1.1. Related Literature and Contribution

Two contemporaneous working papers also unmask IPs using EDGAR but focus on large institutional investors filing Form 13-F filings, while we focus on hedge funds. Chen, Cohen, Gurun, Lou, and Malloy (2017) examine search behavior of 13-F filers, predominantly mutual funds. They find that mutual fund managers follow trades of company insiders for a particular set of firms, that the set is highly persistent, and that investor trades related to these “tracked” firms are more informative than other trades for future stock performance. Dyer (2017) studies whether local institutional investors (13-F filers) use public information to generate an information advantage to make profitable trades in local stocks. Investors are more likely to acquire information for local investments and make more profitable trades when also acquiring public information.

In contrast to these two working papers, our focus on hedge fund usage of filings allows us to directly link search behavior to subsequent *institution-level* performance, rather than *trade-level* performance. Chen, Cohen, Gurun, Lou, and Malloy (2017) study the latter since 13-F holdings are aggregated across a large number of funds for large institutional investors (e.g., Vanguard). Dyer (2017) considers trading profitability in local stocks. Hedge funds, on the other hand, generally run only a few related funds, allowing us to link search activity with fund manager performance. Chen et al. (2017) focus predominantly on views of Form 4 insider trade filings. While our analysis includes Form 4, we consider more general hedge fund information acquisition of firm-specific information (10-K/10-Q/8-K) and other investors’ filings (13-F/13-G). The results of the papers are complementary. Taken together,

they provide strong evidence that institutional investors, more or less heavily regulated, make regular and beneficial use of financial disclosures through EDGAR and that public information usage may complement private signals.

Apart from Chen et al. (2017) and Dyer (2017), several other recent papers have determined the identities of EDGAR users. Unlike our work, the focus of these papers is not on the performance of asset managers. Bozanic, Hoopes, Thornock, and Williams (2017) show that the IRS uses EDGAR to acquire public financial information over time. Li, Lind, Ramesh, and Shen (2017) show that the Federal Reserve and FDIC use EDGAR more intensively in worse macroeconomic states and usage is linked significantly with disclosures from the financial sector. Gibbons, Iliev, and Kalodimos (2018) show evidence that analyst forecast errors are lower and recommendations exhibit abnormal returns when they coincide with EDGAR usage from the analysts' brokerage house. The notion that analysts provide fundamental information using public filings is consistent with a related working paper by Chen, Kelly, and Wu (2018). They show that hedge funds trade more aggressively in stocks with less analyst coverage resulting from brokerage house closures and these trades are more profitable. Chen, Kelly, and Wu (2018) provide suggestive evidence that EDGAR traffic for these firms' filings is higher in geographical areas closer to hedge funds, but they do not directly link EDGAR activity to hedge funds' IP addresses.

More broadly, the paper is related to recent research that utilizes EDGAR logs to proxy for information acquisition. Drake, Roulstone, and Thornock (2015, 2016) and Drake, Quinn, and Thornock (2017) describe determinants of EDGAR usage. Drake et al. (2016) documents that requests for historical reports are positively associated with shocks to firm value. Drake et al. (2017) finds that education is an important determinant of EDGAR usage. Loughran and McDonald (2017), after screening out robot requests, finds that the average firm has their annual report requested only 28.4 times, suggesting investors are not generally researching fundamental information. Interestingly, we show that a sizable fraction of hedge funds use EDGAR and that such activity is related to abnormal excess returns, suggesting such activity

is economically meaningful. Li and Sun (2017) examine the expected return of information embedded in investors' information acquisition activity. They find that abnormal numbers of IPs searching for firms' financial statements strongly predict future stock returns. Their paper largely falls in the abnormal attention and stock performance literature, which includes papers measuring attention using Google Search, abnormal trading volume, and abnormal returns.

Finally, our work contributes to the broader literature assessing which activities and characteristics of hedge funds are associated with performance. A variety of studies show that hedge fund managers appear to possess stock picking skills and are able to identify stock-level mispricing (Brunnermeier and Nagel, 2004; Jiao, Massa, and Zhang, 2016; Agarwal, Jiang, Tang, and Yang, 2013). However, there is relatively little work on how hedge funds generate performance. Hedge funds delay disclosure for some positions which subsequently outperform (Agarwal, Jiang, Tang, and Yang, 2013; Aragon, Hertz, and Shi, 2013; Shi, 2017). Hedge funds also use derivatives. Funds using derivatives exhibit less overall firm risk (Chen, 2011) and their option positions predict volatility and returns for underlying stocks (Aragon and Martin, 2012). Li, Zhang, and Zhao (2011) show that manager characteristics, such as undergraduate university quality, are related to performance. Lu, Ray, and Teo (2016) find that managers going through divorces underperform. Funds with more distinctive strategies perform better (Sun, Wang, and Zheng, 2012). Titman and Tiu (2011) find that funds with lower R -squareds exhibit improved performance, suggesting these funds are better informed. This interpretation is disputed by Bollen (2013) who argues that low- R -squared funds are likely exposed to an omitted risk factor. Massoud, Nandy, Saunders, and Song (2011) find that hedge funds trade on private information from syndicated loan participation. Gao and Huang (2016) show that funds connected to lobbyists outperform on their political holdings. Teo (2011) finds a relation between liquidity risk and fund performance, and Jame (2017) finds that part of hedge fund performance is due to liquidity provision. Gargano, Rossi, and Wermers (2017) show that funds trade profitably on information from the FDA that can be

obtained through Freedom of Information Act requests.

The remainder of our paper is organized as follows: Section 2 details the creation of the dataset of SEC filings’ usage. Section 3 summarizes the extent and determinants of financial filing usage. Section 4 examines how public information acquisition is related to subsequent fund performance and why. Section 5 concludes.

2. Data and Sample Construction

An innovation of this paper is to identify hedge fund usage of the SEC EDGAR database and relate usage to performance. The sample uses data from four primary sources: the SEC’s EDGAR server log files, the American Registry for Internet Numbers (ARIN), MaxMind, and the Hedge Fund Research (HFR) database. We now detail the steps used to construct several panel datasets used in the analysis.

2.1. Step 1: Identify IP addresses of hedge funds

Our sample of hedge funds is initially determined by all hedge funds in the HFR database, which includes alive and dead funds. HFR provides each fund’s monthly returns and often the monthly AUM of the fund. We also have details about the strategies and fees of the funds.

For each sample hedge fund, we search three sources for associated IP addresses. The first two sources are cross-sectional snapshots in 2014 and 2017 of the ARIN WHOIS database. We match records using the organization’s name in ARIN and the hedge fund’s name in HFR. A challenge is that the ARIN WHOIS database only provides bulk snapshots of the current IP registration landscape. Since hedge funds likely enter and exit the industry from 2003 to 2017, survivorship bias induced by the use of ARIN is a potential concern. In an attempt to mitigate this issue, we use a third “IP address book” from MaxMind that provides historical mappings of organizations to IP ranges for 2006 to 2017.³ Name-matching between

³We choose not to use the MaxMind panel alone because of potential data quality concerns. In particular,

HFR and MaxMind produces additional hedge fund-IP matches. Another way to resolve the survivorship issue is to limit the sample to the 2014-2017 period for which we have two snapshots of the ARIN WHOIS database. We find similar results using this subsample.

The above procedure results in a number of potential hedge fund-IP address matches resulting either from the (1) 2014 ARIN, (2) 2017 ARIN, or (3) MaxMind IP address registrars. For each potential match, we use the ARIN WHOIS database to determine the dates a hedge fund used a particular IP. ARIN WHOIS provides historical information about the ownership of particular IP addresses.⁴ Thus, for each hedge fund-IP pair, we have the registration start date and end date and restrict our study to IP-related activity between these registration dates.

2.2. Step 2: Identify hedge fund use in EDGAR logs

For each HFR hedge fund and associated ARIN IP, we examine the IP's activity on the SEC EDGAR server. The EDGAR server tracks all usage, including which files were downloaded and which indexes of files were viewed.⁵ Each record includes the IP address of the user and precise timestamps of the IP's activity on the server. The server reports the CIK of the firm being examined by the IP. For example, an investor studying CIK 21344 would be requesting files related to Coca-Cola Enterprises. We also have an accession number, which uniquely identifies each filing on EDGAR. If the investor clicks on a file with accession number 21344-17-000026, the investor would be looking at the Coca-Cola 10-Q filed on 2017-07-27. The data also provides the file name, which allows us to see whether the investor is accessing the 10-Q or one of the various exhibits. We can use the file extension to identify whether the requested file is a text or html file readable by humans or is an xml files that is machine readable. We obtained the server log files for the SEC EDGAR database for

there is an abnormal temporary spike of registrations in 2011, and some funds that are matched using ARIN are present for only a couple of months in 2011 in MaxMind.

⁴We cannot use ARIN WHOIS for the initial match to HFR because there are no bulk downloads of ARIN WHOIS. Historical registration details for IPs can only be accessed one IP at a time.

⁵We do not count index views towards information acquisition totals in our subsequent analyses.

all days from January 1, 2003 to March 31, 2017, with the exception of September 24, 2005 to May 11, 2006. The SEC did not retain log data for these days (Bauguess et al., 2013).

One challenge is that the last three-digits of the IP addresses provided by the SEC EDGAR server are obfuscated (e.g., 191.191.191.abc). To resolve this challenge, we link IP activity on the EDGAR server to hedge funds in the constructed HFR-ARIN sample using the first three sections of an IP address. Oftentimes, hedge funds register the full 0 to 255 possible IPs available in the fourth section of the IP address (xxx.xxx.xxx.0 to xxx.xxx.xxx.255). Even if the fund only registers a portion of the 0 to 255 range, the other registered owners are frequently un-related to the financial industry.

2.3. Step 3: Create panel datasets for analysis

The analyses in the subsequent sections make use of several panel datasets. The “HF-Month” panel consists of monthly statistics on returns, AUM, and EDGAR usage at the hedge fund level (aggregating across IPs to determine hedge fund EDGAR usage). This panel results from combining HFR data with monthly aggregate EDGAR usage statistics. We determine funds’ abnormal returns using the Fama and French (2015) five factor model augmented with the Carhart (1997) momentum factor and limit the sample to funds with at least 36 months of returns data.⁶ While the HFR database is at the fund level, EDGAR usage is at the management-company level; thus, we aggregate fund-level returns to the institution level by weighting by assets under management in each fund.

We also filter the sample based on fund strategy. Unsurprisingly, macro funds access firm-specific filings much less frequently than equity long-short funds. We exclude macro hedge funds and fund-of-funds from our sample, limiting the sample to “Event-Driven,” “Equity-Hedge,” and “Relative Value” funds.

We also merge firm characteristics from Compustat to hedge fund information acquisition

⁶We use this benchmark because we restrict the sample to equity-oriented funds. The results are similar when using the risk factors developed in Fung and Hsieh (2001) to capture more esoteric hedge fund strategies. We report these results in the Internet Appendix.

statistics to create a “Stock-Year” panel. This merge uses the CIK identifier.

3. Public Information Acquisition by Hedge Funds

3.1. Sample summary statistics

We restrict the sample to funds reporting returns information to HFR. Panel A of Table 1 reports summary statistics of the 557 hedge funds in the sample. The median fund has \$615 million in assets under management. The median management fee is 1.5% and the median incentive fee is 20%. The median market beta from the six factor benchmark is 0.35, indicating that funds are hedged relative to market risk to some extent. The average fund is in the sample for 6.5 years. Figure 1 plots the number of funds in the sample each month. The sample starts with a cross-section of about 100 funds. Most months contain around 300 funds.

Panel B of Table 1 reports information acquisition statistics of the fund-month panel used in our subsequent performance analysis. In the median fund-month, the total number of downloads is only 4 filings. For the 90th percentile fund-month, the total number of downloads is 217. Panel B of Table 1 also reports the relative proportion of form types accessed in a given fund-month. On average, a third of the filings are annual or quarterly financial reports, i.e. 10-Ks or 10-Qs. The next most common filing accessed is the disclosure of unscheduled material events reported in the 8-K, which accounts for 18% of the average fund-month’s downloads. Insider trading filings (Form 4s) are the next most accessed filing—they account for 5% of downloads for the average fund-month.

3.2. Information acquisition by notable hedge funds

Figure 2 reports information usage for some prominent hedge funds—Renaissance Technologies, PanAgora, and AQR—based on the IP addresses we link to these firms. The figure reports the time series of total downloads as well as time series of downloads of various company reports (10-K/Q, 8-K) and investor reports (4, 13-D/F/G). Some of these funds have

relatively consistent use of EDGAR over their time in the sample. For instance, AQR relatively consistently downloads 400-500 filings each month, apart from a handful of months when the firm downloads over 100,000 filings in a single month. In the early part of the sample, downloads by Renaissance Technologies also number in the hundreds before jumping higher in late 2011. This increase is driven by increases in downloads of Form 4s and 10-K/Qs. PanAgora has increasingly accessed public financial information from the SEC since 2008. Its use of various forms is somewhat episodic. For instance, for a period from mid-2010 to 2012, the firm downloaded over 500 10-Qs each month, but then the use of quarterly reports fell in 2013. Similarly, PanAgora’s use of 8-Ks is also more pronounced in the latter part of the sample.

3.3. Heterogeneity of public information acquisition

Table 2 reports the top 30 hedge fund users of EDGAR in our sample.⁷ Renaissance Technologies, PanAgora, and Blackrock are the top users, although the types of filings they view are not the same. The largest fraction of filings viewed by PanAgora are due to company financial reports and disclosures (10-K/Qs and 8-Ks), while the bulk of Renaissance and Blackrock’s downloads consist of disclosures of trading by insiders (Form 4s). Clearly there is some heterogeneity in the types of information that asset managers acquire from public filings.

To explore this further, Figure 3 plots the time series of public information acquisition by hedge funds. The figure reports the fraction of the cross section accessing filings in a given month (left column) as well as the cross-sectional median and the 25th, 75th, and 90th percentiles of EDGAR usage, conditional on a fund downloading a given firm type. Panels (a) and (b) report these time series for all downloads. The remaining panels report statistics

⁷It is worth noting that some of the investment companies presented in this table also have mutual funds. We cannot identify from the logs whether a view comes from hedge fund or mutual fund managers. Therefore these totals represent the entire management company. This measurement error does not affect subsequent performance analyses since our returns are specific to the hedge funds. Any noise induced by capturing mutual fund views should bias any coefficients in the performance regressions toward zero.

for various firm and investor filings.

The fraction of the cross-section accessing at least one file is fairly static over the sample. About two-thirds of the funds in the sample access a filing each month. The intensive margin of EDGAR use, on the other hand, rises over the sample period (panel (b)) for all but the lowest percentiles of users. For funds accessing some filing, the median number of filings downloaded increases from 10-20 per month to around 40 per month. The increase is more dramatic for the 75th and 90th percentiles of use. The 90th percentile of use rises from about 50 filings per month in 2003 to around 1,000 per month in 2017.

Panels (c)-(f) show that similar patterns hold for the major corporate financial filings (10-K/Q and 8-K). Hedge funds are accessing more regularly-scheduled filings (10-K/10-Q) and intermittent disclosures (8-Ks) over time. For 8-Ks, funds have increased the use of 8-Ks on the extensive margin as well. In 2003, around 30% of funds accessed 8-Ks. By the end of the sample, about half of the funds access 8-Ks in a given month.

Some hedge funds access information on what other investors are doing. Panels (g)-(n) report hedge fund attention to various investor filings. Form 4 reports trading by company insiders. Forms 13-D, 13-F, and 13-G report holdings of activists, institutional investors, and passive investors, respectively. About 1 in 5 hedge funds in the sample access trades by company insiders in a given month, and this fraction has been relatively stable over the sample period. On the other hand, there has been increased interest by hedge funds in filings made by other institutional investors. Panels (i), (k), and (m) show that few hedge funds viewed forms 13-D, 13-F, or 13-G in 2003, but the fraction of the sample viewing these forms increased to around a quarter of funds. Conditional on use, the first quartile of usage of investor filings is quite modest, usually an access of a single filing. There are some hedge funds, however, that view these filings more extensively. For example, the 90th percentile fund's use of form 13-F (conditional on use) rises from around 10-20 filings per month in 2003 to over 50 filings per month in 2017.

Funds in the sample also exhibit heterogeneity in the intensity with which they use differ-

ent types of filings. Recall from Table 2 that PanAgora primarily accesses company financial disclosures while Renaissance primarily accesses trade disclosures by company insiders. To examine this more generally, we calculate the fraction of a fund’s total downloads that are due to a particular type of filing. Figure 4 plots the cross-sectional distributions of this metric for common filings. Unsurprisingly, the forms that generally comprise the most common public information acquisition by funds are the annual and quarterly 10-K/Q reports, as well as the more timely 8-K disclosures of material events. However, there are sizable fractions of the cross-section which never access these reports. For instance, just over 5% of funds never access a 10-K or 10-Q over the sample period while 10% of funds never access an 8-K. The filing accounting for the next largest fraction of filings is Form 4. This form comprises at least 2.5% of all filings downloaded for about half of the sample, but it rarely accounts for more than 10% of a fund’s overall public information acquisition. The use of other investor filings like 13-D/F/G comprises an even smaller fraction of funds’ total downloads for the vast majority of hedge funds.

3.4. Determinants of public information acquisition by hedge funds

Hedge funds’ use of EDGAR varies systematically both over time and with fund characteristics. Table 3 reports regressions of a monthly indicator of public information acquisition by form type. The explanatory variables include fund characteristics as well as the lagged market excess return, a time trend, calendar month fixed effects, and the number of new filings on EDGAR that month of each type. The firm characteristics are fund age (log days since inception), log assets under management, incentive and management fees, lagged fund return, and factor loadings from the Fama-French Five Factors plus Momentum, estimated using the fund’s full sample of returns. Consistent with the plots in Figure 3, the probabilities of accessing 8-Ks, 13-Ds, 13-Fs, and 13-Gs are increasing with time. There is also evidence that funds are more likely to access filings of all types following months with negative market excess returns. In terms of firm characteristics, larger funds are more likely to access filings in a given month. Funds with greater exposure to the market (MKTRF Beta)

tend to request fewer filings. When there is more public information of a particular type as proxied by more filings, funds are more likely to access information in annual/quarterly financials and disclosures of trading by insiders.

3.5. Determinants of which stocks are viewed

What types of companies do hedge funds acquire public information about? We examine the characteristics of firms whose filings are downloaded by hedge funds. Specifically, we regress an indicator equal to 1 if any hedge fund accessed a given firm's filings in a given year, and 0 otherwise, on characteristics of that firm (as measured at the end of the prior year). We do this separately for each firm's 10-K/Qs, 8-Ks, and Form 4s. We include all Compustat firm-months available over our sample period. In calculating views, we exclude views from hedge fund companies that have side-by-side mutual funds, since, as noted above, we cannot determine whether the hedge funds specifically were accessing the files.⁸ The results are reported in Table IA.1 in the Internet Appendix. On average, hedge funds are more likely to view filings associated with higher leverage firms and growth firms. These characteristics have the largest economic magnitudes, where a standard-deviation increase in market to book is associated with a 200% increase in the probability of a view and a standard-deviation increase in leverage is associated with a 300% increase. Several other characteristics are statistically significant determinants of views, but have a more modest economic impact. For example, smaller firms and firms that have recently issued equity are more likely to be viewed, while firms with higher idiosyncratic volatility and more tangible assets are less likely to be viewed. In general, these determinants are consistent across all three filing types examined.

⁸This concern does not affect subsequent analyses since the performance analyses use hedge fund views as an independent variable. Any noise induced by capturing mutual fund views should bias any coefficients in the performance regressions toward zero.

4. Public Information Acquisition and Performance

4.1. Hedge Fund Usage of EDGAR and Fund Returns

This section shows that public information acquisition by hedge funds is positively related to the funds' subsequent abnormal performance. We first consider the extensive margin—that is, how does the decision to seek any public information relate to subsequent abnormal performance? Table 4 reports that funds with any download activity exhibit higher abnormal returns in the subsequent month, where abnormal return is calculated using the Fama-French Five Factor model plus momentum. For columns 2-7, the explanatory variables are indicator variables for whether a fund accessed any filing of the specified type in a given month. The improved performance is statistically significant for funds accessing both scheduled and unscheduled financial disclosures (10-K/Q, 8-K) as well as various investor filings (Forms 13-D, 13-F, and 13-G). The resulting annualized abnormal performance ranges from 0.8% per year (13-F) to 1.4% per year (10-K/Q, 8-K, 13-D). This indicates that fundamental information from the firms or activists in the firms seems to be the most profitable piece of public information. The performance relation of downloads of disclosures of trades by firm insiders (Form 4s) is insignificant and smaller in magnitude at 0.6% per year.

We next consider whether more intense acquisition behavior is associated with higher performance. Table 5 reports regressions of abnormal returns on an indicator variable for whether a fund accessed any filing of the specified type in a given month and an indicator for whether the fund downloaded more than the median fund's downloads that month. The median is evaluated cross-sectionally among funds that acquired any information of a given filing type that month. Funds that acquire more information than the median acquiring fund exhibit better subsequent abnormal performance when considering aggregated downloads (across filing types) and 10-K/Q filings. For 10-K/Q filings, above median users outperform non-using funds by almost 2% per year. For 8-Ks and 13-Fs, the point estimates are also economically meaningful, but statistically insignificant. Overall, the results suggest usage intensity is positively related to subsequent performance, which is consistent with hedge

funds deriving value from public information.

4.2. Do differences in hedge fund type explain the profitability of public information?

One concern with the analysis thus far is that public information acquisition may just proxy for differential investment abilities across funds. A given investment fund may possess superior information processing technology or superior private information acquisition ability, and the observable public information acquisition may proxy for these differences across funds. To address this possibility, we consider whether within-fund variation in hedge funds' public information acquisition predicts subsequent within-fund performance. Table 6 shows that this is indeed the case. We model within-fund variation in two ways: (1) using a fund fixed effect to absorb average fund ability, and (2) allowing for time-varying strategies for each fund. On the extensive margin, the annualized point estimate of the value of public information acquisition for next-month's returns only drops from 1.5% per year to 94 bps per year with the addition of fund fixed effects (moving from column (1) to (2) in Panel A). Public information acquisition is thus associated with a non-trivial performance differential *even within-fund*, suggesting that differences in processing ability or private information acquisition across funds do not account for the observed profitability. When considering both the extensive and intensive margins within-fund (column 4), the point estimates suggest that any information acquisition correlates with about 85 bps per year in subsequent performance and that when funds are more intense users, their performance is almost 40 bps higher per year, within-fund, although these estimates are noisy.

It is possible that fund types may not be fixed through time. That is, a fund's information acquisition strategy may vary through time due to new ideas or time-varying efficiency of prices. To account for this, we consider a more flexible way to account for within-fund variation in Panel B of Table 6. Specifically, we identify changes in a fund's public-information-acquisition strategy and relate these changes to subsequent abnormal returns. A fund's abnormal usage is calculated as a z-score of downloads in month t relative to the distribution of the fund's usage during the previous 24 months. The fixed effects

specification used in Panel A essentially compares a fund’s use to its full sample average. In contrast, the results in Panel B use a rolling window of the past 24 months as the firm’s baseline for comparison.

Abnormal public information acquisition is positively related to performance (column (1) of Panel B). Because information-gathering strategies can shift to more or less intensive strategies, column (2) reports separate estimates for abnormal usage falling below the 25th percentile and above the 75th percentile, relative to the prior 24 months. High abnormal usage is positively related to future performance and the relation is highly significant. In the month following increased public information acquisition, funds with abnormally high usage exhibit approximately 1.3% higher returns (annualized). In contrast, low abnormal usage is not significantly related to future performance. This indicates that discontinuing information acquisition is not associated with declines in performance and is consistent with funds rationally taking into account the costs and benefits of information acquisition when considering whether to acquire additional information.

4.3. Why is public information profitable?

As discussed in the introduction, public information may be profitable if hedge funds are skilled information processors. Alternatively, private information may be more valuable when used in conjunction with public information. In this section, we show evidence suggesting that the predominant channel is the latter complementary private information mechanism.

4.3.1. Profitability and Filing Characteristics

To explore these possible channels, we interact the intensive margin of usage with various filing characteristics that help disentangle the processing channel from the complementary private information channel. We focus on 10-K/Q filings because these filings are most commonly requested by funds and file attributes vary significantly across file types, such as file size. We consider the following filing characteristics: the median age (in days) of the accessed 10-K/Qs, the median file size (KB) of the accessed 10-K/Qs, the median number

of other hedge funds viewing the accessed 10-K/Qs, the median idiosyncratic volatility of the filer’s equity returns, the median intensity with which the hedge fund tracked the filer (measured as number of downloads of firm filings in prior months), the median Amihud (2002) illiquidity of the 10-K/Q filer, and the median level of textual uncertainty in the accessed 10-K/Qs (measured using the proportion of uncertain words determined using the Loughran and McDonald (2011) lexicon). In each case, the median is taken within firm-month across all 10-K/Qs accessed by the fund.

Older filings are less likely to contain non-priced information and are likely to be less valuable when used in conjunction with timely private information. Table 7 provides evidence consistent with this intuition. The interaction between above median downloads and filing age has a negative coefficient and is marginally significant, suggesting hedge funds profit more when downloading more recent filings. Drake et al. (2016) show that accessing historical reports seems to be related to shocks to firm value and is important for the overall information environment for investors. Our results suggest that the most valuable information for hedge funds stems from more recent filings, which is likely used in conjunction with private information.

A variety of studies relate the size of filings to market activity. Evidence suggests that greater 10-K file size is associated with more post-filing volatility (Loughran and McDonald, 2014). The interpretation is that more-difficult-to-process filings require longer for the market to process. Sophisticated hedge fund managers may have a relative advantage at processing longer filings and earn greater abnormal returns than less-skilled rival funds. Column (2) of Table 7 shows that the value of public information in financial reports is mitigated when funds access larger filings. The coefficient on the interaction between above median downloads and standardized file size is negative and economically important. This suggests that the information processing channel is less important in explaining overall profitability of public information acquisition. In general, funds accessing longer, more complex corporate filings earn less rather than more.

We use the number of filing views by other hedge funds as a proxy for the extent of complementary private information about underlying investments in the hedge fund cross-section. Column (3) of Table 7 shows that more intensive public information acquisition is significantly more profitable when other hedge funds also view a filing. Since the number of other hedge fund views is generally small (the median is a single other view), this is consistent with a few hedge funds receiving correlated private signals that are complementary to public information.

Firms with greater amounts of idiosyncratic volatility have more firm-specific information. This could induce greater value to information processing as well as private information acquisition. As such, its cross-sectional implications for the channel is ambiguous. While not statistically significant, the point estimate suggests that viewing firms with higher levels of firm-specific information is associated with a subsequent performance difference of 42 basis points (annualized).

Some funds regularly view the filings of certain firms. Chen et al. (2017) show that trades following tracked Form 4 filings are more profitable, arguing that investors are likely to also possess complementary private information about these tracked filers. We test whether tracking certain filers results in more profitable use of public information found in financial statements. Column (5) of Table 7 shows that a one-standard deviation increase in the median tracking status of viewed 10-K/Qs corresponds to about 55 basis points difference (annualized) in fund performance and the effect is marginally significant within-firm. This provides additional support that the predominant mechanism behind the profitability of public information is the complementary private information channel.

One commonly used measure of private information in financial markets is the Amihud (2002) price impact measure. It measures how much prices change relative to trading volume. It is likely that hedge funds possess more complementary private information for filings whose stocks exhibit higher levels of price impact. The estimate in Column (6) of Table 7 indicates that a standard deviation change in illiquidity is associated with 35 bps annual-

ized performance differential. This is consistent with public information acquisition being profitable due to complementary private information, although this result is not statistically significant.

The last filing characteristic we consider is the textual uncertainty of the filing. We retrieve from the WRDS SEC Analytics Suite the proportion of uncertain words in SEC filings, determined using the Loughran-McDonald lexicon (Loughran and McDonald, 2011). Textual uncertainty may arise when company management is unsure of the implications of the financial results or when there exists uncertainty concerning future cash flows of the firm. In these cases, it is likely that complementary private information obtained by other market participants is more valuable, but it is also possible that information processing is also more valuable. Column (7) of Table 7 reports that textual uncertainty in the financial reports viewed by hedge funds is strongly related to subsequent fund returns. The magnitude is quite large. A one-standard deviation difference in textual uncertainty is associated with subsequent performance improvements of about 1% per year.

On balance, the cross-sectional characteristics suggest that public information acquisition by hedge funds is profitable because they also possess (or acquire) complementary private information.

4.3.2. Extreme Information Processing: Robotic Information Acquisition

Another way to test the prevalence of the information processing channel for the profitability of public information is to focus on a subset of funds whose information acquisition is less likely to be related to complementary private information.

One such set of investors is hedge funds using robotic means to acquire SEC filings. We classify a fund-month as robotic if the fund accessed more than 50 filings in a single day and the median time between downloads that day was less than 30 seconds. Figure 5 plots the number of funds classified as robots each month. If a fund has more than one robotic month,

we label the fund as a “Scraper.”⁹ Table 8 reports regressions of abnormal returns on the “Scraper” designation as well as its interaction with the extensive and intensive margin of public information acquisition. “Scraper” funds significantly outperform non-scraper funds. Scrapers earn 1.5% higher abnormal annualized returns when compared to non-scrapers. The profitability of scraping is not immediate though, as its interactions with both the extensive and intensive margins are negative and offset the estimated profitability found for the full sample. Rather, it seems that scraping is related to the development of new strategies and information processing that creates value on a longer horizon than a single month. Controlling for these extreme information processors, modest users of public financial information continue to outperform non-EDGAR users by at least 1.1% as evidenced by the significant estimate on the “Any” indicators in all specifications. While the results suggest funds crawling EDGAR outperform managers that manually retrieve and process filings, the latter group of managers still outperforms non-users. Thus, extreme information processing is not the channel explaining the full sample’s usage-return relation.

4.3.3. Filing Characteristics and Robotic Information Acquisition

To further disentangle the information processing and the complementary private information channels, we consider how filing characteristics affect the profitability of public information for the subsample of funds using robotic information acquisition. The filing characteristics that are more profitable for such funds are likely related to information processing rather than private information. Table 9 reports how the profitability of public information acquisition of 10-K/Q filings varies with filing characteristics across fund-months in which funds use robots to access 10-K/Qs. Processing of larger 10-K/Q filings, which are more complex and costly to process, is related to an increase in profitability. Large-scale processing of 10-K/Q filings is also more profitable when funds have not previously processed information related to a firm; the coefficient on tracked firms is large and significantly negative.

⁹There are some funds that may access filings automatically in real time as filings are posted. If the intervals between filings are longer than 30 seconds, we would not capture such funds as “Scrapers.”

The results suggest that large-scale novel information processing of more complex public information is most profitable for robotic information acquirers. The lack of profitability for scrapers with respect to processing the filings of tracked firms supports the notion that tracking may be a proxy for having access to private information, as conjectured by Chen, Cohen, Gurun, Lou, and Malloy (2017).

4.3.4. Financial statement analysis specialists

The most prevalent type of public information accessed by hedge funds is that found in the annual and quarterly financial statements. Some funds appear to specialize in processing the information in these forms. Whether the public information-return relation is stronger or weaker for these information processing specialists compared to funds that exhibit a more generalist public information acquisition strategy can help determine whether the full sample profitability of public information is due to public information processing or not. We identify financial statement analysis specialists as funds for which both (1) the proportion of their total EDGAR usage that is due to 10-K/Q filings¹⁰ and (2) the total number of 10-K/Q filings they access (scaled by months in the sample) are above the cross-sectional medians for each measure. All other funds are classified as generalists.

Table 10 reports estimates of the relation between the extensive and intensive margins of public information acquisition and subsequent fund returns for financial statement specialists and generalists. We report specifications with and without fund fixed effects. Financial statement analysis specialists do not exhibit statistically significantly higher performance following months in which they view financial reports. The point estimates for above median use is positive and corresponds to a 52-75 basis point increase in performance (annualized) depending on whether the specification includes fund fixed effects.

In contrast, the performance of generalist funds is significantly more related to financial statement views, especially for the extensive margin. Generalist funds that access at least

¹⁰10-K/Qs comprise about 30% of all filing views for the median fund.

one filing exhibit 1.1% higher subsequent returns (annualized). This estimate increases to 1.4% using within-fund variation. The greater sensitivity of performance to 10-K/Q filing usage for generalists but not for specialists is more consistent with hedge funds profitably using public information in conjunction with private signals, rather than hedge funds simply having processing advantages.

5. Conclusion

Hedge funds profit from public information that they obtain directly from the SEC web site. We investigate whether this profitable use of public information is due to hedge funds being superior information processors or if it is due to them using public information in conjunction with private signals about firm values. Several analyses suggest the latter channel is the predominant source of the profits. The use of public information is more valuable when funds access filings with characteristics that should be associated with complementary private information. Also, the use of public information results in less subsequent abnormal performance for funds that specialize in information processing. Both robotic information acquisition funds and financial statement analysis specialists do not significantly profit from accessing financial reports, although the former group of funds outperforms unconditionally. Overall, our results show that hedge funds use and profit from public information, and the evidence is most consistent with this profitability stemming from the complementarity of public and private information.

References

- Agarwal, Vikas, Wei Jiang, Yeuhua Tang, and Baozhong Yang, 2013, Uncovering hedge fund skill from the portfolio holdings they hide, *The Journal of Finance* 68, 597–635.
- Agarwal, Vikas, Kevin A Mullally, Narayan Y Naik, et al., 2015, The economics and finance of hedge funds: A review of the academic literature, *Foundations and Trends in Finance* 10, 1–111.
- Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.
- Aragon, George O, Michael Hertzel, and Zhen Shi, 2013, Why Do Hedge Funds Avoid Disclosure? Evidence from Confidential 13F Filings, *Journal of Financial and Quantitative Analysis* 48, 1499–1518.
- Aragon, George O., and J. Spencer Martin, 2012, A unique view of hedge fund derivatives usage: Safeguard or speculation?, *Journal of Financial Economics* 105, 436–456.
- Bauguess, Scott, John Cooney Jr., and Kathleen Weiss Hanley, 2013, Investor Demand for Information in Newly Issued Securities, *Working Paper* SEC, Texas Tech, and Lehigh.
- Bollen, Nicolas, 2013, Zero- R^2 Hedge Funds and Market Neutrality, *Journal of Financial and Quantitative Analysis* 48, 519–547.
- Bozanic, Zahn, Jeffrey L Hoopes, Jacob R Thornock, and Braden M Williams, 2017, IRS attention, *Journal of Accounting Research* 55, 79–114.
- Brunnermeier, Markus K, and Stefan Nagel, 2004, Hedge Funds and the Technology Bubble, *The Journal of Finance* 59, 2013–2040.
- Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57–82.

- Chen, Huaizhi, Lauren Cohen, Umit Gurun, Dong Lou, and Christopher J Malloy, 2017, IQ from IP: Simplifying Search in Portfolio Choice, *Working Paper* Harvard, UTD, LSE.
- Chen, Yong, 2011, Derivatives Use and Risk Taking: Evidence from the Hedge Fund Industry, *Journal of Financial and Quantitative Analysis* 46, 1073–1106.
- Chen, Yong, Bryan Kelly, and Wei Wu, 2018, Sophisticated Investors and Market Efficiency: Evidence from a Natural Experiment, *Working Paper* Texas A&M and Yale University.
- Drake, Michael S, Phillip J Quinn, and Jacob R Thornock, 2017, Who uses financial statements? a demographic analysis of financial statement downloads from edgar, *Accounting Horizons* 31, 55–68.
- Drake, Michael S., Darren T. Roulstone, and Jacob R. Thornock, 2015, The Determinants and Consequences of Information Acquisition via EDGAR, *Contemporary Accounting Research* 32, 1128–1161.
- Drake, Michael S., Darren T. Roulstone, and Jacob R. Thornock, 2016, The usefulness of historical accounting reports, *Journal of Accounting and Economics* 61, 448–464.
- Dyer, Travis, 2017, Does Public Information Acquisition Level the Playing Field or Widen the Gap? An Analysis of Local and Non-Local Investors, *Working Paper* University of North Carolina.
- Fama, Eugene F, 1970, Efficient capital markets: A review of theory and empirical work, *The Journal of Finance* 25, 383–417.
- Fama, Eugene F, 1991, Efficient capital markets: Ii, *The Journal of Finance* 46, 1575–1617.
- Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1 – 22.
- Fung, William, and David a. Hsieh, 2001, The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers, *The Review of Financial Studies* 18, 313–341.

- Gao, Meng, and Jiekun Huang, 2016, Capitalizing on Capitol Hill: Informed trading by hedge fund managers, *Journal of Financial Economics* 121, 521–545.
- Gargano, Antonio, Alberto G Rossi, and Russ Wermers, 2017, The freedom of information act and the race toward information acquisition, *The Review of Financial Studies* 30, 2179–2228.
- Gibbons, Brian, Peter Iliev, and Jonathan Kalodimos, 2018, Analyst Information Acquisition via EDGAR, *Working Paper* Penn State and Oregon State.
- Grossman, Sanford J, and Joseph E Stiglitz, 1980, On the impossibility of informationally efficient markets, *The American Economic Review* 70, 393–408.
- Jame, Russell, 2017, Liquidity provision and the cross section of hedge fund returns, *Management Science* forthcoming.
- Jiao, Yawen, Massimo Massa, and Hong Zhang, 2016, Short selling meets hedge fund 13F: An anatomy of informed demand, *Journal of Financial Economics* 122, 544–567.
- Kacperczyk, Marcin, and Amit Seru, 2007, Fund manager use of public information: New evidence on managerial skills, *The Journal of Finance* 62, 485–528.
- Kim, Oliver, and Robert E Verrecchia, 1994, Market liquidity and volume around earnings announcements, *Journal of Accounting and Economics* 17, 41–67.
- Kim, Oliver, and Robert E Verrecchia, 1997, Pre-announcement and event-period private information, *Journal of Accounting and Economics* 24, 395–419.
- Li, Edward, Gary Lind, K. Ramesh, and Min Shen, 2017, The Federal Reserve’s Reliance on SEC Filings, *Working Paper* Baruch and Rice.
- Li, Frank Weikai, and Chengzhu Sun, 2017, Information Acquisition and Expected Returns: Evidence from EDGAR Search Traffic, *Working Paper* SMU and HKUST.

- Li, Haitao, Xiaoyan Zhang, and Rui Zhao, 2011, Investing in Talents: Manager Characteristics and Hedge Fund Performances, *Journal of Financial and Quantitative Analysis* 46, 59–82.
- Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *The Journal of Finance* 66, 35–65.
- Loughran, Tim, and Bill McDonald, 2014, Measuring readability in financial disclosures, *Journal of Finance* 69, 1643–1671.
- Loughran, Tim, and Bill McDonald, 2017, The use of EDGAR filings by investors, *Journal of Behavioral Finance* 18, 231–248.
- Lu, Yan, Sugata Ray, and Melvyn Teo, 2016, Limited attention, marital events and hedge funds, *Journal of Financial Economics* 122, 607–624.
- Massoud, Nadia, Debarshi Nandy, Anthony Saunders, and Keke Song, 2011, Do hedge funds trade on private information? Evidence from syndicated lending and short-selling, *Journal of Financial Economics* 99, 477–499.
- Shi, Zhen, 2017, The impact of portfolio disclosure on hedge fund performance, *Journal of Financial Economics* 126, 36–53.
- Stulz, René M, 2007, Hedge funds: Past, present, and future, *Journal of Economic Perspectives* 21, 175–194.
- Sun, Zheng, Ashley Wang, and Lu Zheng, 2012, The road less traveled: Strategy distinctiveness and hedge fund performance, *Review of Financial Studies* 25, 96–143.
- Teo, Melvyn, 2011, The liquidity risk of liquid hedge funds, *Journal of Financial Economics* 100, 24–44.
- Titman, Sheridan, and Cristian Tiu, 2011, Do the best hedge funds hedge?, *Review of Financial Studies* 24, 123–168.

Table 1: Summary Statistics

Panel A reports distributional statistics of the cross-section of hedge funds in the sample. The Fama-French Five Factor betas plus a Momentum beta are estimated using a hedge fund's full time series of fund-level returns net of fees reported to the Hedge Fund Research database. Hedge funds may have multiple funds; we collapse returns to the firm level and present AUM-weighted returns. Panel B reports distributional statistics of the monthly download activity of hedge funds in the sample. "Downloads" is the number of downloads from the SEC's EDGAR database by a given hedge fund in a month. We then report the number of downloads by form-type and, conditional on downloading in a given month, the proportion of monthly downloads by form type. We also report, conditional on downloading, the median characteristics of the form 10-K and 10-Q files examined by a hedge fund that month, including the number of other unique hedge funds viewing the same filing (Competing Views), the age of the filing in days since the filing date (Age), and the size of the filing in KB (File Size).

<i>Panel A. Hedge Fund Characteristics</i>							
	Mean	Std. Dev	25th	Median	75th	90th	99th
Mgt Company Assets (mm)	28838.60	260067.13	132.00	615.00	3851.00	22400.00	490700.00
Months in HFR	149.91	79.04	88.00	132.63	203.29	257.27	373.00
Months in Sample	79.16	43.95	43.00	74.00	115.00	146.00	160.00
Age in Months	149.76	78.87	88.00	132.63	203.29	257.00	373.00
VW Excess Ret	0.41	0.72	0.12	0.42	0.70	1.07	2.37
VW Abnormal Return	0.05	0.63	-0.21	0.09	0.38	0.65	1.63
Incentive Fee	18.97	4.43	20.00	20.00	20.00	20.00	28.55
Management Fee	1.50	0.45	1.00	1.50	2.00	2.00	2.50
Market Beta	0.37	0.35	0.13	0.35	0.56	0.85	1.34
SMB Beta	0.10	0.26	-0.01	0.08	0.20	0.39	0.74
HML Beta	-0.03	0.23	-0.14	-0.02	0.09	0.19	0.61
RMW Beta	-0.06	0.32	-0.18	-0.04	0.07	0.24	0.65
CMA Beta	-0.05	0.37	-0.17	-0.03	0.11	0.26	0.85
MOM Beta	-0.01	0.24	-0.09	-0.01	0.06	0.16	0.42
Observations	557						
<i>Panel B. Public Information Acquisition by Hedge Funds</i>							
	Mean	Std. Dev	25th	Median	75th	90th	99th
Downloads	672	19476	0	4	47	217	5464
Downloads of 10-K or 10-Q	101	3945	0	1	14	71	620
Downloads of 8-K	140	7340	0	0	7	38	587
Downloads of Form 4	290	16601	0	0	1	6	93
Downloads of 13-D	11	833	0	0	0	3	29
Downloads of 13-F	49	2068	0	0	0	4	68
Downloads of 13-G	17	1820	0	0	0	3	30
% 10-K/Q	32	28	8	29	50	70	100
% 8-K	18	21	0	13	25	41	100
% 4	5	15	0	0	3	12	97
% 13-D	2	8	0	0	1	5	33
% 13-F	4	13	0	0	2	9	80
% 13-G	2	8	0	0	1	5	36
% Other	37	29	14	32	51	85	100
Competing Views 10-K/Q	1	2	0	1	2	3	9
Filing Age 10K/Q	324	648	68	133	271	668	3695
Filing Size 10K/Q (KB)	631	831	99	340	978	1429	3562

Table 2: Top 30 Users of EDGAR since 2003

The table reports total download statistics and download activity by form type for the top 30 institutional users that match to the HFR database, sorted by total downloads (distinct within a month). We also report the end-of-sample assets under management.

Firm Name	Total Downloads	Form Type						AUM (MM)
		10-K/Q	8-K	4	13D	13F	13G	
Renaissance Technologies Corp.	4,016,439	3.5%	2.0%	85.8%	0.2%	0.2%	0.9%	50,941
PanAgora Asset Management, Inc	3,969,668	24.2%	21.6%	10.9%	4.9%	15.0%	7.2%	42,798
Blackrock	3,704,596	3.5%	0.7%	92.4%	0.1%	0.1%	0.2%	5,689,273
Hutchin Hill Capital, LP	3,044,555	3.7%	4.7%	8.8%	0.0%	0.2%	0.0%	3,300
Tradeworx Inc	2,068,017	11.2%	5.6%	51.4%	5.0%	11.0%	15.2%	61
First Pacific Advisors, LLC	2,003,707	0.0%	73.8%	0.0%	0.0%	26.0%	0.0%	30,800
AQR Capital Management	1,944,446	28.0%	0.2%	0.1%	0.0%	8.5%	0.0%	194,900
Jennison Associates LLC	1,857,445	4.4%	87.7%	1.5%	1.0%	0.0%	0.3%	167,000
Schroder Investment Management Ltd	1,590,526	39.4%	55.1%	0.0%	0.0%	4.7%	0.0%	490,700
Zack's Investment Management	1,540,158	1.8%	0.3%	96.8%	0.0%	0.2%	0.0%	4,736
Ten Asset Management	1,044,195	26.8%	7.2%	27.1%	8.3%	19.8%	4.2%	36
Neuberger Berman	915,603	3.9%	1.7%	89.3%	0.9%	0.3%	0.3%	270,728
Bailard	804,940	1.1%	0.2%	98.0%	0.0%	0.0%	0.0%	2,421
LIM Advisors Limited	424,800	1.6%	5.8%	76.8%	0.7%	0.7%	2.4%	1,800
Benchmark Capital Advisors	309,561	0.1%	14.4%	84.3%	0.0%	0.0%	0.0%	250
Weiss Asset Management	236,780	24.3%	52.1%	0.1%	0.2%	0.2%	0.3%	1,807
Numeric Investors LLC	230,576	41.8%	0.6%	3.0%	0.1%	0.1%	0.1%	30,367
AllianceBernstein L.P.	199,882	40.2%	18.2%	2.9%	0.7%	3.8%	1.4%	497,875
BlueCrest Capital Management LLP	197,330	36.9%	11.7%	0.0%	0.0%	50.9%	0.0%	14,000
Wellington Management Company, LLP	159,994	40.4%	17.9%	1.3%	0.7%	0.9%	1.9%	1,018,744
Marshall Wace LLP	157,094	13.6%	24.6%	2.7%	2.9%	0.7%	1.8%	22,000
Thornburg Investment Management	133,468	29.0%	8.4%	32.2%	0.9%	5.7%	0.9%	52,805
Ivory Investment Management, LLC	113,709	17.5%	26.5%	2.5%	2.3%	0.7%	1.8%	2,733
First Trust Advisors, L.P.	109,219	61.1%	4.0%	1.7%	0.2%	3.3%	0.8%	111,774
Oaktree Capital Management, LLC	102,302	41.3%	25.4%	2.3%	1.3%	1.1%	1.4%	99,260
Clinton Group, Inc.	88,109	5.0%	2.9%	0.8%	0.6%	85.1%	0.2%	650
Bronson Point Management	80,586	2.4%	0.9%	0.2%	0.4%	92.9%	0.1%	245
Alpha Equity Management LLC	73,095	14.7%	26.8%	4.4%	3.1%	0.5%	2.1%	177
HG Vora Capital Management, LLC	64,955	16.7%	27.0%	2.9%	4.6%	0.7%	2.0%	3,400
Calamos Investments	58,490	33.0%	16.0%	2.8%	0.4%	1.2%	1.4%	19,089

Table 3: Determinants of Information Acquisition

The table reports regressions of whether a hedge fund downloaded a particular form in a given month on various fund and macroeconomic characteristics. The dependent variable is an indicator equal to 1 if the fund downloaded the filing type indicated at the top of the column, and 0 otherwise. Fund variables include the natural log of fund age since inception, the natural log of AUM, incentive fees, management fees, factor betas from the Fama-French 5-factor model plus Momentum, and lagged abnormal returns. Macroeconomic variables include the past month's market return and the number of SEC filings by type. Fixed effects for the calendar month are included to control for seasonality, and a time trend is included to test for increased usage of EDGAR over time. All independent variables are standardized. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	(1) Any	(2) 10-K/Q	(3) 8-K	(4) 4	(5) 13D	(6) 13F	(7) 13G
Log Age	0.02 (1.21)	0.01 (0.77)	0.01 (0.74)	0.01 (1.16)	0.01 (0.60)	0.00 (0.29)	0.01 (0.50)
Log AUM	0.07*** (5.06)	0.10*** (6.38)	0.11*** (6.98)	0.08*** (5.65)	0.09*** (6.24)	0.08*** (5.84)	0.09*** (6.27)
Incentive Fee	-0.02 (-1.52)	-0.03* (-1.81)	-0.02 (-1.32)	-0.01 (-1.08)	-0.02* (-1.85)	-0.02* (-1.90)	-0.02* (-1.76)
Mgt Fee	0.01 (0.73)	0.01 (0.78)	0.02 (1.18)	0.01 (0.97)	0.01 (1.07)	0.00 (0.32)	0.01 (0.71)
MKTRF Beta	-0.04** (-2.37)	-0.04** (-2.55)	-0.04*** (-2.81)	-0.03*** (-3.11)	-0.02** (-2.29)	-0.02** (-2.27)	-0.03*** (-2.62)
SMB Beta	0.01 (1.10)	0.02 (1.18)	0.01 (0.73)	0.00 (0.02)	-0.00 (-0.43)	-0.00 (-0.50)	0.00 (0.21)
HML Beta	-0.00 (-0.35)	-0.00 (-0.13)	-0.00 (-0.19)	-0.00 (-0.27)	-0.00 (-0.02)	-0.00 (-0.47)	-0.00 (-0.14)
RMW Beta	-0.02 (-1.42)	-0.02 (-1.56)	-0.02 (-1.27)	-0.01 (-0.53)	-0.01 (-0.80)	0.00 (0.03)	-0.01 (-0.87)
CMA Beta	0.01 (0.71)	0.02 (1.28)	0.01 (1.08)	0.01 (0.98)	0.01 (1.23)	0.00 (0.26)	0.01 (1.06)
MOM Beta	-0.01 (-0.51)	0.00 (0.18)	0.00 (0.22)	0.01 (0.63)	0.01 (0.51)	0.01 (1.23)	0.01 (0.68)
RET(t-1)	0.01* (1.87)	0.01** (2.03)	0.00 (1.50)	0.00 (0.27)	0.00* (1.93)	0.00* (1.81)	0.00* (1.73)
MKTRF(t-1)	-0.01*** (-2.86)	-0.01*** (-2.75)	-0.01*** (-3.29)	-0.01*** (-6.28)	-0.01*** (-3.39)	-0.00*** (-2.65)	-0.01*** (-5.79)
Time	0.01 (0.85)	0.00 (0.31)	0.03** (2.44)	-0.01 (-1.40)	0.04*** (4.34)	0.03*** (4.19)	0.01 (1.57)
New SEC Filings	-0.00 (-0.17)						
New SEC Filings-10K/Q		0.08*** (4.05)					
New SEC Filings-8K			-0.00 (-0.08)				
New SEC Filings-4				0.03*** (3.77)			
New SEC Filings-13D					0.01 (1.16)		
New SEC Filings-13F						-0.01 (-0.69)	
New SEC Filings-13G							-0.01 (-0.69)
Calendar Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.04	0.06	0.07	0.05	0.06	0.06	0.06
Number Firms	554	554	554	554	554	554	554
Observations	43287	43287	43287	43287	43287	43287	43287

Table 4: Information Acquisition and Performance: Extensive Margin

The table reports the relation between download activity by a hedge fund in month t and the fund's abnormal return (measured in percent) in month $t + 1$. The monthly abnormal returns are calculated using the Fama-French Five Factor Model plus Momentum. Download activity for each form type ("Any") is an indicator variable for whether the fund accessed any forms of the indicated type in month t . AUM is standardized for interpretation. Regressions contain year-month fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any Download	0.1258*** (3.32)						
Any 10K/Q		0.1150*** (3.39)					
Any 8K			0.1157*** (3.58)				
Any 4				0.0511 (1.46)			
Any 13D					0.1196*** (3.65)		
Any 13F						0.0697** (1.98)	
Any 13G							0.0858*** (2.77)
AUM (t)	0.0164 (0.83)	0.0144 (0.72)	0.0131 (0.66)	0.0222 (1.11)	0.0156 (0.79)	0.0206 (1.03)	0.0188 (0.94)
Abn Ret (t)	0.1042*** (8.75)	0.1043*** (8.76)	0.1043*** (8.74)	0.1044*** (8.75)	0.1042*** (8.74)	0.1044*** (8.75)	0.1044*** (8.75)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Number Firms	557	557	557	557	557	557	557
Observations	43435	43435	43435	43435	43435	43435	43435

Table 5: Information Acquisition and Performance: Intensive Margin

The table reports the relation between download activity by a hedge fund in month t and the fund's abnormal return (measured in percent) in month $t + 1$. The monthly abnormal returns are calculated using the Fama-French Five Factor Model plus Momentum. Download activity for each form type ("Any") is an indicator variable for whether the fund accessed any forms of the indicated type in month t . "Above Median" is an indicator variable for whether the fund's download activity falls above month t 's cross-sectional median number of downloads for a form type (conditional on any downloads of the form type). AUM is standardized for interpretation. Regressions contain year-month fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any Download	0.0918** (2.12)						
Above Median Download	0.0733** (2.00)						
Any 10K/Q		0.0784** (1.99)					
Above Median 10K/Q		0.0796** (2.11)					
Any 8K			0.0904** (2.43)				
Above Median 8K			0.0554 (1.42)				
Any 4				0.0356 (0.81)			
Above Median 4				0.0337 (0.68)			
Any 13D					0.1256*** (3.04)		
Above Median 13D					-0.0142 (-0.28)		
Any 13F						0.0388 (0.92)	
Above Median 13F						0.0694 (1.60)	
Any 13G							0.0905** (2.36)
Above Median 13G							-0.0110 (-0.20)
AUM (t)	0.0101 (0.51)	0.0089 (0.44)	0.0100 (0.50)	0.0217 (1.08)	0.0159 (0.79)	0.0199 (0.99)	0.0190 (0.95)
Abn Ret (t)	0.1042*** (8.74)	0.1042*** (8.74)	0.1042*** (8.74)	0.1044*** (8.75)	0.1042*** (8.73)	0.1044*** (8.75)	0.1044*** (8.75)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Number Firms	557	557	557	557	557	557	557
Observations	43435	43435	43435	43435	43435	43435	43435

Table 6: Information Acquisition and Performance: Within-Fund Variation

The table presents the within-fund relation between download activity in month t and a fund's abnormal return (measured in percent) in month $t + 1$. Abnormal returns are calculated using the Fama-French Five Factor Model plus Momentum. Panel A uses hedge-fund fixed effects. "Any Download" indicates whether the fund accessed any filings in month t . "Above Median" indicates whether the fund falls above month t 's cross-sectional median number of downloads (conditional on any downloads). Panel B uses a measure of abnormal downloads. For each fund-month, a standardized trailing download measure is calculated as the fund-month's downloads in excess of the fund's trailing 24 month average downloads, divided by the standard deviation of the fund's download activity over the trailing 24-month period. If there is no variation in download activity over the prior 24 months, the standardized trailing download measure is set to zero if the month's number of downloads is zero, or is set to an arbitrarily large (small) number if the month's number of downloads is greater (less) than the fund's trailing average monthly download. "Abnormal Downloads" is the p -value resulting from applying the standard normal distribution function to the standardized trailing download measure. Thus, "Abnormal Downloads" takes values from 0 to 1. "High (Low) Abnormal Downloads" is an indicator variable for Abnormal Downloads taking a value greater than 0.75 (less than 0.25). AUM is standardized for interpretation. All regressions contain year-month fixed effects, and columns 2 and 4 in Panel A contain fund fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Panel A: Fund Fixed Effects

	Dependent Variable: Abnormal Return (t+1)			
	(1)	(2)	(3)	(4)
Any Download	0.126*** (3.32)	0.078* (1.76)	0.092** (2.12)	0.070 (1.51)
Above Median Download			0.073** (2.00)	0.031 (0.80)
AUM (t)	0.016 (0.83)	-0.479*** (-8.13)	0.010 (0.51)	-0.481*** (-8.14)
Abn Ret (t)	0.104*** (8.75)	0.075*** (6.25)	0.104*** (8.74)	0.075*** (6.25)
Fund FE	No	Yes	No	Yes
Date FE	Yes	Yes	Yes	Yes
Adjusted R ²	0.08	0.09	0.08	0.09
Number Firms	557	556	557	556
Observations	43435	43434	43435	43434

Panel B: Abnormal Time-Varying Strategies

	Dependent Variable: Abnormal Return (t+1)	
	(1)	(2)
Abnormal Downloads(t)	0.1063** (2.07)	
High Abnormal Downloads(t)		0.1118*** (2.72)
Low Abnormal Downloads(t)		0.0447 (1.08)
AUM (t)	0.0312 (1.61)	0.0294 (1.53)
Abn Ret (t)	0.1004*** (7.69)	0.1004*** (7.69)
Date FE	Yes	Yes
Fund FE	No	No
Adjusted R ²	0.08	0.08
Number Firms	555	555
Observations	42321	42321

Table 7: Information Processing or Complementary Private Information?

The table reports the relation between download activity by a hedge fund in month t and the fund's abnormal return (measured in percent) in month $t + 1$. The monthly abnormal returns are calculated using the Fama-French Five Factor Model plus Momentum. "Any Download" indicates whether the fund accessed any filings in month t . "Above Median" is an indicator variable for whether the fund falls above the month's cross-sectional median number of downloads (conditional on any downloads of the form type). "Above Median" usage is interacted with the following filing characteristics: the median age (in days) of the accessed 10-K/Qs, the median file size (in KB) of the accessed 10-K/Qs, the median number of other hedge funds viewing the accessed 10-K/Qs, the median idiosyncratic volatility (IVol) of the accessed 10-K/Q firms (measured as the average squared daily abnormal return of the filer over the preceding year), the median intensity with which the hedge fund tracked the filer (measured as number of downloads of firm filings in prior months), the median Amihud (2002) illiquidity of the accessed 10-K/Q firms, and the median level of textual uncertainty in the accessed 10-K/Qs (measured using the proportion of uncertain words determined using the Loughran and McDonald (2011) lexicon). In each case, the median is taken within firm-month across all 10-K/Qs accessed by the fund. All continuous variables, except returns, are standardized. Regressions contain year-month and fund fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any 10K/Q	0.0751** (1.99)	0.0746** (1.97)	0.0757** (2.00)	0.0749** (1.98)	0.0755** (1.99)	0.0744** (1.96)	0.0757** (2.00)
Above Median 10K/Q	0.0050 (0.11)	0.0115 (0.26)	0.0180 (0.40)	0.0054 (0.12)	0.0032 (0.07)	0.0048 (0.11)	0.0049 (0.11)
- x Age	-0.0509 (-1.61)						
- x File Size		-0.0615* (-1.66)					
- x Other HF Views			0.0737** (2.04)				
- x IVol				0.0348 (1.03)			
- x Tracked					0.0459 (1.61)		
- x Amihud						0.0292 (0.86)	
- x Uncertainty							0.0867** (2.11)
AUM (t)	-0.4813*** (-8.14)	-0.4813*** (-8.14)	-0.4820*** (-8.16)	-0.4805*** (-8.14)	-0.4826*** (-8.18)	-0.4799*** (-8.13)	-0.4818*** (-8.17)
Abn Ret (t)	0.0749*** (6.26)	0.0749*** (6.25)	0.0749*** (6.25)	0.0749*** (6.25)	0.0749*** (6.25)	0.0749*** (6.26)	0.0749*** (6.25)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fund FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.09	0.09	0.09	0.09	0.09	0.09	0.09
Number Firms	556	556	556	556	556	556	556
Observations	43434	43434	43434	43434	43434	43434	43434

Table 8: Robotic vs. Human Information Acquisition

The table reports the relation between download activity by a hedge fund in month t and the fund's abnormal return (measured in percent) in month $t + 1$. A fund's monthly download activity is determined to be robot-initiated if a fund downloaded 50 or more filings in a day and the median time interval between downloads was less than 30 seconds that day. If a fund's activity is robot-initiated in more than one month in the sample, then the fund is deemed a "Scraper." The monthly abnormal returns are calculated using the Fama-French Five Factor Model plus Momentum. "Any Download" indicates whether the fund accessed any filings in month t . "Above Median" is an indicator variable for whether the fund falls above the month's cross-sectional median number of downloads of a form type (conditional on any downloads of the form type). AUM is standardized for interpretation. Regressions contain year-month fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)			
	(1)	(2)	(3)	(4)
Scraper	0.124*** (3.15)	0.307*** (4.08)	0.308*** (4.10)	0.308*** (4.10)
Any Downloads	0.099** (2.55)	0.120*** (2.82)	0.104** (2.29)	0.093** (1.98)
Any Downloads x Scraper		-0.210*** (-2.62)	-0.231*** (-2.90)	-0.135 (-1.54)
Above Median Downloads			0.045 (1.20)	0.078* (1.72)
Above Median Downloads x Scraper				-0.138* (-1.78)
AUM (t)	0.006 (0.31)	0.007 (0.36)	0.005 (0.24)	0.004 (0.21)
Abn Ret (t)	0.104*** (8.73)	0.104*** (8.73)	0.104*** (8.73)	0.104*** (8.73)
Date FE	Yes	Yes	Yes	Yes
Adjusted R ²	0.08	0.08	0.08	0.08
Number Firms	557	557	557	557
Observations	43435	43435	43435	43435

Table 9: Robots and Information Processing Ability

The table reports the relation between the fund's next-month abnormal return (measured in percent) and characteristics of filings downloaded by hedge funds using robotic information acquisition in month t . The sample is constrained to months in which "Scraper" funds as defined in Table 8 employ robotic acquisition. By definition, all months contain above median download activity. Therefore, the coefficients on filing characteristics in this table are comparable to the coefficients on the interactions of filing characteristics with above median downloads shown in Table 7. The monthly abnormal returns are calculated using the Fama-French Five Factor Model plus Momentum. The filing characteristics of the form 10-K/Qs accessed by a fund in month t are: the median age (in days) of the accessed 10-K/Qs, the median file size (in KB) of the accessed 10-K/Qs, the median number of other hedge funds viewing the accessed 10-K/Qs, the median idiosyncratic volatility (IVol) of the accessed 10-K/Q firms (measured as the average squared daily abnormal return of the filer over the preceding year), the median intensity with which the hedge fund tracked the filer (measured as number of downloads of firm filings in prior months), the median Amihud (2002) illiquidity of the accessed 10-K/Q firms, and the median level of textual uncertainty in the accessed 10-K/Qs (measured using the proportion of uncertain words determined using the Loughran and McDonald (2011) lexicon). In each case, the median is taken within firm-month across all 10-K/Qs accessed by the fund. All variables, except returns, are standardized. Regressions contain year-month fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	-0.0250 (-0.34)						
File Size		0.1640** (2.25)					
Other HF Views			0.1080 (0.66)				
IVol				-0.0030 (-0.01)			
Tracked					-0.2114* (-1.85)		
Amihud						-0.0827 (-0.42)	
Uncertainty							-0.0307 (-0.17)
AUM (t)	0.1289 (1.10)	0.1894* (1.70)	0.1184 (1.02)	0.1397 (1.24)	0.1557 (1.27)	0.1422 (1.25)	0.1401 (1.26)
Abn Ret (t)	0.0976 (1.52)	0.0923 (1.28)	0.0987 (1.49)	0.1018 (1.48)	0.0902 (1.25)	0.1040 (1.52)	0.1018 (1.47)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.03	0.04	0.03	0.03	0.04	0.03	0.03
Number Firms	45	45	45	45	45	45	45
Observations	304	304	304	304	304	304	304

Table 10: Financial Statement Analysis Specialists vs. Generalists

This table compares the usage-performance relations of funds that specialize in processing form 10-K/Q filings and those that do not. A fund is classified as a Form 10-K/Q specialist or a generalist. Financial statement analysis specialists are defined as funds for which both (1) the proportion of their total EDGAR usage that is due to 10-K/Q filings and (2) the total number of 10-K/Q filings they access (scaled by months in the sample) are above the cross-sectional medians for each measure. Generalists are non-10-K/Q specialists. For specialists and generalists, the table reports the relation between the fund's abnormal return (measured in percent) in month $t + 1$ and both the extensive margin of download activity and the intensive margin of download activity (above median downloads of 10-K/Q filings in given month conditional on viewing a 10-K/Q filing) in month t . The monthly abnormal returns are calculated using the Fama-French Five Factor Model plus Momentum. AUM is standardized for interpretation. All regressions contain year-month fixed effects, and columns 2 and 4 contain fund fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)			
	10K/Q Strategy		Generalist Strategy	
	(1)	(2)	(3)	(4)
Any 10K/Q	0.014 (0.22)	-0.015 (-0.23)	0.093* (1.89)	0.113** (2.47)
Above Med 10K/Q	0.063 (1.28)	0.043 (0.72)	0.115** (1.98)	0.008 (0.11)
AUM (t)	0.040 (1.40)	-0.334*** (-4.45)	-0.015 (-0.51)	-0.550*** (-6.76)
Abn Ret (t)	0.137*** (5.87)	0.107*** (4.43)	0.092*** (6.83)	0.063*** (4.70)
Date FE	Yes	Yes	Yes	Yes
Fund FE	No	Yes	No	Yes
Adjusted R ²	0.11	0.12	0.07	0.08
Number Firms	182	181	375	375
Observations	15430	15429	28005	28005

Figure 1: Sample composition over time

The figure plots the number of hedge funds in the panel each month. The panel is formed by matching hedge fund names in the HFR database to an address book of IP ranges. The sample contains any hedge fund that accesses EDGAR at least once from 2003 to 2017 using one of the matched IP addresses. The panel has a gap in 2005-6 due to missing SEC server log files.

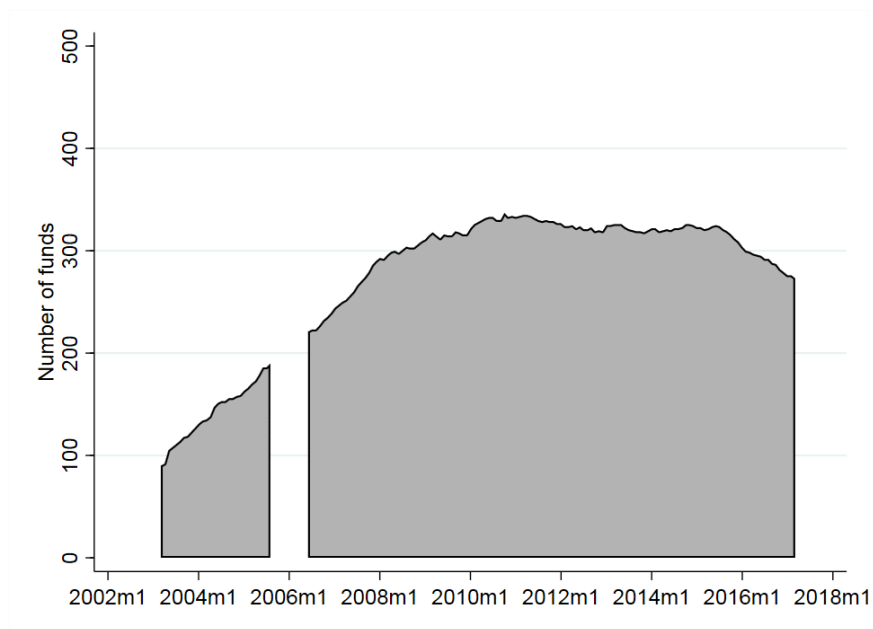
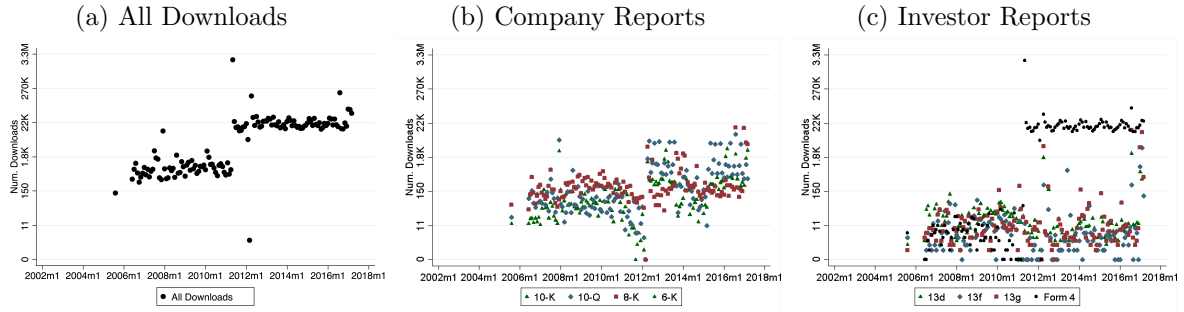


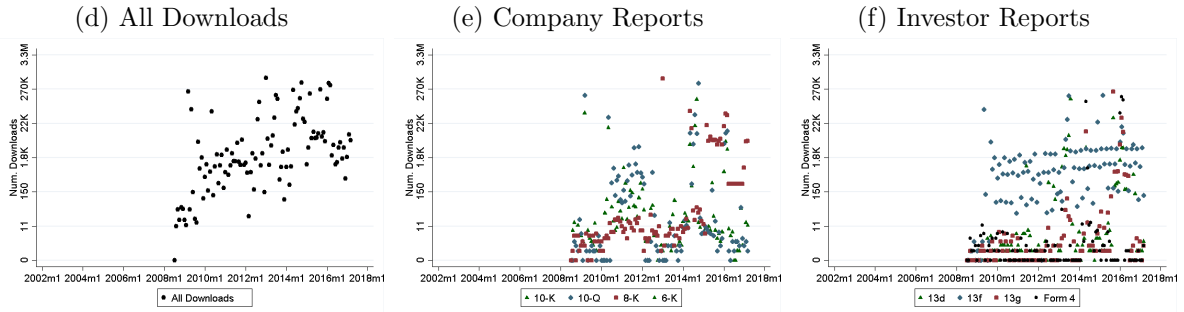
Figure 2: Download activity of notable investment funds

The figure plots the time series of EDGAR usage by some notable hedge funds. The left-hand column plots total downloads. The center column plots downloads of filings containing corporate financial information. The right column plots downloads of filings concerning investor reports (including firm insiders). Downloads are plotted on a log scale. The time series have gaps in 2005-6 due to missing SEC server log files.

Panel A. Renaissance Technologies



Panel B. PanAgora



Panel C. AQR

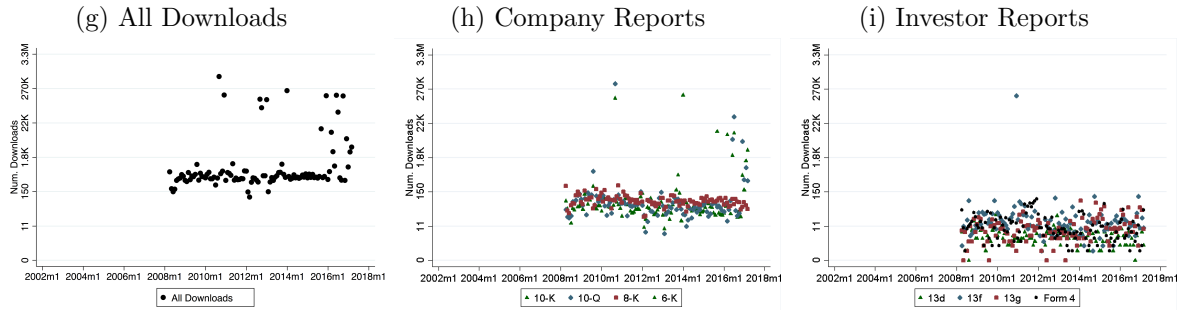


Figure 3: Time series of usage by form

The figure plots the time series of EDGAR usage by hedge funds. The left-hand column plots the fraction of the cross-section accessing a given file type. The right-hand column plots the time series of the cross-sectional 25th, 50th, 75th, and 90th percentiles of download activity of the indicated form type, conditional on a fund downloading that form type. Downloads are plotted on a log scale. The panel has a gap in 2005-6 due to missing SEC server log files.

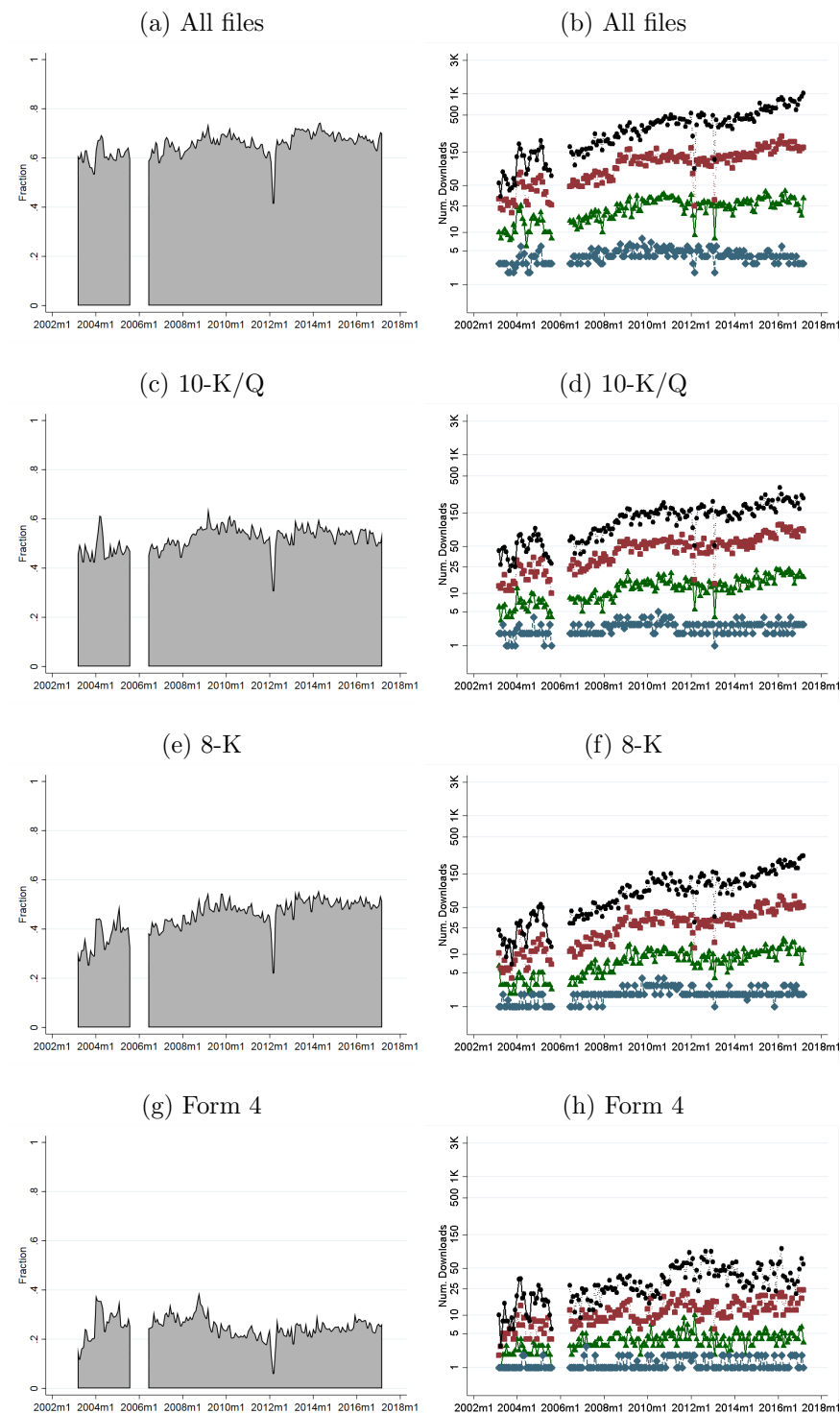
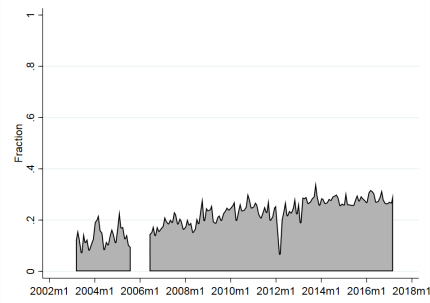
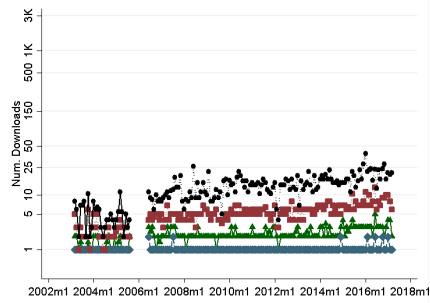


Figure 3: (continued) Time series of usage by form

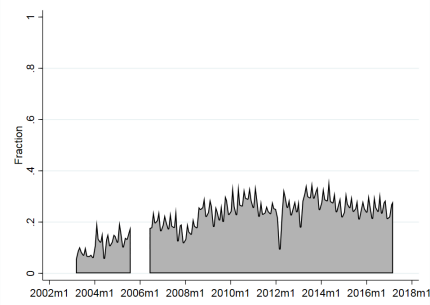
(i) 13-D



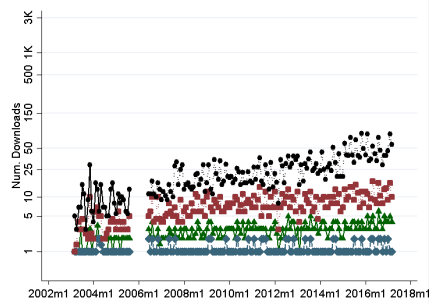
(j) 13-D



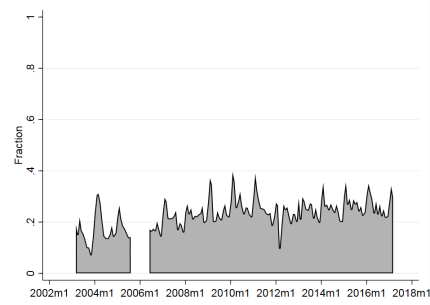
(k) 13-F



(l) 13-F



(m) 13-G



(n) 13-G

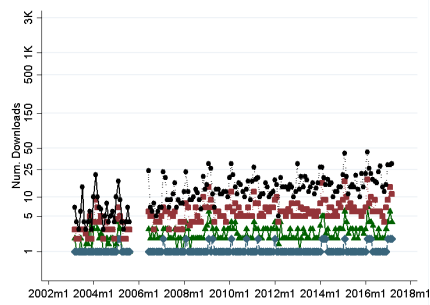


Figure 4: Cross-sectional distributions of the fraction of a fund's total usage due to form type

The figure plots the cross-sectional distributions of the fraction of all download activity for a fund that is due to a given form type. The vertical axis maximum value is 15% for panels (a)-(b) and 80% for the other panels.

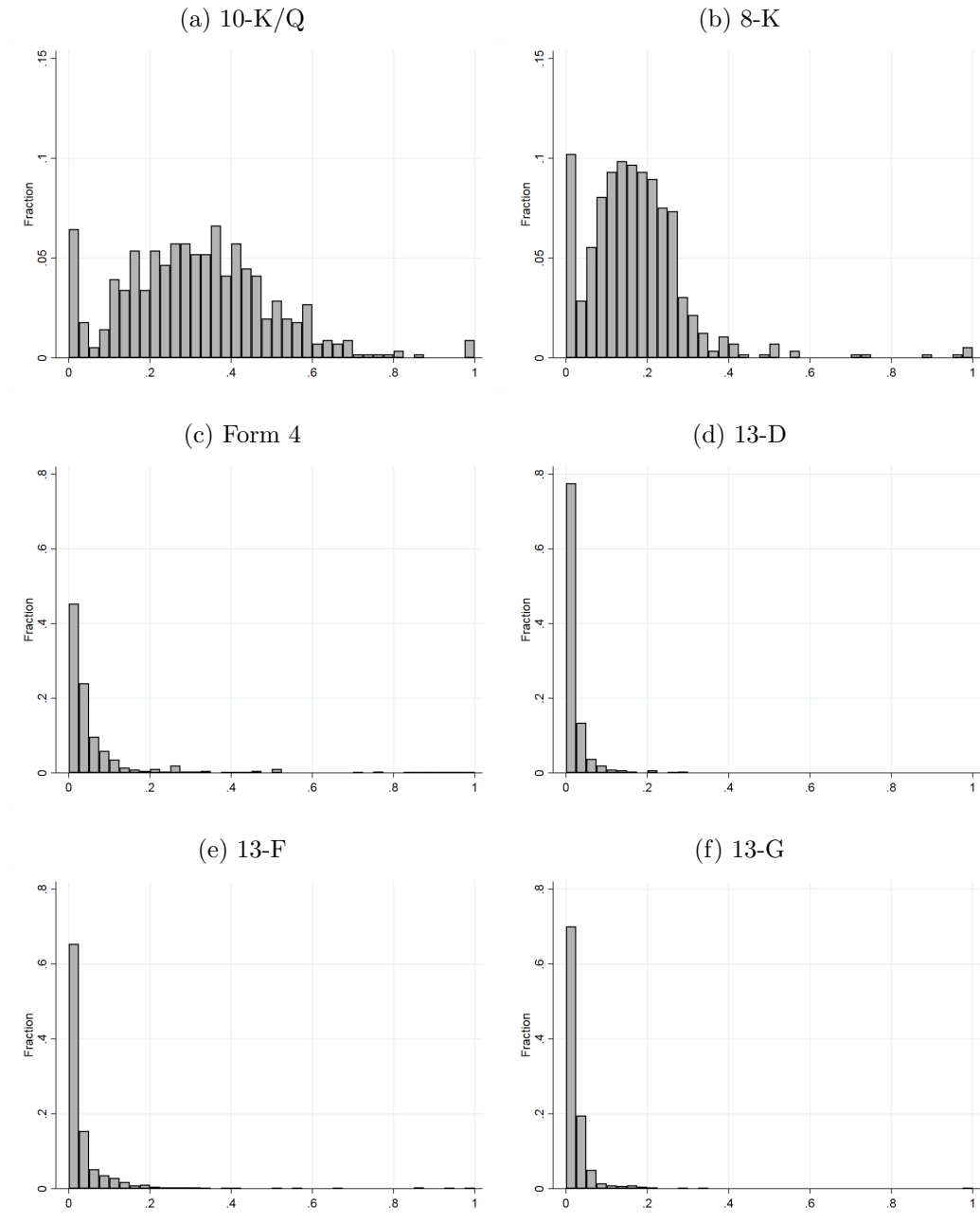
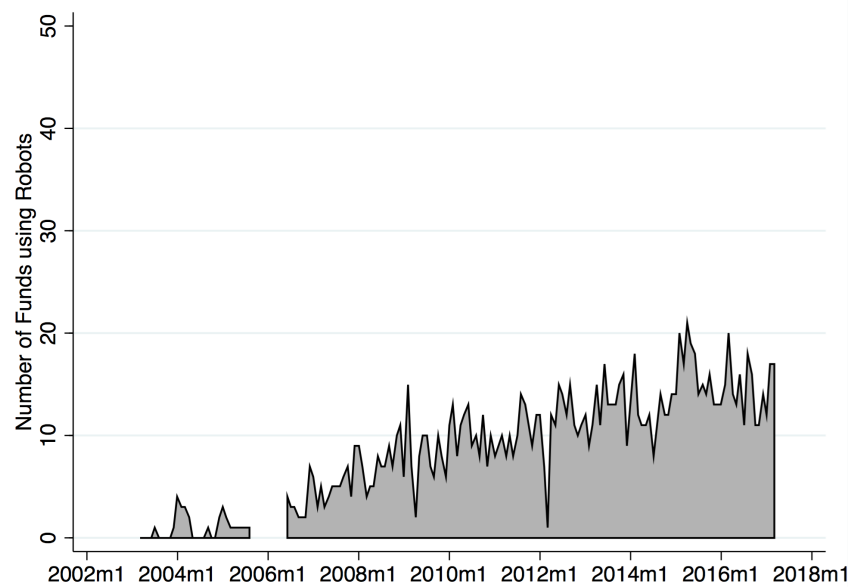


Figure 5: Number of hedge funds using robots over time

The figure plots the number of hedge funds who have activity on the EDGAR server classified as robotic in the panel each month. We classify a fund-month as robotic if the fund accessed more than 50 filings in a single day and the median time between downloads that day was less than 30 seconds.



Internet Appendix

This Internet Appendix contains supplementary tables. These include the following:

1. Table IA.1 presents analysis of firm level determinants of hedge fund filing views.
2. Table IA.2 replicates Table 4 from the paper under the Fung-Hsieh risk adjustments.
3. Table IA.3 replicates Table 5 from the paper under the Fung-Hsieh risk adjustments.
4. Table IA.4 replicates Table 6 from the paper under the Fung-Hsieh risk adjustments.
5. Table IA.5 replicates Table 7 from the paper under the Fung-Hsieh risk adjustments.
6. Table IA.6 replicates Table 8 from the paper under the Fung-Hsieh risk adjustments.
7. Table IA.7 replicates Table 9 from the paper under the Fung-Hsieh risk adjustments.
8. Table IA.8 replicates Table 10 from the paper under the Fung-Hsieh risk adjustments.

Table IA.1: Determinants of Firm Filing Views

The table reports regressions of whether a firm's filings were downloaded by any hedge fund on the firm's characteristics in the prior year. The dependent variable is an indicator equal to 1 if the firm's filing had a hedge fund view in that year, and 0 otherwise. Independent variables are scaled by the pooled standard deviation. Firm characteristics are measured the fiscal year values from the prior calendar year. Standard errors are clustered by firm and reported in parenthesis with significance represented according to: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

	(1)	(2)	(3)
Ln(Market Equity)	-0.04*** (0.01)	-0.05*** (0.01)	-0.04*** (0.01)
Book Leverage	3.08*** (0.53)	3.19*** (0.52)	2.84*** (0.51)
ROA	0.45 (0.46)	-0.03 (0.44)	0.28 (0.44)
Market to Book	2.15*** (0.47)	2.05*** (0.48)	1.78*** (0.53)
Z-score	1.35 (1.23)	0.69 (1.26)	1.32 (1.14)
Tangibility	-0.06*** (0.01)	-0.06*** (0.01)	-0.06*** (0.01)
Dividend Payer	-0.02* (0.01)	-0.01 (0.01)	-0.00 (0.01)
Pos. Net Equity Issue	0.03*** (0.01)	0.04*** (0.01)	0.05*** (0.01)
Idio. Vol	-0.11*** (0.01)	-0.10*** (0.01)	-0.17*** (0.01)
Constant	0.74*** (0.02)	0.60*** (0.02)	0.46*** (0.02)
Observations	41921	41921	41921
Adjusted R^2	0.09	0.14	0.10
Form Type	10-K	8-K	4
Year FE	Yes	Yes	Yes

Table IA.2: Information Acquisition and Performance: Extensive Margin

The table reports the relation between download activity by a hedge fund in month t and the fund's abnormal return (measured in percent) in month $t + 1$. The monthly abnormal returns are calculated using the Fung-Hsieh seven factor model. Download activity for each form type ("Any") is an indicator variable for whether the fund accessed any forms of the indicated type in month t . AUM is standardized for interpretation. Regressions contain year-month fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any Download	0.0945** (2.58)						
Any 10K/Q		0.0955*** (2.93)					
Any 8K			0.0998*** (3.17)				
Any 4				0.0466 (1.31)			
Any 13d					0.0880*** (2.65)		
Any 13f						0.0598* (1.69)	
Any 13g							0.0803** (2.50)
AUM (t)	0.0155 (0.79)	0.0130 (0.66)	0.0115 (0.58)	0.0191 (0.96)	0.0151 (0.76)	0.0180 (0.91)	0.0158 (0.80)
Abn Ret (t)	0.0960*** (9.30)	0.0961*** (9.31)	0.0961*** (9.29)	0.0961*** (9.30)	0.0960*** (9.29)	0.0961*** (9.30)	0.0961*** (9.30)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Number Firms	557	557	557	557	557	557	557
Observations	43435	43435	43435	43435	43435	43435	43435

Table IA.3: Information Acquisition and Performance: Intensive Margin

The table reports the relation between download activity by a hedge fund in month t and the fund's abnormal return (measured in percent) in month $t + 1$. The monthly abnormal returns are calculated using the Fung-Hsieh seven factor model. Download activity for each form type ("Any") is an indicator variable for whether the fund accessed any forms of the indicated type in month t . "Above Median" is an indicator variable for whether the fund's download activity falls above month t 's cross-sectional median number of downloads for a form type (conditional on any downloads of the form type). AUM is standardized for interpretation. Regressions contain year-month fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any Download	0.0640 (1.50)						
Above Median Download	0.0659* (1.72)						
Any 10K/Q		0.0595 (1.55)					
Above Median 10K/Q		0.0782** (2.00)					
Any 8K			0.0774** (2.09)				
Above Median 8K			0.0490 (1.19)				
Any 4				0.0394 (0.90)			
Above Median 4				0.0156 (0.31)			
Any 13d					0.0849** (2.12)		
Above Median 13d					0.0074 (0.15)		
Any 13f						0.0256 (0.59)	
Above Median 13f						0.0765 (1.62)	
Any 13g							0.0878** (2.25)
Above Median 13g							-0.0173 (-0.33)
AUM (t)	0.0098 (0.50)	0.0076 (0.38)	0.0087 (0.44)	0.0189 (0.95)	0.0149 (0.75)	0.0171 (0.87)	0.0161 (0.81)
Abn Ret (t)	0.0960*** (9.29)	0.0960*** (9.30)	0.0960*** (9.29)	0.0961*** (9.30)	0.0960*** (9.29)	0.0962*** (9.30)	0.0961*** (9.30)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Number Firms	557	557	557	557	557	557	557
Observations	43435	43435	43435	43435	43435	43435	43435

Table IA.4: Information Acquisition and Performance: Within-Fund Variation

The table presents the within-fund relation between download activity in month t and a fund's abnormal return (measured in percent) in month $t + 1$. Abnormal returns are calculated using the Fung-Hsieh seven factor model. Panel A uses hedge-fund fixed effects. "Any Download" indicates whether the fund accessed any filings in month t . "Above Median" indicates whether the fund falls above month t 's cross-sectional median number of downloads (conditional on any downloads). Panel B uses a measure of abnormal downloads. For each fund-month, a standardized trailing download measure is calculated as the fund-month's downloads in excess of the fund's trailing 24 month average downloads, divided by the standard deviation of the fund's download activity over the trailing 24-month period. If there is no variation in download activity over the prior 24 months, the standardized trailing download measure is set to zero if the month's number of downloads is zero, or is set to an arbitrarily large (small) number if the month's number of downloads is greater (less) than the fund's trailing average monthly download. "Abnormal Downloads" is the p -value resulting from applying the standard normal distribution function to the standardized trailing download measure. Thus, "Abnormal Downloads" takes values from 0 to 1. "High (Low) Abnormal Downloads" is an indicator variable for Abnormal Downloads taking a value greater than 0.75 (less than 0.25). AUM is standardized for interpretation. All regressions contain year-month fixed effects, and columns 2 and 4 in Panel A contain fund fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Panel A: Fund Fixed Effects

	Dependent Variable: Abnormal Return (t+1)			
	(1)	(2)	(3)	(4)
Any Download	0.094** (2.58)	0.055 (1.19)	0.064 (1.50)	0.047 (0.97)
Above Median Download			0.066* (1.72)	0.033 (0.77)
AUM (t)	0.015 (0.79)	-0.463*** (-7.99)	0.010 (0.50)	-0.465*** (-8.02)
Abn Ret (t)	0.096*** (9.30)	0.069*** (6.73)	0.096*** (9.29)	0.069*** (6.73)
Fund FE	No	Yes	No	Yes
Date FE	Yes	Yes	Yes	Yes
Adjusted R ²	0.06	0.08	0.06	0.08
Number Firms	557	556	557	556
Observations	43435	43434	43435	43434

Panel B: Abnormal Time-Varying Strategies

	Dependent Variable: Abnormal Return (t+1)	
	(1)	(2)
Abnormal Downloads(t)	0.0632 (1.24)	
High Abnormal Downloads(t)		0.0826** (2.06)
Low Abnormal Downloads(t)		0.0343 (0.84)
AUM (t)	0.0274 (1.42)	0.0258 (1.34)
Abn Ret (t)	0.0899*** (7.16)	0.0898*** (7.16)
Date FE	Yes	Yes
Fund FE	No	No
Adjusted R ²	0.06	0.06
Number Firms	555	555
Observations	42321	42321

Table IA.5: Information Processing or Complementary Private Information?

The table reports the relation between download activity by a hedge fund in month t and the fund's abnormal return (measured in percent) in month $t + 1$. The monthly abnormal returns are calculated using the Fung-Hsieh seven factor model. "Any Download" indicates whether the fund accessed any filings in month t . "Above Median" is an indicator variable for whether the fund falls above the month's cross-sectional median number of downloads (conditional on any downloads of the form type). "Above Median" usage is interacted with the following filing characteristics: the median age (in days) of the accessed 10-K/Qs, the median file size (in KB) of the accessed 10-K/Qs, the median number of other hedge funds viewing the accessed 10-K/Qs, the median idiosyncratic volatility (IVol) of the accessed 10-K/Q firms (measured as the average squared daily abnormal return of the filer over the preceding year), the median intensity with which the hedge fund tracked the filer (measured as number of downloads of firm filings in prior months), the median Amihud (2002) illiquidity of the accessed 10-K/Q firms, and the median level of textual uncertainty in the accessed 10-K/Qs (measured using the proportion of uncertain words determined using the Loughran and McDonald (2011) lexicon). In each case, the median is taken within firm-month across all 10-K/Qs accessed by the fund. All continuous variables, except returns, are standardized. Regressions contain year-month and fund fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any 10K/Q	0.0729* (1.90)	0.0725* (1.89)	0.0734* (1.92)	0.0726* (1.90)	0.0733* (1.91)	0.0725* (1.89)	0.0735* (1.92)
Above Median 10K/Q	0.0167 (0.34)	0.0250 (0.51)	0.0280 (0.57)	0.0173 (0.35)	0.0148 (0.30)	0.0174 (0.35)	0.0163 (0.33)
- x Age	-0.0415 (-1.21)						
- x File Size		-0.0854** (-2.13)					
- x Other HF Views			0.0647* (1.77)				
- x IVol				0.0143 (0.41)			
- x Tracked					0.0427 (1.48)		
- x Amihud						0.0047 (0.14)	
- x Uncertainty							0.0881** (2.22)
AUM (t)	-0.4657*** (-8.02)	-0.4658*** (-8.01)	-0.4663*** (-8.05)	-0.4653*** (-8.03)	-0.4669*** (-8.05)	-0.4654*** (-8.03)	-0.4662*** (-8.05)
Abn Ret (t)	0.0691*** (6.74)	0.0691*** (6.73)	0.0691*** (6.73)	0.0691*** (6.73)	0.0691*** (6.73)	0.0691*** (6.74)	0.0692*** (6.74)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fund FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Number Firms	556	556	556	556	556	556	556
Observations	43434	43434	43434	43434	43434	43434	43434

Table IA.6: Robotic vs. Human Information Acquisition

The table reports the relation between download activity by a hedge fund in month t and the fund's abnormal return (measured in percent) in month $t + 1$. A fund's monthly download activity is determined to be robot-initiated if a fund downloaded 50 or more filings in a day and the median time interval between downloads was less than 30 seconds that day. If a fund's activity is robot-initiated in more than one month in the sample, then the fund is deemed a "Scraper." The monthly abnormal returns are calculated using the Fung-Hsieh seven factor model. "Any Download" indicates whether the fund accessed any filings in month t . "Above Median" is an indicator variable for whether the fund falls above the month's cross-sectional median number of downloads of a form type (conditional on any downloads of the form type). AUM is standardized for interpretation. Regressions contain year-month fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)			
	(1)	(2)	(3)	(4)
Scraper	0.122*** (3.18)	0.271*** (3.66)	0.272*** (3.66)	0.272*** (3.67)
Any Downloads	0.068* (1.81)	0.085** (2.06)	0.073 (1.63)	0.062 (1.33)
Any Downloads x Scraper		-0.170** (-2.15)	-0.186** (-2.36)	-0.090 (-1.01)
Above Median Downloads			0.035 (0.87)	0.067 (1.41)
Above Median Downloads x Scraper				-0.137* (-1.77)
AUM (t)	0.005 (0.27)	0.006 (0.32)	0.004 (0.22)	0.004 (0.20)
Abn Ret (t)	0.096*** (9.28)	0.096*** (9.28)	0.096*** (9.28)	0.096*** (9.28)
Date FE	Yes	Yes	Yes	Yes
Adjusted R ²	0.06	0.06	0.06	0.06
Number Firms	557	557	557	557
Observations	43435	43435	43435	43435

Table IA.7: Robots and Information Processing Ability

The table reports the relation between the fund's next-month abnormal return (measured in percent) and characteristics of filings downloaded by hedge funds using robotic information acquisition in month t . The sample is constrained to months in which "Scraper" funds as defined in Table 8 employ robotic acquisition. By definition, all months contain above median download activity. Therefore, the coefficients on filing characteristics in this table are comparable to the coefficients on the interactions of filing characteristics with above median downloads shown in Table 7. The monthly abnormal returns are calculated using the Fung-Hsieh seven factor model. The filing characteristics of the form 10-K/Qs accessed by a fund in month t are: the median age (in days) of the accessed 10-K/Qs, the median file size (in KB) of the accessed 10-K/Qs, the median number of other hedge funds viewing the accessed 10-K/Qs, the median idiosyncratic volatility (IVol) of the accessed 10-K/Q firms (measured as the average squared daily abnormal return of the filer over the preceding year), the median intensity with which the hedge fund tracked the filer (measured as number of downloads of firm filings in prior months), the median Amihud (2002) illiquidity of the accessed 10-K/Q firms, and the median level of textual uncertainty in the accessed 10-K/Qs (measured using the proportion of uncertain words determined using the Loughran and McDonald (2011) lexicon). In each case, the median is taken within firm-month across all 10-K/Qs accessed by the fund. All variables, except returns, are standardized. Regressions contain year-month fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	-0.0593 (-0.93)						
File Size		0.1140 (1.52)					
Other HF Views			0.1652 (1.27)				
IVol				0.0750 (0.33)			
Tracked					-0.0996 (-0.82)		
Amihud						-0.0387 (-0.22)	
Uncertainty							-0.1079 (-0.69)
AUM (t)	0.1924* (1.74)	0.2523** (2.11)	0.1849* (1.71)	0.2174* (1.85)	0.2249* (1.80)	0.2179* (1.82)	0.2184* (1.86)
Abn Ret (t)	0.1637** (2.21)	0.1664** (2.08)	0.1673** (2.22)	0.1705** (2.22)	0.1702** (2.18)	0.1749** (2.30)	0.1736** (2.24)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fund FE	No	No	No	No	No	No	No
Adjusted R ²	0.09	0.10	0.10	0.09	0.09	0.09	0.09
Number Firms	45	45	45	45	45	45	45
Observations	304	304	304	304	304	304	304

Table IA.8: Financial Statement Analysis Specialists vs. Generalists

This table compares the usage-performance relations of funds that specialize in processing form 10-K/Q filings and those that do not. A fund is classified as a Form 10-K/Q specialist or a generalist. Financial statement analysis specialists are defined as funds for which both (1) the proportion of their total EDGAR usage that is due to 10-K/Q filings and (2) the total number of 10-K/Q filings they access (scaled by months in the sample) are above the cross-sectional medians for each measure. Generalists are non-10-K/Q specialists. For specialists and generalists, the table reports the relation between the fund's abnormal return (measured in percent) in month $t + 1$ and both the extensive margin of download activity and the intensive margin of download activity (above median downloads of 10-K/Q filings in given month conditional on viewing a 10-K/Q filing) in month t . "Any 10K/Q" indicates whether the fund accessed any 10K/Q filings in month t . The monthly abnormal returns are calculated using the Fung-Hsieh seven factor model. AUM is standardized for interpretation. All regressions contain year-month fixed effects, and columns 2 and 4 contain fund fixed effects. Standard errors are clustered by fund. t statistics are in parentheses, and statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

	Dependent Variable: Abnormal Return (t+1)			
	10K/Q Strategy		Generalist Strategy	
	(1)	(2)	(3)	(4)
Any 10K/Q	-0.007 (-0.11)	-0.033 (-0.44)	0.074 (1.60)	0.114** (2.55)
Above Med 10K/Q	0.073 (1.31)	0.063 (0.94)	0.099* (1.70)	0.012 (0.17)
AUM (t)	0.035 (1.35)	-0.283*** (-3.44)	-0.012 (-0.41)	-0.551*** (-7.05)
Abn Ret (t)	0.116*** (6.42)	0.088*** (4.82)	0.088*** (7.14)	0.062*** (5.04)
Date FE	Yes	Yes	Yes	Yes
Fund FE	No	Yes	No	Yes
Adjusted R ²	0.09	0.10	0.06	0.07
Number Firms	182	181	375	375
Observations	15430	15429	28005	28005